

REPUBLIQUE DU CAMEROUN
Paix-Travail-Patrie

UNIVERSITE DE YAOUNDE I

FACULTE DES ARTS, LETTRES ET
SCIENCES HUMAINES

CENTRE DE RECHERCHE ET DE
FORMATION DOCTORALE EN ARTS
LANGUES ET CULTURES

UNITE DE RECHERCHE ET DE
FORMATION DOCTORALE EN
LANGUE ET LITTERATURE

DEPARTEMENT DE LANGUES,
LITTERATURES ET CIVILISATIONS
GERMANIQUES



REPUBLIC OF CAMEROON
Peace-Work-Fatherland

THE UNIVERSITY OF YAOUNDE I

FACULTY OF ARTS, LETTERS AND
SOCIAL SCIENCES

POST GRADUATE SCHOOL FOR
ARTS, LANGUAGES AND CULTURES

DOCTORAL RESEARCH UNIT FOR
LANGUAGES AND LITERATURES

DEPARTMENT OF GERMANIC
LANGUAGES, LITERATURE AND
CIVILISATIONS

COMPUTER- UND TEXTLINGUISTIK
ENTWICKLUNG EINES SYSTEMS FÜR DIE AUTOMATISCHE
AUFLÖSUNG VON PRONOMINALADVERBIENANAPHERN : Ein Beitrag
zur theoretischen und angewandten Computerlinguistik

Thèse présentée en vue de l'obtention d'un Doctorat/ Ph.D en études germaniques

Option : Linguistique allemande

par :

Nelson Leroi SOFAC FOGO

MASTER EN ÉTUDES GERMANIQUES, OPTION LANGUE



Travail présenté devant un jury ainsi constitué :

Président: SIMO David, Professeur, Université de Yaoundé1

Rapporteur: MBONGUE Joseph, Maître de Conférences, Université de Yaoundé1

Membres: PEPOUNA Souleymane, Maître de Conférences, Université de Yaoundé1
NDEFFO Alexandre, Maître de Conférences, Université de Buéa
MASSOCK Georges, Maître de Conférences, Université de Yaoundé 1

MAI 2025

WIDMUNG

Meinem verstorbenen Vater **SOFAC Gilbert**

Meiner lieben Mutter **SOFAC Ide Bernadette geboren AFOUAZE**

Dem Deutschen Akademischen Austauschdienst (DAAD) und der Universität Potsdam

DANKSAGUNG

Die Fertigstellung dieser Arbeit wäre ohne den Beitrag einiger Personen nicht möglich. An der ersten Stelle der Liste stehen natürlich Familienangehörigen. In dieser Gruppe möchte ich insbesondere nicht nur meiner lieben Mutter, der ich die Arbeit schon gewidmet habe, danken, sondern auch meiner Tante **Feulefack Sabine** und meinen beiden großen Schwestern **Sofac Edwig & Sofac Nina**, ohne die ich höchstwahrscheinlich kein universitäres Studium angeschaltet hätte. Weiterhin möchte ich mich bei Personen meines Studiums-Kreises bedanken. In dieser zweiten Gruppe möchte ich **meinem Betreuer, Herrn Professor Mbongue** einen besonderen Dank aussprechen. Seine ständige Bereitschaft (rund um die Uhr, sei es per WhatsApp oder per Mails) und wertvolle Beratungen, sowohl bei der Gestaltung als auch bei der Durchführung dieses Forschungsprojekts, halfen mir bei jeder getroffenen Schwierigkeit Überwindungswege zu finden. Außerdem bedanke ich mich auch ganz herzlich bei **dem DAAD und vor allem beim Prof. Dr. Manfred Stede vom Lehrstuhl für angewandte Computerlinguistik am Departement für Linguistik der humanwissenschaftlichen Fakultät der Universität Potsdam in Deutschland**. Dieser Letzte gab mir In der Tat die Chance seine Arbeitsgruppe zu integrieren und somit eine neue Forschungsphilosophie zu erlernen und anschließend eine neue Forschungsumwelt zu entdecken. Dabei konnte ich dank regelmäßiger stattfindenden Seminare; Beratungsgespräche und Ausstattung der Forschungseinrichtung, wissenschaftliche Erfahrungen sammeln, welche mir dann zum Ausbau dieser Dissertationsarbeit geholfen und geführt haben. Letztendlich möchte ich **allen Angestellten** (D

ozenten, Sekretären und Bibliothekaren) der Deutschabteilung der Universität von Jaunde 1 für die Ausbildung, die sie mir während meines akademischen Aufenthalts an der Universität von Jaunde 1 gegeben haben, danken.

RÉSUMÉ

Les présents travaux de recherche, effectués sous le thème « **Linguistique computationnelle et textuelle : Développement d'un logiciel pour la résolution automatique de l'adverbe pronominal anaphorique** » partent du constat selon lequel, les logiciels (Systèmes) de résolution d'expressions anaphoriques (anaphores) existants n'étaient pas capables d'identifier l'antécédent d'une anaphore (Résolution de l'anaphore) lorsque celle-ci est un adverbe. Partant de ce constat, nous nous sommes donc intéressés aux raisons pouvant expliquer ce manquement. Et chemin faisant, nous nous décidâmes donc de contribuer à combler ce déficit en proposant un Système dédié à la résolution automatique de l'adverbe pronominal anaphorique. Ainsi cette thèse de doctorat contribue à l'expansion d'une branche de la linguistique computationnelle dénommée NLP (Natural Language Processing) / TALN (Traitement Automatique du Langage Naturel).

Ces travaux de recherche se sont déroulés essentiellement en 4 phases. Nous avons en effet, lors de la première phase (Chapitre 1&2), rassembler un maximum d'informations sur la nature et la portée respective des deux grands domaines de la linguistique appliquée qui constituent le socle de base de notre recherche. Une fois ces informations structurelles gagnées, nous nous sommes intéressés, dans un second temps, aux différentes propriétés (morphologiques, sémantiques et anaphoriques (Pragma-linguistiques)) de l'adverbe pronominal. Et une fois ces connaissances sur la nature et le fonctionnement anaphorique de l'adverbe pronominal en notre possession, nous avons directement mis sur pied un algorithme provisoire de résolution de l'adverbe pronominal anaphorique (chapitre 3). Par la suite, nous avons pu rendre cet algorithme définitif et cela notamment grâce aux connaissances acquises lors de la recherche d'informations sur la conception et la structuration de systèmes informatiques dédiés à la réalisation ou automatisation de telle tâche (Chapitre 4). Une fois notre algorithme revisité au prisme des principes informatiques qui régissent les pratiques dans ce domaine, nous avons fait appel à un spécialiste (Programmeur) pour encoder notre algorithme en langage compréhensible par un ordinateur. Au sortir de cette phase, notre système de résolution de l'adverbe pronominal anaphorique dénommé (PAAR) était mis sur pied (Chapitre 5). Ainsi, il ne nous restait plus, à travers un Corpus méthodologiquement conçu, qu'à vérifier sa performance (Chapitre 6).

De cela, il en ressort notamment que le Système PAAR réalise une précision de 94% dans l'identification des adverbes pronominaux ; une performance de 83% dans la classification de ces adverbes selon leur nature respective (Anaphore, cataphore ou Deixis) et pour finir une performance de 73% dans la résolution proprement dite des adverbes pronominaux anaphoriques. Ce pendant nous avons tout de même, après analyse du fonctionnement des différents modules qui constituent le Système (PAAR), identifié quelques leviers d'amélioration des performances de PAAR. Ceux-ci sont listés de manière exhaustive à la fin de ce travail. Mais nous pouvons, entre autre, mentionner l'amélioration de la qualité des deux Parsers (Phrases et mots).

ABSTRACT

The present research work, carried out under the theme "**Computational and textual linguistics: Development of software for the automatic resolution of the anaphoric pronominal adverb**" is based on the observation that software (systems) for the resolution of anaphoric expressions (anaphors) existing ones were not able to identify the antecedent of an anaphor (Resolution of the anaphor) when this one is an adverb. Based on this observation, we are therefore interested in the reasons that may explain this failure. And along the way, we therefore decided to contribute to fill this deficit by proposing a System dedicated to the automatic resolution of the anaphoric pronominal adverb.

This research was essentially carried out in 4 phases. In fact, during the first phase (Chapter 1&2), we gathered as much information as possible on the nature and respective scope of the two major fields of applied linguistics which constitute the basic foundation of our research. Once this structural information gained, we are interested, in a second time, in the various properties (morphological, semantic and anaphoric (Pragma-linguistics)) of the pronominal adverb. And once this knowledge on the nature and the anaphoric functioning of the pronominal adverb in our possession, we directly set up a provisional algorithm for resolving the anaphoric pronominal adverb (chapter 3). Afterwards, we were able to make this algorithm definitive and this in particular thanks to the knowledge acquired during the search for information on the design and structuring of computer systems dedicated to the realization or automation of such a task (Chapter 4). Once our algorithm was revisited through the prism of computer principles that govern practices in this field, we called on a specialist (Programmer) to encode our algorithm in a language understandable by a computer. At the end of this phase, our resolution system for the anaphoric pronominal adverb named (PAAR) was set up (Chapter 5). Thus, all that remained was for us, through a methodologically designed Corpus, to verify its performance (Chapter 6). we called on a specialist (Programmer) to encode our algorithm in a language understandable by a computer. At the end of this phase, our resolution system for the anaphoric pronominal adverb named (PAAR) was set up (Chapter 5). Thus, all that remained was for us, through a methodologically designed Corpus, to verify its performance (Chapter 6).

From this, it emerges in particular that the PAAR System achieves an accuracy of 94% in the identification of pronominal adverbs; a performance of 83% in the classification of these adverbs according to their respective nature (Anaphora, cataphore or Deixis) and finally a performance of 73% in the actual resolution of anaphoric pronominal adverbs. However, after analyzing the operation of the various modules that make up the System (PAAR), we have identified some levers for improving the performance of PAAR. These are listed exhaustively at the end of this work. But we can, among other things, mentioned the improvement of the quality of the two Parsers (Phrases and words).

INHALTSVERZEICHNIS

WIDMUNG	2
DANKSAGUNG	3
RÉSUMÉ	4
ABSTRACT	5
INHALTSVERZEICHNIS	6
TABELLENVERZEICHNIS	12
ABBILDUNGSVERZEICHNIS	14
KAPITEL0: EINFÜHRUNG.....	1
ERSTER TEIL: TERMINOLOGISCHE ANNÄHRUNGEN (Computerlinguistik & Textlinguistik)	19
KAPITEL 1: COMPUTERLINGUISTIK (CL)	20
a. Gegenstandbestimmung	20
i. Computerlinguistik als Fachbereich und Computerlinguistische Fachgesellschaften	26
b. Methodische und theoretische Grundlage der Computerlinguistik: Regelbasierte und statistische Verfahren.....	43
c. Computerlinguistische Systeme und Ihre alltäglichen Anwendungen.....	47
i. Texttechnologische Anwendungen.....	48
ii. Dialogsysteme (Chatboots)	51
iii. CL& DIDAKTIK	52
iv. CL &WIRTSCHAFT: Berufsfelder von Computerlinguisten.....	53
KAPITEL 2: TEXTLINGUISTIK UND DIE ANAPHORIK.....	55
2.1. Zu Begriffen <i>Text und Textlinguistik</i>	55

2.1.2. Die Textlinguistik.....	58
2.2. Zum Begriff <i>Textgrammatik</i>	61
2.2.1. Die Textkohäsion.....	62
2.2.2. Textkohärenz und Thematische Progression.....	64
2.3. Die Anaphorik.....	67
2.3.1. Anfänge der Forschungen zu Verweisphänomenen in der Textlinguistik.....	70
2.3.2. Die expliziten Anaphern.....	74
2.3.2.1. Explizite pronominale Anaphern.....	74
2.3.2.2. Die Nullanapher.....	77
2.3.2.3. Explizite Nominalphrasenanaphern.....	78
2.3.2.4. Pluralanaphern	79
2.3.2.5. Komplexe Anaphern.....	79
2.3.3. Die impliziten Anaphern.....	80
ZWEITER TEIL: THEORETISCHE ANNÄHRUNGEN (PRONOMINALADVERBIEN & AR-SYSTEME)	85
KAPITEL 3: PRONOMINALADVERBIEN (PAn) UND ANAPHORIK:	
Das Pronominaladverb (PA) als Anapher.....	86
3.1. Adverb oder Adverbial: Gegenstandbestimmung.....	86
3.2. Klassifikation von Adverbien: Inhaltlichen und Funktionalen Kriterien	88
3.2.1. Nach dem inhaltlichen Faktor.....	88
3.2.2. Nach dem Funktionalen Faktor.....	91
3.3. Pro-Adverbien und Pronominaladverbien	94
3.3.1. Pro-Adverbien.....	95

KAPITEL 4: ANSÄTZE UND THEORIEN ZUR AUTOMATISCHEN ANAPHERNAUFLÖSUNG: Algorithmen, AR-Ansätze und AR-Systeme.....142

4.1. Historische und thematische Darstellung von bestehenden AR-Ansätzen; - Algorithmen und -Systemen (Von 1960 bis 1990).....	143
4.1.1. Das System STUDENT von Bobrow (1964); SHERDLU von Winograd (1972); LUNAR von Woods & Co (1977) und der Algorithmus von Hobbs (1977).....	143
4.1.2. Der BFP Algorithmus von Brennan, Friedman & Pollard (1987): Eine Implementation der Centeringtheorie.....	147
a. Computerlinguistische Ansätze, Algorithmen & Systeme zur Auflösung von koreferenten Pronominalanaphern (Personal- Possessiv- und Reflexivpronomina)	155
i. Auf Heuristiken basierende AR-Algorithmen und oder AR-Systeme für Pronomen im Deutschen und im Englischen.....	156
1. Das AR-System von Shalom Lappin und Herbert J. Leass (1994).....	156
2. Der wissensarme Ansatz von Kennedy & Bougarrev (1996): Die sogenannte KP-RAP.....	166
3. Die Übertragung von RAP ins Deutsche: German-RAP (G-RAP).....	167
a. Der Ansatz von Hinrichs & Co. (2005)	167
b. G-RAP nach Hollger Wunsch (2010).....	169
c. Der Ansatz von Ludger Paschen (2013).....	172
ii. Auf Technik des Maschinenlernens basierende AR-Systeme: Einsatz der Künstlichen Intelligenz bei der automatischen Auflösung von Anaphern.....	174
2.3.3.1. Künstliche Intelligenz (KI) und Maschinenlernen (ML).....	174
2.3.3.1.1. Intelligenz und Intelligenzarten.....	175
2.3.3.1.2. Zum Begriff Künstliche Intelligenz.....	178
2.3.3.1.3. Das maschinelle Lernen (Machine Learning).....	182
2.3.3.2. Einsatz des maschinellen Lernens (ML) bei der Auflösung von Anaphern.....	188

4.3. Entworfenen Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern.....	192
--	------------

DRITTER TEIL: EMPIRISCHER TEIL (HERSTELLUNG; EINSETZUNG UND AUSWERTUNG DES PAAR).....	196
--	------------

KAPITEL 5: UMWANDLUNG DES AR-ALGORITHMUS INS AR-SYSTEM: ENTSTEHUNG DES PRONOMINALADVERBIENANAPHERNREOLVERS (PAAR).....	197
---	------------

5.1. Algorithmen zur Ausschließung von kataphorischen/deiktischen PA.....	197
---	-----

5.1.1. Der Filter für kataphorische hier- und da-PA.....	197
--	-----

5.1.2. Der Filter für kataphorische/deiktische wo-PA.....	198
---	-----

5.2. Angaben über die Auswahl des Parsers.....	199
--	-----

5.3. Angaben über den Ausbau eines geeigneten morpho-orthographischen Taggers.....	200
--	-----

5.4. Angaben über die Kollaboration mit dem Programmierer: Entehrung des PAARS.....	200
---	-----

5.5. Aufbereitung der Datensätze und Einsatz des PAAR.....	205
--	-----

5.5.1. Aufbau eines Textkorpus.....	205
-------------------------------------	-----

5.5.1.1. Bestehende Textkorpora für korpuslinguistische Untersuchungen im Deutschen.....	206
--	-----

5.5.2. Einsatz des PAAR aufs PAK (Pronominaladverbien-Korpus)	212
---	-----

KAPITEL 6: AUSWERTUNG DES PAAR-EINSATZES: ANALYSEN, INTERPRETATIONEN, ERGEBNISSE UND HYPOTHESENPRÜFUNG.....	231
--	------------

6.1. Analyse und Interpretation des Einsatzes vom PAAR auf das Textkorpus.....	231
--	-----

6.1.1. Einsatz des PAAR auf den Text N°2.....	231
6.1.2. Einsatz des PAAR auf den Text N°3.....	232
6.1.3. Einsatz des PAAR auf den Text N°4.....	234
6.1.4. Einsatz des PAAR auf den Text N°5.....	235
6.1.5. Einsatz des PAAR auf den Text N°6.....	236
6.1.6. Einsatz des PAAR auf den Text N°7.....	238
6.1.7. Einsatz des PAAR auf den Text N°8.....	239
6.1.8. Einsatz des PAAR auf den Text N°9.....	240
6.1.9. Einsatz des PAAR auf den Text N°10.....	241
6.1.10. Einsatz des PAAR auf den Text N°11.....	242
6.1.11. Einsatz des PAAR auf den Text N°12.....	244
6.1.12. Einsatz des PAAR auf den Text N°13.....	245
6.1.13. Einsatz des PAAR auf den Textabschnitt (N°14-N°20)	246
6.1.14. Einsatz des PAAR auf den Textabschnitt (N°21-N°30)	249
6.1.15. Einsatz des PAAR auf den Textabschnitt (N°31-N°50)	255
6.1.16. Einsatz des PAAR auf den Textabschnitt (N°51-N°75)	264
6.1.15. Einsatz des PAAR auf den Textabschnitt (N°76-N°100)	264
6.1.15. Einsatz des PAAR auf den Textabschnitt (N°101-N°154)	264
6.2. Ergebnisse der Einsatzaktion vom PAAR aufs PAK: Berechnung der Präzision von PAAR.....	265
6.3. Hypothesenprüfung, Schlussbetrachtung und Ausblicke.....	272
7. LITERATURVERZEICHNIS.....	276
a. BÜCHER.....	277
b. ARTIKEL.....	283
c. VORLESUNGEN UND SKRIPTEN.....	292
d. LEXIKA UND WÖRTERBÜCHER.....	293
e. ABSCHLUSSARBEITEN.....	293
<i>f. DISSERTATIONEN.....</i>	<i>293</i>
g. ERZÄHLUNGEN, ROMANE UND ZEITUNGEN.....	294
h. INTERNETQUELLEN.....	294

8. ANHANG.....	295
i. Das mit der DGD-Team (IDS) geführte Gespräch.....	1A
j. Weitere Angaben zum Aufbau des PAK.....	3A
k. Einsatz des PAAR auf den Textabschnitt (N°51-75)	21A
l. Einsatz des PAAR auf den Textabschnitt (N°76-100)	35A
m. Einsatz des PAAR auf den Textabschnitt (N°101-154)	51A
n. TDR.....	74A
o. PAAR auf einem CD-ROM/ einer D-Karte/ einem USB-Stick.....	79A

TABELLENVERZEICHNIS

<u>Tabelle 1:</u> Werte der Recence (Vgl ; Ebd : 233).....	11
<u>Tabelle 2:</u> Werte von syntaktischen Funktionen.....	11
<u>Tabelle3:</u> Eine Taxonomie der Studien- und Forschungsorte für Computerlinguistik in Deutschland.....	40
<u>Tabelle 4:</u> Liste der Pronominaladverbien (Vgl. Dudengrammatik, 2009: 580).....	104
<u>Tabelle 5:</u> Syntaktische Eigenschaften von DAMIT als Subjunktion und als PA.....	115
<u>Tabelle6:</u> Deiktische/kataphorische und anaphorische Erscheinungen von Wo-Verbindungen.....	134
<u>Tabelle 7:</u> Die Technik des Mustervergleichs im AR-System STUDENT von Bobrow (1964).....	144
<u>Tabelle 8:</u> Beispiel einer Anapherresolution durch die CT ohne Rücksicht auf (CB _{Si} & CB _{Si+1})-Übergänge.....	150
<u>Tabelle 9:</u> Häufig auftretende (CB _{Si} & CB _{Si+1})- Übergänge bei anaphorischen Beziehungen.....	150
<u>Tabelle 10:</u> Beispiel einer Anaphernresolution durch die CT mit Rücksicht auf (CB _{Si} & CB _{Si+1})-Übergänge.....	151
<u>Tabelle 11:</u> Salienzfaktoren und Gewichtungen bei RAP (Vgl. Holger Wunsch 2010:106)...	163
<u>Tabelle 12:</u> Erweiterung von Salienzfaktoren und deren jeweiligen Gewichtung bei Lappin & Leass 1994 (Vgl. L.Paschen 2013: 60)	164
<u>Tabelle 13:</u> <i>Features and their weights in RAP-G, Hinrichs et al. (2005)'s adaption of Lappin and Leass (1994)'s approach to English third person pronoun resolution.</i>	168
<u>Tabelle 14:</u> Morpho-orthographischer Tagger.....	194
<u>Tabelle 15:</u> Syntaktischer Parser.....	194
<u>Tabelle 16:</u> Filter zur Identifizierung von kataphorischen Da- und Hier-PA.....	195
<u>Tabelle 17:</u> Filter zur Identifizierung von deiktischen/kataphorischen Wo-PA.....	195
<u>Tabelle 18:</u> Konstituenten des PAK.....	210
<u>Tabelle 19:</u> Distribution der PA im PAK.....	212
<u>Tabelle 20:</u> Performanz vom PAAR bei Erkennung von PA im Text N°1.....	228
<u>Tabelle 21:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°1.....	228
<u>Tabelle 22:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°1..	228
<u>Tabelle 23</u> Performanz vom PAAR bei der Erkennung von PA im Text N°2.....	230
<u>Tabelle 24:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°2.....	230
<u>Tabelle 25:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°2..	230
<u>Tabelle 26</u> Performanz vom PAAR bei der Erkennung von PA im Text N°3.....	231
<u>Tabelle 27:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°3.....	231
<u>Tabelle 28:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°3..	231
<u>Tabelle 29</u> Performanz vom PAAR bei der Erkennung von PA im Text N°4.....	232
<u>Tabelle 30:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°4.....	233
<u>Tabelle 31:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°4..	233
<u>Tabelle 32</u> Performanz vom PAAR bei der Erkennung von PA im Text N°5.....	234
<u>Tabelle 33:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°5.....	234
<u>Tabelle 34:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°5..	234
<u>Tabelle 35</u> Performanz vom PAAR bei der Erkennung von PA im Text N°6.....	235
<u>Tabelle 36:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°6.....	235
<u>Tabelle 37:</u> Performanz vom PAAR bei Erkennung von PA im Text N°7.....	236
<u>Tabelle 38:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°7.....	237
<u>Tabelle 39:</u> Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°7..	237
<u>Tabelle 40:</u> Performanz vom PAAR bei Erkennung von PA im Text N°8.....	238
<u>Tabelle 41:</u> Performanz vom PAAR bei Erkennung der Natur von PA im Text N°8.....	238

<u>Tabelle 42:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°8..	238
<u>Tabelle 43:</u>	Performanz vom PAAR bei Erkennung von PA im Text N°9.....	239
<u>Tabelle 44:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Text N°9.....	239
<u>Tabelle 45:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°9..	239
<u>Tabelle 46:</u>	Performanz vom PAAR bei Erkennung von PA im Text N°10.....	240
<u>Tabelle 47:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Text N°10.....	240
<u>Tabelle 48:</u>	Performanz vom PAAR bei der Auflösung anaphorischer PA im Text N°10.....	240
<u>Tabelle 49:</u>	Performanz vom PAAR bei Erkennung von PA im Text N°11.....	241
<u>Tabelle 50:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Text N°11.....	241
<u>Tabelle 51:</u>	Performanz vom PAAR bei der Auflösung anaphorischer PA im Text N°11.....	241
<u>Tabelle 52:</u>	Performanz vom PAAR bei Erkennung von PA im Text N°12.....	242
<u>Tabelle 53:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Text N°12.....	242
<u>Tabelle 54:</u>	Performanz vom PAAR bei der Auflösung anaphorischer PA im Text N°12.....	243
<u>Tabelle 55:</u>	Performanz vom PAAR bei Erkennung von PA im Text N°13.....	244
<u>Tabelle 56:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Text N°13.....	244
<u>Tabelle 57:</u>	Performanz vom PAAR bei der Auflösung anaphorischer PA im Text N°13.....	244
<u>Tabelle 58:</u>	Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°14-20).....	247
<u>Tabelle 59:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°14-20).....	247
<u>Tabelle 60:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°14-20)	248
<u>Tabelle 61:</u>	Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°21-30).....	253
<u>Tabelle 62:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°21-30)	253
<u>Tabelle 63:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°21-30)	253
<u>Tabelle 64:</u>	Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°31-50)...	262
<u>Tabelle 65:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°31-50)	263
<u>Tabelle 66:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°31-50)	263
<u>Tabelle 67:</u>	Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°51-75).....	34A
<u>Tabelle 68:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°51-75).....	34A
<u>Tabelle 69:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°51-75).....	35A
<u>Tabelle 70:</u>	Performanz vom PAAR bei Erkennung von PA im Textabschnitt(N°76-100)..	50A
<u>Tabelle 71:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°76-100)	50A
<u>Tabelle 72:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°76-100)	51A
<u>Tabelle 73:</u>	Performanz vom PAAR bei Erkennung v PA im Textabschnitt(N°101-154).....	73A
<u>Tabelle 74:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°101-154).....	73A
<u>Tabelle 75:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°101-154).....	74A
<u>Tabelle 76:</u>	Performanz vom PAAR bei Erkennung von PA im PAK.....	263
<u>Tabelle 77:</u>	Performanz vom PAAR bei Erkennung der Natur von PA im PAK.....	266
<u>Tabelle 78:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK.....	268

ABBILDUNGSVERZEICHNIS

Abbildung 1: Die globale Struktur von RESUMAN ₀ (Vgl. Selmi.A 2019: 218).....	7
Abbildung 2: Die Segmentierungsfunktion (Vgl. Ebd.: 221).....	9
Abbildung 3: Morphosyntaktische Etikettierung durch Fips (Vgl. Ebd: 224).....	9
Abbildung 4: Verrechnung der Salienz durch das AR-System RESUMAN ₀	12
Abbildung 5: Schema des Algorithmus RESUMAN ₀ (Vgl. Selmi, 2019: 228).....	13
Abbildung 6: Vorübergehender Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern.....	141
Abbildung 7: <i>Distribution of referential relations of third person pronouns in the first 1 188 articles of the TuBa-D/Z treebank. The relations in dark gray" are handled by G-RAP.</i> (Vgl. Holger Wunsch 2010: 114).....	169
Abbildung 8: <i>Flow chart of German RAP</i> (Vgl. Holger Wunsch 2010: 122)	170
Abbildung 9: <i>Computation of salience value.</i> (Vgl. Ebd.: 123)	171
Abbildung 10: <i>Übertragung von RAP (Lappin & Leass 1994) ins Deutsche nach Ludger Pacher 2013</i> (Vgl. Ludger Paschen 2013: 63).	173
Abbildung 11: Allgemeine Architektur von AR-Systemen (Heuristische Systeme)	174
Abbildung 12: Endgültiger Algorithmus zur Auflösung von Pronominaladverbienanaphern.....	193
Abbildung 13: Algorithmen zur Filterung von kataphorischen da-/hier-PA.....	197
Abbildung 14: Algorithmen zur Filterung von deiktischen und kataphorischen wo-PA (Prototyp 1).....	198
Abbildung 15: Algorithmen zur Filterung von deiktischen und kataphorischen wo-PA (Prototyp 2).....	198
Abbildung 16: Globale Architektur des PAAR.....	204
Abbildung 17: Archivierungsnomenklatur (Wechselseitige Beziehungen zwischen IDS-Korpora; DeReKo und COSMAS II.....	209
Abbildung 18: Startseite der Software ANACONDA NAVIGATOR (1/3).....	212
Abbildung 19: Startseite der Software ANACONDA NAVIGATOR (2/3).....	213
Abbildung 20: Startseite der Software ANACONDA NAVIGATOR (3/3).....	213
Abbildung 21: Startseite der Python-Entwicklungsumgebung SPYDER.....	214
Abbildung 22: Input des Textes N°1 in der Exekutionsstelle von SPYDER (1/2).....	215
Abbildung 23: Input des Textes N°1 in der Exekutionsstelle von SPYDER (2/2).....	215
Abbildung 24: Aufruf des Moduls Preprocessing vom PAAR (1/4).....	216
Abbildung 25: Aufruf des Moduls Preprocessing vom PAAR (2/4).....	217
Abbildung 26: Aufruf des Moduls Preprocessing vom PAAR (3/4).....	217
Abbildung 27: Aufruf des Moduls Preprocessing vom PAAR (4/4).....	217
Abbildung 28: Aufruf der Klasse „Preprocessing_Language“ und der Methode „sentence_tokenizer“ (1/2).....	218
Abbildung 29: Aufruf der Klasse „Preprocessing_Language“ und der Methode „sentence_tokenizer“ (2/2).....	219
Abbildung 30: Aufruf der Methode „word_tokenizer“ (1/3).....	219
Abbildung 31: Aufruf der Methode „word_tokenizer“ (2/3).....	220
Abbildung 32: Aufruf der Methode „word_tokenizer“ (3/3).....	220
Abbildung 33: Aufruf des Moduls „Part-of-Speech“ (1/3).....	221
Abbildung 34: Aufruf des Moduls „Part-of-Speech“ (2/3).....	221
Abbildung 35: Aufruf des Moduls „Part-of-Speech“ (3/3).....	221
Abbildung 36: Aufruf der Klasse „find_pa“.....	222
Abbildung 37: Aufruf des Moduls „Anaphora_filter“ (1/8).....	223
Abbildung 38: Aufruf des Moduls „Anaphora_filter“ (2/8).....	223
Abbildung 39: Aufruf des Moduls „Anaphora_filter“ (3/8).....	223
Abbildung 40: Aufruf des Moduls „Anaphora_filter“ (4/8).....	224

<u>Abbildung 41:</u>	Aufruf des Moduls „Anaphora_filter“(5/8).....	224
<u>Abbildung 42:</u>	Aufruf des Moduls „Anaphora_filter“(6/8).....	225
<u>Abbildung 43:</u>	Aufruf des Moduls „Anaphora_filter“(7/8).....	225
<u>Abbildung 44:</u>	Aufruf des Moduls „Anaphora_filter“(8/8).....	226
<u>Abbildung 45:</u>	Aufruf der Klasse „anaphora_filter“ und der Methode „Da_or_Hier_PA_classifier“.....	224
<u>Abbildung 46:</u>	Aufruf des Hauptmoduls: Aufruf vom PAAR.....	225
<u>Abbildung 47:</u>	Einsatz vom PAAR zur Auflösung der erfundenen PA (1/3).....	226
<u>Abbildung 48:</u>	Einsatz vom PAAR zur Auflösung der erfundenen PA (2/3).....	226
<u>Abbildung 49:</u>	Einsatz vom PAAR zur Auflösung der erfundenen PA (3/3).....	226
<u>Abbildung 50:</u>	Einsatz des PAAR auf den Text N°2.....	230
<u>Abbildung 51:</u>	Einsatz des PAAR auf den Text N°3.....	231
<u>Abbildung 52:</u>	Einsatz des PAAR auf den Text N°4.....	232
<u>Abbildung 53:</u>	Einsatz des PAAR auf den Text N°5.....	233
<u>Abbildung 54:</u>	Einsatz des PAAR auf den Text N°6.....	235
<u>Abbildung 55:</u>	Einsatz des PAAR auf den Text N°7.....	236
<u>Abbildung 56:</u>	Einsatz des PAAR auf den Text N°8.....	237
<u>Abbildung 57:</u>	Einsatz des PAAR auf den Text N°9.....	238
<u>Abbildung 58:</u>	Einsatz des PAAR auf den Text N°10.....	239
<u>Abbildung 59:</u>	Einsatz des PAAR auf den Text N°11.....	241
<u>Abbildung 60:</u>	Einsatz des PAAR auf den Text N°12.....	242
<u>Abbildung 61:</u>	Einsatz des PAAR auf den Text N°13.....	243
<u>Abbildung 62:</u>	Einsatz des PAAR auf den Textabschnitt (N°14-20) (1)	245
<u>Abbildung 63:</u>	Einsatz des PAAR auf den Textabschnitt (N°14-20) (2)	246
<u>Abbildung 64:</u>	Einsatz des PAAR auf den Textabschnitt (N°14-20) (3)	246
<u>Abbildung 65:</u>	Einsatz des PAAR auf den Textabschnitt (N°14-20) (4)	247
<u>Abbildung 66:</u>	Einsatz des PAAR auf den Textabschnitt (N°21-30) (1)	249
<u>Abbildung 67:</u>	Einsatz des PAAR auf den Textabschnitt (N°21-30) (2)	250
<u>Abbildung 68:</u>	Einsatz des PAAR auf den Textabschnitt (N°21-30) (3)	251
<u>Abbildung 69:</u>	Einsatz des PAAR auf den Textabschnitt (N°21-30) (4)	251
<u>Abbildung 70:</u>	Einsatz des PAAR auf den Textabschnitt (N°21-30) (5)	251
<u>Abbildung 71:</u>	Einsatz des PAAR auf Texte N°21-30 (6)	252
<u>Abbildung 72:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (1)	256
<u>Abbildung 73:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (2)	257
<u>Abbildung 74:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (3)	258
<u>Abbildung 75:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (4)	258
<u>Abbildung 76:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (5)	259
<u>Abbildung 77:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (6)	259
<u>Abbildung 78:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (7)	260
<u>Abbildung 79:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (8)	261
<u>Abbildung 80:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (9)	261
<u>Abbildung 81:</u>	Einsatz des PAAR auf den Textabschnitt (N°31-50) (10)	262
<u>Abbildung 82:</u>	Einsatz des PAAR auf Texte N°51-75 (1)	24A
<u>Abbildung 83:</u>	Einsatz des PAAR auf Texte N°51-75 (2)	25A
<u>Abbildung 84:</u>	Einsatz des PAAR auf Texte N°51-75 (3)	26A
<u>Abbildung 85:</u>	Einsatz des PAAR auf Texte N°51-75 (4)	26A
<u>Abbildung 86:</u>	Einsatz des PAAR auf Texte N°51-75 (5)	27A
<u>Abbildung 87:</u>	Einsatz des PAAR auf Texte N°51-75 (6)	27A

<u>Abbildung 88:</u> Einsatz des PAAR auf Texte N°51-75 (7)	28A
<u>Abbildung 89:</u> Einsatz des PAAR auf Texte N°51-75 (8)	28A
<u>Abbildung 90:</u> Einsatz des PAAR auf Texte N°51-75 (9)	28A
<u>Abbildung 91:</u> Einsatz des PAAR auf Texte N°51-75 (10)	29A
<u>Abbildung 92:</u> Einsatz des PAAR auf Texte N°51-75 (11)	30A
<u>Abbildung 93:</u> Einsatz des PAAR auf Texte N°51-75 (12)	30A
<u>Abbildung 94:</u> Einsatz des PAAR auf Texte N°51-75 (13)	31A
<u>Abbildung 95:</u> Einsatz des PAAR auf Texte N°51-75 (14)	31A
<u>Abbildung 96:</u> Einsatz des PAAR auf Texte N°51-75 (15)	32A
<u>Abbildung 97:</u> Einsatz des PAAR auf Texte N°51-75 (16)	32A
<u>Abbildung 98:</u> Einsatz des PAAR auf Texte N°51-75 (17)	33A
<u>Abbildung 99:</u> Einsatz des PAAR auf Texte N°51-75 (18)	33A
<u>Abbildung 100:</u> Einsatz des PAAR auf Texte N°51-75 (19)	34A
<u>Abbildung 101:</u> Einsatz des PAAR auf den Textabschnitt (N°51-75) (20).....	34A
<u>Abbildung 102:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (1)	39A
<u>Abbildung 103:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (2)	39A
<u>Abbildung 104:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (3)	40A
<u>Abbildung 105:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (4)	40A
<u>Abbildung 106:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (5)	41A
<u>Abbildung 107:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (6)	41A
<u>Abbildung 108:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (7)	42A
<u>Abbildung 109:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (8)	43A
<u>Abbildung 110:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (9)	43A
<u>Abbildung 111:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (10)	43A
<u>Abbildung 112:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (11)	44A
<u>Abbildung 113:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (12)	44A
<u>Abbildung 114:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (13)	45A
<u>Abbildung 115:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (14)	45A
<u>Abbildung 116:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (15)	46A
<u>Abbildung 117:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (16)	46A
<u>Abbildung 118:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (17)	47A
<u>Abbildung 119:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (18)	47A
<u>Abbildung 120:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (19)	48A
<u>Abbildung 121:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (20)	48A
<u>Abbildung 122:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (21)	49A
<u>Abbildung 123:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (21)	49A
<u>Abbildung 124:</u> Einsatz des PAAR auf den Textabschnitt (N°75-100) (22)	50A
<u>Abbildung 125:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (1)	58A
<u>Abbildung 126:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (2)	59A
<u>Abbildung 127:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (3)	59A
<u>Abbildung 128:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (4)	60A
<u>Abbildung 129:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (5)	60A
<u>Abbildung 130:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (6)	60A
<u>Abbildung 131:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (7)	61A
<u>Abbildung 132:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (8)	61A
<u>Abbildung 133:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (9)	62A
<u>Abbildung 134:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (10)	63A
<u>Abbildung 135:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (11)	63A
<u>Abbildung 136:</u> Einsatz des PAAR auf den Textabschnitt (N°101-154) (12)	64A

<u>Abbildung 137:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (13)	64A
<u>Abbildung 138:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (14)	64A
<u>Abbildung 139:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (15)	65A
<u>Abbildung 140:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (16)	66A
<u>Abbildung 141:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (17)	66A
<u>Abbildung 142:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (18)	67A
<u>Abbildung 143:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (19)	67A
<u>Abbildung 144:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (20)	68A
<u>Abbildung 145:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (21)	68A
<u>Abbildung 146:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (22)	69A
<u>Abbildung 147:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (23)	69A
<u>Abbildung 148:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (24)	70A
<u>Abbildung 149:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (25)	70A
<u>Abbildung 150:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (26)	71A
<u>Abbildung 151:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (27)	71A
<u>Abbildung 152:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (28)	72A
<u>Abbildung 153:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (29)	72A
<u>Abbildung 154:</u>	Einsatz des PAAR auf den Textabschnitt (N°101-154) (30)	73A
<u>Abbildung 155:</u>	Performanz vom PAAR bei der Erkennung von PA im PAK.....	264
<u>Abbildung 156:</u>	Performanz vom PAAR bei der Erkennung der Natur von im PAK erhaltenen PA.....	267
<u>Abbildung 157:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK.....	269
<u>Abbildung 158:</u>	Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK bezüglich aller aufgelösten PA	270

ABKÜRZUNGSVERZEICHNIS

Abs.	Absatz
AR-Algorithmen	Anaphora Resolution-Algorithmen
AR-Ansätze	Anaphora Resolution-Ansätze
AR-Systeme	Anaphora Resolution-Systeme
Aufl.	Auflage
Bearb	Bearbeitet
Bd	Band
BFP- Algorithmus	Brennan, Friedman & Pollard- Algorithmus
bzw.	Beziehungsweise
CAR	Computational Anaphora Resolution
COSMAS II	Corpus Search, Management and Analysis System II
CL	Computerlinguistik
CR	Coreference resolution
CT	Centerintheorie
DeReKo	Deutsche Referenzkorpus
DGD	Datenbank für Gesprochenes Deutsch
d.h.	das heißt
DR	Diskursreferent
DTA:	Deutsches Textarchiv
DWDS	Das digitale Wörterbuch der deutschen Sprache
Ebd.	Ebenda Ebenda
Evtl.	Eventuell
ggf.	gegebenenfalls
G-RAP	German- Resolution of Anaphora

ID	Identifikationsnummer
IDE	Integrierte Open-Source- Entwicklungsumgebung
IDS	(Leibniz)-Institut für Deutsche Sprache
Kap.	Kapitel
KB-RAP	Kennedy und Bougarrev-RAP
KI	Künstliche Intelligenz
MARS	Mitkov's Anaphora Resolution System
ML	Maschinenlernen
ML-AR-System	Maschinenlernen- Anaphora Resolution-Systeme
MUC	Message Understanding Conference
NP	Nominalphrase
PA	Pronominaladverb
PAn	Pronominaladverbien
PAA	Pronominaladverbienanapher
PAAAn	Pronominaladverbienanaphern
PAAR	Pronominaladverbienanaphernreolvers
PAK	Pronominaladverbien-Korpus
PDW	Projekt Deutscher Wortschatz
PP	Präpositionalphrase
PP	Personalpronomen
Randnr:	Randnummer
RAP	Resolution of Anaphora Procedure
S.	Seite
u.a.	unter anderen
usw.	und so weiter

KAPITEL 0: EINFÜHRUNG

In diesem ersten Kapitel wird ein Überblick über das Forschungsvorhaben dieser Arbeit gegeben. Von der Motivation über den Forschungsstand bis hin zum Arbeitsplan werden alle wichtigsten Aspekte dieser Arbeit in Kürze erläutert. Dabei werden auch kerne Sektionen, wie etwa der theoretische und methodologische Rahmen; die Problematik (Fragestellungen & Hypothesen) und die Zielsetzungen dieser Arbeit berücksichtigt.

1. MOTIVATIONEN UND VORSTELLUNG DES THEMAS

• Motivationen

Die Entstehung des vorliegenden Dissertationsprojekts ist zum einen meinem akademischen Werdegang aber auch dem aktuellen Wissensbedarf in dem Forschungsgebiet der automatischen Verarbeitung der Sprache zu danken. Was meinen akademischen Werdegang betrifft, habe ich, in der Tat, in den letzten fünf Jahren Germanistik an der Universität von Jaunde 1 in Kamerun studiert. Dennoch hatte ich zuvor dank eines im Jahr 2011 erworbenen naturwissenschaftlichen Abiturs die Möglichkeit gehabt, Informatik an der Fachhochschule Dortmund zu studieren. Dies liegt zwar hinter mir her, aber dank einigen im Gebiet der Programmierungssprache gewonnenen Kenntnissen ist es für mich möglich gewesen, dieses Forschungsvorhaben auf die Beine zu stellen. So betrachtet, lässt sich die Wahl dieses Themas durch das Streben motiviert, zwei persönlichen Erfahrungen zu kombinieren. Erfahrungen, die ich sowohl im Bereich der germanistischen Linguistik als auch im Bereich der Informatik (Programmierung in Java-Sprache) gesammelt habe.

Dieses Dissertationsprojekt gilt als eine mögliche Fortsetzung meiner Abschlussarbeit. In der Tat liegt der Schwerpunkt meiner am Ende des Master-2-Studiengangs absolvierte Arbeit auf die wissenschaftliche Beschreibung des Funktionierens von anaphorischen Konstruktionen in der Textsorte PRÜFUNGSARBEIT. Zielsetzung dieser Abschlussarbeit war es, das Funktionieren von Gattungsanaphern zu beschreiben. Daraus ergibt es sich, dass Gattungsanaphern in diesem Texttyp nicht häufig auftreten und falls sie auftreten, werden sie eben durch ein gemeinsames Wissen zwischen Produzenten und Rezipienten aufgelöst (Sofac 2021: 100) und nicht nur lediglich durch besondere bzw. institutionalisierte Eigenschaften des Texttyps wie es Gérin (2014) für den Texttyp ERZÄHLUNGEN festgestellt hat.

Nach der erfolgreichen Verteidigung meiner Masterarbeit wollte ich meine Kenntnisse über das Forschungsgebiet der Anaphern-Resolution¹ immer weiter ausdehnen. Es gab zwar Möglichkeiten dieses Gebiet weiterzuverarbeiten ohne es mit einem anderen zu verknüpfen. Jedoch passten all diese beliebigen Forschungsmöglichkeiten nicht mit meinem Streben etwas sowohl **Wissenschaftliches** als auch **Technisches** aktuell zu leisten. Denn, wenn man das Funktionieren der heutigen Gesellschaft betrachtet, dann merkt man, dass die neuen Technologien von Kommunikationen das Miteinander von Menschen meistern. Und diese Herrschaft ist vor allem dank eines immer mehr wachsenden Einflusses des Computers auf die Verarbeitung und Vermittlung von Informationen (**Sprache**) möglich gewesen. So gesehen stand ich also vor einer großen Herausforderung: Wie könnte ich das linguistische Phänomen (**Anaphern-Resolution**) mit dem technischen Werkzeug (**Computer**) zusammenbringen?

Wie schon oben erläutert, beschäftigte ich mich im Rahmen meiner Abschlussarbeit mit der Frage der Auflösung von Gattungsanaphern. Es war also für mich natürlich, dass ich mich im Rahmen eines möglichen Dissertationsprojekts direkt mit diesem Anapherntyp auseinandersetze. Aber, weil ich die Auseinandersetzung etwas dynamisieren wollte, fiel es mir sofort ein, dass ich tatsächlich prüfen kann, ob und wie dieser Anapherntyp von Computerprogrammen bzw. maschinellen Übersetzungsprogrammen aufgelöst und übersetzt werden kann. Doch, ich musste bald feststellen, dass obwohl es Wissenslücke in diesem Bereich gibt; mir allerdings die nötigen Fach Erfahrungen bzw. Fachkenntnissen über das Thema fehlten, um ein Forschungsprojekt eines solchen Umfangs auf die Beine zu stellen. Würde ich mich also mit der Verarbeitung von Gattungsanaphern innerhalb eines automatischen Übersetzungstools beschäftigen, dann hätte ich dies unter zwei Perspektiven erledigen können. Ich konnte entweder prüfen, wie performant die existierenden Übersetzungstools sind (Performanz-Check) und bei gegebenen Fällen, diese Performanz auch verbessern; indem ich entweder die geprüften Algorithmen² verbessere oder eher neue Algorithmen theorisiere und implementiere.

Und Genau da stieß ich während der Elaboration des Projekts auf die erste Schwierigkeit. Denn, der Entwurf eines Algorithmus für die maschinelle Übersetzung, dem die Zielsetzung geschrieben wird, nicht nur sprachliche Einheiten von einer Sprache in eine andere zu

¹ Prozess der Bestimmung de Antezedens einer Anapher, der sich in drei Phasen gliedern lässt: Die Identifizierung der Anapher; das Finden von möglichen Antezedenten und die Auswahl eines Antezedens.

² Verfahren mit bestimmten Handlungsschritten zur Lösung eines Problems.

übersetzen, sondern auch zugleich indirekte anaphorische Konstruktion aufzulösen (und dies sowohl vor als auch nach der Übersetzung) verlangt nicht nur eine Zusammenbringung bzw. gleichzeitige Verarbeitung von Zwei unterschiedlichen Sprachen (Semantik und Pragmatik insbesondere) sondern auch eine Elaboration von anaphorischen Auflösungstools in jeder der beiden Sprachen, wobei noch einen Bindestrich zwischen diesen beiden in jeder Sprache elaborierte Tools gefunden werden soll. All dies führt mich mein Bestreben etwas zu vermindern und dieses abgebrochene Projekt vielleicht doch erst für nächste Forschungsarbeiten zu bewahren.

Über all diese Gründe hinaus entschied ich mich nicht mehr dem Bereich der maschinellen Übersetzung hinzugeben, sondern eher dem Bereich der automatischen Resolution der Anaphorik. Im Gegensatz zum ersten Streben erweist dieser zweite Versuch den Vorteil, dass es sich nur um die Verarbeitung einer einzigen Sprache geht. Anders formuliert, während es bei der ersten abgebrochenen Forschungsmöglichkeit darum geht, sowohl Anaphern automatisch aufzulösen als auch sie automatisch richtig zu übersetzen, handelt es sich hier in diesem zweiten Versuch, ein Dissertationsprojekt aufzubauen, um die automatische Auflösung von Anaphern durch ein spezielles selbst generiertes Tool.

- **Vorstellung des Themas**

Das Thema dieser Arbeit lautet **„Computer- und Textlinguistik: Entwicklung eines Systems für die automatische Auflösung von Pronominaladverbienanaphern. Ein Beitrag zur theoretischen und angewandten Computerlinguistik.“**. Bei der Näherbetrachtung dieser Überschrift, fällt es einem deutlich ein, dass es sich tatsächlich um die Kombination von zwei Forschungsgebieten. Zum einen haben wir die Referenzlinguistik und zum anderen die Computerlinguistik. Was den referenzlinguistischen Aspekt dieser Arbeit angeht, handelt es sich, wie eben oben schon hingewiesen, um das rückverweisende Phänomen der **Anaphorik**. Das vorliegende Forschungsvorhaben fokussiert sich also nicht mehr auf die menschliche Verarbeitung von Anaphern, sondern auf die Identifizierung, die Beschreibung und das Verstehen von Mechanismen, die eingeschaltet werden, wenn Computers mit anaphorischen Konstruktionen konfrontiert sind. Anders ausgedrückt, wird es versucht zu wissen, wie Computers Anaphern auflösen. Und bei diesem Versuch werden dann Elemente gesammelt, die danach zum Aufbau eines neuen Algorithmus dienen werden, welcher dann bei seiner

Implementierung Computer helfen wird, anaphorische Konstruktionen adverbialerart während der automatischen Verarbeitung der Sprache aufzulösen. Zu merken ist allerdings, dass Computers solche Algorithmen bedürfen, wenn sie zum Beispiel gefordert sind, eine Information in einem Text zu suchen (**Textverstehen; Informationsstruktur; Informationsextraktion**); oder noch, wenn sie gefordert sind, auf Internet eine Suche zu gestalten (**Maschinelle Suche**); oder noch, wenn sie vor Aufgaben stehen wie **Sprachgenerierung** (zum Beispiel, wenn Maschinen mit menschlichen Benutzern kommunizieren wollen) oder noch **Sprachsynthese** (Wenn zum Beispiel menschliche Äußerungen von Maschinen analysieren und verstehen werden müssen)

Allerdings, obwohl es zu merken ist, dass zahlreiche Algorithmen für die automatische Auflösung von Anaphern entwickelt und auch mit unterschiedlichen Erfolgsraten implementiert worden sind (Sevcikovac.R:2005). Es lässt sich trotzdem in der Forschungsliteratur feststellen, dass aufgrund der unterschiedlichen Strukturen von natürlichen Sprachen immer neue Ansätze entwickelt werden müssen um die Herausforderungen zu überwinden, die sich bei dem Umgang mit jeder einzelnen Sprache vorstellen. Weiterhin ist der Faktor der Anzahl von natürlichen Sprachen nicht der einzige, der zu überwinden ist. Man rezensiert auch innerhalb einzelner Sprache mehrere sprachliche Einheiten, die anaphorisch gebraucht werden können und benötigen somit jeweils auch ihren eigenen Algorithmus (Selmi.A: 2019). In dieser Hinsicht ist zum Beispiel, wie (Sevcikovac.R 2005: 3) in ihrer Arbeit hingewiesen hat, zu merken, dass Algorithmen für die Auflösung von Adverbialanaphern kaum vorhanden sind.

Die Behandlung von anderen Anapherntypen, wie Adverbanaphern; Verbalanaphern (VP-Anaphern) und Null-Anaphern wird aufgrund der eher raren Lösungsansätze nicht näher beleuchtet. (Sevcikovac.R 2005: 3)

Aus diesem Grund war es uns wichtig, einen Algorithmus für die automatische Auflösung von Adverbialanaphern in der deutschen Sprache zu generieren, dann zu implementieren und schließlich zu evaluieren.

2. Forschungsstand

Die folgenden Zeilen dieses Abschnitts werden uns dabei helfen, einen umfassenderen Blick über das computerlinguistische Phänomen **der automatischen Auflösung von Anaphern** zu

verschaffen. Wie schon in dem vorangehenden Teil dieser Arbeit hingedeutet, beabsichtigt diese Studie einen Algorithmus zu entwickeln, der **die maschinelle Verarbeitung von Adverbialenaphern** ermöglichen wird. In dieser Hinsicht scheint uns also von Relevanz zu zeigen, wie weit die Forschungen in diesem Bereich ausgestreckt sind; dabei wird insbesondere der Fokus auf **die Natur, die Struktur, Ziele und Performanz der jeweiligen ausgewählten Systeme** gesetzt. Das erste System zur automatischen Resolution von Anapher (**AR-System**), das wir besprechen wollen, ist dieses, das von Afef SELMI (2019) im Rahmen seiner Dissertation zum Thema *L'ambiguïté anaphorique et la résolution automatique de l'anaphore pronominale Propositions méthodologiques à partir d'un corpus fermé de résumés d'œuvres littéraires en français (RESUMAN)*, an der Universität BOURGOGNE FRANCHE-COMTE entwickelt worden ist.

Wie das Thema uns es schon verrät, beabsichtigt der Autor ein System namens RESUMAN₀ zu entwickeln, das in der Lage sein wird, die anaphorischen Personalpronomen der dritten Person (il, ils, elle, elles) automatisch aufzulösen. Dafür konstituiert er zunächst einen Korpus, der aus Zusammenfassungen von über 100 narrativen Texten besteht. Dieser Korpus nennt er RESUMAN_c und kümmert sich direkt danach um den Aufbau des systemzugrundliegenden Algorithmus, den er später mit Hilfe einer informatischen Studentin seiner Universität implementieren wird (Umwandlung des theoretischen Algorithmus in ein automatisches System) (Selmi 2019: 185). Doch bevor er seinen eigenen Algorithmus entwirft, macht er einen breiten Überblick über die Algorithmen bzw. Systeme, die seinem vorangehen. Darunter zählt man beispielweise den Algorithmus von Hobbs (1976); den von Lppin/Leass (1994); den Ansatz von Grosz & co. (1995), den Algorithmus von Kennedy/Boguraev (1996); das System CogNIAC von Baldwin (1997) und die Arbeiten von Mitkov (1999). Er bezeichnet diese Arbeiten als stark linguistisch ausgeprägt; damit meint er, dass die Heuristiken bzw. Regeln, worauf sich diese Algorithmen stützen nicht situationsgemäß veränderbar sind. (Vgl. Ebd.: 188) Dieser Gruppe Gegenüber stellt er die Systeme, die auf statistische Weise funktionieren; ihm nach bestehe die Besonderheit dieser Letzten darin, dass die Heuristiken/Parameter mit deren Hilfe sie Anaphern Auflösen, sich im Laufe der Zeit (Anzahl der schon bearbeiteten Texte/anaphorischen Konstruktionen) verbessern können; hier redet man also von der Schulung des Algorithmus (der Algorithmus baut sich, aufgrund statistischen Mitteln, eigene Wege zur Auflösung von Anaphern). Unter diesen Systemen zitiert er unter anderem das System von Connolly & co (1994) und das berühmte System BART (2010). Darüber kommen wir noch im Weiterverlauf des Forschungsstands zurück. (Vgl. Ebd.: 210)

Die Forschungsarbeit von Afef SELMI (2019) gliedert sich in drei Teilen. Der erste Teil kümmert sich um die Vorstellung des theoretischen Rahmens, hierbei sind hauptsächlich zwei Kapitel bearbeitet worden, in denen es jeweils geht, das Phänomen Anaphorik aus unterschiedlichen Forschungsperspektiven (Syntaktische, textlinguistische und kognitive) zu erläutern einerseits, und andererseits theoretische Kenntnisse über den Forschungsbereich Korpuslinguistik zu liefern. Dabei ist den Fokus daraufgesetzt worden, die Wahl des Korpus zu rechtfertigen. Im zweiten Teil seiner Arbeit berichtet er weiter über theoretische Kenntnisse, die zum besseren Verständnis der Anaphorik relevant sind. Dabei beschränkt er sich nur auf die Erläuterung der textlinguistischen Auffassung der Anaphorik und stellt zugleich klar welche Grenze es zwischen den Begriffen *Kohäsion* und *Kohärenz* gibt. Weiterhin erläutert der Autor, nachdem er die Rolle der Anapher in der Kohärenz eines Textes erörtert hat, die Begriffen **Sallience³ und Accessibilité/Zugang eines Diskursreferenten**. Dabei stützt er sich auf etablierte Theorien, wie etwa, die von Givón (1983); Ariel (1990); Gundel et al. (1993). Somit gelingt es ihm die Parameter zu verzeichnen, die dazu beitragen, dass ein Antezedent (Bezugsausdruck einer Anapher) zugängliche bzw. salienter als ein anderer gekennzeichnet wird. Unter diesen Parametern berücksichtigt er insbesondere **die morphologische Kongruenz** zwischen Antezedenten und anaphorischem Ausdruck; **der syntaktische Parallelismus**; **der thematische oder topische Faktor**; **die Distanz zwischen Anaphern und Antezedenten** und **der Pragmatische Faktor**. (Vgl. Ebd.: 125). Über all diese Faktoren werden wir noch im Weiterverlauf unserer Arbeit ganz und breit zurückkommen, denn sie sind die Grundsäule der Auflösung von anaphorischen Konstruktion.

Direkt nach der Besprechung über die Faktoren, die eine Reibungslose Auflösung von Anaphern ermöglichen, greift er zum wesentlichen Teil seiner Arbeit: Die **automatische**

³ « **La saillance** : La saillance est en corrélation directe avec l'accessibilité. Plus le référent est saillant et plus il est accessible. Ariel met **l'accessibilité** au même niveau que la topicalité. Ainsi, plus le référent se retrouve en position de sujet, plus il est accessible. [...] Par la suite, la saillance a été plutôt définie comme une notion incluant elle-même plusieurs facteurs (Schnedecker 2009, Landragin 2012). [...] Tout d'abord, ceux de **types syntaxiques** où l'on regroupe plusieurs entités afin de mettre en saillance un élément du discours (soit en tête, soit en fin de phrase par exemple). Ensuite, **l'ordre** (la position, la symétrie) et **la fréquence** (les répétitions) des mots [...] Un troisième facteur serait la **fonction grammaticale** avec une hiérarchie où le sujet se trouve devant le complément d'objet direct, lui-même devant le complément d'objet indirect. Il y a d'autres facteurs, incluant la totalité des niveaux d'interprétation et d'analyse ainsi que les dimensions du langage, qui définissent la saillance comme **les facteurs de types lexical et sémantique** (par exemple : la catégorie d'un mot, la sémantique de la phrase et le thème ou le topique selon le contexte du discours.). Cette multiplicité fait alors de **la saillance un phénomène multifactoriel**. » (Selmi 2019 : 124)

Dimension der Anaphorik. Hier zeigt er zunächst einmal, welchen Gebrauch man von AR-Systemen in der automatischen Verarbeitung der Sprache machen kann. Dazu sagt er zum Beispiel, dass diese Systeme dem Computer dabei helfen werden, Texte zu verstehen (**Textverstehen**); Informationen zu extrahieren (**Informationsextraktion**) und Texte zusammenzufassen (**Textzusammenfassungen**) u.a. gebraucht. Über die Nützlichkeit und Wichtigkeit von AR-Systemen in einer digitalisierten Gesellschaft werden wir noch später, im Hauptteil unserer Arbeit, zurückkommen. Im dritten und letztem Teil seiner Arbeit begibt er sich letztendlich, das Hauptziel seiner Arbeit zu füllen. Wie schon oben erwähnt, besteht dieses darin, ein AR-System zur Auflösung von vieldeutigen Personalpronomen zu entwickeln. Doch vor dem konkreten Einstieg zur Erreichung dieses Ziels erläutert er zunächst den Begriff *künstliche Intelligenz* und veranschaulicht zugleich welche Zusammenhänge zwischen *ihm, der Linguistik und der Kognition* bestehen. Erst im Anschluss daran entwickelt er, wie oben schon vermerkt, sein eigenes System, dessen globale Struktur sich durch das folgende Schema skizzieren lässt.

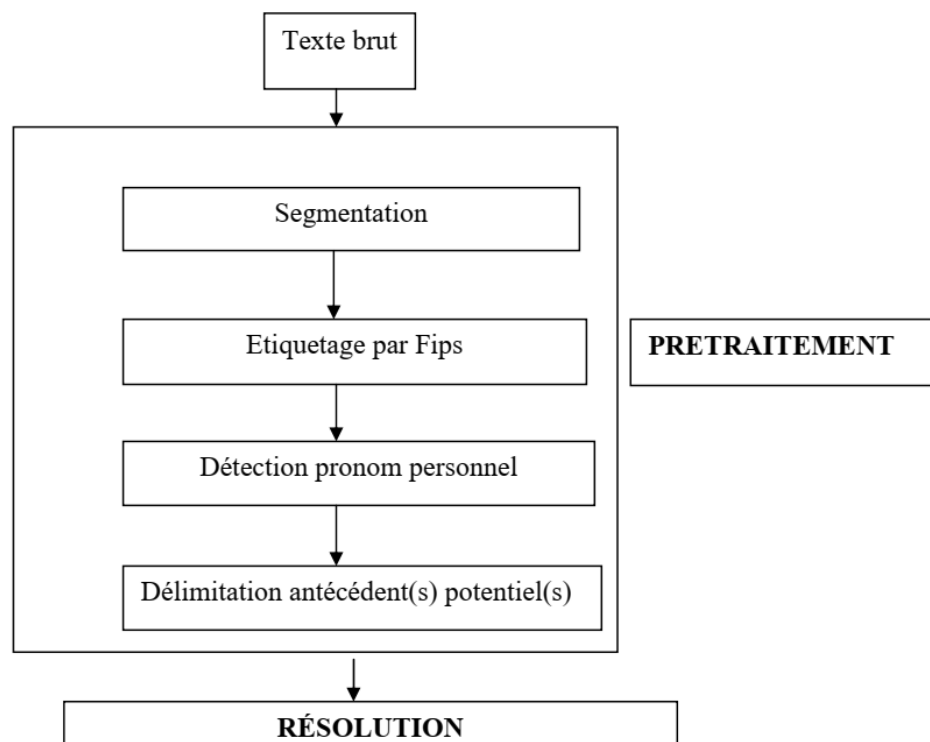


Abbildung 1: Die globale Struktur von RESUMAN₀ (Vgl. Selmi.A 2019: 218)

Die nähe Betrachtung dieses obig stehenden Schemas verrät uns, dass das AR-System RESUMAN₀ hauptsächlich aus zwei Modulen besteht: Einem, in dem Vorverarbeitungen über den ausgewählten Text geführt werden; und einem anderen, in dem Wege zur Identifizierung des geeigneten Antezedenten angeschaltet werden. So betrachtet, kümmert sich das erste Modul insbesondere um die Etappen **der Textauswahl, der Segmentierung, und der Herstellung**

eines morphosyntaktischen Etiketts für jede NP und jedes Pronomen des verarbeiteten Textes.

In der Tat erfolgt die Segmentierung eines im AR-System eingegebenen Textes auf der Basis zwei Kriterien: die Satzzeichen (das Komma ist ausgeschlossen) und die Präsenz eines Pronomens (il, ils, elle, elles). Und sobald die Segmente konstituiert sind, findet nun mit Hilfe eines syntaktischen Analysators (der Autor hat der im Internet freizugängliche Parser Fips⁴ gebraucht, der von Laenzlinger et Wehrli (1991) im ATALA⁵ elaboriert wurde.) die morphosyntaktische Aufzeichnung von identifizierten Pronomen so statt, dass das System nachhinein eine klare Grenze, zwischen Pronomen (il, ils, elle, elles), die unpersönlich gebraucht worden sind und denen, die eher als Personalpronomen fungieren, ziehen kann. Diese Etappe ist von großer Relevanz, wenn man ein performantes AR-System aufbauen möchte. Denn eben nur Personalpronomen können anaphorisch sein. Dazu sagt der Autor nämlich, dass

L'élimination des pronoms impersonnels est une étape importante pour éviter toute confusion avec le pronom personnel il. Nous rejoignons Danlos (2005 : 390) dans le sens où: Un système de résolution des anaphores doit être capable de repérer les occurrences des pronoms impersonnels avant de s'attaquer aux pronoms anaphoriques et aux autres anaphores. (Selmi 2019 :225)

Die folgenden Bilder zeigen uns wie diese beiden ersten Etappen (Segmentierung und morphosyntaktische Etikettierung) konkret im System aussehen:

⁴ FIPS ist kostenlos erhältlich unter <http://latlapps.unige.ch/Parser> ; <http://alpage.inria.fr/iwpt09/atala/fips.pdf>

⁵ Association pour le Traitement Automatique des Langues : <http://www.atala.org/>

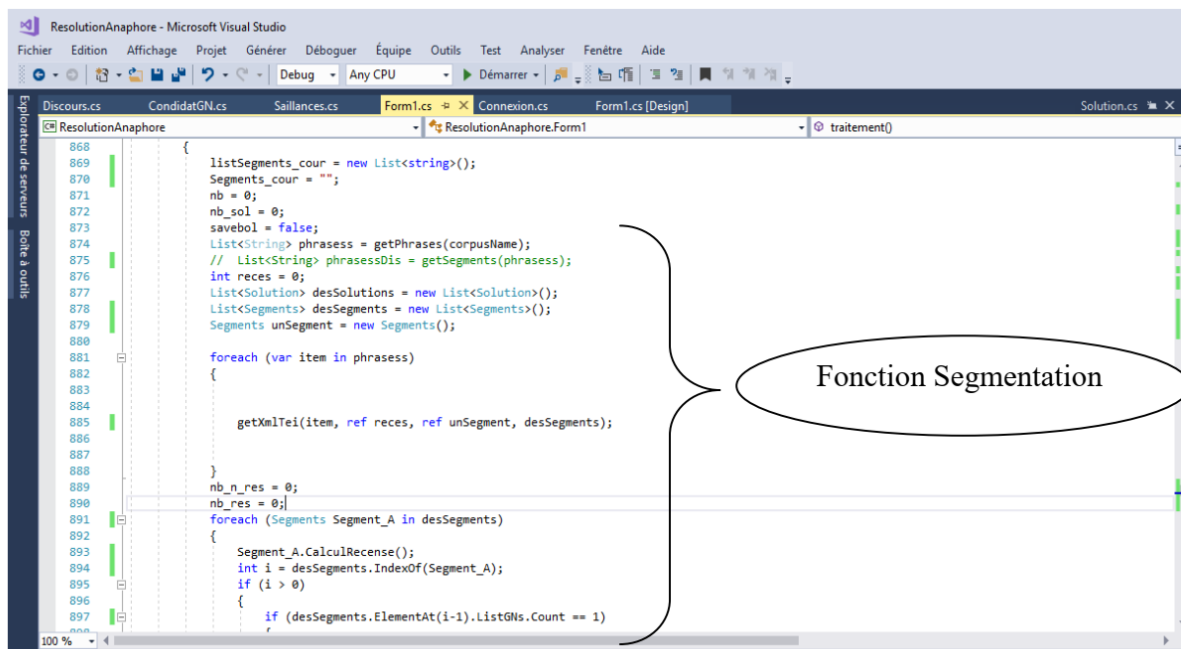


Abbildung 2: Die Segmentierungsfunktion (Vgl. Ebd.: 221)

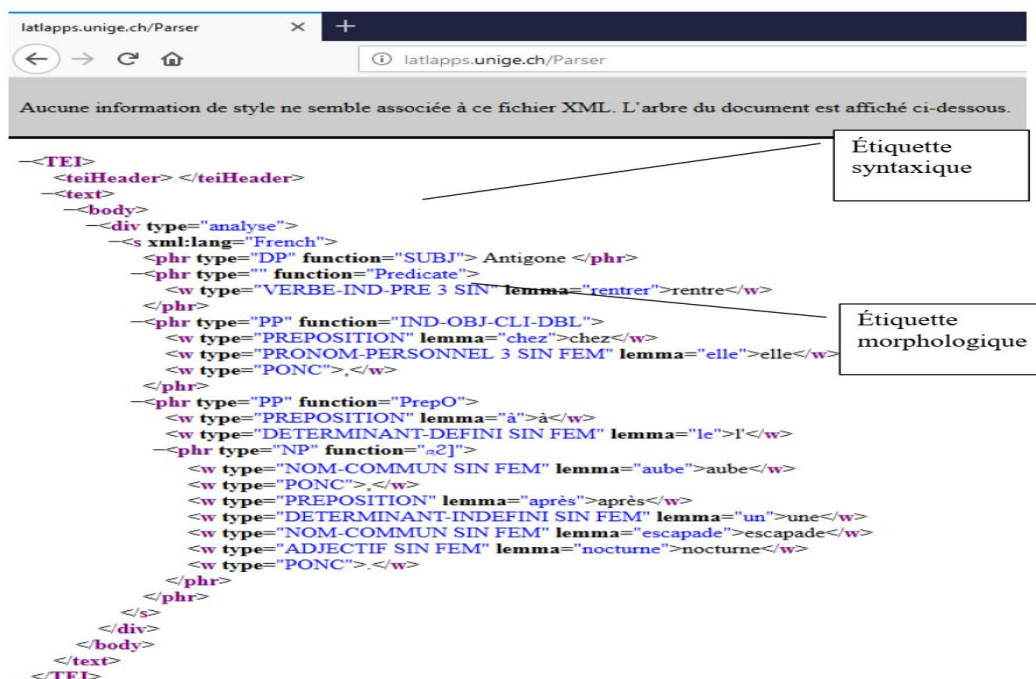


Abbildung 3: Morphosyntaktische Etikettierung durch Fips (Vgl. Ebd: 224)

Sobald die beiden ersten Etappen ausgeführt sind, ist es nun für das System möglich Personalpronomen von unpersönlichen Pronomen zu unterscheiden, hierbei werden die Personalpronomen in einer besonderen Liste verzeichnet. Ab diesem Moment kann nun das zweite Modul des AR-Systems eingeschaltet werden. In diesem Modul geht's insbesondere um das Auffinden des richtigen Antezedens für jedes verzeichnete anaphorische Pronomen. Noch als **Anaphemresolution** kennzeichnet, streckt sich dieser Vorgang in vielen Etappen aus. Die erste Etappe ist zunächst **die Selektion von allen möglichen Antezedens-kandidaten** eines anaphorischen Ausdrucks. Dies erfolgt, indem das System alle Nomen und NP, die sowohl im Satz, zu dem der anaphorische Ausdruck gehört, als auch zu denjenigen Sätzen, die diesen Stammsatz vorangehen, auszeichnet. Diese Aufgabe wird dann für jede identifizierte Anapher gemacht und erst danach kommt die zweite Etappe dieser Phase.

In dieser zweiten Etappe handelt es sich um die Auswahl des passenden Antezedens; dabei werden, wie schon oben eingeführt, mehrere Heuristiken gebildet und ausgeführt. Die erste Heuristik ist **die Genus-Numerus-Kongruenz** zwischen potentiellen Antezedens-kandidaten und der Anapher. Bleibt nach dem Einsatz dieser Regel nur ein passender Kandidat, dann ist die Auflösung gelungen; genügt aber dieser Parameter nicht, dann wird ein weiterer ins Spiel gebracht: Der **Aktualitätsfaktor (Recency Faktor)** dieser besagt, dass das dem anaphorischen Ausdruck am nächsten stehende Antezedens den Vorrang auf die anderen potentiellen Antezedenzen hat. Falls bis jetzt das System noch kein richtiges Antezedens gefunden hat, wird **die Heuristik der syntaktischen Funktion** eingeführt, hier gilt zum Beispiel die Tatsache, dass der Vorrang dem Antezedens gegeben wird, das die Position des Subjekts hat. Bleibt nach dem Einsatz dieser Heuristik immer noch viele potentielle Antezedens-Kandidaten, dann muss das System die Salienz einzelner Kandidaten rechnen. Und nach der Verrechnung der Salienz wird dann der Kandidat ausgewählt, der die meisten Punkte hat. Zu der Verrechnung der Salienz kommen wir gleich zurück. Gerät man jedoch in einen Fall, in dem mehrere Kandidaten die gleichen Punktzahlen haben, dann wird, dem Autor nach, weitere Regel, wie etwa **der syntaktische Parallelismus**⁶ eingesetzt. Genauso in diesem Aspekt unterscheidet sich sein System von den anderen Pronomenresolutionssystemen; denn während

⁶ Als Erweiterung der Heuristik **syntaktische Funktion** gilt zum Beispiel **der Faktor des syntaktischen Parallelismus**. Dieser besagt, dass das AR-System, bei zweideutigen Fällen, den Antezedens-Kandidaten auswählen soll, der mit der Anapher dieselbe syntaktische Funktion hat. Ist also die Anapher in Position des Subjekts wird also der Vorrang dem Antezedens-kandidaten gegeben, der dieser Funktion in dem Satz hat, in den er eingebettet ist.

zum Beispiel die Verrechnung der Saliens in anderen Systemen als letzter Entscheidungsfaktor eingesetzt wird, ist es in dem System von Selmi (2019) nicht der Fall. Dazu sagt er Folgendes :

La résolution de l'anaphore ne pose pas de problème si au final le processus sélectionne un seul élément. Dans le cas où plusieurs éléments sont sélectionnés, alors la saillance est prise en compte. Si la saillance ne résout pas le problème alors le parallélisme syntaxique entre en jeu, et si le conflit n'est toujours pas résolu alors c'est l'élément le plus proche du pronom qui sera sélectionné. (Vgl. Ebd.: 233)

In dieser Hinsicht scheint es also wichtig einiges über das Kalkül der Saliens zu verraten; dafür werden folgende Angaben gebraucht:

Position du groupe nominal et du pronom	Valeur de la récence
Dans la même phrase	100
Dans la phrase précédente	Diminution de 10 à chaque retour en arrière
Dans la 10 ^{ème} phrase précédant l'anaphore	0 (valeur de récence supprimée)

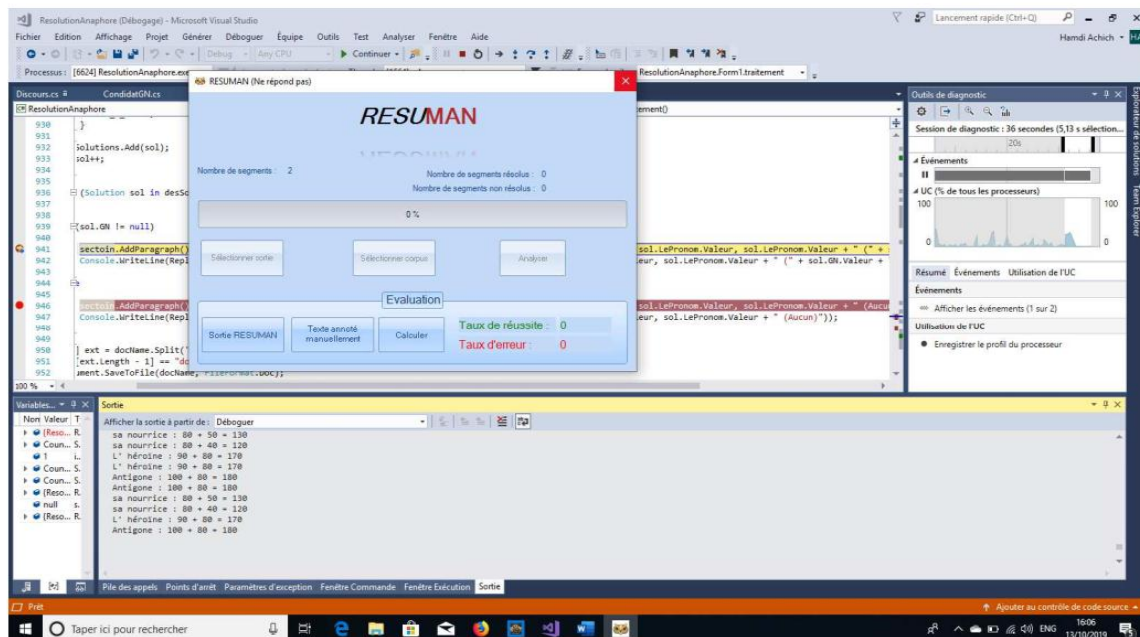
Tabelle 1: Werte der Recence (Vgl ; Ebd : 233)

Fonction	Poids grammatical
Sujet	80
COD	50
COI	40
Attribut	30
Subordonnée	10

Tabelle 2: Werte von syntaktischen Funktionen

Nous montrons comment RESUMAN₀ calcule la saillance des GN antécédents :

- Calcul de la Saillance (récence + poids grammatical)



```
sa nourrice : 80 + 50 = 130
sa nourrice : 80 + 40 = 120
L' héroïne : 90 + 80 = 170
L' héroïne : 90 + 80 = 170
Antigone : 100 + 80 = 180
Antigone : 100 + 80 = 180
sa nourrice : 80 + 50 = 130
sa nourrice : 80 + 40 = 120
L' héroïne : 90 + 80 = 170
Antigone : 100 + 80 = 180
```

Abbildung 4: Verrechnung der Saliens durch das AR-System RESUMAN₀

Das Kalkül der Saliens in diesem AR-system erfolgt also lediglich anhand des Addierens von zwei Parametern, etwas das, wie schon zuvor erwähnt, bei anderen AR-Systemen (wie die von Lappin et al. (1994) und Mitkov (2002) u.a.) nicht der Fall ist. (Sevcikova.R 2005: 32-41). Wie in den obigen Tabellen festgelegt, werden potentiellen Kandidaten Werte zugeschrieben, welche dann später in der Verrechnung der Saliens, wie es in der Abbildung 4 zu sehen ist, addiert werden. Nur das Antezedens mit dem besten Punktwert wird am Ende des Verfahrens bewahrt. Somit wird dann die Anapher automatisch aufgelöst. Zusammengefasst lässt sich der Algorithmus von Selmi (2019) wie folgt schematisieren.

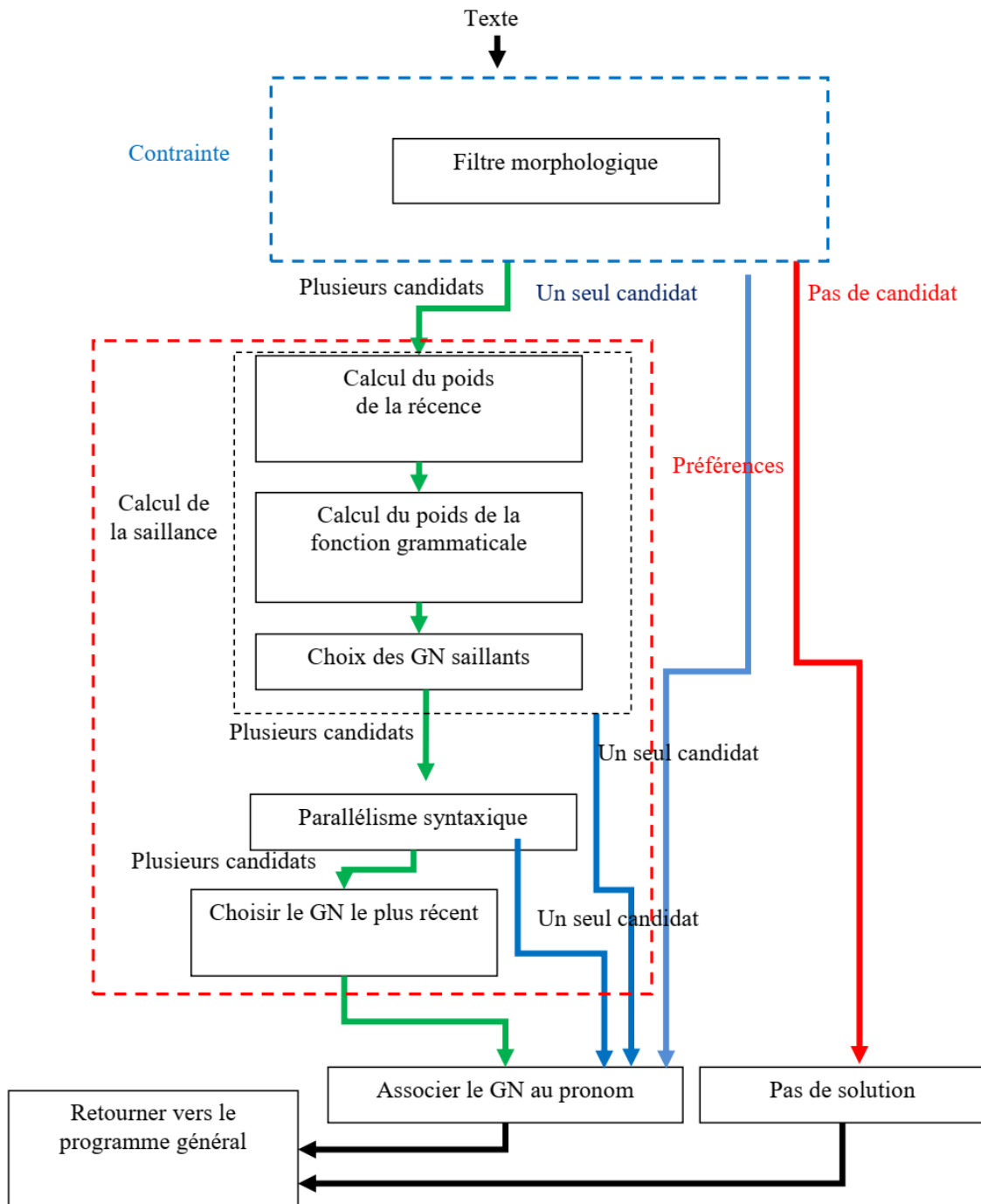


Abbildung 5: Schema des Algorithmus RESUMAN₀ (Vgl. Selmi, 2019: 228)

Wie Sie es gemerkt haben, habe ich mich beim Berichten über die Lage der Forschungen im Gebiet der automatischen Verarbeitung von Anaphern ganz stark auf das System RESUMAN₀ von Selmi (2019) konzentriert. Grund dafür liegt an der Tatsache, dass dieses System einerseits, das aktuellste, das verfügbar ist und andererseits, rekuriert dieses System auf **Vorgehensweisen, Mechanismen und Techniken**, die beim Aufbau älterer Systeme zu

beobachten sind. Jedoch unterscheidet es sich von diesen älteren Systemen, beispielweise an dem Aspekt der Berechnung der Salienz. Hierzu lässt sich vor allem merken, dass die **Formulierung und Hierarchisierung (Reihenfolge des Einsatzes der verschiedenen Regeln) der Heuristiken** je nach dem Autor bzw. dem zu erreichenden Ziel unterschiedlich ist. Dies ist zum Beispiel der Fall, wenn man **das AR-System von Lappin/Leass (1994)** betrachtet.

Noch als **RAP (Resolution of Anaphora Procedure)** bekannt, stützt sich dieses AR-System selbst auf einen Algorithmus zur Pronomenresolution, der von **Hobbs im Jahre 1976** entwickelt wurde. Dieser Algorithmus von Hobbs wurde damals zwar nur noch manuell durchgeführt, aber die Ergebnisse (Eine Performanz von fast 92%), die es liefert haben dazu beigetragen, dass zukünftige Arbeiten schon ein trauriges Muster hatten, worauf sie sich stützen konnten. Dies taten eben Lappin und Leass (1994) als sie ihr System RAP entwickelten. Zu der reinsyntaktischen Dimension, die der Hobbs Algorithmus präsentiert, haben sie eine semantische Dimension bei der Berechnung der Salienz hinzugefügt, um die Personalpronomen der dritten Person (he,she,it) und Reflexivpronomen (himself, herself, themselves, itself) in der englischen Sprache auflösen zu können. (Vgl. **Sevcikova**.R 2005: 32).

Darüber hinaus unterscheidet man noch etliche Pronomenresolutionssysteme, dazu zählen wir zum Beispiel: das System CogNIAC von Baldwin (1997); das System von Dember (1998); das AR-System namens MARS von Mitkov (2002a) und das System von Siddharthan (2003) u.a. (Vgl. Ebd.: 32-73). Über diese Systeme hinaus, die für die Resolution von Anaphern pronominaler Art sind, bestehen auch Systeme, die für **die automatische Verarbeitung von NP-Anaphern** zuständig sind. Hier werden wir nicht im Einzelnen jedem System nachgehen, sondern sie nur erwähnen und ihre jeweiligen Hauptmerkmale verraten. Das erste System, das wir betrachten möchten ist der Einsatz von **Viera et al. (2000)**: dieses System versucht zwischen **nicht-anaphorischen definiten NPs, Identitäts-Anaphern und assoziativen Anaphern** zu unterscheiden (Vgl. Ebd.: 71). Somit lässt sich verstehen, dass dieser Einsatz sowohl für direkte (koreferente) als auch für indirekte (nicht-koreferente) NP-Anapher konzipiert wurde. In dieser Hinsicht wird uns diese Arbeit sehr behilflich sein; denn sie gibt uns Hinweise und Angaben, wie man mit Komplexanaphern umgeht, also mit Anaphern, die als Antezedens meistens einen ganzen Satz oder einen ganzen Textabschnitt, wie es eben bei **Pronominaladverbien** der Fall ist, haben. Ebenso wie das System von Vierra et al (2000) verarbeitet das AR-System von **Postolache (2004)** namens **RARE (Robust Anaphora Resolution Engine)** auch assoziativen Anapher aber mit anderen Techniken. In der Tat wurde dieses System entwickelt um die

Schwierigkeiten der Auflösung von assoziativen Anaphern aus unterschiedlichen Perspektiven zuzudecken. Zu den gebrauchten Techniken zählen wir zum Beispiel den Gebrauch von Axiomen aus Korpora; von vektorbasierten Methoden; von Ähnlichkeitskriterien; von den Techniken des Maschinenlernen; von fokussierbaren Techniken u.a. (Vgl. Ebd.: 54-64). Diese Techniken sowie Algorithmen, die jeweils von Salmon-Alt (2004), und Bittar (2006) entwickelt worden sind, werden wir noch im Weiterverlauf unserer Arbeit tiefer angehen.

3. Problemstellung, Fragestellungen und Hypothesen

Gemerkt wurde eben im Laufe des Forschungsstandes, dass alle entwickelten Systeme zur Anaphernaufflösung vor dem semantischen Aspekt eine Biegung nehmen. Das heißt unter den Werkzeugen (Heuristiken), die zur maschinellen Verarbeitung von Anaphern gebraucht worden sind, werden semantischen und pragmatischen Heuristiken vermisst. Genauso dieser Mangel erklärt (vielleicht doch nur zum Teil) die Rarität von AR-Systeme zur automatischen Auflösung von Adverbialanaphern. Denn dieser Anapherotyp lässt sich vorwiegend durch semantische Inferenzen (Kongruenzen) auflösen (d.h. die morphosyntaktischen Kongruenzen; die Oblichkeitshierarchie; die Nähe-Fern-Regel und der syntaktische Parallelismus sehen ihre Wichtigkeit beim Aufbau eines AR-Systems zur Auflösung von Adverb-Anaphern sinken). In dieser Hinsicht ist also unsere Arbeit von großer Relevanz denn wir beabsichtigen ein AR-System zu entwickeln, das unter anderem ein semantisches Modul beinhaltet wird.

Wie wir es gerade gesehen haben, lässt sich die automatische Verarbeitung der Sprache schon seit dem Anfang des 21. Jahrhunderts in der Wissenschaft thematisieren. Doch es bleibt trotzdem Aspekte, die entweder unberührt geblieben sind oder die eher nicht ausgeforscht sind. Zu diesen Aspekten zählen zum Beispiel den Entwurf von Algorithmen, die dafür zuständig sind, Adverbien, die anaphorisch eingesetzt sind, aufzulösen. Gemerkt wurde nämlich, dass die meisten vorhandenen AR-Systeme nur für die Auflösung von Pronomen entwickelt worden sind, wobei es sich nur um Personal- und Reflexivpronomen handelt. Relativ- und Possessivpronomen wurden ebenso wie Adverbien wenig geforscht. Dazu sagt Sevcikova.R (2005) im Rahmen ihrer Arbeit Folgendes;

[...] In Kapitel 4 wird auf die Probleme und Lösungsansätze in Bezug auf den Typ der Anapher eingegangen. Der Schwerpunkt liegt hierbei auf Anaphern, die als nicht-lexikalische Proformen oder als vollständige NPs realisiert sind. Die Behandlung von

anderen Anapherntypen wie **Adverbanaphern**, Verbalphrasenanaphern (VP-Anaphern) und Nullanaphern wird aufgrund **der eher raren Lösungsansätze** nicht näher beleuchtet. (Vgl. **Sevcikova**.R 2005: 3).

Woran könnte dieser Mangel liegen? Darüber erfahren wir mehr während der vorliegenden Forschungsarbeit. Somit lässt sich die Hauptfrage dieser Arbeit so formulieren: **Wie lässt sich ein AR-System für die Auflösung von anaphorischen Adverbien entwickeln?** Zur Beantwortung dieser Hauptfrage werden wir die folgenden Zwischenfragen gebrauchen:

- Unterscheidet sich die automatische Resolution von pronominalen Anaphern stark von der der Adverbialanaphern? Anders formuliert; wird der Aufbau von Algorithmen zur Auflösung von Adverbialanaphern der gleichen Vorgehensweise bzw. Struktur, wie der von pronominalen Anaphern folgen? Werden die gleichen Parameter /Heuristiken gehalten?

Zu dieser Frage wollen wir direkt die folgende Hypothese formulieren: Obwohl es ersichtlich ist, dass Pronominalanaphern, semantische betrachtet, fast leer sind (etwas das nicht der Fall bei Adverbien ist), wollen wir vermuten, dass ihre automatische Auflösung trotzdem in einigen Punkten der der adverbialen Anaphern ähnelt. Denn die verschiedenen Etappen, die dem konkreten Verlaufspunkt, in dem die Anapher aufgelöst wird, vorangehen, sehen bei allen Algorithmen fast ähnlich aus.

Weiterhin könnte man sich zur Weitererläuterung der Problematik die folgende Frage stellen

- Werden die gleichen Präferenzen und Restriktionen, die bei der automatischen Auflösung von Pronominal- und NP-Anaphern berücksichtigt worden sind, auch bei der automatischen Auflösung Adverbialanaphern gelten?

Als vorläufige Antwort zu dieser Frage würde zum Beispiel die folgende Aussage formulieren: Für die Adverbialanaphern, die Nomina oder NPn als Antezedens haben, wird es höchst wahrscheinlich sein, dass ihre Auflösung fast in allen Aspekten der der NP/Pronominalen-Anaphern ähnelt (das ist zum Beispiel der Fall bei **Lokaladverbien**). Denn anaphorische Lokaladverbien haben eben wie anaphorische Pronomen und NP-Anaphern eben nur Nomen oder nominale Phrase als Antezedens. Im Gegensatz dazu wird den Aufbau von neuen Regeln und Vorgehensweisen benötigt um Adverbialanaphern, die Sätze oder Abschnitte als Antezedent haben. (das ist zum Beispiel der Fall bei **Pronominaladverbien**), auflösen zu können.

Eine weitere Hypothese zu der gleichen Frage wäre zum Beispiel: Angenommen, dass Adverbialanaphern genauso wie NPs-Anaphern keine leeren linguistischen Einheiten sind, vermuten wir, dass ihre automatische Resolution fast ähnlich ablaufen wird. (Es werden zwar sowohl Heuristiken als auch restriktive Modifikationen gebraucht dennoch eine schlichte Übertragung wäre nicht plausibel)

Im Gegensatz zu den ersten beiden Forschungsfragen, die eher theoretisch orientiert sind, wollen wir nun eine technische Frage formulieren? Diese wäre zum Beispiel in Bezug auf die unterschiedlichen informatischen Mittel (Programmiersprache, Parser-Tools und weitere Computerlinguistische Tools). Die Frage wäre zum Beispiel,

- ob man bei der Implementierung von einem Algorithmus zur Auflösung von Adverbialanaphern dieselben informatischen Werkzeuge gebrauchen wird, wie diejenigen, die bei der Implementierung von Pronominalen-und NP-Anaphern gebraucht worden sind?

Zu dieser Frage lässt sich hauptsächlich zwei Hypothesen formulieren: Angenommen, dass alle Systeme auch mit der Aufgabe der syntaktischen bzw. strukturellen Analyse des Textteils beschäftigen, in dem sich die Anaphorik befindet; wollen wir sagen, dass auf dieser Ebene der Analyse die gleichen Tools (wie etwa Fips) angewendet werden. Jedoch wird, wenn es sich um die Wahl **der Programmierungssprache** (Java oder C++) handelt, eine **Gegenhypothese** formulieren. Denn je nach den Spezifitäten der herausgebildeten Heuristiken werden wir zwischen all den existierenden Programmiersprachen eine geeignete für **die Implementierung unseres Algorithmus wählen**. Hierzu werden wir bei Gelegenheit die Hilfe von Experten (Informatikern) holen.

4. Theoretische und methodische Vorgehensweise

Zielsetzung dieser Arbeit ist es, einen Algorithmus bzw. ein informatisches System zur maschinellen Verarbeitung von Adverb-Anaphern zu entwickeln. Dafür werden wir uns auf schon vorhandene AR-Systeme stützen und somit auch auf Theorien zugreifen, worauf sich diese AR-Systeme selbst stützen. Darunter zählen wir Beispielsweise zur Verrechnung der Salienz: die Zugänglichkeitstheorie von Ariel (2001); Gundel et al. (1993); Givón (1983). Allerdings muss hier jedoch gemerkt werden, dass hinsichtlich der Vielfältigkeit der Aspekte, die berührt werden sollen, werden höchstwahrscheinlich im Laufe der Arbeit weitere Theorien einfließen

müssen, mit welchen wir geeigneten Werkzeugen zur Lösung der Problematik aufarbeiten werden.

Was die methodische Vorgehensweise angeht, beabsichtigen wir sobald der Algorithmus konzipiert und anhand informatischer Mittel implementiert worden ist, Beispieltex te aus deutschen Online-Zeitungen heranzuziehen, um unser System zu prüfen und somit seine Performanz zu evaluieren.

5. Aufbau der Arbeit

Nach den üblichen Teilen, wie die Widmung, die Danksagung und die Zusammenfassung wird die Einleitung folgen. Dieser Teil besteht aus mehreren unteren Teilen: der Motivation, der Themavorstellung, dem Forschungsstand, den Problem- und Fragenstellungen, den Hypothesen, den Zielsetzungen, dem methodischen Rahmen und dem Aufbau der Arbeit.

Nach der Einleitung folgt der theoretische Teil der Arbeit, der aus drei Kapiteln besteht. Das erste Kapitel kümmert sich um die Bestimmung von relevanten Begriffen, wie etwa Computerlinguistik; Algorithmus; Informationsextraktion; Adverb-Anapher; Anaphernresolution und Heuristik u.a. Das zweite Kapitel wird Kenntnisse über das Forschungsgebiet der automatischen Auflösung von Anaphern liefern; dabei werden insbesondere eine historische Entwicklung dieser Tätigkeit von 1970 bis heute wiedergegeben, wobei jeweils der Schwerpunkt auf die für uns interessanten Aspekte gesetzt wird. Das dritte Kapitel dieses Teils wird dem theoretischen Rahmen dieser Untersuchung reserviert. Es wird nämlich alle Theorien, Ansätze und vor allem AR-Systeme, die zum Entwurf unseres eigenen Protokolls für die automatische Verarbeitung von Adverbialanaphern helfen werden, erläutern. Und im Anschluss daran werden wir einen Algorithmus zur Auflösung von Adverbialanaphern entwerfen.

Der zweite Teil dieser wissenschaftlichen Auseinandersetzung besteht, eben wie der erste aus drei Kapiteln. Das erste Kapitel dieses Teiles widmet sich der Implementierung des im dritten Kapitels entwickelten Algorithmus. In der Tat werden wir mit Hilfe eines PROLOG oder einer geeigneten informatischen Anwendung/Tools unseren Algorithmus codieren. Am Ende dieses Kapitels soll also ein endgültiges AR-System vorhanden sein. Dieses AR-System wird dann in dem danach folgenden Kapitel anhand Online Texten aus deutschen Zeitungen eingesetzt. Hierbei wird über die Performanz des Systems berichtet. Im letzten Kapitel dieses Teils werden

dann die Hauptergebnisse der Forschungsarbeit zusammengefasst, Hypothesen geprüft und die entsprechenden Schlussfolgerungen gezogen

ERSTER TEIL: TERMINOLOGISCHE ANNÄHRUNGEN (**Computerlinguistik** und **Textlinguistik**)

Wie der Titel es schon klarstellt, wird es in diesem ersten Teil die Rede sein, die für die erfolgreiche Ausarbeitung unserer Problematik relevanten Fachkenntnisse über die Forschungsgebiete Computer- und Textlinguistik zu enthüllen. Dafür wollen wir diesen Teil in zwei Kapitel gliedern. Das erste Kapitel wird sich, einerseits, darum kümmern, den Begriff **Computerlinguistik**(CL) anhand seiner geschichtlichen Gegenstandsentwicklungen und Forschungsschwerpunkte in universitären Einrichtungen Deutschlands und Kameruns zu bestimmen. Andererseits wird es auch Aufgabe des ersten Kapitels sein, die theoretischen und methodischen Rahmen, worauf sich jede Auffassung der Computerlinguistik stützt, zu erläutern. Anschließend darauf werden die sowohl für die Wissenschaft als auch für die Gesellschaft herausragenden Anwendungen vorgestellt. Was das zweite Kapitel betrifft, wird es die Rede sein, den Begriff **Textlinguistik** zu erläutern. Dabei wird fast die gleiche Vorgehensweise, wie im ersten Kapitel, verfolgt. Dennoch, im Gegensatz zum ersten Teil, in dem der Fokus auf den Kernbegriff **Algorithmus** gesetzt wurde, wird im zweiten Kapitel der Fokus eher auf das sprachliche Phänomen **Anaphorik** gesetzt.

KAPITEL 1: COMPUTERLINGUISTIK (CL)

Dieses Kapitel wird hauptsächlich in drei Abschnitten eingeteilt werden. Der erste kümmert sich um die Bestimmung des Begriffs ‚Computerlinguistik‘ (Abschnitt 1.1.); der zweite um die Erläuterung der methodologischen und theoretischen Grundlagen der CL (Abschnitt 1.2.); und der dritte um die Anwendungen der CL im alltäglichen Leben (Abschnitt 1.3.).

1.1. Gegenstandsbestimmung

Der Begriff *Computerlinguistik* verweist, rein morphologisch betrachtet, auf ein **Determinativkompositum**⁷, das aus *Computer-* (**Determinans**/ untergeordnetem Begriff) und *-Linguistik* (**Determinatum**/ dominantem Begriff) besteht. Davon ausgehend, lässt sich die *Computerlinguistik* als eine Disziplin der Linguistik bzw. der angewandten Linguistik, **die die Möglichkeiten zur Verarbeitung natürlicher Sprachen mit dem Computer untersucht**.⁸ Anders gesagt, soll es bei so einer Wissenschaft darum gehen, **die natürlichen Sprachen** (wie etwa Deutsch, Englisch, Französisch, Yemba) algorithmisch bzw. automatisch zu verarbeiten (und zwar sowohl rezeptiv: **Sprachanalyse** als auch produktiv: **Sprachgenerierung**) (vgl. A. V. Abyzov, A. G. Rys 2017: 148). Dennoch lässt sich bei dieser Auffassung des Gegenstandes der *Computerlinguistik* etwas nuancieren, denn auch **Computersprachen** (Programmiersprachen und Formalsprachen) sind auch mittlerweile⁹ im Zentrum des Forschungsgegenstandes gerückt. Aus dieser Metamorphose des Untersuchungsgegenstandes

⁷**Ein Determinativkompositum** ist eine (Wort)-Zusammensetzung, bei welcher der erste Teil, das untergeordnete Determinans (in diesem Fall **Computer-**), den Zweiten, das dominante Determinatum (in diesem Fall **-Linguistik**), in seiner Bedeutung näher bestimmt einschränkt. (Wiktionary ; unter der Lizenz CC BY-SA 3.0 License)

⁸ Auch **Wikwörterbuch** genannt ist das **Wiktionary** ein freies Wörterbuch und ein Wikimedia-Projekt, das am 1. Mai 2004 startete. Es umfasst derzeit 1.029.295 Einträge und zielt darauf ab, die Erfassung des gesamten deutschen Wortschatzes und des aller anderen Sprachen frei zugänglich zu ermöglichen.

⁹**Die Computerlinguistik** ist eine relativ junge wissenschaftliche Disziplin, die, im heutigen Sinne, erst nach den **80er Jahren** betrieben wurde. Dennoch wird um die geschichtliche Entwicklung dieser Forschungsdisziplin herumschon Spuren des Gebrauchs dieses Konzepts in den **60er Jahren** aufgezeichnet. Dies war beispielweise der Fall in den Arbeiten von Noam Chomsky (1957) über die syntaktische Struktur. In diesen Arbeiten entwarf er schon Ansätze zur Formalisierung der syntaktischen Struktur einer natürlichen Sprache. In dieser Hinsicht galten seine Arbeiten als Vorreiter für **eine automatisierte Sprachübersetzung**, denn diese Anwendung ist stark von den IC-Strukturen geprägt. Und erst mit der Verbreitung von Arbeitsplatzrechnern (Personal Computer) und mit dem Aufkommen des Internets hat sich der Umfang der Computerlinguistik erweitert (vgl. A. V. Abyzov, A. G. Rys 2017: 148).

sind weitere Begriffe zur Bezeichnung dieser Disziplin entwickelt worden. Davon unterscheidet man folgende: *Maschinelle Sprachverarbeitung (MSV)* *Automatische Sprachverarbeitung (ASV)*¹⁰; *Linguistische Informatik*; *Linguistische Datenverarbeitung (LDV)*; *Informationslinguistik*; *Sprachtechnologie*; *Computerphilologie*; *Texttechnologie* (vgl. Kai-Uwe Carstensen & Co, 2010: 2; Michael Hess.WS. 2004/2005: 20)¹¹. In dieser Hinsicht könnte man den Begriff *Computerlinguistik*, im weiteren Sinne, als eine Schnittstelle zwischen **Linguistik, Ingenieurwissenschaften** (Informatik und Künstliche Intelligenz) **und Kognitionswissenschaften** (Psycho- und Neurolinguistik) auffassen und im engeren Sinne, als ein Teilgebiet der modernen Linguistik, **das berechenbare Modelle menschlicher Sprache entwirft, implementiert und untersucht**. So betrachtet, wird der Computerlinguistik die Funktion zugeschrieben, natürlich sprachliche Phänomene so zu modellieren, dass sie von Computern verstanden werden können. Dazu schreibt noch Metzger (2002) beispielweise Folgendes:

[...] sie kann verstanden werden als Verbindung einer linguistischen Teildisziplin mit informatischen und informationstechnologischen Methoden der Formalisierung von sprachlichen Daten, von sprachlichen Strukturen und Prozessen, von sprachbezogenen Regeln und Regularitäten, sei es in einer **Grundlagenforschungs-,** sei es in einer **Anwendungsperspektive**“ (Metzger 2002:203).

Allerdings hat die Proliferation verschiedener Fachbegriffe zur Bezeichnung dieser jungen wissenschaftlichen Disziplin nicht nur **den Gegenstandsentwicklungen** zu danken, sondern auch **den Produktionsverfahren und den Nutzungsperspektiven bzw. Gebrauchsperspektiven** von den in den jeweiligen Teildisziplinen produzierten Wissen. Davon unterscheidet man jeweils die folgenden Gegenüberstellungen: **Symbolische, logikbasierte** vs. *statistische, probabilistische Methoden*; **Anwendungsorientierung** vs. *Theorieorientierung*; **Algorithmisierung als Proof-Of-Concept** vs. *effiziente (kommerziell einsetzbare) Systeme*; **Hilfswissenschaft** vs. *eigenständige Forschung*;

¹⁰ *Es wird von maschinellen oder automatischen Sprachverarbeitung* die Rede sein, wenn es sich darum handelt, natürlich-sprachliche Phänomene auf dem Computer zu realisieren. Diesem entgegen wird es die Rede von **Linguistischen Datenverarbeitung**, wenn es sich darum handelt, linguistische Programme zu entwickeln, die dafür eingesetzt werden große Mengen von linguistischen Daten (etwa wie Textkorpora) im Computer zu verarbeiten (Textaufbereitung/Annotation; Informationssuche und -extraktion...). Gehören diese Texte zu einer literarischen Gattung, dann wird die Bezeichnung **Computerphilologie** gebraucht; andernfalls wird eher der Begriff **Texttechnologie** eingesetzt. Was nun die Begriffe **Sprachtechnologie und Informationslinguistik** angeht, handelt es sich insbesondere um die Entwicklung von Sprachsoftware, die dann in dem alltäglichen Leben Einsatz finden; (wie etwa bei Kommunizieren: Entwicklung von neuen Informations- und Kommunikationstheorien).

¹¹ **Verwandte englische Fachbezeichnungen:** Linguistic and Literary Computing (LLC); Natural Language Processing (NLP); (Natural) Language Engineering (NLE); Human Language Technology (HLT); Computational Philology; Corpus Linguistics; Speech Processing (**Vorlesung Wintersemester 2004/2005 Einführung in die Computerlinguistik (I) Universität Zürich Institut für Computerlinguistik Prof. Dr. Michael Hess**).

Psychologische/Neurologische Plausibilität vs. *Ingenieurmäßige Lösung* (vgl. Simon Clematide WS 2006/2007:13). Weiterhin lässt sich auch diese Feststellung über den Einfluss der in jedem Fachterminus verborgenen Schwerpunktbildung durch die folgende Äußerung von Michael Hess 2004 bekräftigt:

[...] **Ihre Ausprägungen** unterscheiden sich entlang der folgenden Dimensionen (was zu **dieser Terminologie-Vielfalt** geführt hat): Institutionelle Verankerung; Breite des Gebietes; Theorie- vs. Anwendungsorientiertheit; Simulation vs. Modellierung (Michael Hess.WS. 2004/2005:25).

Die Bestimmung der Computerlinguistik über ihren Gegenstandsbereich hat gerade gezeigt, dass es starke Uneinheitlichkeit je nach dem, was man als Gegenstand (Formale/natürliche Sprache) in Betracht zieht. Würde man nun Jedoch die Bestimmung dieses Begriffs eher über ihr Werkzeug und Medium, den Computer (**Computer als Werkzeug zur linguistischen Arbeit**), bestimmen, dann würde man den Gebrauch all dieser oben zitierten Begriffe rechtfertigen können.

Grob betrachtet, lassen sich hauptsächlich zwei Termini (*Computerlinguistik und Maschinelle Sprachverarbeitung*)¹² für fast alle Teilbereiche und Forschungsschwerpunkte einsetzen. Allerdings wird der CL hauptsächlich zwei Teilgebieten zugeschrieben, welche sich unter den Bezeichnungen **theoretische Computerlinguistik** und **praktische Computerlinguistik** aufzeichnen lassen (vgl. Kai-Uwe Carstensen& Co, 2010: 6). Daher kann man die CL sowohl als **Angehöriger der geistwissenschaftlichen Disziplinen** (theoretische Computerlinguistik) als auch als **Angehöriger der naturwissenschaftlichen Disziplinen** (Praktische Computerlinguistik) betrachten. Während die theoretische CL darauf abzielt, **formal-rechnerische Erklärung für ein besonderes linguistisches oder psycholinguistisches Phänomen zu erarbeiten**¹³, ist es in der praktischen CL eher die Rede von der Entwicklung von technologischen Komponenten eines sprachbasierten Systems, das die Mensch-Maschine-Kommunikation (MMK) erleichtert und die Verarbeitung von Texten in natürlicher Sprache zu verschiedensten Zwecken ermöglicht (vgl. A. V. Abyzov; A. G. Rys 2017: 148).

Will man nun die *Computerlinguistik* aus der Perspektive der Geisteswissenschaften betrachten, dann ergibt sich, wie oben schon erwähnt, dass dieser Teilbereich sich insbesondere darum kümmert, **linguistische Modelle** zur Erkennung und Erklärung sprachlicher Phänomene

¹²In der englischsprachigen Literatur und Informatik wird der Begriff **natural language processing (NLP)** gebraucht.

¹³<https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/> (14.01.2022 um 14h)

zu entwerfen, die dann eine gewisse **Adaptation bei der maschinellen Verarbeitung** erweisen (sich leicht maschinell verarbeiten lassen; dies impliziert, dass sprachliche Modelle in Hinsicht auf bestimmte computerlinguistische Anwendungen konzipiert werden müssen). Genau in diesem Teilbereich befindet sich diese Dissertationsarbeit, denn sie beabsichtigt **ein Modell zur Auflösung eines sprachlichen Phänomens (Anaphorik) zu entwerfen, welches dann bei konkreten computerlinguistischen Anwendungen eingesetzt wird** (Aufbau des AR-Systems und Integrierung dieses Systems in noch konkreteren computerlinguistischen bzw. sprachtechnologischen Anwendungen, wie etwa automatische Textzusammenfassungssysteme). Das Ziel der CL hier ist also die automatischen Werkzeuge zur Auflösung von anaphorischen Konstruktionen zu erweitern. So gesehen können wir sagen, dass die theoretische Computerlinguistik entwirft, implementiert und untersucht die Modelle mit dem Ziel, das Verständnis der zugrundeliegenden linguistischen und psychologischen Theorien zu verbessern. Diese oben beschriebene Auffassung des Konzepts **Computerlinguistik** wird durch das folgende Zitat von **Kai-Uwe Carstensen, Susanne Jekat und Ralf Klabunde** bekräftigt:

Computerlinguistik als Teildisziplin der Linguistik (wie Psycholinguistik, Soziolinguistik usw.), die sich, in der Regel theoriegeleitet, mit berechnungsrelevanten Aspekten von Sprache und Sprachverarbeitung beschäftigt (vgl. auch den englischen Terminus für Computerlinguistik, computational linguistics), unabhängig von ihrer tatsächlichen Realisierung auf dem Computer. Die Entwicklung von Grammatikformalismen ist ein Beispiel für diese Auffassung von Computerlinguistik (Carstensen, S. Jekat und R. Klabunde 2010: 2)

Betrachtet man nun die Computerlinguistik aus dem praktischen Blickwinkel, dann merkt man, dass es sich um die Entwicklung von Computeranwendungen handelt, die Texte und gesprochene Sprache verarbeiten zu können. Hinrich Schütze 2019 fasst die CL wie folgt zusammen auf:

Angewandte Computerlinguistik entwirft, implementiert und untersucht die Modelle mit dem Ziel, Softwareanwendungen zu ermöglichen, die über eine (eingeschränkte) Beherrschung menschlicher Sprache verfügen (Hinrich Schütze 2019: 45).

Allerdings lässt sich die Entwicklung von Softwareanwendungen bis zum Ende der 30er und Anfang der 40er Jahre zurückverfolgen. Damals handelte es sich insbesondere um numerische Problemstellungen, wie etwa, die Berechnung von ballistischen Kurven. Manche Aufgaben bestanden aber auch darin, verschlüsselte Nachrichtentexte zu dechiffrieren (Dechiffrierung von Symbolen); aus diesen Aufgabentypen wurde dann die maschinelle Übersetzung (MÜ) entwickelt, die dann damals als Spezialfall einer Dekodierungsaufgabe verstanden wurde. Deswegen gilt sie auch heutzutage als emblematische sprachtechnologische Anwendung (vgl. Stefan Müller 2013: 17).

Der Begriff **Sprachtechnologie** wird meistens gebraucht, um auf die Umsetzung von linguistisch gewonnenen Erkenntnissen in technologischen verwendbaren Anwendungen (vgl. Detmar Meurers 2019: 10). Dabei sind folgende Fragestellungen von Relevanz: Wie lassen sich natürliche Sprache in der Praxis verarbeiten? Welche praktisch gesellschaftlich orientierten Anwendungen lassen sich (er)finden und für die der aktuelle Sprachtechnologiewissenstand hinreichend gut ist? Zu diesen Sprachsoftwares gehören Beispielsweise die Umwandlung von gesprochener in geschriebene Sprache (Spracherkennung) umwandeln, z.B. Diktiersysteme (Sprachsynthese) und Dialogsysteme; Internet-Suchmaschinen; die Übersetzungssysteme; Rechtschreibprüfung und Grammatikprüfung; Informationssuche und -extraktion in/aus Texten; Automatische Zusammenfassung (Summarisation) (vgl. Kai-Uwe Carstensen & Co, 2010: 23-25; Vera Demberg WS 2014/15: 20-25). Genauso bei solchen Anwendungen und bei allen anderen Sprachsoftwares im Allgemeinen werden der im Rahmen dieser Dissertationsarbeit zu entwickelnde Algorithmus und das darauf anschließende praktische Modul Einsatz finden. **Denn er/es wird dem Computer bzw. schon bestehenden Sprachsoftwares die Möglichkeiten geben, ihrer Fähigkeit zum besseren Textverständnis ermöglichen.** Das am Ende entwickelte System könnte also entweder individuell (zur Auflösung von Anaphern) oder eher als Komponente eines noch wichtigeren bzw. umfangreichen Sprachsystems (wie etwa Übersetzungssysteme; Dialogsysteme oder noch Informationssuch- und Informationsextraktionssysteme) eingesetzt werden.

Weiterhin lässt sich merken, dass im Gegensatz zur theoretischen Computerlinguistik, die, wie oben schon hingewiesen worden ist, Grundlagen für ein Verständnis natürlicher Sprache durch Rechner bereits gestellt werden; beabsichtigt **die praktische Computerlinguistik** eher menschliche Nutzer bei allen Formen des Umgangs mit elektronischen Sprachbeständen aktiv zu unterstützen. Als besonderer Umgang mit elektronischen Sprachbeständen gilt zum Beispiel **die Umwandlung gesprochener oder gedruckter Sprache in ein elektronisches Format.** Etwas, das einen großen Beitrag in der Informationstechnologie leistet, ein Zweig, der in der heutigen Gesellschaft von hoch Relevanz ist; denn wir leben in einer Informationsgesellschaft (Gesellschaft, die ohne Informationsaustausch nicht funktionieren könnte) (vgl. A. V. Abyzov; A. G. Rys 2017: 149).

Von all diesen Überlegungen ausgehend, ergibt es sich, dass die Computerlinguistik, sei es theoretisch oder praktisch betrachtet, sich als eine Zusammenbringung unterschiedlichen

Disziplinen bezeichnen lässt.¹⁴ In dieser Hinsicht soll also der Computerlinguist als ein Ingenieur verstanden werden, der eine gestellte Aufgabe lösen will. Daher ist es ihm wichtiger:

Algorithmen und Mechanismen zu finden, die es erlauben, konkrete Ziele im Bereich der Sprachverarbeitung zu erreichen. Man kann daher Computerlinguistik zwar als eine Verbindung von Linguistik, Informatik und Logik zu einer Disziplin betrachten. Man muß dabei jedoch beachten, daß die spezielle Ausrichtung des Faches auf die Konstruktion funktionierender Mechanismen, **eine eigenständige Dynamik hervorbringt**, die zu Resultaten führt, die von den zugrundeliegenden Fächern nur selten oder gar nicht erreicht werden (vgl. Hinrich Schütze 2019: 60).

Will man nun diese *eigenständige Dynamik* der Computerlinguistik betrachten, dafür lassen sich am Besten die Forschungsschwerpunkte, die der CL in jeder deutschen Universität zugeschrieben wurde, betrachten.

1.1.1. Computerlinguistik als Fachbereich und computerlinguistische Fachgesellschaften

Die Erläuterung *der eigenständigen Dynamik*, die die Computerlinguistik mit sich bringt, könnte sich, wie oben vermutet, vielleicht anhand der fachlichen Orientierung des an den verschiedenen computerlinguistischen Studienorten angebotenen Lehr- und Forschungspakets veranschaulichen lassen. Diese computerlinguistischen Studienorte lassen sich je nach Schwerpunkte hauptsächlich in drei Kategorien einordnen: **TEXTTECHNOLOGIE**;

¹⁴Darunter zählen **die Linguistik** (Wissenschaft, die sich mit menschlicher Sprache aus unterschiedlichen Perspektiven (Phonetik/Phonologie, Morphologie, Syntax, Semantik, Pragmatik; Korpuslinguistik) beschäftigt; **die Informatik** (Algorithmen, Datenstrukturen, Software Engineering); die **Kognitionswissenschaften** (Sprachbeherrschung ist spezieller Teilbereich der kognitiven Fähigkeiten des Menschen); die **Künstliche Intelligenz** (Wissensrepräsentation; Nachdenken und Selbstlernen von Maschinen); **Philosophie** (Verbindung von Sprache, Denken und Handeln; Relation zu außersprachlichen Gegebenheiten) und die **Mathematik** (Logik, Wahrscheinlichkeitstheorie, Statistik, Graphentheorie). Außerdem gibt es auch andere Zweige der angewandten Linguistik, die auch **elektronische Texte** im Zentrum ihrer Forschungsgegenstände haben; das ist beispielweise der Fall **der Korpuslinguistik**. In der Tat, werden insbesondere Werkzeuge und Methoden, die von **Computerlinguisten** entwickelt worden sind, verwendet, um manche korpuslinguistische Problemstellungen zu klären. Darunter zählen wir beispielweise **den Parser** (automatisches Werkzeug zur syntaktischen Segmentierung von Korpora) oder noch das **Tagging** (automatisches Werkzeug zur morphologischen Etikettierung von sprachlichen Einheiten eines Korpus) (vgl. Kai-Uwe Carstensen & Co, 2010: 280-328). Zusammengefasst, wird also der **CL**, je nach Aufgabenstellungen, unterschiedliche Ziele zugeschrieben. Darunter zählt man beispielsweise **die Erstellung, Verwaltung und Aufbereitung von Korpora**; Die Bereitstellung anderer textzentrierten Wissenschaften **Werkzeuge, Methode und Techniken**; die Lösung praktischer Probleme durch die Entwicklung realistischer und technologisch fortschrittlicher (**Sprachtechnologie**) zu entwickeln und die Entwicklung und Evaluierung sprachverarbeitender Systeme und Komponenten (vgl. Ebd.:1-25).

DIGITALHUMANITIES (DH) und INFORMATIONSWISSENSCHAFT (IW). Mit dem Begriff *Texttechnologie* ist ein Teilgebiet der Computerlinguistik gemeint, das die geschriebene Sprache (Texte) als Forschungsgegenstand hat. Handelt es sich bei computerlinguistischer Forschung um historische bzw. philologische Texte, dann wird die Bezeichnung *Digitalhumanities* (Computerphilologie¹⁵) lieber gebraucht. Was die *Informationswissenschaft und –Informationsretrivial* betrifft, handelt es sich insbesondere vielmehr um eine fachliche Orientierung der CL, in der computerlinguistische Methoden und Theorien gebraucht werden, um Informationssysteme und –prozesse zu verarbeiten (entwerfen und untersuchen) (Fotis Jannidis, 1999: 39–60).

Die historische Entwicklung der Lehre und Forschung der Computerlinguistik in Deutschland verrät uns, dass die ersten Computerlinguistik-Studiengänge an der Universität des Saarlandes und in Stuttgart (Baden-Württemberg) eingerichtet wurden¹⁶. Damals wurde sie noch als kleines Fach eingestuft, denn nur Bachelor-und Master-Studienabschlüsse waren möglich. Mittlerweile ist die Auffassung dieses Faches deutschlandweit gewachsen; man stellt nämlich fest, dass die Computerlinguistik heutzutage in mehr als 25 Universitätsstandorte (staatliche Angebote) betrieben wird. Darüber hinaus werden auch seit dem letzten Jahr (2021) auch neue Professuren für Computerlinguistik in theologischen und privaten Hochschulen verzeichnet (Michael Hess.WS. 2004/2005: 30)¹⁷. Im Allgemeinen zählen zu den bekanntesten Angeboten die Studiengänge der Universität des Saarlandes, der Ludwig-Maximilians-Universität München, der Universität Bielefeld, **der Universität Potsdam**, der Ruprecht-Karls-Universität Heidelberg und der Universität Trier (vgl. Ebd.: 30). Weiterhin lässt sich jedoch merken, dass obwohl der Studiengang *Computerlinguistik* in manchen Studienorten explizit nicht angeboten wird (als eigenständiger Studiengang nicht verfügbar), werden trotzdem computerlinguistische Module als Lehrveranstaltungen für andere engverwandten Studiengänge freigeschaltet. Dies ist beispielweise der Fall von der technischen Universität (TU) Dresden, in der einige sprachtechnologisch orientierte **Lehrveranstaltungen**¹⁸, wie etwa *Signalverarbeitung*,

¹⁵Der Begriff *Computerphilologie* ist ein Sammelbegriff für die Einsatzmöglichkeiten des Computers in der Literaturwissenschaft. Das Wort ist Anfang der 1990er Jahre als Parallelbildung zu dem älteren Begriff Computerlinguistik entstanden *Die Computerphilologie* ist ein Teil der *Digital-Humanities*. Fotis Jannidis: Was ist Computerphilologie? In: Karl Eibl, Volker Deubel, Fotis Jannidis (Hg.): Jahrbuch für Computerphilologie 1. Paderborn: mentis 1999, S. 39–60. <https://de.m.wikipedia.org/wiki/Computerphilologie> (21.02.2022 um 18h03)

¹⁶<https://de.m.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03)

¹⁷Vorlesung Wintersemester 2004/2005 Einführung in die Computerlinguistik (I) Universität Zürich Institut für Computerlinguistik Prof. Dr. Michael Hess.

¹⁸Solche sprachtechnologischen Lehrformaten sind auch in den folgenden universitären Einrichtungen erhältlich: Universität **Regensburg**; **Karlsruher Institut für Technologie**; Universität **Hildesheim**; **RWTH Aachen** und Universität **Bremen**. Außerdem merkt man auch das gleiche Phänomen sowohl bei digitalhumanistischen (DH)

Spracherkennung und Sprachsynthese für Computerlinguistik engverwandte Studiengänge freigeschaltet sind. Solche Gegebenheiten erschweren **eine vernünftige Klassifikation des Profils von CL-Absolventen** herauszuarbeiten. Und zwar nämlich, weil es nicht durchsichtlich ist, welche Fähigkeit(en) ein Computerlinguist von einem anderen distanziert bzw. unterscheidet. (Zu dieser Diskussion kommen wir noch gleich zurück); lassen wir uns aber zunächst die Studien-und Forschungslandschaft der CL in Deutschland in Einzelheiten betrachten.

	UNIVERSITÄTSNAME	LOKALE FACH-/STUDIENGANGBEZEICHNUNG	INSTITUTIONELLE VERANKERUNG
TEXTTECHNOLOGIE	Universität Bielefeld	<i>Angewandte- und Theoretische Computerlinguistik</i> ¹⁹	<u>FAKULTÄT FÜR LINGUISTIK UND LITERATURWISSENSCHAFT</u> <u>2Studiengänge</u> (Bachelor in Texttechnologie und Computerlinguistik als Nebenfach) Master of Arts in Sprachtechnologie
	Ruhr-Universität Bochum	<i>Computerlinguistik</i> ²⁰	<u>FAKULTÄT FÜR PHILOLOGIE</u> <u>SPRACHWISSENSCHAFTLICHES INSTITUT</u> <u>2 Professuren:</u> -Stefanie Dipper (Computerlinguistik) ; -Tibor Kiss (Allgemeine und

orientierten CL-Studiengängen (**Universität Bamberg; Universität Passau; Universität Dortmund; Universität Würzburg; Universität Göttingen; Universität Münster**) als auch bei IW-orientierten CL-Studiengängen (**Universität Hildesheim ; Technische Universität Dortmund; Bauhausuniversität Weimar; Universität Kaiserslautern; U. Hamburg**) <https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/> (29.01.2022 um 10h).

¹⁹<https://www.uni-bielefeld.de/fakultaeten/linguistik-literaturwissenschaft/studium-lehre/faecher/texttechnologie/> (20.02.2022 um 13h).

²⁰<https://www.linguistics.ruhr-uni-bochum.de/> (20.02.2022 um 13h).

			vergleichende Sprachwissenschaften unter besonderer Berücksichtigung der multimedialen Computerlinguistik)
	Universität Erlangen-Nürnberg	Korpus-und Computerlinguistik ²¹	<u>FACULTY OF HUMANITIES, SOCIAL SCIENCES AND THEOLOGY</u> <u>DEPARTMENT GERMANISTIK UND KOMPARISTIK</u> <u>1 Studiengang:</u> Bachelor of Arts (Linguistische Informatik) <u>1 Professur</u> -Stephanie Evert (Korpuslinguistik)
	Justus-Liebig-Universität Gießen	Angewandte Sprachwissenschaft und Computerlinguistik ²²	<u>FACHBEREICH FÜR SPRACHE LITERATUR UND KULTUR</u> <u>INSTITUT FÜR GERMANISTIK</u> <u>1Professur</u> -Nicht Besetzt (Angewandte Sprachwissenschaft und Computerlinguistik)
	Ruprecht-Karls-Universität Heidelberg	Computerlinguistik ²³	<u>FAKULTÄT FÜR NEUPHILOLOGIE</u> <u>INSTITUT FÜR COMPUTERLINGUISTIK</u>

²¹<https://www.linguistik.phil.fau.de/> (20.02.2022 um 13h).

²²<https://www.uni-giessen.de/fbz/fb05/germanistik/iprof/asclhome> (20.02.2022 um 13h).

²³<https://www.cl.uni-heidelberg.de/> (20.02.2022 um 13h).

			<u>2Studiengänge</u> Computerlinguistik (Bachelor of Arts Zwei-Fach-Bachelor) (Bachelor of Arts Ein-Fach-Bachelor)) (Master of Arts) <u>3 Professuren bzw. Arbeitsgruppen</u> -Annette Frank (Computerlinguistik) -Katja Markert (Computerlinguistik) -Stefan Riezler (Linguistische Informatik)
	Universität Potsdam	<i>Computerlinguistik</i> ²⁴	<u>PHILOSOPHISCHE FAKULTÄT</u> <u>DEPARTMENT LINGUISTIK</u> <u>Studiengang</u> -Computerlinguistik (Bachelor of Science/ Ein-Fach-Bachelor) <u>2Professuren</u> David Schlangen (Grundlagen der Computerlinguistik) Manfred Stede (Angewandte Computerlinguistik)
	Eberhard Karls Universität Tübingen	<i>Theoretische Computerlinguistik, Allgemeine Sprachwissenschaft / Computerlinguistik.</i> ²⁵	<u>PHILOSOPHISCHE FAKULTÄT</u> <u>SEMINAR FÜR SPRACHWISSENSCHAFT</u>

²⁴<https://www.uni-potsdam.de/en/ling/index> (20.02.2022 um 13h).

²⁵<https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft?context=ascl> (20.02.2022 um 13h).

			<u>Studiengänge</u> Bachelor of Arts in CL Master of Arts in CL <u>-2Professuren</u> Erhard Hinrichs (Allgemeine Sprachwissenschaft und Computerlinguistik) Walt Detlar Meurs (Theoretische Computerlinguistik)
	Universität Trier	<i>Linguistische Datenverarbeitung²⁶</i>	<u>FACHBEREICH SPRACH-, LITERATUR- UND MEDIENWISSENSCHAFTEN</u> INSTITUT FÜR COMPUTERLINGUISTIK UND DIGITALHUMANITIES <u>Studiengänge</u> Bachelor of Arts in CL Master of Arts in CL Master (Ein-Fach) of Arts in CL Master of Science in Natural Language Processing <u>Professur</u> Achim Rettinger (Computerlinguistik)
	Universität des Saarlandes	<i>Computerlinguistik²⁷</i>	<u>PHILOSOPHISCHE FAKULTÄT</u> <u>FACHRICHTUNG SPRACHWISSENSCHAFT</u>

²⁶<https://www.uni-trier.de/index.php?id=1175> (20.02.2022 um 13h).

²⁷<https://www.uni-saarland.de/fachrichtung/lst.html> (20.02.2022 um 13h).

			<u>HAFT UND SPRACHTECHNOLOGIE</u> <u>Studiengang</u> Bachelor of Science in CL Master of Science <u>Professuren</u> Matthew Crocker (Psycholinguistik) Alexander Koller (Computerlinguistik)
DIGITALHUMANITIES	Technische Universität Darmstadt	<i>Computerphilologie / Computer- und Korpuslinguistik²⁸</i>	<u>FACHBEREICH 2 GESELLSCHAFTS- UND GESCHICHTSWISSENSCHAFT</u> <u>INSTITUT FÜR SPRACH- UND LITERATURWISSENSCHAFT</u> <u>FACHGEBIET UBIQUITÄRE WISSENSVERARBEITUNG</u> <u>Studiengänge</u> -Master of Arts <u>Professuren</u> -Nicht besetzt (Computerphonologie und Corpuslinguistik) -Iryna Gurevych (Ubiquitous Knowledge Processing (UKP))
	Humboldt-Universität zu Berlin	<i>Korpuslinguistik²⁹</i>	<u>SPRACH UND LITERATURWISSENSCHAFTLICHE FAKULTÄT</u> <u>FACHCHAFT GERMANISTIK</u>

²⁸<https://www.linglit.tu-darmstadt.de/institutlinglit/index.de.jsp> (20.02.2022 um 13h).

²⁹<https://www.linguistik.hu-berlin.de/de> (20.02.2022 um 13h).

			<u>INSTITUT FÜR DEUTSCHE SPRACHE UND LINGUISTIK</u> <u>1Professur</u> -Anke Lüdeling (Korpuslinguistik und Morphologie)
	Johannes Gutenberg- Universität Mainz	<i>Allgemeine und Angewandte Sprachwissenschaft sowie Translationstechnologie³⁰</i>	<u>FACHBEREICH: TRANSLATIONS- SPRACH- UND KULTURWISSENSC HAFT</u> <u>ARBEITSBEREICH ALLGEMEINE SPRACHWISSENSC HAFT SOWIE TRANSLATIONSTE CHNOLOGIE</u> <u>Professur</u> -Christoph Rösener (Allgemeine und Angewandte Sprachwissenschaft sowie Translationstechnologie)
	Johann Wolfgang- Goethe- Universität Frankfurt am Main	<i>Computational Humanities</i>	<u>FACHBEREICH INFORMATIK UND MATHEMATIK</u> <u>INSTITUT FÜR INFORMATIK</u> <u>APPLIED COMPUTATIONAL LINGUISTIK LAB</u> <u>TEXTTECHNOLOG Y LAB</u>

³⁰<https://astt.fb06.uni-mainz.de/> (20.02.2022 um 13h).

			<u>2Professuren</u> Alexander Mahler (Computational Humanities/ Text Technology) Christian Chiarcos (Angewandte Computerlinguistik)
	Friedrich-Schiller-Universität Jena	<i>Computerlinguistik³¹</i>	<u>PHILOSOPHISCHE FAKULTÄT</u> <u>INSTITUT FÜR GERMANISTISCHE SPRACHWISSENSCHAFT</u> <u>1PROFESSUR</u> Udo Hahn Germanistische Sprachwissenschaft (Computerlinguistik)
	Universität Kassel	<i>Computerphilologie³²</i>	<u>FABEREICH 2 GEISTES UND KULTURWISSENSCHAFT</u> <u>INSTITUT FÜR ANGLISTIK UND AMERIKANISTIK</u> <u>1Professur</u> Josef Wallmannsberger (Anglistik/Linguistik-Computerphilologie)
	Universität Konstanz	<i>Allgemeine Sprachwissenschaft</i>	

³¹<https://www.gw.uni-jena.de/germanistische-sprachwissenschaft> (20.02.2022 um 13h).

³²<https://www.uni-kassel.de/fb02/> (20.02.2022 um 13h).

		<i>und Computerlinguistik³³</i>	<u>GEISTWISSENSCHAFTLICHE SEKTION</u> <u>FACHBEREICH LINGUISTIK</u> <u>Professur</u> -Miriam Butt (Allgemeine Sprachwissenschaft mit Schwerpunkt Computerlinguistik)
	Universität zu Köln	<i>Sprachliche Informationsverarbeitung³⁴</i>	<u>PHILOSOPHISCHE FAKULTÄT</u> <u>INSTITUT FÜR DIGITAL HUMANITIES</u> <u>Studiengang</u> Terminologie und Sprachtechnologie (Master of Arts) <u>1Professur</u> Nils Reiter (Digital Humanities – Sprachliche Informationsverarbeitung)
	Universität Leipzig	<i>Computerlinguistik³⁵</i>	<u>PHILOLOGISCHE FAKULTÄT</u> <u>INSTITUT FÜR LINGUISTIK</u> <u>ARBEITSGRUPPE TEMIR</u> <u>Professuren</u> Grygory Kobele (Computationelle/expe

³³<https://www.ling.uni-konstanz.de/> (20.02.2022 um 13h).

³⁴<https://dh.phil-fak.uni-koeln.de/> (20.02.2022 um 13h).

³⁵<https://www.philol.uni-leipzig.de/institut-fuer-linguistik> (20.02.2022 um 13h).

			rimentelle Zugänge zur Grammatik)
	Universität Stuttgart	Maschinelle Sprachverarbeitung³⁶	<u>FAKULTÄT FÜR INFORMATIK, ELEKTRONIK UND INFORMATIONSTECHNIK</u> <u>INSTITUT FÜR MASCHINELLE SPRACHVERARBEITUNG</u> <u>Studiengang</u> Master of Science in computational Linguistics <u>Professuren</u> Jonas Kuhn (Grundlage der CL) Sebastian Pado (Theoretische CL) Thang Vu (Digitale Phonetik) Carina Silberer (Computerlinguistik)
INFORMATIONSWISSENSCHAFT UND INFORMATIONSR ETRIEVIAL	Heinrich-Heine-Universität Düsseldorf	Computerlinguistik³⁷	<u>DIE PHILOSOPHISCHE FAKULTÄT</u> <u>INSTITUT FÜR SPRACHE UND INFORMATION</u> <u>Studiengänge</u> Computerlinguistik (Bachelor of Arts) (Master of Arts) <u>Professur</u> Laura Kallmeyer (Computerlinguistik)

³⁶<https://www.ims.uni-stuttgart.de/> (20.02.2022 um 13h).

³⁷<https://www.isi.hhu.de/> (20.02.2022 um 13h).

	Ludwig Maximilians- Universität München	Informations- und Sprachverarbeitung/Co mputerlinguistik³⁸	<u>FAKULTÄT FÜR SPRACH UND LITERATURWISSE NSCHAFT</u> <u>CENTRUM FÜR INFORMATIONS- UND PRACHVERARBEIT UNG</u> <u>Studiengänge</u> -Computerlinguistik (Bachelor of Science) -Computerlinguistik (Master of Science mit Nebenfach) --Computerlinguistik (Master of Science mit Profildbereich) <u>2Professuren</u> -Klaus U. Schulz (Informations- und Sprachverarbeitung) -Hinrich Schütze (Computerlinguistik)
	Otto-von- Guericke- Universität Magdeburg	Wissensbasierte Systeme und Dokumentverarbeitung³⁹	<u>FAKULTÄT FÜR INFORMATIK</u> <u>INSTITUT FÜR WISSENS-UND SPRACHVERARBEI TUNG</u> -Arbeitsgruppe (Wissensbasierte Systeme und Dokumentverarbeitun g) -Arbeitsgruppe (ComputationalIntelli gence) -Arbeitsgruppe (Theoretische Informatik)

³⁸<https://www.cis.lmu.de/> (20.02.2022 um 13h).

³⁹<http://iks.cs.ovgu.de/iws/> (20.02.2022 um 13h).

			-Arbeitsgruppe (intelligenten Systeme)
	Universität Koblenz- Landau	<i>Computerlinguistik</i> ⁴⁰	<u>FACHBEREICH INFORMATIK</u> <u>INSTITUT FÜR COMPUTERVISUAL ISTIK</u> <u>2Studiengänge</u> Computervisualistik (Bachelor of Science) Computervisualistik (Master of Science) <u>3 Professuren bzw. Arbeitsgruppen</u> -Karin Harbusch (Naturlichsprachliche Künstliche Intelligenz) -Stefan Müller (Computergrafik) Dietrich Paulus (Aktives Sehen)

Tabelle 3. Eine Taxonomie⁴¹ der Studien- und Forschungsorte für Computerlinguistik in Deutschland⁴²

Die obenstehende Tabelle, die anhand der an Webseiten der jeweiligen Universitäten erhobenen Daten aufgefasst worden ist, zeigt uns klar, dass die CL in deutschen universitären Einrichtungen sich schwer von einem Studienort zu einem anderen unterscheidet. In der Tat, lassen sich die verzeichneten Unterschiede nicht nur auf der Ebene der angeschriebenen Schwerpunkte (**TEXTTECHNOLOGIE, DH, IW**), aber auch auf der Art des jeweiligen **Bildungsformats**⁴³ und der **institutionellen Verankerung** jeder Universitäten veranschaulichen. In dieser Hinsicht wurde nämlich bemerkt, dass der Studiengang

⁴⁰<https://www.uni-koblenz-landau.de/de/koblenz/fb4/icv> (20.02.2022).

⁴¹https://www.kleinefaecher.de/kartierung/kleine-faecher-von-a-z.html?tx_dmdb_monitoring%5BdisciplineTaxonomy%5D=26&cHash=016cd76da667e9ee63e6eefeb4acfb4#overview (20.02.2022).

⁴²<https://www.studieren-studium.com/master/studieren/Computerlinguistik-Deutschland#resulthttps://www.studieren-studium.com/studium/studieren/Computerlinguistik-Deutschland#result> (04.03.2022 um 14h30).

⁴³Während die CL bei manchen Universitäten nur als Hauptfach eines zwei-/drei-Fächer-Typs Studiengang verfügbar ist, ist sie in einer anderen Universität nur als Neben Fach eines zwei-/drei-Fächer-Typs zugänglich.

Computerlinguistik meistens unter den Studiengängen einer Philosophischen Fakultät eingeordnet bzw. angeboten wird. Nur wenige Universitäten verzeichnen diesen Studiengang als Angehöriger einer Informatischen oder Mathematischen Fakultät (Das ist der Fall von den Universitäten **Magdeburg, Kloblenz-Landau; Stuttgart und Frankfurt**). Dieses Faktum gibt uns schon Hinweise über die globale Orientierung, die in der Forschung und Lehre der Computerlinguistik herrscht. Es handelt sich also, wie wir es bei der Erläuterung des Begriffs zuvor erwähnt haben, zunächst um **Linguistik** und erst nachher um **Informatik**.

Allerdings ist mittlerweile diese Auffassung gewachsen. Denn, obwohl der Studiengang **Computerlinguistik** in einer philosophischen Fakultät einer gegebenen Universität angeboten wird, werden den CL-Absolventen aber anstatt ein Bachelor/ Master of Arts eher **ein Bachelor/ Master of Science** als Abschlussdiplom angeboten. Das ist beispielweise der Fall von der Ludwig-Maximilians-Universität München. Diese Feststellung gibt uns schon klare Hinweise über die Natur **der Lehrinhalte und Forschungsschwerpunkte**, die in solchen Studienorten angeboten werden. Vermittelt werden nämlich naturwissenschaftliche Kenntnisse (Mathematik & Informatik) und dies, obwohl wir uns in einer philosophischen Fakultät befinden. Solche Gegebenheiten zeigen immerhin nicht nur, dass, wenn man von Computerlinguistik redet, der Vorrang der Linguistik gegeben wird, sondern auch, dass sich **eine gewisse Dynamik** bei der Strukturierung der Lehrinhalte und Forschungsgruppen auszeichnen lässt.

Außerdem lässt sich diese Dynamik außerhalb des Merkmals **Institutionelle Verankerung** herausarbeiten. Es wurde nämlich beobachtet, dass die am Anfang dieses Teils erwähnten allgemeinen Orientierungen des Fachs/Studiengangs **Computerlinguistik (Texttechnologie; DH und IW)** nicht immer geradlinig sind. Es wurde nämlich festgestellt, dass während manche Studienorte sich sowohl in der Rubrik Texttechnologie als auch in der Rubrik DH einordnen lassen können (**das ist beispielweise, der Fall von der Friedrich-Schiller-Universität Jena**); lassen sich andere Studienorte eher sowohl in der Rubrik DH als auch in der Kategorie IW einordnen (**das ist beispielweise, der Fall von der Heinrich-Heine-Universität Düsseldorf**). Ein Erklärungsversuch dieser Gegebenheit stützt sich einerseits, auf die Tatsache, dass die Grenze zwischen *Texttechnologie* und *Digitalhumanities* nicht so ganz dicht ist. (Handle es sich um **Theorien** oder auch um **Methoden**) (vgl. Kai-Uwe Carstensen & Co, 2010: 2-18); und andererseits auf die Tatsache, dass die **Konfiguration der inhaltlichen Merkmale** (der beim Aufbau der Rubrik berücksichtigenden Aspekte) jeder Rubrik nur an hauptvermittelte Kompetenzen beschränkt ist bzw. gestarrt hat.

Allerdings lässt sich weiterhin trotzdem feststellen, dass sich eine besondere Logik in dieser Klassifizierung herauszeichnet. Denn, wenn man von der Ausrichtung Sprachtechnologie zur Digitalhumanities geht, merkt man, dass das Studium von Computerlinguistik von **einer philosophischen Fakultät zu einer geisteswissenschaftlichen (philologischen oder noch kulturwissenschaftlichen) Fakultät** abrutscht.⁴⁴ Und von den beiden Ausrichtungen (Texttechnologie & DH) zu der letzten (IW) ausgehend, wird ein Abrutsch von den **Geisteswissenschaften zu den Naturwissenschaften** beobachtet. In dieser Hinsicht wurde eben auch gemerkt, dass die meisten Universitäten, in denen der Studiengang **Computerlinguistik** als Angehöriger **einer informatischen Fakultät** fungiert, in der Rubrik Informationswissenschaft verzeichnet wurden. Und dies ist so vielleicht, weil es in dieser Ausrichtung viel mehr darum geht, neue Informations- und Kommunikationstechnologien zu entwickeln.

All dies führt dazu bei, dass das Profil **des CL-Absolventen**, obwohl es aus einem DH-orientierten Fach stammt, sich stark von einer Uni zu einer anderen unterscheidet. Dies ist sowohl für die Forschung⁴⁵ (Studierenden und Wissenschaft) als auch für Unternehmen (Berufswelt) unangenehm, denn diese Letzten müssen jedes Mal, wenn sie CL-Absolventen rekrutieren möchten; zusätzlich noch verifizieren welche **Grund- und Vertiefungswissen** dieser besitzt (vgl. Udo Hahn, Matthias Heyn & Co, 1988: 86).

Wie oben schon zuvor erwähnt, lässt sich die Überschaubarkeit und Vergleichbarkeit der Studienangebote in unterschiedlichen computerlinguistischen Standorten aufgrund **der divergierenden fachlichen und institutionellen Einbindung jeder Universität** schwer nachvollziehen. Dies erschwert ebenfalls ein einheitliches **Profil von CL-Absolventen** zu

⁴⁴Gemeinsam bleibt aber der Fakt, dass das Studium fast immer in einem sprachwissenschaftlichen Institut stattfindet.

⁴⁵ Diese Feststellung lässt sich mit **zwei Erklärungsansätzen** kommentieren, welche sich sowohl distanzieren als auch kombinieren lassen. In der Tat, wird eine solche Diskrepanz in dem Fachgebiet Computerlinguistik gemerkt, denn die CL ist eine relative junge wissenschaftliche Disziplin, die sich erst in den letzten 50 Jahren etabliert hat. Und im Gegensatz zu anderen Disziplinen, die schon Uralt sind, **versucht sie momentan ein eigenständiges und einheitliches Profil aufzubauen**. Allerdings lässt sich dieser Ausstieg von den alten Kategorisierungen/Denkweisen von was man unter **wissenschaftlicher Disziplin** verstanden hat, nicht nur auf der Ebene der CL beobachten, sondern auch auf der Ebene der neuen erstandenen Wissenschaften feststellen. Denn es herrscht heutzutage in der Wissenschaft **ein Paradigma der Abschaffung der Grenzen zwischen wissenschaftlichen Disziplinen**. Methoden einer Disziplin wird in einer anderen eingesetzt und Problemstellung einer Fachrichtung wird in einer anderen untersucht. Zusammengefasst können wir also sagen, dass dieser Charakter des Fachgebiets Computerlinguistik seinem Zeitgeist gemäß ist.

skizzieren, etwas, das den Studienanfängern keine klare Ausbildungs- und Berufsperspektive ermöglicht. In dieser Hinsicht sind in den letzten Jahren viele **Fachgesellschaften** entstanden, die sich das Ziel gesetzt haben, Fachleute zusammenzubringen⁴⁶, die über eine bestimmte Thematik (Schwerpunkt) arbeiten. **Das angestrebte Ziel war also Ordnung in der Vielfalt der Fragestellungen und Forschungsschwerpunkte zu bringen.** Als zentrale Fachgesellschaften der Computerlinguistik unterscheidet man die **Association for Computational Linguistics (ACL)**⁴⁷; die **Deutsche Gesellschaft für Sprachwissenschaft (DGfS)** / Sektion Computerlinguistik; **die Gesellschaft für linguistische Datenverarbeitung (GLDV)** sowie **die Gesellschaft für Sprachtechnologie und Computerlinguistik (GSCL)**. Als weitere Gesellschaften haben wir die AFNLP (Asian Federation of Natural Language Processing Associations); Österreichische Gesellschaft für Artificial Intelligence (ÖGAI)/ Bereich Sprachverarbeitung (vgl. M. Hess 2005: 35).

Gemerkt wurde eben auch, dass die Mitglieder dieser Gesellschaften sich auch regelmäßig treffen. Als regelmäßig stattfindende Tagungen, zählen beispielsweise: die jährlich stattfindende Konferenz der „Association of Computational Linguistics (ACL)“; die in zweijährigem Abstand stattfindende Jahrestagung der „Gesellschaft für linguistische Datenverarbeitung (GLDV)“⁴⁸; die in zweijährigem Abstand stattfindende Konferenz zur Verarbeitung natürlicher Sprache (KONVENS), die abwechselnd von den Gesellschaften ÖGAI, DGfS-CL und GSCL organisiert ist. Weiterhin lassen sich auch viele andere Tagungen verzeichnet; manche sind auch von Studierenden selbst invitiert worden. Dies ist beispielsweise der Fall von der jährlich stattfindenden „Studentische Tagung Sprachwissenschaft (StuTS)“; die „Tagung der

⁴⁶Auf Anregung der Arbeitskreise *"Ausbildung und Berufsperspektiven"* sowie *"LDV und Nachbarn"* haben daher Beirat und Vorstand der GLDV im März 1986 Arbeitstreffen interessierter Mitglieder initiiert, deren Ziel es war, etablierte und im Aufbau begriffene Studiengänge, deren **Inhalte, Ausbildungsprofile** sowie die sich daraus ergebende Art von **Berufsqualifikation** zu analysieren. Langfristige Zielvorstellungen für diese Arbeit waren, - **das Profil des Faches CL** genauer zu konturieren und **damit die Berufsqualifikation eines CL-Absolventen** beschreibbar zu machen, - **inhaltliche Minimalvorstellungen** (je Studiengangtypus) zu entwickeln, - **Lehr- und Lernziele einzelner Lehrveranstaltungen zu beschreiben**, - **die "Durchlässigkeit"** (Studienort-Wechsel) zwischen einzelnen Studiengängen zu erhöhen oder überhaupt erst zu ermöglichen, - **die Überschaubarkeit für die Studienberatung zu verbessern**. Für die gestellte Aufgabe erwies es sich als zweckmäßig, einen Diplomstudiengang Computerlinguistik (**ohne eine institutionelle Einbindung**) als "tertiumcomparationis" (s. Kap. 3) anzusetzen, zu dem die existierenden oder geplanten Studiengänge dann in Beziehung gesetzt werden können. Einig war man sich auch über **die Fachbezeichnung "Computerlinguistik" (CL)** (vgl. Udo Hahn, Matthias Heyn, Ludwig Hitzenberger Bernhard Kelle, Hans-Dieter Lutz, Klaus Wothke 1988: 86).

⁴⁷<https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/> (21.02.2022 um 17h10)

⁴⁸ Seit 2008 wird die Bezeichnung „Gesellschaft für Sprachtechnologie und Computerlinguistik (GSCL) an der Stelle von (GLDV) gebraucht. Als weite Konferenzen unterscheidet man: Eine internationale Konferenz namens „COLING“, die in zweijährigem Abstand stattfindet; Die „International Joint Conference on Natural Language Processing (IJCLP)“, die seit 2004 im asiatischen Raum in unregelmäßigen Abständen stattfindet. <https://de.m.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03).

Computerlinguistik-Studierenden (TaCoS)“ deutschsprachiger Universitäten, die seit 1992 jährlich an jeweils an einer anderen Universität stattfindet (vgl. Udo Hahn, Matthias Heyn & Co, 1988: 88-90).

In diesem Abschnitt war es also die Rede **diese eigenständige Dynamik der Computerlinguistik** anhand der Gestaltung ihres Fachbereiches zu charakterisieren. Es ergibt sich dass, obwohl die Studiengänge heutzutage dank des LMD Systems ganz anders strukturiert sind (Hauptfach & Nebenfach)⁴⁹, merkt man trotzdem, dass jede computerlinguistische Ausrichtung (sei es texttechnologischer Art oder im Bereich Digitalhumanities noch im Bereich Informationswissenschaft verankert) darauf abzielt, bestimmte Profile von morgigen Computerlinguisten auszubilden. Allerdings erweist sich, dass je nachdem, ob der CL-Absolvent eine Ausrichtung oder eine andere gewählt hat, muss er vor allem **Grundwissen** in Bereichen der Linguistik und Informatik erweisen und dazu noch **Vertiefungswissen** in einem Spezialisierungsbereich wie etwa: Maschinengestützte Übersetzung; Dialogsysteme; Maschinengestützte Lexikographie; Maschinelle Textverarbeitung (Bürokommunikation); Textzusammenfassung (Informationssuche; Informationsextraktion und –Synthese); Computerunterstützter Unterricht. Davon ausgehend könnte also eine überschaubare Klassifikation des Profils des CL-Absolventen eher anhand der Natur des erworbenen Vertiefungswissens, denn alle CL-Absolventen eines Spezialitätsbereichs immer das gleiche Grundwissen erweisen.

Alles in Allem lässt sich also feststellen, dass Die Verzeichnung **der eigenständigen Dynamik**, die die Computerlinguistik mit sich bringt, sich nicht ganz deutlich anhand der institutionellen Gestaltung jeder Universität (Strukturierung des Fachbereichs) veranschaulichen lässt. Daher werden wir einen anderen Ansatz versuchen, indem wir diese Dynamik eher unter **den Perspektiven des Vergleichs von Theorien und Methoden, die in den zwei zuvor erwähnten Hauptgebieten der Computerlinguistik (Theorie und Praxis) angewendet werden.**

⁴⁹ [...] Welche Inhalte in welcher Form aus welchen Veranstaltungen des Rahmens für einen Diplomstudiengang (vgl. Kap. 3) gewählt werden, hängt einerseits von den örtlichen Gegebenheiten (vor allem von den Forschungsschwerpunkten) ab und andererseits von dem gewählten Hauptfach. CL ist als Nebenfach zu einem Hauptfach Linguistik bzw. sprachwissenschaftlich ausgerichtete Einzelphilologie anders zu strukturieren als Nebenfach zu einem Hauptfach Informatik.[...] (vgl. Udo Hahn, Matthias Heyn, Ludwig Hitzemberger Bernhard Kelle, Hans-Dieter Lutz, Klaus Wothke 1988: 87).

1.2. Methodische und theoretische Grundlage der Computerlinguistik

Wie oben schon demonstriert, lässt sich **die eigenständige Dynamik, die die Computerlinguistik** mit sich bringt, nicht nur am Beispiel der Gegenstandsentwicklung, sondern auch am Beispiel der institutionellen Gestaltung der Studien- und Forschungsorten erläutern. Nun beabsichtigen wir in diesem Abschnitt mit der Erläuterung **dieser Dynamik** fortzufahren, indem wir auf einige in diesem Forschungsgebiet eingesetzten **Theorien und Methoden** fokussieren. Dabei wollen wir uns hauptsächlich dafür einsetzen, zu zeigen, dass die theoretischen und methodologischen Vorgehensweisen, die in dieser Disziplin angewendet sind, sind keine schlichte Entlehnung von Theorien und Methoden aus den Nachbardisziplinen, wie etwa der Linguistik oder noch der Informatik,⁵⁰ sondern vielmehr eine eigenständige Farbe erweisen. Dazu schreiben nämlich **Kai-Uwe Carstensen, Susanne Jekat und Ralf Klabunde** Folgendes:

Die Computerlinguistik ist aber nicht nur ein Abnehmer linguistischer Theorien und Sachverhalte, sondern sie kann auch ein Stimulus für Erkenntnisgewinn und die Erarbeitung neuer Ansätze [...] sein (Carstensen, S. Jekat und R. Klabunde 2010: 3).

Gehen wir davon aus, dass die Computerlinguistik grundsätzlich aus zwei Hauptteilen besteht: **Der theoretische Teil**, der, wie seine Bezeichnung schon Auskunft gibt, ein Teil ist, in dem nur Ansätze bzw. Modelle zur Erklärung/Beschreibung eines Sachverhalts elaboriert werden, welche dann letztendlich eine Anwendung in **dem praktischen Teil** finden werden. Dazu müssen wir jedoch hinweisen, dass eine schlichte Differenzierung zwischen den beiden anerkannten Hauptteilen, mittlerweile nicht mehr funktionsfähig bzw. Aktuell ist. Denn es wurde nämlich im Laufe der Zeit beobachtet, dass manche Phänomene (**Forschungsgegenstände beliebiger Untersuchungen**) stark technisch gebunden sind und lassen sich somit dann völlig nur in der sogenannten **praktischen Computerlinguistik** abarbeiten (vgl. Ebd.:6). In dieser Hinsicht sehen sich die Grenzen zwischen geisteswissenschaftlichen und sprachtechnologischen Perspektiven umrunden. Denn, obwohl

⁵⁰ Das Hauptziel der CL ist es die Sprache zu beschreiben und dieser Beschreibung dem Computer zugänglich zu machen. Dabei werden methodische und theoretische Prinzipien der Linguistik und Informatik meistens eingesetzt. Was die Linguistik angeht, werden meistens Methoden und Erkenntnisse aus empirischen Arbeiten miteinbezogen; Theorien und Modelle zur Phonologie, Morphologie, Syntax, Semantik, Pragmatik werden auch ausgewertet. Was die Informatik angeht, werden sowohl rein mathematische als auch rein informatische Prinzipien gebraucht. Davon unterscheiden wir: Algebra, Graphentheorie, formale Logik, Modelltheorie, Statistik für die Mathematik und Rechnerarchitektur, Compilerbau, Betriebssysteme, Programmiersprachen, Software Engineering, Datenstrukturen, Algorithmentheorie; **Theorie formaler Sprachen, Automatentheorie u.a. für die** für die Informatik (vgl. Zur Konturierung des Faches Computerlinguistik Udo Hahn, Matthias Heyn, Ludwig Hitzenberger Bernhard Kelle, Hans-Dieter Lutz, Klaus Wothke LDV-Forum Bd. 5, Jg. 1987/88, Nr. 4, September 1988)

die Abarbeitung solcher Gegenstände stark technologisch verläuft, werden trotzdem Theorien bzw. Ansätze entwickeln, um an das Phänomen ran zu kommen. Gemerkt wurde also, dass die Erklärung bzw. Beschreibung solcher Gegebenheiten anhand neuer erstellten Theorien erfolgt, die keinerlei ihrer Entstehung der sogenannten **theoretischen Computerlinguistik** zu danken haben, sondern vielmehr der sogenannten **praktischen Computerlinguistik**. Denn diese neuen erstellten Theorien sind stark technologisch orientiert, weil sie zumindest auf technologische Grundstrukturen basieren, etwas, dass nicht der Fall bei den anderen Kategorien von Theorien, nämlich bei denen, die innerhalb *der theoretischen Computerlinguistik* entstanden sind (Udo Hahn 2019:28-30).

Als direkte Folge und konkrete Illustration einer solchen Wende gilt beispielsweise, die Tatsache, dass der Abruf von Theorien aus Nachbardisziplinen, wie etwa der Linguistik, der Psycho- und Neurolinguistik, u.a., nicht mehr zur Verarbeitung computerlinguistischer Phänomene unabdingbar waren. Denn es herrschte vielmehr ein Paradigma der selbst entwickelten Methoden und Theorien. Dazu schreibt nämlich Hinrich Schütze im Rahmen eines seiner im Jahre 2019 gehaltenen Vorträge Folgendes:

*Computational linguistics is the scientific study of models and methods for automatic processing of natural language. Computational linguistics is an interdisciplinary field that shares a large part of its subject matter with computer science and linguistics. However, computational linguists also work **on theories, models and methods** that are not part of core linguistics or core computer science. (Hinrich Schütze WS. 2018-2019: 15)*

Von diesem Zitat von H. Schütze ausgehend, lässt sich auch außerdem merken, dass die in der Computerlinguistik eingesetzten Theorien und Methoden keine schlichte Zusammensetzung von Methoden und Theorien aus verwandten Disziplinen sind,⁵¹ sondern vielmehr eine Art

⁵¹Bei computerlinguistischen Beschreibungen werden beispielsweise **linguistische Grammatikmodelle** und **formale Grammatikmodelle** der Informatik unter einem Hut gebracht. Eine andere Kombinationsmöglichkeit wäre zum Beispiel, wenn **Wissensrepräsentationssprachen**, wie etwa die DRT (Linguistik und Logik) zusammen mit semantischen **Repräsentationsformalismen** (Informatik) verarbeitet werden. Eine andere Illustration dieser Zusammensetzung von linguistischen und informatischen Methoden zeigt sich im Bereich der automatischen Verarbeitung von syntaktischen Verfahren (Parsing); dabei werden Automatenmodelle der Informatik eingesetzt. Außer den Bereichen **Informatik** und **Logik** werden auch theoretische und methodische Prinzipien aus dem Bereich **Mathematik** eingesetzt. Es werden nämlich probabilistische Techniken zur Unterstützung unterschiedlicher Formen des maschinellen Lernens (sogenannte **statistische Verfahren**) (Udo Hahn 2019:28-30)

Anpassung an gegebenen computerlinguistischen Realitäten/Einschränkungen⁵². Daher lässt sich die Computerlinguistik manchmal als Innovationstreiber für die Fächer, die einen engen Zusammenhang mit ihr zeigen. Dies ist beispielweise der Fall mit der Programmiersprache PROLOG. Das Werkzeug wurde innerhalb der Computerlinguistik bzw. für die Lösung computerlinguistischer Probleme entwickelt und fand letztendlich seine Nutzung sowohl bei der Modellierung der Sprache (**Informatik**) als auch beim Beweisen von Theorien (**Logik**) (vgl. Ebd.: 20).

Alles in allem unterscheidet man hauptsächlich zwei Methodengruppen: **Die regelbasierten Vorgehensweisen und die statistischen Verfahren** (vgl. Kai-Uwe Carstensen & Co, 2010: 168-328). Mit der Regelbasierten Vorgehensweise ist es die Rede, wenn man zur Lösung eines computerlinguistischen Problems Algorithmen aufbaut, die auf **Heuristiken** stützt, die stakt auf Regel der Sprache (Theorien/Theorie geleitet) basiert sind. In dieser Hinsicht unterscheidet man je nach der Natur des verarbeiteten Phänomens unterschiedlicher Heuristiken (Die Heuristiken, die zur Erlösung eines phonologischen Problems gebraucht werden, können nicht bei der Erlösung eines syntaktischen Problems wiederaufgenommen werden)⁵³. Solche Verfahren werden meistens in der sogenannten theoretischen Computerlinguistik gebraucht; denn sie sind theoriegeleitete Vorgehensweisen und werden somit nur in dem Teilbereich behandelt, in dem man mit Algorithmen handelt, die *auf absoluten Generalisierungen* mit Hilfe Heuristiken basieren. In dem anderen Teil der Computerlinguistik, dem Praktischen CL, ist es vielmehr die Rede von dem Aufbau von Algorithmen auf Basis von statistischen Verfahren. In diesem Fall wird der Algorithmus nicht von Anfang an auselaboriert, sondern wird so programmiert, dass er sich selbst im Laufe der Zeit verbessert, und dies Je nach der Art/ Natur, der Qualität und Quantität der Daten, die er während seines Gebrauchs verarbeitet (vgl. Ebd.: 168-328). Das ist

⁵²Zu solchen Einschränkungsmodellen sagt Hinrich Schütze (2019) zum Beispiel Folgendes: *Die Ingenieurwissenschaft hat das Problem des Fliegens gelöst, indem sie das Flugzeug konstruierte, ohne **das biologische Modell des Vogelflugs** nachgeahmt zu haben. Ähnlich lässt sich das Problem der maschinellen Übersetzung in speziellen Bereichen lösen, ohne jemals **einen menschlichen Übersetzer mit allen seinen Fähigkeiten** simulieren zu können. So wie das Flugzeug in einigen Eigenschaften die biologischen Modelle weit übertrifft (etwa in der Geschwindigkeit), so kann ähnliches auch für computerlinguistische Verfahren (etwa bei der Übersetzung von Massentexten) zutreffen* (Hinrich Schütze WS. 2018-2019: 25).

⁵³Für jede sprachliche Beschreibungsebene gibt es passende Methoden: In **der Phonetik** werden beispielsweise Verfahren zur Spracherkennung (Rezeption) oder zur Sprachsynthese (produktiv) gebraucht. In **der Morphologie** werden Verfahren zur Wortzerlegung, Wortartenbestimmung (bei solchem Verfahren redet man von **Tagging**). In **der Syntax** werden Verfahren zur automatischen syntaktischen Analyse (man redet auch von **Parsing**). In der **Semantik** werden Verfahren zu Wissensdatenbanken und zur automatischen semantischen Analyse entwickelt. Und in der Pragmatik werden Verfahren zur **Koreferenzresolution** und Kontextmodellierung u.a. entwickelt. Zu merken ist jedoch die Tatsache, dass diese Verfahren, je nach Zielsetzungen der Aufgabe, entweder **regelbasiert** oder **statistisch** aufgebaut werden können (vgl. Kai-Uwe Carstensen & Co, 2010: 168-328).

meistens das Verfahren, das wir während dem Aufbau von computerlinguistischen Systemen und Anwendungen (Tools) beobachten, die diese Letzten integrieren (**Detmar Meurers 2009: 17**). In dieser Hinsicht stellen wir also fest, dass während regelbasierte Methoden eine Art richtige Vorhersage prognostiziert (**Linguisten sehen die Zukunft vorher!**), basieren statistische Methoden eher auf Schulungen des Algorithmus (**Statistik lernt aus der Vergangenheit für die Zukunft!**).

Die geschichtliche Entwicklung des Gebrauchs von Methoden zeigt uns, dass **statistische Verfahren** ganz kurz am Anfang der Geschichte zur maschinellen Sprachverarbeitung eingesetzt wurden (vgl. Udo Hahn, Matthias Heyn & Co, 1988: 88-90). Wie schon zuvor erwähnt, war die Übersetzung von großen Mengen von Texten in einer sehr begrenzten Zeit die erste Notwendigkeit, die dazu geführt hat, dass man sich für die maschinelle Sprachverarbeitung interessierte. Allerdings wurde es bald gemerkt, dass der Gebrauch von solchen **Prozessen (Maschinellen Übersetzungssoftware, die auf statistische Verfahren basieren)** nicht genug effizient waren, denn die damalige Hardware (Computer) nicht genug/optimal ausgerüstet waren um große Menge von Textdaten zu verarbeiten (vgl. Ebd.: 88-90). Daher merkte man, schon kurz nach dem Anfang der maschinellen Übersetzung, etwa ab den 60er Jahren, ein Abkehr von statistischen Methoden zur Gunst von **regelbasierten Methoden**. Diese Letzten wurden eben für eine lange Zeit gebraucht und erst ab dem 21. Jahrhundert wurden sie abgeschafft. Denn es wurden der **wissenschaftlichen Community** Rechner bereitgestellt, die hohe Performanz zeigen (Arbeitsspeicher; Prozessor). Ab diesem Moment stellt man eine Rückkehr **von statistischen Verfahren** (vgl. Bernhard Fisseni 2010: 87).

Allerdings erweisen auch solche Verfahren außer ihren Vorteilen auch einige Nachteile. Es wurde nämlich festgestellt, dass, im Gegensatz zu regelbasierten Methoden, ist die Probabilität korrekte Ergebnisse⁵⁴ zu erzielen weniger gut, denn ihre Aufbau basiert auf pragmatische Vorgehensweise (d.h. Systeme werden anhand konkreter Situationen geschult und nicht auf

⁵⁴Angenommen, dass **die linguistisch regelbasierten und basierten Systeme** meistens, wie schon zuvor erwähnt, darauf abzielen, das untersuchte Phänomen zu verallgemeinern, können wir sagen, dass sie bei der Lösung einer bestimmten Aufgabe besser abschneiden werden als die **probabilistischen Verfahren**. Anders ausgedrückt, will man beispielweise eine einzige und selbe Aufgabe durch die beiden Methoden auflösen. Dann muss man zunächst, was die theoriegeleitete Vorgehensweise angeht, alle Parametern zur Auflösung dieser Aufgabe im Kauf nehmen, um dann später eine Anfrage auf das entsprechende Regelsystem zu starten. Im Gegensatz dazu haben wir beispielweise bei den statistischen Verfahren eher eine Umschulung des Systems, welches dann auf Reflex die gestellte Aufgabe lösen soll. Und was wir sagen ist, dass die **Probailität bessere Ergebnisse** zu haben, lässt sich dann besser bei der Anwendung von regelbasierten Methoden, und dies sofern die Erlösung des angestrebten Phänomens zuvor programmiert wurde; andersfall ist es besser die statistischen Methoden zu gebrauchen.

Basis schon vorprogrammierten Heuristiken gelaufen). Das ist zum Beispiel der Fall von einem **Information Retrieval-System (z. B. Google)**. Es wird in der Tat bei einer Suchanfrage auf Google sowohl relevante als auch irrelevante Dokumente angezeigt. So gesehen sind solche Systeme nur in eingeschränkter Masse vorhersehbar. In dieser Hinsicht schreibt zum Beispiel Bernhard Fisseni (2010) Folgendes:

[...] Das soll nicht heißen, dass man es in sprachtechnologischen Anwendungen unterlässt, die Fehleranzahl zu minimieren; es ist jedoch an vielen Stellen nicht zu vermeiden, dass Fehler auftreten, und die Systeme sind so konstruiert, dass eine gewisse Anzahl Fehler toleriert wird (Bernhard Fisseni 2010: 87).

Zusammengefasst können wir also sagen, dass **Sprachdaten-orientierte (statistische Verfahren) und Theorie-orientierte CL-Methoden (regelbasierte Methoden)** grundsätzlich verschieden sind, aber können dennoch zusammen eingesetzt werden. Dies ist dann der Fall, wenn sie zusammen in einer einzigen und selben Anwendung auftreten. Bei solchen Fällen redet man von sogenannten hybriden Ansätzen.

Spätestens seit **Verbmobil** ist es üblich, **Hybrid-Systeme** zu bauen. Darin arbeiten sowohl formallinguistisch motivierte **sog. tief verarbeitende Systeme** als auch informatisch-statistische **sog. flache Komponenten zusammen und gegeneinander**. Wie schon angedeutet, ergänzen sie sich, denn tiefe Verarbeitung ist oft zeitaufwändiger als flache, dafür im Erfolgsfall meist korrekter. Es ist oft vorzuziehen, irgendein Ergebnis zu haben, als keines zu haben - aber warum nicht das mutmaßlich korrekteste nehmen, wenn möglich? **Natürlich können flache und tiefe Komponenten auch kooperieren und einander ergänzen (vgl. Bernhard Fisseni 2010: 88).**

Alles in allem, unterscheidet man, je nach computerlinguistischer Anwendung, unterschiedliche Vorgehensweisen, aber gemeinsam weisen alle diese Anwendungen die folgende Grundstruktur während der Sprachanalyse auf: **Spracherkennung; Tokenisierung⁵⁵; die Morphologische Analyse; Syntaktische Analyse; Semantische Analyse Dialog- und Diskursanalyse**. Man redet auch von dem **Pipelinemodell (Sprachlicher Input – zur Bedeutung)**(Nitin Indurkha, Fred J.Damerau, 2010).

1.3.Computerlinguistische Systeme und Ihre alltäglichen Anwendungen

Nachdem wir in den vorherigen Unterkapiteln konzeptuelle, theoretische und methodische Prinzipien der CL erläutern haben, beabsichtigen wir in diesem Abschnitt, die CL unter einer

⁵⁵**Tokenisierung** Die Buchstabenkette wird in Wörter, Sätze etc. segmentiert.

alltäglichen Perspektive darzustellen. Dafür werden wir verschiedene Anwendungen beschreiben, die ohne Erkenntnisse der CL nicht zum Stande gekommen wären.

Chronologisch betrachtet, lassen sich die maschinellen Übersetzungssysteme (MÜ) als erste Generation von computerlinguistischen Anwendungen verzeichnen. Sie wurden zum ersten Mal in den späten 50er Jahren aus geopolitischen Motivationen in den USA entwickelt. Gefordert und gefördert wurde eine schnelle und vollautomatische Übersetzung zwischen dem Russischen und dem Englischen (**Entstehung der Computerlinguistik**). Zwanzig Jahre später wurden **Dialogsysteme** (zweite Generation von CL-Systemen) entwickelt. Als berühmte Dialogsysteme zählen beispielsweise **ELIZA (1966)**, von Joseph Weizenbaum, und **SHRDLU (1970)** von T. Winograd. Während das erste mit einfachen **Mustervergleichs-Techniken (Pattern Matching)**, also ohne den Einsatz von Wissen arbeitet, ist das Zweite ein **wissensbasiertes System**. Im Unterschied zu ELIZA verwendet SHRDLU bei der Dialogführung unterschiedliche Arten von Wissen (Sprachliches/linguistisches Wissen; Kontext- und Weltwissen) (vgl. Kai-Uwe Carstensen & Co, 2010: 642-656). In den letzten fünfzig Jahren sind ganz viele Anwendungen/Systeme entwickelt worden. Ihre Ausbreitung geht von der intelligenten Textverarbeitung (Textklassifikation, Textextraktion, Textzusammenfassung, Textmeinung usw.) über textbezogenen Informationstechnologien bis hin zum computerunterstützten Sprachenlernen.

1.3.1. Texttechnologische⁵⁶ Anwendungen

Bei texttechnologischen Anwendungen handelt es sich insbesondere um Systeme, die in der Lage sind, Informationen (oder Textphänomene) in Texten zu identifizieren und sie maschinell, je nach Zielsetzung der Aufgabe, weiterzuverarbeiten. Im Allgemeinen werden solchen Systemen das Ziel zugeschrieben, Informationen/Texte/Dokumente innerhalb von kurzer Zeit und mit einer hoch Genauigkeit aufzufinden, zu extrahieren und zu klassifizieren. Innerhalb von texttechnologischen Anwendungen unterscheiden wir beispielsweise, Anwendungen, die auf Systeme basieren, die nur darauf abzielen Informationen in Texten zu suchen. In solchen Fällen redet man von **Informationserschließungssysteme** (auf Englisch Information-

⁵⁶**Die Texttechnologie** ist ein Teilgebiet der Computerlinguistik, das sich der Verarbeitung geschriebener Sprache widmet. Dazu gehören u.a. folgende Anwendungsfelder: Maschinelle Übersetzung; Machine Learning für NLP (Natural Language Processing); Dialogsysteme (Chatbots); Medieninhaltsanalyse (Media Content Analysis); Informationsextraktion (Text Mining); Sentiment-Analyse (Opinion Mining); Automatische Textvereinfachung und Textzusammenfassung; E-Accessibility; Texttechnologie für historische Texte (Digital Humanities) <https://www.cl.uzh.ch/de/texttechnologies/what-is-cl.html> (12.02.2022 um 17h).

Retrieval (IR))^(vgl. Kai-Uwe Carstensen & Co, 2010: 576-614). Neben dieser Gruppe von Anwendungen unterscheidet man auch eine andere Familie von texttechnologischen Anwendungen, die darauf abzielen, Texte in großen Mengen von Texten (Dokumenten) zu suchen. Bei solchen Fällen redet man von **Texterschließungssystemen** (auf Englisch *Textretrieval-Systems*) **oder von Dokumentenzugriffssystemen** (*Dokumentenretrieval-Systems*). Als weltweit bekanntestes Informationserschließungssystem gilt GOOGLE. Diese texttechnologische Anwendung ermöglicht den Benutzer bestimmte Informationen (**Suchtermini**) in einem Häufen von Dokumenten (**Indextermini**) aufzufinden. In der Tat können GOOGLE⁵⁷ und alle anderen Suchmaschinen, wie being; Yahoo, oder noch msn etc. Texte auffinden, in denen die Suchtermini auftauchen, deswegen werden sie auch als Texterschließungssysteme aufgezeichnet. In der Tat wird eine Suche als erfolgreich aufgezeichnet, wenn es dem Benutzer gelingt indexierte Termini als Suchanfrage zu gebrauchen; denn bei der Aufzeichnung von Dokumenten (Texten) in sprachlichen Datenbanken werden nur einige Wörter oder Ausdrücke (Meistens Titels oder häufig im entsprechenden Text gebrauchte Ausdrücke) indexiert. Andersfalls wird die Suche nicht erfolgreich sein; und der Benutzer könnte zum Beispiel glauben, die/der gesuchte Information/Text nicht im Netz vorhanden ist, und dies obwohl diese/dieser tatsächlich unter einem anderen Suchterminus⁵⁸ abzurufen ist. Viel spezifischer können auch Anwendungen zur Informations- und Texterschließung aussehen. In der Tat könnte eine Firma für ihre Angestellten ein texttechnologisches System bestellen, das dafür eingesetzt wird, beispielweise auf dem Rechner

⁵⁷ Die gleichen Überlegungen gelten auch für YOUTUBE und alle anderen Videos- und Audiosuchsysteme. Hier werden aber Videos/Audios gesucht, aber allerdings immer mit einer Suchanfrage, welche auch eben auf sprachliche Angabe angewiesen ist. Es ist auch hier die Stelle zu erörtern, dass diese sprachlichen Angaben entweder schriftlich oder mündlich also akustisch erfolgen können. In Einzelheiten kommen wir gleich während der Verarbeitung von Spracherkennung-; Sprachsynthese-; und Dialogsysteme zurück.

⁵⁸ Ich will nur an dieser Stelle am Beispiel einer persönlichen Erfahrung darauf hindeuten, wie wichtig solche Kenntnisse über das Funktionieren von Suchmaschinen bei der Dokumentsuche ist. In der Tat hat mir mein Betreuer, als er merkte, dass ich im Gegensatz zu allen anderen seiner betreuten Studierenden keine Sorge über den Zugang zu Dokumenten ausdrückte (ich fand, in der Tat, auf Internet fast alle Dokumente, die ich für die Forschung brauchte) aufgefordert, ein Seminar in unserer Forschungsgruppe darüber zu führen. Ich nutzte mir dann die Gelegenheit meinen Forschungskollegen zu erklären, wie man mit Suchtermini in Suchanfragen umgehen soll. Ich erklärte Ihnen nämlich, dass diejenigen, die Schwierigkeiten bei der Informationssuche haben, sind diejenigen, die bei der Suchanfrage keine **indexierten Termini** eingeben. Und sie fragten mich zurück, wie sie an indexierte Termini kommen könnten. Und ich schlug ihnen vor eine erste Suchanfrage mit einem beliebigen potentiellen Terminus, der für die gesuchte Information eingesetzt werden könnte. Und beim Durchlesen der Suchergebnisse darauf aufmerksam sein, wie bzw. mit welchen sprachlichen Ausdrücken die gesuchte Information häufig vorkommt. Und erst nachher eine neue Suchanfrage starten und dieses Mal aber mit dem neuen Terminus, welcher höchstwahrscheinlich (auf etwa 80%) der indexierte Terminus für die gesuchte Information sein könnte. Am Ende des Seminars waren sie alle dankbar und sagten, sie hatten keine Ahnung über solche Suchmechanismen gehabt.

der Sekretärin, unnötige E-Mail-Meldungen, (wie etwa Werbungen) zu filtern. Man redet bei solchen Fällen von **Informationsfiltrierungssystem**⁵⁹ (vgl. Ebd: 576-614). Solche Informationsfiltrierungssysteme können entweder auf Stichworterkennung basieren oder eher auf Textmining funktionieren. **Das Textmining** ist eine Technik, die von Computerlinguisten entwickelt wurde, um den inhaltlichen Gehalt von Texten zu extrahieren. Gezielt wird beispielsweise zu wissen, ob ein Text positive oder negative Bewertungen ausdrückt. Solche Systeme werden meistens bei Firmen gebraucht, die, beispielsweise, große Menge von Feedbacktexten ihrer Kunden analysieren wollen und somit ganz schnell erschließen, wie sie ihre Dienste verbessern können (Müller 2013: 34).

Allerdings, lässt sich jedoch die Tatsache merken, dass je nach der Art des Bedürfnisses einer computerlinguistischen Anwendung werden von Computerlinguisten angepasste Lösungen entwickelt (Computerlinguistik als Ingenieurwissenschaft). Wir werden hier, im Rahmen dieser Dissertation, nicht alle diesen potentiellen Anwendungen bzw. Systeme erläutern können. Denn dies wäre weder möglich noch zielorientiert. Allerdings müssen wir darauf hinweisen, dass **der Algorithmus bzw. das AR-System**, das wir zu aufbauen beabsichtigen, dazu beitragen wird, dass Computerlinguisten beim Aufbau von **Lösungen zum Textverständnis, -Zusammenfassung, -Klassifikation** (sowie auch beim Aufbau aller anderen Lösungen, in denen es um die maschinelle Extraktion der Bedeutung eines Textes⁶⁰ geht), eine breite Auswahl an Tools/Werkzeuge zur Verfügung besitzen. Und dies ist umso wichtiger, denn jede CL-Lösung ist eigenartig. Das heißt, der Computerlinguist wird bei einer Problemstellung nicht immer mit den gleichen Teilsystemen arbeiten, die er sich zuvor bedient hat. Er wird sich stattdessen die für die Problemstellung bestmöglichen Teilsysteme aussuchen und diese dann einsetzen, um eine bestmögliche effiziente Lösung zu entwerfen.

Wirft man nun einen geübten Blick auf den alltäglich persönlichen Gebrauch von Rechnern und Handys, dann merkt man, dass **texttechnologische Anwendungen** ständig präsent sind.

⁵⁹**Informationsfiltrierungssystem** können auch bei Firmen, die auf enormen Eingangsmails ausgesetzt sind, eingesetzt werden. Ziel eines solchen Systems wäre, beispielsweise, diese Mails an die richtige Firmenabteilung automatisch weiterzuleiten. Eine weitere Ausprägung solcher Systeme lässt sich, zum Beispiel, anhand des Funktionierens des SPAM-Filters in E-Mail-Diensten (wie Gmail oder noch Yahoo) noch veranschaulichen.

⁶⁰ Der Algorithmus und das daraus ableitende System, das wir zu entwickeln beabsichtigen, werden dafür eingesetzt, Texte auf semantische Vorgehensweise zu analysieren. Dem Computer wird nämlich das Verständnis von Texten ermöglicht, **indem semantische Verknüpfungen, die durch anaphorische Ketten entstehen, aufgelöst und somit verständlich gemacht werden**. Darüber hinaus könnte auch dieser Algorithmus bzw. dieses AR-System als Komponente eines großen Systems zur Identifizierung und Extraktion von bestimmten Informationen in Texten behilflich sein.

Das ist beispielsweise der Fall, wenn man Texte in einer Editoranwendung, wie etwa (Wordpad, Microsoftword oder Ähnliches) tippt. Einem wird meistens **Rechtschreibung- und Grammatikkorrekturen** automatisch angeboten. Betrachtet man weiterhin das Schreiben von SMS auf Handy, dann merkt man, dass bevor ein Wort zu Ende geschrieben wird, bittet das System mögliche passende Wörter (solche Gegebenheiten funktionieren auf Basis von **Lexikographie-Systemen** und diese Letzten sind meistens auf statistische Verfahren angewiesen; d.h. Je mehr ein Wort schon in einem gegebenen sprachlichen Kontext eingesetzt worden ist, desto wahrscheinlicher ist die Chance, dass das Lexikographie-System es als potentielles Schreibziel in einem selben oder ähnlichen Kontext anbietet. Weiterhin ist es üblich für jeden Smartphon-Besitzer ein Wörterbuch-App auf dem Handy zu haben. Solche Anwendungen und alle anderen auch, in denen die Sprache zusammen mit rechnerischen Methoden verarbeitet werden, sind Früchte der Computerlinguistik. Dazu gehören u.a. Maschinelle Übersetzungssysteme (MÜ); Question-Ansering-Systeme; Chatboots/ Dialogsysteme (**Carstensen, S. Jekat und R. Klabunde 2010: 462**).

1.3.2. Dialogsysteme (Chatboots)

Von Dialogsystemen oder Chatboots ist die Rede, wenn Befehle oder Fragen den Maschinen (nicht nur Computer oder Handys aber auch Roboter und alle anderen Apparaten, die mit einem Spracherkennungssysteme gerüstet sind) gegeben werden kann und diese Letzten letztendlich darauf reagieren oder beantworten können. Natürlich-sprachliche Dialogsysteme erlauben also Kommunikation zwischen Computer und Menschen mittels menschlicher Sprache. Diese Kommunikation kann entweder einseitig oder beidseitig geführt werden (vgl. Kai-Uwe Carstensen& Co, 2010: 624-632). Bei dem ersten Fall redet man von **Kommandowortsystemen** (Einseitige Steuerung von elektronischen Geräten durch sprachliche Äußerungen deren Benutzer). In dieser Kategorie befindet sich die sogenannten Question-Answering-Systeme). Und bei dem zweiten Fall von **Conversational User Interface (CUI)** (Dialoge ist von Menschen und Maschine gesteuert (mixed-initiative dialogue). In dieser Kategorie befindet sich die neulich entwickelte Systeme wie etwa das deutsche System SIRI von der T-Mobile; das Google-Smartphon-Assistent PIXEL und ALEXA ein Dialogsystem, das von Amazon entwickelt worden ist (vgl. ebd.: 624-632).

Solche Dialoge Systeme haben immer als Teilkomponente **Spracherkennung- und Sprachgenerierungssysteme**. Von Spracherkennungssysteme (Speech Recognition System)

ist es die Rede, wenn von dem akustischen Signal gesprochener Sprache eine textuelle Darstellung erzeugt wird. Sie wurden ursprünglich erzeugt, um Computern Befehle geben zu können. Somit zielen sie u.a. auch darauf ab, dem akustischen Signal von gesprochener Sprache eine (partielle) Bedeutung zuzuschreiben. Bei solchen Systemen handelt es sich also darum, Computer die Fähigkeit zu geben, gesprochene Sprache zu erkennen, zu verarbeiten und zu interpretieren (vgl. ebd.: 624-632). **Sprachsynthesen- oder Sprachgenerierungssysteme** ihrer Seite haben für ursprüngliche Funktion, ein akustisches Signal aus einer Zeichenkette (Text) zu erzeugen. (Im Englischen redet man von Text-To-Speech Systemen TTS). Bei TTS-Systemen handelt es sich also darum, die Erforschung und Bereitstellung von Lösungen, die Maschinen dabei helfen werden, geschriebene Sprache (in Form von Texten) in gesprochene Sprache automatisch zu umwandeln. Alltäglich betrachtet, lassen sich die beiden Systeme in verschiedenen Anwendungen einsetzen: von der **Mobiltelefonie und Informationssysteme** über die Anwendungen für kognitiv beeinträchtigte Personen, wie Seh- und Schwerbehinderte bis hin zur Sprachdidaktik sind ihre Einflüsse spürbar. In der **(Mobil-)Telefonie und Informationssysteme** wie beispielweise **Auskunftssysteme (Tourismus, Kinoprogramm, Wetter, Börsenkurse usw.) und Navigationssysteme im Auto, E-Mail-Vorlesesysteme im Auto usw.)** werden natürlich sprachliche Texte automatisch von Maschinen generiert (vgl. Michael Hess.WS. 2004/2005: 55). Weiterhin werden beispielweise mit Hilfen von den beiden Systemen Anwendungen für Sehbehinderte entwickelt, die ihnen den Umgang mit Computer erleichtern sollen, indem alles worauf die Mause steht, sprachakustisch umgewandelt wird. (Dies ist beispielweise der Fall bei Anwendungen wie Talback und VoiceOver). Immer in dieser Richtung blickend, werden beispielweise Anwendungen entwickelt, die künstliche Stimme für Sprechbehinderte ermöglichen⁶¹.

1.3.3. CL& DIDAKTIK

Die CL kann unteranderen gesellschaftlichen Anwendungsgebieten auch bei der Sprachdidaktik behilflich sein. Je nach der Ausprägung dieser Hilfe kann man von CALL (Computer Assisted Language Learning); ICALL = Intelligent CALL oder von **Intelligenter Tutorieller Systeme (ITS)** reden (vgl. ebd.: 59). Während der Lerner im Rahmen des CALL-Systems selbst bestimmt was, wann und wie er lernen möchte, ist es bei den ICALL und IT-Systemen, der Computer, der bestimmt, einerseits, wie der Kurs durchgearbeitet werden soll,

⁶¹<http://seattlepi.nwsource.com/business/217802grammar28.asp> (29.01.2022)

und andererseits erkennt, lokalisiert und kategorisiert die Fehler der Lernenden um darauf entsprechend reagieren zu können (Tutor). Dafür werden beispielsweise RECHTSCHREIB- UND GRAMMATIKKORREKTUR-Programme, Diktier-Systeme und andre Systeme als Teilkomponente eingesetzt (vgl. France Santi 2015: 23-25).

1.3.4. CL & WIRTSCHAFT : Berufsfelder von Computerlinguisten

Der Gebrauch computerlinguistischer Erkenntnisse in wirtschaftlichen Sektoren ist nicht mehr zu beweisen. Webportale für Service, Verkauf und Handel, zum Beispiel, werden auch auf Basis computerlinguistischen Erkenntnisse elaboriert und programmiert (vgl. Simon. C 2007: 17).

Dies ist zum Beispiel der Fall, wenn Webdokumente, Angebote und Produktbeschreibungen so mit texttechnologischen Anwendungen aufbereitet werden, dass der Kunde sie schnell auffinden kann. Außerdem merkt man auch in manchen wirtschaftlichen Sektoren, wie die Medizin (am Beispiel von teilautomatisierten Analysen von Patientenreports) und die Versicherung (am Beispiel von teilautomatischen Beantwortungen), dass es der Trend wird, Arbeitsprozesse zu digitalisieren. Was die möglichen Berufsfelder von Computerlinguisten angeht, wollen wir zuerst das folgende Zitat von Clematide (2006/2007) erläutern:

Große Softwarehäuser sowie Verlage beschäftigen Computerlinguisten, die an der Erschließung der Sprachdatenbestände (z.B. für elektronische Taschenwörterbücher) sowie an der Übersetzung oder Verarbeitung von Dokumenten arbeiten. Interessante Arbeitsmöglichkeiten bieten auch Universitäten in der gesamten Welt, an denen die Computerlinguistik mittlerweile einen festen Platz einnimmt. (Simon Clematide WS 2006/2007: 17).

Von diesem Zitat ausgehend, ergibt es sich, dass Computerlinguisten insbesondere in allen Bereichen arbeiten können, in denen die Aufbereitung sowohl von gesprochener als auch von geschriebener Sprache als Ressourcen verlangen ist. Dies ist beispielsweise der Fall bei Verlagshäusern, Übersetzungsbüros, bei Verwaltungen von Ämtern und in akademischem Bereich (Lehren & Forschen). Zusammengefasst stellt man also fest, dass Firmen, die sprachtechnologischen Software entwickeln, vor einer großen Herausforderung stehen: **Die Strukturierung und Steuerung des digitalen menschlichen Wissens.** Und diese Herausforderung wird man nur erheben können, indem man Kenntnisse im Bereich der Computerlinguistik ständig produziert.

Zusammenfassend ging es in diesem Teil die relevanten Fachkenntnisse über die Forschungsgebiete Computer- und Textlinguistik zu erläutern. Das nächste Kapitel thematisiert die Textlinguistik und die Anaphorik.

KAPITEL 2: TEXTLINGUISTIK UND DIE ANAPHORIK

Das Hauptanliegen dieses Kapitels ist es, das im Mittelpunkt dieser Dissertationsarbeit stehende Sprachphänomen in seine Forschungsdomäne einzubetten. In der Tat, handelt es sich, in diesem Kapitel, das Phänomen der **ANAPHORIK** als ein Gegenstand der Textlinguistik einzuordnen. Hierfür wird also zunächst der Begriff *Textlinguistik* erläutert und zwar indem, einerseits, eine forschungsorientierte historische Entwicklung der Textbegriffsauffassungen wiedergegeben wird. Und andererseits, indem die Begriffe *Textgrammatik*, *Textkohäsion* und *Textkohärenz* in der Reflexion miteinbezogen werden (Abschnitt 2.1 & 2.2). In Anschluss daran wird dann das *textlinguistische* bzw. *textgrammatische* Phänomen **ANAPHORIK** ausführlich erläutert: Von den Anfängen der wissenschaftlichen Untersuchungen zu diesem Phänomen über die verschiedenen Arten von anaphorischen Verweisen (Abschnitt 2.3) bis hin zur speziellen Verarbeitung des Anapher-Typs **PRONOMINALADVERBIALANAPHER** (Gegenstand dieser Arbeit).

2.1. Zu Begriffen Text und Textlinguistik

2.1.1. Der Text

Etymologisch betrachtet, stammt das Wort **Text** aus dem lateinischen Namen *Textum* (Gewebe), welches aus dem lateinischen Verbs *texere* (verkneten, etwas kunstvoll zusammenfügen) abgeleitet wurde (vgl. **Wiktionary**⁶²). So gesehen, *kann der Text als ein Gewebe aus kunstvoll verketteten Sätzen*, betrachtet werden (Duden 2003, zitiert nach Schindler⁶³). Doch wissenschaftlich betrachtet, wurde der Begriff **Text** zum ersten Mal als Fachausdruck im Rahmen der *rationalistischen Hermeneutik* im 18. Jahrhundert verwendet (Gerhard Helbig 2002: 276). Erst im 20. Jahrhundert wurde er als Erkenntnisinteresse der Sprachwissenschaft eingerückt. Seitdem wurde ihm, je nach dem Ausgangspunkt, unterschiedliche Bestimmungen zugeschrieben. In der *Textlinguistik*, beispielweise, unterscheidet man insbesondere drei Textbegriffe: **einen statischen, einen dynamischen und einen prozeduralen** (vgl. Ebd.: 276-289).

⁶² Auch **Wikwörterbuch** genannt ist das **Wiktionary** ein freies Wörterbuch und ein Wikimedia-Projekt, das am 1. Mai 2004 startete. Es umfasst derzeit 1.029.295 Einträge und zielt darauf ab, die Erfassung des gesamten deutschen Wortschatzes und des aller anderen Sprachen frei zugänglich zu ermöglichen

⁶³<http://wolfgang-schindler.userweb.mwn.de/skripte/Textling10-06.pdf> (Texte) Abruf am (24.10.2020)

Der statische Textbegriff noch als textgrammatische Auffassung gekennzeichnet, ist an sprachliche Strukturen im Text orientiert. Diese Auffassung wurde von Autoren wie **Harweg (1968)**, **Weinrich (1993)** und **Isenberg (1970)** u.a. vertreten. Diese Letzten definieren jeweils den Textbegriff wie folgt:

Ein Text ist eine durch ununterbrochene pronominale Verkettung konstituierter nacheinander sprachlicher Einheiten (Harweg 1968: 148).

Der Text ist eine Folge von Sätzen, die durch Verkettungsmittel miteinander verknüpft sind (Isenberg 1970: 1).

Texte sind sinnvolle Verknüpfungen sprachlicher Zeichen in zeitlich-linearer Abfolge (Weinrich 1993: 17).

Aus diesen Zitaten ergibt es sich, dass Harweg (1968), Isenberg (1970) und Weinrich (1993) den **Text** als eine Folge bzw. Verkettung von sprachlichen Einheiten bestimmen. Außerdem merkt man jedoch divergierende Gesichtspunkte, was die **Verkettungsmittel** angeht. Während Harweg (1968) sich exklusiv auf die **Pronomina** beschränkt, schließt Isenberg (1970) auch alle weitere Verkettungsmittel (**Konnektoren, Rekurrenz und andere Arten der Wiederaufnahme**) ein. Aus der Betrachtung dieser drei Bestimmungen, lässt sich feststellen, dass sie die **sprachsystematisch orientierte Textlinguistik** (erste Entwicklungsphase) als Ausgangspunkt haben. Denn es werden in diesen Bestimmungen nur textinterne bzw. syntaktische Merkmale von Texten berücksichtigt. Etwas, das, bei **der dynamischen Auffassung des Textbegriffs**, nicht der Fall ist (vgl. Kallmeyer u. a. 1974: 45 & Schmidt 1976: 150). In diesem Letzten stehen eher **die semantischen und pragmatischen Merkmale von Texten** im Vordergrund. Dies lässt sich durch die folgenden Zitate von **Schmidt (1976)** und **Kallmeyer (1974)** illustrieren.

Ein Text ist jeder geäußerte Bestandteil eines Kommunikationsaktes in einem kommunikativen Handlungsspiel, der thematisch orientiert ist und eine erkennbare kommunikative Funktion erfüllt, d. h. ein erkennbares Illokutionspotential realisiert (Schmidt 1976: 150).

Text ist die Gesamtmenge der in einer kommunikativen Interaktion auftretenden kommunikativen Signale (Kallmeyer u. a. 1974: 45).

Von diesen Zitaten ausgehend, ergibt es sich, dass der **Text**, einerseits, als **ein Bestandteil einer kommunikativen Handlung** aufgefasst wurde und andererseits als **die Gesamtmenge kommunikativer Signale eines kommunikativen Austauschs**. Während sich Schmidt (1976) jedoch exklusiv auf sprachliche Bestandteile beschränkt, schließt die Auffassung Kallmeyers (1974) auch alle anderen Formen der Kommunikation ein. Kallmeyers Auffassung nach

könnten auch **Bilder und Gesten zu Textbestandteilen gehören**. Weiterhin betonte auch Isenberg die **Handlungsbezogenheit** von Texten. Er begründet in seinem Artikel betitelt „Text versus Satz“ **die Vorzüge einer dynamischen Auffassung gegenüber einer propositionalen Auffassung** und sieht daher den „*Text als zeitliche Abfolge von sprachlich-kommunikativen Handlungen*“ (Isenberg 1977: 144). Somit schließt er sich an die Textauffassung Schmidts (1976) ein.

Zusammenfassend kann behauptet werden, dass die dynamische Textauffassung die kommunikativorientierte Textlinguistik (2te Phase) als Ausgangspunkt hat.

Eben wie die dynamische Auffassung des Textbegriffs, **hat der prozedurale Ansatz** auch die kommunikativorientierte *Textlinguistik* als Ausgangspunkt. Dieser Ansatz, der zum ersten Mal **von Beaugrande und Dressler (1981)** vertreten wurde, definiert den Text als eine „**kommunikative Okkurrenz**“ und geht davon aus, dass ein sprachliches Gebilde als Text bezeichnet wird, nur bei dem Fall es einige Kriterien berücksichtigt, die sowohl einen **sprachsystematischen als auch kommunikativ-pragmatischen Charakter** erweisen. Diese verschiedenen Kriterien der **Texthaftigkeit** wurden als **Textualitätskriterien** genannt. Davon unterscheidet sieben (*Die Kohäsion, die Kohärenz, die Intentionalität, die Akzeptabilität, die Informativität, die Situationalität und die Intertextualität*) (vgl. **Beaugrande und Dressler 1981**). So gesehen, werden alle sprachlichen Gebilde, die nur eines dieser Kriterien nicht berücksichtigt als **Nicht-Text** aufgefasst.

Weiterhin ist es auch notwendig, darauf aufmerksam zu machen, dass der Fokus bei dem prozeduralen Ansatz zunehmend vom Text selbst (als Objekt) auf **die Produktion und Rezeption von Texten** (d.h. in den Kopf des Sprechers und Hörer) verlagert wird (vgl. Helbig 2002: 265). Somit rezensiert man **eine vierte Textauffassung**, die sowohl **prozedural** als auch **kognitiv** motiviert ist. Prozedural, weil die sieben Kriterien der Texthaftigkeit immer von Relevanz sind und kognitiv, weil der Begriff Kohärenz nicht mehr als textinterne Eigenschaft von Texten, sondern als Resultat eines kognitiven Prozesses verstanden wird. Dazu sagt Spooren (1999) nämlich Folgendes:

*Der Zusammenhang in einem Text ist nicht in erster Linie auf linguistische Ausdrücke selbst gegründet, sondern entsteht letztendlich durch **konzeptuelle Verknüpfungen** (Spooren 1999: 187, Zit. nach Gerhard Helbig 2002: 264).*

Ziel der vorliegenden Forschungsarbeit ist es, einen Algorithmus und bei ggf. ein [AR-System](#) zu entwickeln, dem die Zielsetzung geschrieben wird, **direkte Pronominaladverbialanaphern automatisch auflösen zu können**. Von der Tatsache ausgehend, dass anaphorische Konstruktion sich nicht nur rein **syntaktisch** beschreiben lassen, sondern auch **inhaltlich** und manchmal auch **kognitiv** zu analysieren sind, wollen wir im Rahmen dieser Untersuchung eine integrative Bestimmung des **Textbegriffs** gebrauchen. Daher haben wir die folgende Auffassung K. Brinkers (2001) ausgewählt: „**Der Terminus ‚Text‘ bezeichnet eine begrenzte Folge von sprachlichen Zeichen, die in sich kohärent ist und die als Ganzes eine erkennbare kommunikative Funktion signalisiert**“ (vgl. Brinker 2001: 17).

2.1.2. Die Textlinguistik

Das Wort **Textlinguistik** ist ein Determinativkompositum aus den Wörtern **Text** und **Linguistik**. Dem **Wiktionary**⁶⁴ nach, verweist der Begriff **Determinativkompositum** auf eine Wortzusammensetzung, bei welcher das erste Wort dem zweiten untergeordnet ist. Versteht man nun den Begriff **Textlinguistik** tatsächlich nach diesem Prinzip, dann ergibt es sich, dass dieser Begriff tatsächlich auf die Linguistik also auf die Sprachwissenschaft verweist. Man würde sich jedoch fragen wieso ein neuer Begriff? War der ältere nicht umfassend genug? Oder handelt es sich eher auf einen neuen bzw. besonderen Typ von Sprachwissenschaften? Würde sich tatsächlich um eine neue Auffassung der Sprachwissenschaft, dann würde man sich auch fragen, was für Objekte, Gegenstände oder sprachliche Phänomene in diesem neuen Rahmen zu verarbeiten sind, welche nicht in dem alten werden konnten? Anders formuliert warum die Textlinguistik? Die Antworten auf diese Fragen werden uns helfen zu verstehen, worum es sich tatsächlich in diesem Sprachforschungsgebiet handelt. Allerdings können wir von Anfang an schon behaupten, dass, weil das Wort **-Linguistik** im Determinativkompositum das dominant ist und das Wort **Text-** das dominierte (also das untergeordnete) ist, die **Sprachwissenschaft unter einer textuellen Perspektive** untersucht wird. Dies wird voll heißen, dass der Gegenstand dieser besonderen Linguistik nicht mehr zwischen dem Laut, dem Morphem und dem Satz zu finden ist. Aber wo denn?

⁶⁴ Auch **Wikwörterbuch** genannt ist das **Wiktionary** ein freies Wörterbuch und ein Wikimedia-Projekt, das am 1. Mai 2004 startete. Es umfasst derzeit 1.029.295 Einträge und zielt darauf ab, die Erfassung des gesamten deutschen Wortschatzes und des aller anderen Sprachen frei zugänglich zu ermöglichen.

Fangen wir zur Beantwortung dieser verschiedenen Fragen mit der Erörterung des Begriffs **Linguistik** an. Die *Linguistik* wird beispielweise vom Wikiwörterbuch als *wissenschaftliche Disziplin, deren Ziele und Aufgabe in der Beschreibung und Erklärung der menschlichen Sprache, ihrer inneren Zusammenhänge, ihrer Funktion und Rolle in der Gesellschaft liegen (vgl. Wiktionary)*. In ihrer ersten Entwicklungsphase hat sich die **Sprachwissenschaft** auf die Beschreibung und Erklärung des Funktionierens „elementarer“ sprachlicher Einheiten wie etwa Laute, Morpheme, Lexeme, Wortgruppen und Sätze (vgl. Gerhard Helbig 2002: 276). Jedoch merkt man **ab den 70er Jahren** das Blühen neuer linguistischen Ansätze, die sich auf die Idee stützen, **dass sprachliche Kommunikationen nur in Form von Texten erscheinen**. Daher wurde der **Text** in den Gegenstandskreis der linguistischen Untersuchungen miteinbezogen (vgl. K. Brinker 2010). So gesehen versteht man die **Textlinguistik** als ein Zweig der Linguistik, der den Text als Gegenstand hat. Diese Auffassung des Begriffs teilt auch N. Lurjidge (2007). Dazu sagt er nämlich folgendes:

*Textlinguistik ist eine moderne Wissenschaftsdisziplin, die sich innerhalb der Linguistik in der zweiten Hälfte der 60er und Anfang der 70er Jahre herausgebildet hat, um den Untersuchungsgegenstand **Text in seinen inhaltlichen und formellen Facetten** zu erfassen und zu beschreiben* (N. Lurjidge 2007: 4).

Allerdings ist es mit einer Einheitlichkeit der Bestimmung des Begriffs **Textlinguistik** in den Fachliteraturen nicht zu rechnen (vgl. Gerhard Helbig 2002: 259-266). Dies liegt zumeist an der Tatsache, dass dieser Disziplin keinen klar eingegrenzten Gegenstand erkannt wird (Harweg 1990: 22). Betrachtet man die Zweige der Sprachwissenschaft, dann merkt man, dass jeder dieser Zweige einen gut erkennbaren Gegenstand besitzt. **Der Laut** für die Phonologie und die Phonetik, **das Morphem** für die Morphologie und **der Satz** für die Syntax u.a. Diese Einheitlichkeit bezüglich der **Gegenstandsbestimmung** ist jedoch in der Textlinguistik nicht festzustellen. Denn von der rein sprachsystematisch orientierten Textlinguistik bis zu der kommunikativorientierten Textlinguistik werden unterschiedliche Gegenstände (**Textgrammatik, Textsemantik, Textstruktur, Textfunktion, Textsorten, Texthaftigkeit, Textproduktion und -rezeption, usw.**) bearbeitet (vgl. Gerhard Helbig 2002: 259-266). Deswegen behauptet Harweg (1990), einer der Vorläufer dieser Disziplin, **dass die Textlinguistik, bedingt durch die Komplexität ihres Gegenstandes, eine sehr heterogene, aber auch ungewöhnlich interessante und zukunftsreiche Disziplin sei** (vgl. Harweg 1990: 22).

Van Dijk (1990) komplettiert diese Meinung indem er die Diversität der in diesem Forschungsgebiet angewendeten Theorien und Methoden thematisiert. Dazu sagt er nämlich:

[. . .] discourse analysis is not a strictly unified discipline, with one or a few dominant theories and methods of research. For some this is a sign of richness, for others a symptom of immaturity (Van Dijk 1990: in TEXT 10 (1/2)).

Im Allgemeinen ist es aus der Literatur zu dem Thema festzustellen, dass die **Forschungsschwerpunkte** der *Textlinguistik* sich immer im Laufe der Zeit entwickelt haben. Am Anfang der Textlinguistik war die **Satzgrammatik**, dann folgte satzübergreifende grammatische Phänomene und letztendlich stand die Analyse der **Textstruktur** im Vordergrund, wobei textpragmatische Aspekte und **Wissensverarbeitung** an Bedeutung gewannen (vgl. Gerhard Helbig 2002: 260). So gesehen, fungiert die **Textlinguistik** als **eine integrative Disziplin**, die sich nicht nur um das Produkt Text kümmert, sondern auch um Wissensverarbeitungsprozesse der Textproduktion und –Rezeption kümmert (vgl. Ebd.:260). Unter diesem Blickwinkel lässt sich die Textlinguistik nicht mehr als eine Teildisziplin der allgemeinen Linguistik auffassen, sondern **als eine eigenständige Disziplin**, die vielmehr zu **den angewandten sprachwissenschaftlichen Disziplinen** gehört, denn sie benutzt Theorien aus unterschiedlichen Bereichen (wie die **Sprechaktheorie** (Handlungstheorie) und Theorie aus der **Kognitionswissenschaft**) um Erkenntnisse zu gewinnen.

Dass die Textlinguistik ein interdisziplinäres Feld sei, wird noch von Tietz (1997) in ihrer Zusammenfassung der Ergebnisse einer Tagung zur Textlinguistik im Jahre 1996 bekräftigt. Sie sagte nämlich, **dass die Textlinguistik der Gegenwart sich zunehmend über Fragestellungen, die sich aus Erfordernissen und Bedürfnissen vielfältiger Bereiche, auch der Praxis ergeben, definiert** (Tietz 1997: 224). Van Dijk (1990) bestätigte auch dieser Hinwendung der Textlinguistik zu einer eigenständigen Disziplin der angewandten Sprachwissenschaft. Er sagte nämlich, dass es nicht zu leugnen sei, dass die vielen Richtungen in der Textlinguistik zu den überraschendsten und innovativsten Annäherungen an Sprache, Kommunikation, **Kognition, Interaktion, Gesellschaft und Politik** führten (vgl. van Dijk 1990: 3). Zusammengefasst können wir behaupten, dass der Begriff *Textlinguistik* nicht eine in sich geschlossene sprachwissenschaftliche Disziplin bezeichnet, sondern erfasst die Gesamtheit der sprachwissenschaftlichen Untersuchungen, die Texte betreffen, unabhängig davon, ob dabei mit Erkenntnissen anderer wissenschaftlichen Disziplinen vorgegangen wurde oder nicht. Wie schon bei der Erläuterung des Textbegriffs hingewiesen, zielt diese Forschung

darauf ab, **dem Computer das Verständnis eines textlinguistischen Phänomens zu ermöglichen**. Es handelt sich also in dieser Arbeit um den Text bzw. die Textlinguistik **aus einer computerlinguistischen Perspektiven** zu untersuchen. Aus diesem Grund wollen wir für den Weiterverlauf unserer Arbeit die folgende Auffassung der *Textlinguistik* behalten.

Heinemann & Viehweger (1991) definieren die Textlinguistik als eine Wissenschaft, die sich ausschließlich auf die Erforschung von Textstrukturen und Textformulierungen bezieht, wobei die Einbettung dieser Texte in kommunikative, allgemein soziologische und psychologische Zusammenhänge ebenfalls berücksichtigt wird (Schindler, 2010:3).

2.2. Zum Begriff Textgrammatik

Dem Begriff *Textgrammatik* wird im Allgemeinen zwei Bedeutungen zugeschrieben: die eine verweist auf **eine spezielle Richtung der Textlinguistik**, in der semantische und syntaktische Strukturen von Texten untersucht werden (vgl. Gerhard Helbig 2002: 290). Und die andere **auf einen linguistischen Ansatz**, der darin besteht, sprachliche Einheiten unterschiedlicher Sprachebene anhand ihrer textuellen Einbettung zu beschreiben und deren Funktionieren zu erklären sind (vgl. ebd.: 290). Noch als Weinrichs Grammatik bekannt, wurde dieser Ansatz schon zuvor bei Autoren wie Halliday (1985), der seine Beschreibung auf die **Satzebene** beschränkte; Dressler (1985), der sich mit den **Morphemen** beschäftigte und Brazil (1985), der seine Versuche auf der **Intonation** im Text (Phonemebene) beschränkte, eingeführt. Das Besondere an Weinrich liegt in der Tatsache, dass er alle diese Vorarbeiten in einem gut strukturierten theoretischen und methodologischen grammatischen Ansatz aufgefasst hat. Dazu sagt er „Diese Grammatik versteht die Phänomene der Sprache **von Texten her**, da eine natürliche Sprache nur in Texten gebraucht wird“ (Weinrich 1993: 17). Es steht somit fest, dass die Zielsetzung Weinrichs Grammatik darin besteht, sprachliche Einheiten und andere sprachliche Phänomene, die in Texten gebraucht werden, anhand ihrer Rolle im gut *Funktionieren* des Textes zu beschreiben.

Die Auffassung, die für die vorliegende Untersuchung von Relevanz ist, ist die, in der die *Textgrammatik* als ein Zweig der Textlinguistik verstanden wird. Dem Metzler Lexikon nach zielt diese Auffassung darauf ab *[...] syntaktische und semantische Regeln zur Erzeugung und Interpretation sowie Kriterien für die Grammatikalität und Wohlgeformtheit von Texten zu ermitteln. [...]* (Metzler Lexikon 2005). Im Mittelpunkt dieser Auffassung stehen die Beschreibung und gegebenenfalls auch die Erklärung des Funktionierens von **Textstrukturen** und sprachlichen Phänomenen, die diese Textstrukturen bilden. Als Textstrukturen werden

sowohl Elemente der **Textgliederung** (Abschnitte & Teiltex-te) als auch Elemente der **Textkonstitution** (Sätze & Wortgruppen) bezeichnet. Als **sprachliche Phänomene**, die diesen Strukturen unterliegen, zählen wir sowohl, kohäsive Phänomene (wie etwa die Wiederaufnahme und die Konnektoren), als auch Phänomene, die zur Kohärenz eines Textes beitragen, wie etwa die Tempusformen und die Thema-Rhema-Gliederung. In dieser Hinsicht kategorisiert Moskalskaja (1984) die Untersuchungsobjekte der Textgrammatik in zwei Hauptgruppen: Die Makrostruktur und die Mikrostruktur.

Bei der **Makrostruktur** werden Aspekte des Gesamttextes (Beziehungen zwischen Abschnitten; thematische Progression) untersucht. Dabei gilt also die Auffassung des Textes als eine thematische bzw. sinnliche und somit kommunikative Einheit (Moskalskaja 1984: 17). Im Gegensatz zur Makrostruktur stehen die syntaktischen und semantischen Beziehungen zwischen Textkonstituenten (Sätzen insbesondere aber auch Syntagmen) im Mittelpunkt der **mikrostrukturellen Untersuchung**. Dabei gilt also die Auffassung des Textes als eine transphrastische Einheit. Und da wir im Rahmen dieser Arbeit einen integrativen **Textbegriff** ausgewählt haben, werden wir also sowohl thematische bzw. sinnliche als auch rein syntaktische Aspekte des Textes berücksichtigen.

Zu merken ist, jedoch, die Tatsache, dass je nachdem, ob man **die Makro- oder die Mikrostruktur** betrachtet, wird man sich entweder im Bereich der **Kohäsion** (insbesondere bei der Beschreibung direkter Anaphern) oder dem der **Kohärenz** (insbesondere bei der Beschreibung indirekter Anaphern) befinden. Aus diesen Gründen scheint es somit sinnvoll den semantischen Inhalt dieser beiden Begriffe für die vorliegende Untersuchung zu begrenzen.

2.2.1. Die Textkohäsion

Etymologisch betrachtet, wurde das Nomen „Kohäsion“ aus dem lateinischen Verb „**cohaerere**“ (beieinander kleben, zusammenstecken, zusammenhaften) abgeleitet (Beaugrande & Dressler 1981: 3). Es wird direkt gemerkt, die Bedeutung dieses Verbs hat einen engen Zusammenhang mit dem lateinischen Verb „**textere**“ (verkneten, verbinden), aus dem das Nomen „**Text**“ abgeleitet wurde. Im Allgemeinen bedeutet der Begriff **Kohäsion** **den formalen Zusammenhalt eines gesprochenen oder geschriebenen Textes**. Wobei dieser Zusammenhang mit kohäsiven Mitteln wie die Konnektoren oder noch die Wiederaufnahme geleistet wird. Diese Auffassung des Begriffs **Kohäsion** wird auch von Autoren wie Beaugrande und Dressler (1981) geteilt. Von der **prozeduralen Auffassung des Textes** aus, definieren sie

die *Kohäsion* als eines der sieben Kriterien, das ein sprachliches Gebilde aufweisen soll, um als ein Text verzeichnet zu werden. Sie schreiben dem Begriff *Kohäsion* folgende Bedeutung:

Es betrifft die Art, wie die Komponenten des OBERFLÄCHENTEXTES, d.h. die Worte, wie wir sie tatsächlich hören oder sehen, miteinander verbunden sind. Die Oberflächenkomponenten hängen durch grammatische Formen und Konventionen voneinander ab, so daß also Kohäsion auf GRAMMATISCHEN ABHÄNGIGKEITEN beruht. (Beaugrande & Dressler 1981: 3).

Von diesem Zitat aus, ergibt sich, dass Beaugrande & Dressler (1981) die Textoberfläche als Gegenstand der *Kohäsion* bezeichnen. Ihnen nach kümmert sich die *Kohäsion* besonders um morpho-syntaktische Kongruenzen von Wörtern innerhalb von Sätzen eines Textes. Jedoch besteht die Textoberfläche nicht nur aus Wörtern, sondern auch aus Sätzen. In dieser Hinsicht erweitert Vater (1994) den Inhalt des Kohäsionsbegriffs und schreibt ihm folgende Bedeutung zu:

Eine interessante Definition von Kohäsion und Kohäsionsbeziehungen liefert Vater (1994: 32), der sie als grammatikalische Beziehung zwischen den verschiedenen Einheiten des Textes definiert, wobei es vorwiegend um satzübergreifende Beziehungen geht. Kohäsionsbeziehungen fallen oft mit syntaktischen Beziehungen zusammen, die er als „x dominiert y und z“, „x kommandiert y“, „x regiert y“, „x wird mit y koinzidiert“ darstellt (Katia, 2010: 9).

Von dieser Definition aus, ergibt es sich, dass die Kohäsion eines Textes also die *Textkohäsion* nicht anders ist als der syntaktische bzw. morpho-syntaktische Zusammenhalt des ganzen Textes ist (*satzübergreifende Beziehungen*.) Somit kann die *Textkohäsion* noch als die *Textsyntax* aufgefasst werden. Die *Textkohäsion* kümmert sich daher auch um die korrekten Satzverknüpfungen und -anknüpfungen innerhalb von Texten. Wobei die Satzverknüpfungen durch **Wiederaufnahme** (Proformen; Rekurrenz...) und die Satzanknüpfungen durch **Konnektoren** (Konjunktionen; Satzadverbien...) auf gesichert sind.

Im Gegensatz zu Beaugrande & Dressler (1981) und Vater (1994), die die *Textkohäsion* in den Bereich der **Syntax** einordnen, bestimmen Halliday & Hasan (1976) ihn unter einem **semantischen Blickwinkel**. Sie behaupten nämlich Folgendes:

The concept of cohesion is a semantic one; it refers to relations of meaning that exist within the text, and that define it as a text. Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another. The one PRESUPPOSES the other, in the sense that it cannot be effectively decoded except by recourse to it. When this happens, a relation of cohesion is set up, and the two elements,

the presupposing and the presupposed, are thereby at least potentially integrated into a text (Halliday & Hasan 1976: 4).

Halliday und Hassan (1976) bearbeiten also den Begriff **Kohäsion bzw. Textkohäsion** (*Cohesion of Discours*) unter einer semantischen Perspektive. Sie kategorisieren somit die *Kohäsion* auf der Ebene der **Textsemantik** und ggf. auch auf der der **Textpragmatik**, denn zur sinnlichen Rekonstruktion einer Äußerung braucht man öfters Zugang zu zusätzlichen Weltwissen, wie etwa dem Kontext der Äußerung oder noch der Beziehung zwischen dem Rezipienten und dem Produzenten. Sie schreiben somit dem Begriff **Textkohäsion** schon Merkmale der **Textkohärenz** zu (vgl. Halliday & Hasan 1976: 4). Jedoch steht es in der heutigen textlinguistischen Forschung einen klaren Unterschied zwischen den beiden Begriffen.

2.2.2. Textkohärenz und Thematische Progression

Etymologisch betrachtet, geht das Nomen „**Kohärenz**“ auf das lateinische Verb „cohaerere“ zurück, was „**zusammenhängen**“ bedeutet (vgl. Beaugrande & Dressler 1981: 14). Zu merken ist jedoch, die Tatsache, dass diese etymologische Bestimmung des Begriffs „Kohärenz“ keine Hinweise über die Natur des Zusammenhangs (**Formel oder Inhaltlich?**) gibt. Zu Beginn der Forschung in der Textlinguistik wurde diese Lücke, jedoch aufgefüllt, indem der Begriff **Textkohärenz** als *eine spezielle Richtung der Textlinguistik, die sich um die semantischen Bezüge zwischen den Sätzen eines Textes bemüht* (vgl. Glottopädia)⁶⁵, bestimmt wurde. Außerdem verweist auch die **Kohärenz** auf die Beziehung zwischen einem sprachlichen Ausdruck (Wort) und dem Inhalt, den er bezeichnet (vgl. Metzler Lexikon Sprache, 2005).

Grob gesehen, wurde dem Begriff *Kohärenz* im Laufe der Forschung insbesondere **2 Bedeutungen** zugeschrieben. Einerseits, wurde er als **ein Textualität-Merkmal** aufgefasst. In dieser Auffassung steht der syntaktisch-semantische Zusammenhalt des Textes im Vordergrund. Andererseits wurde er **als ein logischer Zusammenhang (Sinnzusammenhang)** eines Textes aufgefasst (Duden-Grammatik, 2005). Diese zweite Auffassung wurde insbesondere durch den Einfluss von Untersuchungsergebnissen **der kognitiven Linguistik** motiviert. In dieser Betrachtungsweise wurde die *Kohärenz* **als das Ergebnis der Textverarbeitung kategorisiert**, wobei dieses Ergebnis erst durch eine Reihe von kognitiven

⁶⁵<http://glottopedia.org/index.php/Koh%C3%A4renz> Abruf am (24.11.2020)

Operationen zum Stande gekommen ist (M. Schwarz 1992a: 159). Beaugrande & Dressler bekräftigen in ihrem prozeduralen Ansatz diese Hinsicht. Sie definieren den Begriff **Textkohärenz als Sinn eines Textes** auf:

Beaugrande und Dressler beschreiben Kohärenz als „das Ergebnis der Verbindung von Konzepten und Relationen zu einem NETZ betrachten, die aus WISSENSRÄUMEN mit den Haupt-TOPIKS in deren Mittelpunkt zusammengesetzt ist“ (Beaugrande/Dressler, 1981, zit. nach Lee, 2002, 78) und stellen viel kompliziertere Vernetzungsmuster vor (Kliukaitė 2006: 8).

Ihnen nach verweist also **die Kohärenz auf den Sinn eines Textes** und hat die **Sinnkontinuität** als Grundlage. Außerdem gibt es zu merken, dass von der Sinnkontinuität die Rede ist, wenn sich alle Textteile (Sätze und Abschnitte) in einem vom Textverwender gut erkennbaren logischen Zusammenhang verkettet sind. Dies bestätigen Dressler und Braugrande indem sie behaupten: **Die Sinnkontinuität besteht aus dem Vorwissen und aus der Konstellation der ausgedrückten Konzepte und Relationen** (vgl. Beaugrande/Dressler, 1981). Es ergibt sich also, dass nicht nur der Textinhalt zum sinnlichen Zusammenhang eines Textes beiträgt, sondern auch das Vorwissen des Textverwenders.

Weiterhin merkt man, dass die Sinnkontinuität nach Beaugrande und Dressler einen engen Zusammenhang mit dem Begriff **Thematische Progression** von Daneš (1978), ein Vorläufer der heutigen Textlinguistik, erweist. Er definiert nämlich die thematische Progression als **eine textlinguistische Theorie, die darauf abzielt, den thematischen Aufbau eines Textes zu erklären** (Bystrica, 2005). Während es bei der Sinnkontinuität von der Verbindung von *Konzepten, Relationen* und eigenen Wissen des Textverwenders zur Erzeugung bzw. Erkennung eines Textsinns geht, geht es bei der thematischen Progression **um eine Thema-Rhema-Verbindung zwischen Sätzen eines Textes**; wobei die Verbindung zwischen Thema (das Bekannte) und Rhema (die neue(n) Information(en)) den sinnlichen Zusammenhalt (**Kontinuität**) des Textes sichert (Eroms 1986: 13).

Es steht also fest, dass Daneš (1978), um **den Textaufbau** erklären zu können, Untersuchungen zur Funktionalen Satzperspektive (Thema-Rhema-Gliederung) von der Satzebene auf die Ebene des Textes übertragen hat. Dabei stellte er die Struktur eines Textes als **Sequenzen von Themen** dar und unterschied zwischen fünf Typen **von thematischen Progressionen**: Die einfache lineare Progression; die Progression mit durchlaufendem Thema; die Progression mit Themen, die von einem Hyperthema abgeleitet sind; Entwicklung eines gespaltenen Rhemas

und Progression mit einem thematischen Sprung (vgl. Daneš 1978: 72- 78). Es ergab sich am Ende sein Typologisieren, **dass der Textaufbau formal-syntaktisch (kohäsiv) und gleichzeitig semantisch-inhaltlich (kohärent) analysiert werden kann** (vgl. Daneš 1978: 189 ff). Daher wollen wir **die Anaphorik, ein Mittel der Textkonstitution (Textaufbau)**, nicht nur syntaktisch analysieren und interpretieren, sondern auch als ein Mittel das zur Kohärenz eines Textes beiträgt, betrachten. In dieser Hinsicht werden wir im Rahmen dieser Untersuchung diese beiden Begriffe, *Textkohäsion* und *Textkohärenz*, integrativ behandeln.

Eine integrative Untersuchung von Kohäsion und Kohärenz im Rahmen dieser Arbeit lässt sich weiterhin durch die Tatsache erklären, dass obwohl die vorliegende Arbeit sich im Bereich der *Textgrammatik* und insbesondere in dem der *Textsyntax* einordnen lässt, denn sie hat als Gegenstand die **Verkettung von Sätzen durch anaphorische Pronominaladverbien**, werden wir auch trotzdem in dieser Untersuchung Aspekte der **Textsemantik und -pragmatik** behandeln. Denn die Satzverkettung durch anaphorische Wiederaufnahme sichert sowohl **die Kohäsion als auch die Kohärenz eines Textes auf** (Finkbeiner 2018: 190). Während die Gewährleistung der Kohäsion erfolgt, indem es eine morpho-syntaktische Kongruenz zwischen dem Antezedenten und dem wiederaufnehmenden Ausdruck herrscht (vgl. Ebd.:190), erfolgt die Gewährleistung der Kohärenz durch thematische Progression, denn durch wiederaufnehmenden Ausdruck werden öfters neue Informationen (Rhema) hinzugebracht (insbesondere bei indirekten Anaphora)(vgl. Ebd.:190). Man muss in der Tat bei einer Verkettung durch **indirekte Anaphorik** auf informationelle Gehalte vom Antezedenten-Ausdruck und dessen des wiederaufnehmenden Ausdrucks zugreifen, um die syntaktische Beziehung zwischen den beiden analysieren zu können. Zu merken ist jedoch, dass dieser informationelle Gehalt entweder von **der Text-welt oder Weltwissen des Sprachverwenders** abgeleitet werden kann, daher redet man auch von der Textpragmatik (vgl. M. Schwarz 2000b:118).

Schließlich lässt sich also merken, dass eine strenge Differenzierung von *Textkohäsion* und *Textkohärenz* in den heutigen textlinguistischen Forschungen immer seltener ist. Deswegen plädiert Brinker für einen dreidimensionalen Kohärenzbegriff: **die grammatische Kohärenz (Kohäsion); die thematische und pragmatische Kohärenz (Kohärenz)** (vgl. K. Brinker 1979:6).

2.3. Die Anaphorik

Das Wort **Anaphorik** bezieht sich, etymologisch betrachtet, auf das griechische Verb *ἀναφέρειν*. Dieses griechische Verb hat die gleiche Bedeutung mit den deutschen Verben **herauftragen und sich beziehen (auf etwas)** (Wiktionary⁶⁶). Demnach kann das Wort Anaphorik sowohl für Rückwärts- als auch für Vorwärtsverweise gebraucht werden. Jedoch wird diesem Wort in der Sprachwissenschaft **eine Rückwärtsbedeutung** zugeschrieben (M.Schwarz 1997: 445-456). In dieser Hinsicht wird der Begriff **Anaphorik** in den textlinguistischen Forschungen als **der Verweis** eines Satz- oder Textteiles (**Anapher**) auf einen anderen, vor ihm stehenden Satz- oder Textteil (**Antezedent**) aufgefasst (Konrad Ehlich 2007: 23). Man unterscheidet im allgemeinen zwei Arten von **Anaphorik**: **Die direkte** und **die indirekte**, wobei jeweils zwischen **intrasententialer** (Anapher und Antezedent stehen im selben Satz) und **intersententialer** (Anapher und Antezedent stehen in verschiedenen Sätzen) Anaphora zu unterscheiden ist (Schwarz-Friesel, Monika/Consten, Manfred, 2007: 268). **Eine anaphorische Konstruktion** (Anaphorik) wird als **direkt** genannt, wenn sie durch Anwendung **morphosyntaktischer Regeln** aufgelöst (also interpretiert bzw. verstanden) werden kann. Hingegen wird eine anaphorische Verbindung als **indirekt** genannt, wenn sie außer grammatischen Kenntnissen **noch zusätzliche Wissen (Text-und Weltwissen)** für ihr Verständnis voraussetzt (vgl. M. Schwarz 2000b: 118).

Ein weiterer Begriff, der für die Erörterung unserer Problemstellung von Relevanz ist, ist **der Anapher-Begriff**. Wie schon oben angedeutet, bezeichnet die textgrammatische Auffassung dieses Begriffs **den wiederaufnehmenden Teil einer Rückwärtsverweiskonstruktion** (vgl. Finkbeiner 2018: 190). Jedoch lässt sich in der Forschungsliteratur merken, dass dieser Begriff seinen Ursprung in der klassischen Rhetorik hat, wo er **als eine Redefigur verstanden wird**, die in der Wiederholung eines Wortes oder einer ganzen Phrase besteht (vgl. Wiktionary). Allerdings wird diesem Begriff, je nach Autoren, Forschungsansätzen und -perspektiven, verschiedene Bedeutungen **in den textlinguistischen Forschungen** zugeschrieben.

Die verschiedenen Ansätze sind insbesondere in zwei Hauptkategorien eingeteilt: **Die diskursbasierten** und **die phorischen Ansätze**. Während die diskursbasierten Ansätze

⁶⁶ Auch **Wikwörterbuch** genannt ist das **Wiktionary** ein freies Wörterbuch und ein Wikimedia-Projekt, das am 1.Mai 2004 startete. Es umfasst derzeit 1.029.295 Einträge und zielt darauf ab, die Erfassung des gesamten deutschen Wortschatzes und des aller anderen Sprachen frei zugänglich zu ermöglichen.

Anaphern **über den Diskursstatus des Referenten**⁶⁷ bestimmen, berücksichtigen die phorischen Ansätze eher **die Art der Domäne**⁶⁸, in der der Referent zugänglich ist (vgl. Finkbeiner 2018: 192). Die vorliegende Arbeit lässt sich zu den Untersuchungen zählen, die **den phorischen Ansatz als Ausgangspunkt** haben. Jedoch werden noch innerhalb dieses Ansatzes verschiedene Forschungsperspektiven rezensiert. Darunter zählt man zum Beispiel **die textfunktionale und die satzfunktionale Perspektive** (vgl. ebd.: 194). Während **die Anapher** in der textfunktionalen Perspektive **als Textphänomen** verstanden wird, wird sie in der satzfunktionalen Perspektive **als ein Satzphänomen** betrachtet. Die satzfunktionale Perspektive ist insbesondere von **Noam Chomsky (1981)** vertreten worden. In seiner Untersuchung zur generativen **Bindungstheorie** definiert er **die Anapher** als **eine Wortart**, nämlich als **ein Reflexivpronomen**. Nimmt man den folgenden Textauszug zur Illustration, dann lässt sich feststellen, dass das Reflexivpronomen „**mich**“, um auf den Textsprecher zu referieren, sich zuerst auf das Personalpronomen „**ich**“ bezieht⁶⁹. Gemerkt wird also, die Anapher („**mich**“) kann auf den Textsprecher referieren nur, wenn sie zuvor Bezug auf den Antezedenten („**Ich**“) genommen hat. Ohne diesen Umweg kann der Textrezipient (Derjenige, der die Anapher aufzulösen bzw. zu interpretieren hat) der Anapher keinen Referenten zuschreiben.

(1) „**Ich**₁ setze **mich**₁ im hohen hellen Speisesaal an meinem einsamen runden Tischlein und sah **mir**₁ zugleich zu, wie ich mich setzte [...]“ (Aus dem Buch von Herman Hesse (1977) betitelt *Kurgast*; Seite 112).

Hoffmann (2000) nach, stehen dieser **generativen** Auffassung gegenüber **die formalen und funktionalen Perspektiven**. In der **formalen** Perspektive wird **zwar** die *Anapher* als **ein Textphänomen** untersucht, **jedoch** wird sie, wie bei dem generativen Ansatz (N. Chomsky 1981), auf eine bestimmte Wortartklasse beschränkt. Dieser Perspektive nach, können nur Personalpronomen im Allgemeinen und insbesondere **die der dritten Person** als *Anapher*

⁶⁷ Der Textreferent soll also **im Fokus** des Diskurses stehen. Konrad Ehlich (2007) versteht in dieser Hinsicht die *Anapher* als *solche sprachlichen Einheiten, die innerhalb eines Textes, einer Rede oder Sprechhandlungssequenz oder innerhalb einer Sprechhandlung einen Rückbezug auf schon vorher verbalisierte propositionale Elemente herstellen, die vorgängig bei S und H fokussiert sind und deren vorgängige Fokussierung bei H S bekannt ist, und die so eine Kontinuität der vorgängigen Fokussierung signalisieren* (Konrad Ehlich 2007: 23. Zit. Nach. Finkbeiner 2018: 192).

⁶⁸ Ist der Referent durch den Text, nämlich mit Hilfe von Textreferenten zugänglich, dann liegt **eine Anapher (Textuelle Domäne)** vor. Ist der Referent aber direkt nur mit Bezug auf die außersprachliche Realität zugänglich (**der Referent ist physisch in der Äußerungssituation – oder im Vorstellungsraum vorhanden**), dann liegt **eine Deixis** vor (vgl. Ebd.: 192).

⁶⁹ Es wird auch gesagt, das Reflexivpronomen „**mich**“ (also die Anapher) verweist auf das Personalpronomen „**Ich**“ (den Antezedenten). Dieser Verweis wird dann entweder **ANAPHORIK** oder eher **ANAPHORISCHE KONSTRUKTION** genannt.

auftreten (vgl. Hoffmann 2000, Zit. nach Finkbeiner 2018: 192). Dieser Perspektive entgegen steht **die funktionale Auffassung der Anapher**, in der *Anapher als eine Gebrauchsweise von sprachlichen Einheiten* aufgefasst wird. Somit werden alle sprachlichen Ausdrücke, mit denen ein Rückverweis geleistet werden kann und **als Anapher kategorisiert**. Zu diesen Ausdrücken gehören insbesondere definite Nominalphrasen und verschiedene Pronomen (Consten/Schwarz-Friesel 2007:287). Für die vorliegende Arbeit wird die folgende Auffassung des Anaphernbegriffs gelten:

*Anapher nennt man in der Linguistik **sprachliche Ausdrücke**, mit denen Sprachbenutzer auf **einen bereits erwähnten Referenten** im Text oder Diskurs erneut Bezug nehmen* (M. Schwarz 1997: 445).

Mit der Wahl dieser Auffassung des Anaphernbegriffs steht es fest, dass die vorliegende Untersuchung insbesondere dazu neigt, die Funktion von sprachlichen Einheiten innerhalb der Texte zu untersuchen. **Wobei das Augenmerk sich insbesondere auf diejenigen gerichtet wird, die eine anaphorische Funktion erfüllen.**

Wie oben schon eingeführt, kann der Sprecher bei der Bezugnahme auf den Textreferenten (Antezedent) entweder eine direkte oder indirekte Anaphorik aufbauen (Kubczak, H. 2016: 221-247). **Von der Rede eines direkten Rückverweises** ist es, wenn der wiederaufnehmende Ausdruck (Anapher) und der Textreferent (Antezedent) **auf dasselbe außersprachliche Objekt referieren**. In dieser Hinsicht sagt man, die beiden Ausdrücke (Antezedent und Anapher) **stehen in einer Relation der Koreferenz** (Bußmann 1990:426). Dieser Art des Rückverweises gegenüber steht die indirekte Anaphorik, wobei die Anapher und der Textreferent (**Ankerausdruck**) **nicht auf denselben außersprachlichen Gegenstand (Referent/Objekt) referieren**. Man sagt auch, die beiden sprachlichen Einheiten stehen **in einer nichtreferenzidentischen Relation** (vgl. M.Schwarz 1994: 81-102). Im Folgenden werden diese beiden Arten des textlichen Rückverweises noch näher beschrieben.

Weiterhin ist es notwendig schon von Anfang an zwischen den beiden Verben (**referieren und verweisen**) eine klare Bedeutungsgrenze zu ziehen. Während **referieren** auf diese Verweisbeziehung, **die zwischen einem Objekt (dem Referenten) und der sprachlichen Einheit, die es versprachlicht, hinausdeutet**, bedeutet das Verb **verweisen** eher die Verweisbeziehung, **die zwischen einem Textreferenten (Antezedent/Anker) und seinem wiederaufnehmenden Ausdruck (Anapher) existiert**. Man kann, grob betrachtet, Folgendes

behaupten: Anaphern *verweisen* auf Textreferenten, die ihrerseits über Referenten *referieren*. Dies wird durch das folgende Zitat von **Monika Schwarz (2000a)** bekräftigt:

*„Mit anaphorischen Ausdrücken **verweist** ein Sprecher im Text rückwärts auf ein vorangegangenes Textelement. Dabei ist die Redeweise von der **‘Rückbezüglichkeit’** nicht so zu verstehen, dass der anaphorische Ausdruck auf den Antezedens – Ausdruck **referiert** [...]. Vielmehr benennt der Sprecher mit einem anaphorischen Ausdruck zum **zweiten Male den (außersprachlichen) Referenten**, den er vorher mit dem Antezedens – Ausdruck benannt wird.“* (M. Schwarz 2000a: 53)

2.3.1. Anfänge der Forschungen zu Verweisphänomenen in der Textlinguistik

Die Auseinandersetzung mit Verweisphänomenen in der Sprachwissenschaft wurde mit der **kommunikativ-paradigmatischen Wende** deutlicher (vgl. Gerhard Helbig 2002: 259). Dabei wurden, in der Tat, deiktische, kataphorische und anaphorische Verweisphänomene näher studiert. Während sich **deiktische Phänomene** explizit im Rahmen der **Pragmatik** abarbeiten lassen, werden Rückwärts- und Vorwärtsverweise innerhalb von Texten im Rahmen der **Textlinguistik** untersucht (vgl. Finkbeiner 2018: 186-189). Die vorliegende Arbeit beschäftigt sich exklusiv mit Phänomenen des Rückwärtsverweises.

Wie schon im Abschnitt 2.1.2. eingeführt, zielte die erste Entwicklungsphase der Textlinguistik darauf ab, eine sprachsystematische Beschreibung von Texten, wie es schon mit dem Satz der Fall war, vorzulegen (vgl. Harweg 1990: 22). Die Textlinguistik in dieser Phase hatte also rein grammatische Strebungen, denn sie wollte **sprachsystematische Regeln entwickeln, die die Textkonstituierung (Textkohäsion- und-kohärenz) normieren**. Daher wurde sie in dieser ersten Phase ihrer Entwicklung mit der Textgrammatik gleichgesetzt (vgl. K. Brinker 2006: 25). Als wesentliches Mittel **der Textkonstitution** zählt die Kohäsion und als zentrales Prinzip der Textkohäsion zählt **die Anaphorik** (Rückwärtsverweis).

Die Auseinandersetzung mit Rückwärtsverweisen im Rahmen der Textgrammatik innerhalb des deutschsprachigen Raums ist von unterschiedlichen Forschern geführt worden. **Renate Steinitz (1968)** zum Beispiel prägte in einem ihres erschienenen Artikels unter dem Titel **„Nominale Pro-Formen“**, den Begriff **Pro-Fortführung** (Steinitz 1968, zit. nach Koch 2001: 7). In diesem Artikel interessiert sie sich besonders für **die Natur von sprachlichen Einheiten**, die in einer textlichen Verweisbeziehung auftreten. Dabei unterscheidet sie zwischen drei Arten des Wiederaufnehmens: die Wiederaufnahme **durch Wiederholung desselben Nomens** (Rekurrenz); die Wiederaufnahme **durch Wiederholung eines definiten Personalpronomens**

(Rekurrenz) und **die Wiederaufnahme durch ein anderes Nomen** (vgl. Steinitz 1968, zit. nach Ebd.: 7).

Im Gegensatz zu Steinitz (1968) beschränkt sich **Horst Isenberg (1971)** nicht auf die Erwähnung der Natur von wiederaufnehmenden Ausdrücken und deren Bezugsausdrücken, sondern untersucht er auch **die Typen oder Arten der Verweisbeziehungen** zwischen den beiden Ausdrücken (vgl. Horst Isenberg 1971: 155-172). In seinem im Jahre 1971 erschienenen Artikel betitelt „**Überlegungen zur Texttheorie**“ betrachtet er hauptsächlich die Referenz- und Verweisrelationen von Nomina und prägte dabei den Begriff der **Koreferenz**. In seinen Überlegungen zielte er besonders **darauf ab, einen Kriterienkatalog von Referenzmerkmalen, die beschreiben, ob ein Nomen erstmalig auf einen Gegenstand verweist oder ob und wie weitere Nomina auf diesen Gegenstand sich beziehen** (vgl. Koch 2001: 8). Zu merken ist einerseits die Tatsache, dass er keine strikte Grenze zwischen **Referenzbeziehungen** (Beziehungen zwischen sprachlichen Bezeichnungen und deren jeweiligen außersprachlichen Gegenständen) und **Verweisbeziehungen** (Beziehungen zwischen sprachlichen Einheiten eines selben Textes, wobei zwischen anaphorischen und kataphorischen zu unterscheiden sind), etablierte. Andererseits lässt sich auch merken, dass er auch darauf gezielt hat, die möglichen grammatikalischen oder semantischen Verhältnisse zu beschreiben, die zwischen den beiden Ausdrücken (Bezugsausdruck und wiederaufnehmender Ausdruck) entstehen können (vgl. Ebd.:8).

Zur Diskussion über die Wiederaufnahme als ein textkonstitutives Prinzip trägt **Roland Harweg (1979)** mit seiner Monographie betitelt „**Pronomina und Textkonstitution**“, erschienen im Jahre 1979, bei. Im Gegensatz zu seinen Vorläufern, **Steinitz (1968)** und **Isenberg (1971)**, die sich besonders **um Pronomina als Wiederholung (Rekurrenz)** kümmerten, interessiert er sich ausschließlich **für Pronomina als Substitution (also als Wiederaufnahme)** und somit nicht als Wiederholung, wie es zuvor der Fall war. Anschließend entwirft er den Begriff der **syntagmatischen Substitution** und definiert zugleich das Pronomen **als das für den Prozess der Textkonstitution verantwortliche Ausdrucksmittel** (Harweg 1979: 10, zit. nach Koch 2001: 3). Außerdem untersucht er zugleich **die Arten von Bezugsrelationen**, die zwischen Pronomina und wiederaufgenommenen Ausdrücken bestehen können. Dabei beschränkt er sich nicht nur um die Beschreibung **koreferenter Verweisformen**, sondern auch um diejenigen, wobei Bezugswörter und Pronomina auf

unterschiedliche Referenten zurückweisen. Harweg (1979) unterscheidet drei Typen von Verweisformen:

2. **Die eindimensionale Substitution:** Mit den Begriffen wie *Substituens* (der ersetzende Ausdruck) und *Substituendum* (der ersetzte Ausdruck) bestimmt er *die eindimensionale Substitution* als **ein Verfahren in dem ein Substituens, syntagmatisch betrachtet, nur ein bestimmtes und somit kein anderes Substituendum ersetzen kann** (vgl. Harweg 1979: 26 zitiert nach Koch 2001.). Zu dieser Kategorie zählen insbesondere **Wiederholungen von Eigen- oder Gattungsnamen**. Als Beispiel können wir den folgenden Text nehmen, in dem der Eigenname „Mandela“ nur ihn selbst im vorkommenden Satz ersetzen kann:

(2) **Mandela** war 27 Jahre lang in Haft. **Mandela** blieb trotzdem liebevoll.

3. Erst mit **der zweidimensionalen Substitution** distanziert sich Harweg (1979) von Steinitz (1968). Im Gegensatz zu diesem Letzten, der die **Wiederholung von Pronomina** beschrieb, elaboriert Harweg (1979) eine **Substitutionsform**, in der die Eigennamen (Bezugswort/Substituendum) durch Pronomina (Wiederaufnahme/Substituent) ersetzt werden: Er definiert somit die zweidimensionale Substitution **als die kombinierte Form syntagmatischer und durch sie vermittelter paradigmatischer Substitution** (Harweg 1979: 25, zit. nach Koch 2001.). Davon ausgehend erweisen *Substituentia* (Ersetzungsausdrücke) in einer zweidimensionalen Substitution **nicht nur syntagmatische Eigenschaften**, wie es bei der eindimensionalen Substitution der Fall ist (siehe das Beispiel (2)), **sondern auch paradigmatische Eigenschaften**, denn sie gehören zur gleichen paradigmatischen Klasse mit dem Substituendum (Bezugsausdruck). Nehmen wir als Beispiel den folgenden Text:

(3) **Nelson** ging gestern zur Uni. **Er** musste an einer Vorlesung teilnehmen. In diesem Text stehen das Personalpronomen „er“ und sein Bezugsausdruck „Nelson“ sowohl in einer syntagmatischen als auch in einer paradigmatischen Beziehung. Denn, syntagmatisch betrachtet, kann man „Nelson“ nur durch „Er“ ersetzen und, paradigmatisch betrachtet, können „Er“ und „Nelson“ zur gleichen paradigmatischen Klasse gehören.

4. In seiner dritten Substitutionskategorie beschreibt Harweg (1979) ein Substitutionsverfahren, in dem der Ersetzungsausdruck (Wiederaufnahme) Merkmale eines **einer zwei dimensional Substitution** aufweist, während der Bezugsausdruck eher die Merkmale eines Ersetzungsausdrucks **einer eindimensionalen Substitution** aufweist. Noch als **kontaminierte Substitution** bezeichnet, erfolgt sie durch **eine**

Ersetzung eines eindimensionalen syntagmatischen Substituens durch ein zweidimensionales Substituens (vgl. Harweg 1979: 25, zit. nach Koch 2001: 4). Es scheint also, es handelt sich hier um zwei Wiederaufnahmen: die erste ist eine *eindimensionale Substitution* und die zweite *eine zweidimensionale*. Nehmen wir den folgenden Text als Beispiel:

(4)Mandela war 27 Jahre lang in Haft. **Mandela** blieb trotzdem immer liebevoll. **Er** wurde als erstes schwarzafrikanisches Staatsoberhaupt von Südafrika gewählt.

(Letztendlich scheint es also als, ob Harweg (1979) die beiden ersten dimensional Substitutionen integrativ verarbeiten möchte).

In der Kontinuität Harwegs (1979) Überlegungen über Verweisphänomene unterscheidet Klaus Brinker (2001) zwischen **expliziten und impliziten Verweisformen**. Jedoch beschäftigte er sich, wie die meisten seiner Vorläufer, zumeist **mit anaphorischen Verweisen**. Während der explizite Verweis auf **Referenzidentität** zwischen Bezugswort und Wiederaufnahme berührt, ist bei dem impliziten Verweis eher von der **Rede einer Enthaltensseinsrelation** (Brinker 2001: 36). Noch als *semantische Kontiguität* bezeichnet, entsteht diese *Enthaltensseinsrelation durch logische, ontologische oder kulturell begründete Beziehungen zwischen dem Anker Ausdruck und dem wiederaufnehmenden Ausdruck* (vgl. Brinker 2001: 36; zit. nach Koch 2001: 10). Außerdem bearbeitet Brinker (2001) die Fragen der Konfiguration (Aufbau) und der Auflösung (Verstehen) von Anaphorik. Was **die Konfiguration** angeht, bildet er die Hypothese, dass bei impliziten anaphorischen Konstruktionen, im Allgemeinen, **das Hyponym**, also der untergeordnete Begriff (**Anapher**) **dem Hyperonym**, der Oberbegriff (Antezedent/Anker), in der linearen Kette folgt (vgl. Ebd.: 36). **Was die Frage der Auflösung** von Anaphern angeht, behauptet er, *die Beziehungen zwischen den Bedeutungen* (von Antezedenten/Ankern und Anaphern) *ergeben sich dabei entweder auf der Grundlage des Sprachsystems, [...] oder auf der Basis der Sprachverwendung [...]* (vgl. Brinker 2001: 36; zit. nach Koch 2001: 10). Dies bedeutet die völlige Interpretation einer impliziten Anapher (**Auflösung**) sich entweder mit extrahierten Informationen aus dem entsprechenden Text nachvollziehen lässt oder benötigt dafür eher noch zusätzliche Wissen aus der Sprechsituation, in der der Text zum Stande gekommen ist (Weltwissen).

An der Arbeit Klaus Brinker (2001) schließt sich insbesondere Arbeiten über **Rückwärtsverweise** an. Vorwärtsverweisen wurden auch bearbeitet aber nicht so intensiv, wie Rückwärtsverweisen (vgl. Finkbeiner 2018: 186). Vielleicht liegt dies an der Tatsache, dass

anaphorische Verweise beim Sprachgebrauch üblicher als kataphorische sind. Allerdings zielten die wissenschaftlichen Auseinandersetzungen zu diesem Thema der Anaphorik zunächst, **eine allgemeine Typologie von Rückverweisen zu erstellen**. Weiterhin wurden **Fragen des Verstehens bzw. der kognitiven Verarbeitung von Anaphern** im Mittelpunkt eingerückt, wobei anschließend der **Entwurf von Regeln für den Anapherngebrauch** von besonderer Bedeutung war (vgl. Ebd.: 186-188).

2.3.2. Die expliziten Anaphern

Eine explizite oder direkte Anapher ist **eine sprachliche Einheit**, die eine andere im vorangehenden Textteil stehende sprachliche Einheit, mit der sie in einer **Koreferenzbeziehung** steht, wiederaufnimmt (vgl. Bußmann 1990: 82). Nehmen wir den folgenden Beispieltext:

(5) **Eine Frau**₁ kommt in die Botschaft. **Sie**₁ befindet sich im Warteraum.

Die indefinite Nominalphrase *eine Frau* ist der Antezedent von der Anapher *Sie*. Diese Letzte wird auch als **direkte Anapher** genannt, denn sie referieren zusammen mit ihrem Antezedenten zu dem gleichen außersprachlichen Gegenstand (eine weibliche Person). Man redet auch **von direkter anaphorischer Beziehung** zwischen den beiden Textteilen (*Eine Frau* und *Sie*) (vgl. M. Schwarz, 1996: 399). Solche Arten von Anaphora sind meist offensichtlich und **ihr Aufbau** ebenso wie **ihre Auflösung** erfolgen insbesondere über die Grammatik des Textes. Zu merken ist jedoch, die Tatsache, dass je nach dem Anapherntyp unterschiedliche grammatische Regeln bei der Verarbeitung auftreten. Im Allgemeinen werden Anaphern je nach **ihrer grammatikalischen Natur klassifiziert**. Davon ausgehend, unterscheidet man insbesondere zwischen **pronominalen Anaphern, adverbialen Anaphern, und Nominalphrasen Anaphern** (vgl. ebd.: 394).

2.3.2.1. Explizite pronominale Anaphern

Der anaphorische Verweis durch Pronomen ist die einfachste und zugänglichste Form der direkten Anaphorik (P.F. Pause, 1991: 550). Er kann durch eine Vielfalt von Pronomen vollgezogen werden; darunter zählt man beispielsweise: **Personalpronomina, Possessivpronomina, Relativpronomina, Interrogativpronomina, Reflexivpronomina, Indefinitpronomina und Demonstrativpronomina** (vgl. Finkbeiner 2018: 186). Was die Personalpronomina angeht, sind insbesondere die der dritten Person im Singular und im Plural

häufig Objekt **anaphorischer Konstruktionen**. Personalpronomina der ersten und zweiten Personen im Singular und im Plural erfüllen meistens eher eine deiktische Funktion (sie referieren direkt auf Gegenstände in der kommunikativen Sprechsituation und benötigen dabei nicht auf ein im Vordertext stehendes Wort Bezug zu nehmen) und werden somit nur **in deiktischen Konstruktionen** gebraucht (vgl. ebd.: 188).

Das Besondere an pronominalen Anaphern ist, dass **ihr lexikalischer bzw. semantischer Gehalt so gering ist**, dass der Textrezipient, um sie verstehen bzw. völlig interpretieren zu können, unbedingt auf entsprechende Antezedenten zurückgreifen muss. Dieser Rückgriff auf Antezedenten wird noch als **referentielle Verankerung** genannt und erfolgt durch **grammatische Kongruenzen zwischen Pronomen- und Antezedens-Ausdrücken** im Kopf des Textverwenders (vgl. Schwarz-Friesel/Consten 2007 zit. nach Finkbeiner 2018: 193). Dies bekräftigt Schecker (1996) wenn sie behauptet:

*Pronomina erweisen sich **als semantisch weitgehend leer** und auf was ein Sprecher oder Schreiber mit einem Pronomen **referiert**, ist deshalb nur im Rückbezug auf einen jeweiligen Kontext entscheidbar* (Schecker 1996: 161).

Dementsprechend kann man *die referentielle Verankerung **als ein mentaler Vorgang definieren, der darauf abzielt, einen passenden Antezedenten für die Wiederaufnahme (Anapher) zu finden***. Beim Verstehen direkter Pronominalanaphern erfolgt dieser Vorgang insbesondere durch **Genus-Numerus-Kongruenz**; wobei auch **die Rektion** (Festlegung des Kasus eines Satzgliedes durch ein Verb, eine Präposition oder eine andere Wortart) einen deutlichen Einfluss hat (Dipper und Zinsmeister 2011: 1-2). Jedoch gibt es Fälle, bei denen es, nach grammatischen Kongruenzen, mehr als einen potentiellen Antezedenten gibt. Man redet dann von **referentieller Ambiguität** (H. Kubczak 2016: 231). Bei solchen Fällen müssen weitere Parameter in der Suchstrategie abgerufen werden. Dazu zählen zum Beispiel **pragmatische und konzeptuelle Wissen**, die, normalerweise, insbesondere erst bei der Auflösung indirekter Anaphorik auftreten müssen (H; Zsofia, 2016). Im Folgenden werden einige Fälle beschrieben, in denen unterschiedliche **Pronominaarten, textfunktionell** betrachtet, als Anapher gebraucht werden.

1. Personalpronomen, Relativpronomen, Possessivpronomen und Interrogativpronomen

(6) „[...] **Tonio Kröger₁** las **den frommten Spruch₂**, **der₂** in halb verwischten Lettern über dem Eingang stand. Dann atmete **er₁** auf und ging hinein. Sein Herz schlug ängstlich, denner₁ gewärtigte, sein Vater könnte aus einer der Türen zu ebener Erde, an denen er₁ vorüberschritt, hervortreten, im Kontor-Rock und die Feder hinter Ohr, **ihn₁** anhalten und ihn₁ wegen seines extravaganten Lebens streng **zur Rede₃** stellen, **was₃** er₁ sehr in der Ordnung gefunden hätte
 70.

Beim Lesen des Textes (6) muss der Leser, um den Inhalt mancher Ausdrücke entschlüsseln zu können, öfters auf den Vordertext zugreifen. Dies geschieht zunächst mit dem Wort „er“ im zweiten Satz. Dieses Personalpronomen, das eine anaphorische Textfunktion erfüllt, wird auf der Basis seines lexikalischen Gehaltes, konzeptuell betrachtet, nicht vollständig verstanden: Man weist nur, es handelt sich um **eine männliche Person**, aber nicht genau um welche. Daher muss der Textrezipient das Bezugswort des Personalpronomens suchen. **Die Suchstrategie** erfolgt insbesondere durch **Genus-Numerus-Kongruenz** mit allen in dem Vordertext stehenden Nomina. Im Text (6) zum Beispiel haben wir zwei potentielle Bezugswörter (den Namen *Tonio Kröger* und die Nominalphrase *den frommten Spruch*), die eben wie die Anapher, Maskulin Singular sind. Um die beiden Kandidaten abzutrennen, muss der Leser noch weitere Wissen hervorrufen, die er entweder aus dem Textkontext oder aus seinen Weltwissen herauszieht. Bei dem Fall im Text (6) muss der Leser zum Beispiel dank dem Textkontext wissen, dass der anaphorische Ausdruck „er“ auf ein Lebewesen verweist. Und somit wird er, der Leser, **den Namen Tonio Kröger** als Antezedent für die Pronominalanapher *er* wählen. Außerdem merkt man, die beiden sprachlichen Einheiten (Name und Personalpronomen) **referenzidentisch sind**. Während der Antezedent den außersprachlichen Gegenstand **in den Text einführt, wiedernimmt ihn** die Anapher; daher ist die Rede von direkten Anaphern. Die gleichen Überlegungen gelten für die Auflösung von den weiteren im Text (6) vorhandenen direkten Anaphora: *den frommten Spruch₂* – *der₂* (**Relativpronominale Anaphorik**); *Tonio Kröger₁* – *ihn₁* (**Personalpronominale Anaphorik**) *zur Rede₃* – *was₃* (**Interrogativpronominale Anaphorik**).

2. Indefinitpronomen und Demonstrativpronomen

⁷⁰(Textauszug aus der Erzählung betitelt *Tonio Kröger*, die in dem Buch von Thomas Mann (2007) betitelt *Das Wunderkind und andere Erzählungen* erschienen ist; Seite: 100).

(7) **Die Priester**₁ gehen in den Essraum der katholischen Gemeinde. **Einer**₁ ruft den **Kellner**₂ an. Doch **dieser**₂ kommt nicht. (Ausgedachter Text)

Im Text (7) hat man mit zwei Anaphernarten zu tun. **Eine indefinite "Einer" und eine demonstrative „dieser“**. Die Auflösung der beiden Anaphern erfolgt insbesondere durch referentielle Verankerung an potentiellen Antezedenten. **Der semantische Gehalt der Indefinit-pronominalen** verrät dem Leser, dass es sich um einen Teil eines Ganzen handelt (das Indefinitpronomen bezeichnet einen Teil eines Ganzen); und das einzige Satzglied, dem im Vordertext dieses Merkmals entspricht, ist die Nominalphrase "Die Priester". **Was die Auflösung der demonstrativpronominalen Anapher angeht**, werden, nach Genus-Numerus-Kongruenzen, zwei potentielle Antezedenten "Einer" und "den Kellner" gefunden. Doch **bei referentiellen Ambiguität bei der Auflösung demonstrativpronominaler Anaphern**, wird eine andere Regel in den Auflösungsprozess einspringen, nämlich **„die Nähe/Ferne Regel“** (vgl. H. Kubczak 2016: 231).

Noch als **„Die Proximitäts-Leseart von Demonstrativa“** (vgl. Finkbeiner 2018: 193) bekannt, besagt diese Regel, dass der potentielle Antezedent-ausdruck, **der dem anaphorischen Demonstrativpronomen in der linearen Kette am nächsten steht**, wird bei **Ambiguitätsfall** präferiert, als derjenige, der von der demonstrativen Wiederaufnahme entfernt steht (vgl. Consten/Schwarz-Friesel 2007:276 in ebd.: 193). Im Text (7) zum Beispiel steht die Nominalphrase "den Kellner" der demonstrativpronominalen Anapher am nächsten. Im Text (8) ist eher den Eigennamen "Georg".

(8) „Der Killer war mit einem langen Schritt heran und schlug **Georg**₁ ins Gesicht, dass **dieser**₁ auf die Bank zurückkrachte, und irgendwie schaffte er es im Gegenschwung hoch zu kommen" (Aus dem Buch von Thomas le Blanc (2007), betitelt *die Anderen* Seite 130).

2.3.2.2. Die Nullanapher

Im Gegensatz zu anderen Anapherntypen, in denen der wiederaufnehmende Ausdruck eine sprachliche Einheit ist, wird **die Wiederaufnahme** bei Nullanaphern eher durch eine Leerstelle im Satz bezeichnet (vgl. M. Schwarz: 1997: 449). **Eine Leerstelle**, die jedoch hindeutet, dass es ein Wort fehlt (daher die Bezeichnung *Null*) und, dass trotz der Abwesenheit dieses Wortes der Satz- oder Textteil immer einen Sinn hat. Man stellt sich also die Frage, wieso dies möglich ist? Als Antwort dieser Frage ergibt es sich, dass das fehlende Wort, das meistens **das Subjekt**

des vorangehenden Hauptsatzes ist, wird aufgrund **sprachökonomischer Gründe** im Nebensatz getilgt⁷¹. Dementsprechend können wir **eine Nullanapher** als eine Leerstelle, deren Verstehen durch den Bezug auf dem im Vordertext stehenden Subjekt des Hauptsatzes möglich ist, definieren (vgl. ebd.:449). Nullanaphern gehören somit prinzipiell **zur intrasententialer Anaphemarten**, denn sie treten zumeist innerhalb von Sätzen auf. Im Folgenden Text stehen einige Beispiele davon:

(9) „**Der Polizist₁** breitete sie auf der Pultplatte aus und **Ø₁** begann, darin zu lesen. Auch Herr **Seehaase₂** trat näher herzu und **Ø₂** beteiligte sich an der Lektüre“

(Textauszug aus der Erzählung betitelt *Tonio Kröger*, die in dem Buch von Thomas Mann (2007) betitelt *Das Wunderkind und andere Erzählungen*, erschienen ist; Seite: 106).

2.3.2.3. Explizite Nominalphrasenanaphern

Eine Anapher wird als Nominalphrasenanapher bezeichnet, wenn der wiederaufnehmende Ausdruck **eine definite Nominalphrase** ist (vgl. M. Schwarz: 1997: 449). Im Gegensatz zu den meisten pronominalen Anaphern, die definite Nominalphrasen als Antezedent haben, beziehen sie sich direkte Nominalphrasenanaphern auf **im vorangehenden Textteil stehende indefinite Nominalphrasen**, mit denen sie zusammen **eine Koreferenzbeziehung** teilen (vgl. ebd.: 450). Außerdem tragen sie auch aufgrund ihrer semantischen Inhalte **zur Themaentwicklung** bei, indem sie zusätzliche Informationen über den zu referierenden Gegenstand verraten. Sie distanzieren sich somit **von Pronominalanaphern**, die lediglich nur den durch den Antezedenten eingeführten Referenten wiederaufnehmen, ohne dabei neue Informationen zu verraten (vgl. P.F. Pause 1991: 551). Im folgenden Beispiel tragen alle vier Anaphern im zweiten Satz von (10) zur Themenfortführung bei, aber nur die drei Nominalphrasenanaphern entwickeln das Thema gleichzeitig fort, indem sie neue Informationen über Michael bereitstellen.

⁷¹ Dieses Sprachphänomen ist noch unter dem Namen **Ellipse** bekannt. Jedoch besteht allerdings ein grundsätzlicher Unterschied zwischen den beiden Konzepten. Während es zum Beispiel **bei der Nullanapher** darum geht, wie schon oben hingewiesen, *das Subjekt im Nebensatz* zu tilgen, handelt es sich **bei Ellipsen** nicht nur um die Tilgung des *Nebensatzsubjekts*, sondern um die Tilgung von beliebigen Satzteilen, die dank dem Sprechkontext vom Textrezipienten nachvollgezogen werden. Dementsprechend ist jede Nullanapher eine Ellipse aber jede Ellipse ist automatisch nicht eine Nullanapher.

(10) **Michael₁** wurde gestern von der Klinik entlassen. **Der zweifache Vater₁/ /Der Maries langjährige Geliebte₁ /Der ehemalige brave Doktorand₁/ Er₁** kann zurzeit nicht mehr weiterstudieren (Ausgedachter Text).

M. Schwarz (2000b) spricht bei solchen Nominalphrasenanaphern von **Spezifikationsanaphern**, da sie den wiederaufgenommenen Diskursreferenten zusätzlich genauer charakterisieren. Sie haben also nicht nur thematischen, sondern zugleich auch rhematischen Status (vgl. M. Schwarz 200b: 111-126). Diese Eigenschaft wird auch **den Pluralanaphern** zugeschrieben.

2.3.2.4. Pluralanaphern

Unter dem Begriff **Pluralanapher** versteht man einen wiederaufnehmenden Ausdruck, meistens definite Nominalphrasen, der auf mehr als einen (meistens zwei) **Textreferenten** verweist (vgl. M. Schwarz 1997 & Finkbeiner 2018). Obwohl die Pluralanapher in einer Koreferenz-Beziehung mit ihrem Antezedenten steht, trägt sie nicht nur, wie die anderen oben erwähnten direkten Anaphern, **zur Fortführung des Themas**, sondern bringt auch zusätzliche weitere Informationen mit (vgl. Finkbeiner 2018: 186-197). Somit weist er einige Merkmale **der indirekten Anaphorik** auf (vgl. **ZSOFIA, Haase**, 2016). Im Text(11) zum Beispiel wird der pluralische Antezedent **Bellvanie und Ismael** als eine textreferentielle Einheit durch die spezifizierende Nominalphrase **das Nachbarpaar** wiederaufgenommen. Dabei werden neue Informationen über den Antezedenten verraten, nämlich, dass die beiden Personen Nachbarn sind.

(11) **Bellvanie und Ismael₁** haben heute Abend gut gelernt. **Das Nachbarpaar₁** beabsichtigt den morgigen Test fehlerfrei zu bestehen (Ausgedachter Text).

2.3.2.5. Komplexe Anaphern

Komplexanaphern werden nach **Asher (1993)** als wiederaufnehmende sprachliche Ausdrücke, die nicht auf Nominalphrasen oder Nomen, sondern **auf abstrakte Objekte (Ereignisse, Prozesse oder Zustände)** im vorhergehenden Text, zurückweisen (vgl. M. **Schwarz/C. Manfred/M. Konstanze**, 2004: 67-69). **Komplexanaphern** treten entweder in Form von **neutralen Demonstrativpronomina** (12) oder in Form von **Nominalphrasen** auf und haben **eine propositionale Struktur** (Textteil, Satz oder Verbalphrase) **als Antezedent** (C. Manfred

& M. Schwarz, 2007: 275). Im Text (12) zum Beispiel fungiert der gesamte erste Satz als Antezedent der demonstrativen Pronominalanapher.

(12) [„Ich gebe das nicht nur an“, sagte Tonio Kröger und machte eine nervöse Bewegung mit den Schultern]₁. Dies₁ rief einen gewissen Eindruck hervor.

(Textauszug aus der Erzählung betitelt *Tonio Kröger* erschienen in dem Buch von Thomas Mann (2007) betitelt *Das Wunderkind und andere Erzählungen* Seite 103).

Eben wie Pluralanaphern gehören Komplexanaphern zur Kategorie **der expliziten Anaphern**, denn ihre Auflösung bzw. ihr Verstehen benötigt unbedingt den Rückgriff auf den im Vordertext stehenden Textreferenten, mit welchem sie in **einer ko-referenziellen Beziehung** stehen (vgl. ebd.: 276). Doch im Gegensatz zu anderen direkten Anaphern, aber im Einklang mit Pluralanaphern, tragen sie auch zur Entwicklung des Themas, indem sie zusätzliche Informationen (*In Form von Evaluationen oder Metadiskursiven Einordnung über den wiederaufgenommenen Textreferenten geben*) (vgl. Schwarz-Friesel/Consten/Marx 2004, zit. nach Finkbeiner 2018: 194). Dies wird insbesondere durch diejenigen Komplexanaphern, die **die Form definiter Nominalphrasen** haben, geleistet, wie es im (13) der Fall ist.

(13) „Schweigend standen wir noch eine Weile beisammen, während Peter sich allmählich beruhigte. Nach einer unerträglich langen Zeit sagte er dann [...] „Aber unseren Eltern sagen wir davon nichts?!“₁ [...] Dieses Geheimnis₁ müsste gehütet werden, koste es was es wolle (Aus dem Buch von Thomas le Blanc (2007), betitelt *die Anderen* Seite 120).

2.3.3. Die impliziten Anaphern

Noch als *assoziative Anaphern* (Kleiber 2001a), *Kontiguitätsanaphern* oder noch *inferierte Anaphern* (Pause 1991) genannt, werden *indirekte* oder *implizite* Anaphern von M. Schwarz (2000a) als **definite Nominalphrasen**, die, im Gegensatz zu expliziten Anaphern, **keinen expliziten bzw. direkten Antezedent-ausdruck im Vordertext haben** (vgl. M. Schwarz 200a, Kleiber 2001 & P.F. Pause 1991). An der Stelle eines Antezedent-ausdruckes steht **ein Bezugswort**, das noch als *Anker* bezeichnet wird, mit dem der wiederaufnehmende Ausdruck (Anapher) **in keiner Koreferenzbeziehung** steht (vgl. Finkbeiner 2018: 194). Dieser Ankerausdruck wird noch als **impliziter, indirekter, inhaltlicher Antezedent** genannt, denn

er pflegt mit der Anapher eine systematische (*schematische, semantische oder diskursbasierte*)⁷² Beziehung. (ebd.: 194).

Ebenso wie beim Verstehen direkter Anaphorik, erfordert **die völlige Interpretation einer indirekten Anapher** einen Bezugsausdruck (Kenntnisse über den Textreferenten) (vgl. **ZSOFIA, Haase**, 2016). Jedoch im Gegensatz zur Auflösung direkter Anaphorik, wobei die Identifizierung des Antezedenten insbesondere über grammatische Kongruenzen zwischen potentiellen Bezugswörtern und der direkten Anapher erfolgt, erfordert **die Identifizierung des Ankerausdrucks** von dem Rezipienten den Abruf, durch kognitive Operationen, von zusätzlichen Wissen. Dies geschieht durch die Etablierung **von Brückenannahmen** (vgl. H. **Kubczak**, 2016:221-247). Noch **als Bridging oder Inferenz** genannt, verweist der Begriff *Brückenannahme* **auf die Bildung von Hypothesen, die die möglichen Bezüge zwischen Anaphern und ihrem jeweiligen Ankerausdruck erklären können** (vgl. H. Bußmann 1990: 418).

Die Erschließung des Bezugs, der zwischen Anaphern und Ankern besteht und der zugleich beim Verstehen der indirekten Anaphorik helfen wird, verlangt vom Textrezipienten **inferentiell-pragmatische Kompetenzen** (S. **Dipper & H. Zinsmeister** 2001: 37-52). Dies wird durch das folgende Beispiel veranschaulicht:

(14) Der Aufzug₁ ist außer Betrieb. Das Hängekabel₂ muss vom Techniker gewechselt werden
(Ausgedachter Text).

Beim Verstehen des Textes **(14)** stößt der Textrezipient auf das Problem **der kohäsiven Verbindung** zwischen den beiden Sätzen. Diese ist, grammatisch betrachtet, also **textoberflächlich** nicht erfassbar. Um eine beliebige Verbindung zwischen den beiden Textteilen zu erschließen, könnte der Textverwender zunächst einmal versuchen zu bestimmen, um welches **Hängekabel** es sich tatsächlich handelt. Dafür muss er auf dem Vordertext zugreifen, indem er mit allen dort-stehenden möglichen Bezugswörtern **Zusammenhangshypothesen** schlagen wird. Bei dem Fall **(14)** gibt es nur ein möglicher Bezugsausdruck, nämlich die definite Nominalphrase „Der Aufzug“. Weiterhin muss der Rezipient **den Zusammenhang**, der zwischen den beiden Ausdrücken besteht, identifizieren können. Bei dem Fall **(14)** handelt es sich um **eine Teil-Ganze Relation**: die Anapher ist ein

⁷² Solche Beziehungen stehen also in Opposition zu **morphosyntaktischen Beziehungen**. Denn während diese Letzten sich **auf der Sprachoberfläche** (daher *Explizit*) realisieren, sind die Ersteren, ihrerseits, eher **auf der unteren Ebene der Sprache** (und deswegen auch als *implizite* Rückverweise bezeichnet) zu nachvollziehen.

Teil des Ankers; um genauer zu sein, ist die Nominalphrase „**Das Hängekabel**“ ein **semantischer Teil des Ankerausdrucks „Der Aufzug“**. Die völlige semantische bzw. konzeptuelle Repräsentation des Anaphern-Ausdrucks erfolgt also durch den Rückgriff auf seinen semantischen oberen Teil (den Anker). Es besteht somit **eine Kontiguitäts-beziehung** (hier **die Meronymie**: Sie ist die semantische Teil-von-Relation) zwischen den beiden Ausdrücken (vgl. H. Bußmann 1990: 418).

H. Bußmann (1990) definiert den Begriff *Kontiguität* als **ein semantischer Begriff, welcher sich auf die Relation zwischen Lexemen bezieht, die der gleichen semantischen, logischen, kulturellen oder situationellen Sphäre angehören** (vgl. Bußmann 1990: 418, zit. nach Waldmüller 2006: 9). Unter diesen möglichen Relationen zwischen Lexemen werden hauptsächlich bei indirekten anaphorischen Konstruktionen drei Typen unterschieden: **Die Meronymie**, die als die typischste und häufigste Form der indirekten Anaphorik gilt (Anapher als semantischer Teil des Ankers oder Anker als semantischer Teil der Anapher); **die Metonymie** (Ersetzung eines Begriffs (Ankers) durch einen solchen (Anapher), der in gedanklicher Beziehung zu dem zu bezeichnendem (Anker) steht) und **die Hyponymie** (paradigmatische, semantische Relation der Unterordnung zwischen Ankern und Anaphern) (vgl. Finkbeiner 2018: 194-196 & M. Schwarz 1994: 81-82). Zur Illustration von Kontiguitätsbeziehungen zwischen Ankern und Anaphern können wir die folgenden Beispiele betrachten:

(15) „Das Eis klapperte gegen die Wände **seines Glases₁**, **die Flüssigkeit₂** perlte lange Ketten“(In dem Buch von Thomas le Blanc (2007), betitelt *die Anderen* Seite 103)

(16) Einige Monate Später mussten wir **den Baum₁** fallen. **Seine Äste₂** waren so lang, dass sie die elektrische Spannung fast berührten (Ausgedachter Text).

(17) Ich habe mir **ein gebrauchtes Auto₁** gekauft. **Der Benzintank₂** hat ein dickes Loch (Ausgedachter Text).

Das kognitive Schema, das darin besteht, zu wissen, dass Flüssigkeiten nur im Glas enthalten werden können und nicht im Eis, da dieses Letzte eben gefrorene Flüssigkeit ist, hilft beim Verstehen der indirekten Anaphorik im Text (15). **Diese Brückeninferenz oder -annahme** hilft dem Textrezipienten **dem Text eine Kohärenz zuzuschreiben**. Denn die Nominalphrase „seines Glases“ hilft dem Leser eine referentielle Verankerung zu schlagen, welche für eine auffassende Repräsentation der semantischen Merkmale der definiten Nominalphrase „die Flüssigkeit“ wichtig ist. Etwas, was letztendlich dazu führt, **dem Text eine Kontinuität**

(Kohäsion) und eine thematische Entwicklung (Kohärenz) zu gewährleisten. Das Verstehen dieses Textes erfolgt also erst **auf der inferentiell-pragmatischen Ebene** (vgl. S. Dipper & H. Zinsmeister 2001: 37-52). Das Gleiche wird auch bei Texten (16)) und (17) beobachtet.

Wie oben schon angedeutet, erscheinen indirekte Anaphern zumeist in Form von Nominalphrasen. Jedoch erfasst man auch einige Fälle, in denen sie in Form von Pronomina erscheinen (vgl. H.Kubczak, 2016). Betrachtet man den Text (18) zum Beispiel, dann merkt man, die Pronominalanapher „**Sie**“ teilt mit den beiden potentiellen Ausdrücken „**der Universität**“ und „**meine Einschreibungsprozedur**“ keine direkte Beziehung. In der Tat, werden diese beiden Ausdrücke nach Genus-Numerus-Kongruenzen als mögliche Bezugsausdrücke ausgewählt. Jedoch kann **der Anaphern-Ausdruck** nur einen von den beiden als Textreferent haben. Erst mit **der konzeptuellen Repräsentation**, dass Einschreibungsprozeduren vom Personal der Universität geführt werden, dass der Textverwender die definite Nominalphrase „**Der Universität**“ **als Anker Ausdruck der pronominalen Anapher „Sie“ auswählen wird**. Somit wird die indirekte Anaphorik aufgelöst. Wie schon oben behauptet, referieren Anker-und Anaphernausdrücke auf ganz verschiedene außersprachliche Referenten (vgl. H. Zsofia, 2016). Es handelt sich also um **eine indirekte pronominale Anaphorik**, wobei der wiederaufnehmende Ausdruck als **Subjekt des Satzes**, in dem er auftritt, fungiert. Etwas das nicht der Fall bei dem Beispiel (19), in dem **die indirekte Pronominalanapher** eher die Funktion **eines Akkusativobjekts** trägt. Allerdings erfolgt, wie in dem Text (18), **die referentielle Verankerung** durch das kognitive Schema, dass „ihn“ auf einen Wagen, der, obwohl er keine Parkplatzberechtigung „auf dem Lehrerparkplatz“ besitzt, trotzdem immer dort geparkt wird.

(18) Gestern war ich an **der Universität**₁, um meine Einschreibungsprozedur zu beenden. **Sie**₂ haben gesagt, ich müsste erst in zwei Tagen wiedervorbeikommen (Ausgedachter Text).

(19) *Ich wäre wunschlos glücklich, wenn Sie nicht immer auf dem **LEHRERPARKPLATZ PARKEN** würden. Das nächste Mal lasse ich **ihn** abschleppen.* (Beispiel aus der Seifen-Oper Beverly Hills 90210, RTL, 31.8.95, in Consten, 2001, zit. in Cornish 2007: 23).

Im Anschluss an diese beiden Arten von indirekten pronominalen Anaphern unterscheidet man auch **die indirekten pronominalen Gattungsanaphern** (O. Guérin 2014:3091-3110). Diese Anaphernarten zeichnen sich aus, durch die Tatsache, dass **sie auf die Gattung** referieren, zu

der **der Referent des Ankerausdrucks** gehört (vgl. ebd.: 3099). Die implizite pronominale Gattungsanapher teilt also mit ihrem Anker nicht den gleichen Referenten, denn sie verweist **auf die ganze Gruppe**, während der Anker Ausdruck eher nur **auf ein Objekt/Element dieser Gruppe** verweist (vgl. ebd.: 3100). Zur Illustration gelten die folgenden Beispiele:

(20) Ich habe **eine Toyota** gekauft, weil **sie** robust und preiswert sind. (ein Beispiel aus dem Artikeln von O. Guérin 2014:3100.; eigene Übersetzung (Französisch zu Deutsch)).

(21) Ich habe **eine Katze** adoptiert, weil **sie** liebevoll sind. (ein Beispiel aus dem Artikeln von Guérin 2014: 3101; eigene Übersetzung (Französisch zu Deutsch)).

Nachdem wir das Phänomen der Anaphorik als Gegenstand der Textlinguistik erörtert haben. Beabsichtigen wir nun, uns auf eine besondere Art der expliziten/direkten Anaphern (die Pronominaladverbienanapher), zu konzentrieren.

**ZWEITER TEIL: THEORETISCHE ANNÄHRUNGEN
(PRONOMINALADVERBIEN & AR-SYSTEME)**

KAPITEL 3: PRONOMINALADVERBIEN (PAn) UND DIE ANAPHORIK: Das Pronominaladverb (PA) als Anapher

Dieses dritte Kapitel ist als eine zielorientierte Folge des zweiten Kapitels gedacht und ausgebaut worden. Wie die Überschrift schon hinausweist, handelt es sich hier um die Bearbeitung von einem besonderen Anapherntyp (**die Pronominaladverbienanapher**). Die Notwendigkeit dieses Anapherntyps in einem einzelnen Kapitel⁷³ zu bearbeiten lässt sich, vor allem, dadurch begründen, dass die maximalen Kenntnisse über die Wortart, womit dieser Anapherntyp gebildet wird, vorerst gewonnen werden müssen; bevor man etwas über das Funktionieren des entsprechenden Anapherntyps auffassen kann. In dieser Hinsicht wurde dieses Kapitel in zwei semantische Einheiten eingeteilt. Während die erste uns Informationen über die Bestimmung und Klassifikation vom Wort **Adverb** und dessen semantisches Synonym **Adverbial** wiedergibt (Abschnitt 3.1 & 3.2), hilft uns die zweite semantische Säule dabei, die Wortart **Pronominaladverb** in einer breiten Ansicht zu verarbeiten, wobei der Schwerpunkt auf ihre textlinguistischen bzw. **anaphorischen Eigenschaften** anschließend gesetzt wird (Abschnitt 3.3). Etwas, das uns dabei geholfen hat, schon am Ende dieses Kapitels in der Lage zu sein, einen vorübergehenden Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern aufzubauen (Abbildung 1).

3.1. Adverb oder Adverbial: Begriffsbestimmung

Etymologisch betrachtet, bezieht sich der Begriff **Adverb** (*Plural: die Adverbien*) auf das lateinische Wort **adverbium** und bedeutet, wortwörtlich übersetzt, **was beim Verb steht**. (Hoffmann, 2009:223). Doch interessanter ist auch festzustellen, dass dieses lateinische Wort (*adverbium*) nur eine Übersetzung des griechischen Ausdrucks **epirrhema** (*altgriechisch επίρρημα*) ist und bedeutet, immer wortwörtlich übersetzt, **Beiwort zur Satzaussage (dem ῥήμαρHEMA)** (Gisa Rauh, 2015: 19). So gesehen, wurde der Begriff **Adverb**, ursprünglich betrachtet, nicht exklusiv, wie es etwa der Sinn der Bezeichnung *Ad-Verb* fordert, als Beiwort für Verben aufgefasst, sondern auch als Beiwort für andere **Wortkategorien** (wie etwa Substantiv; Adjektiv und Adverb selbst) und bei ggf. auch für manche **Wortgruppen** (wie etwa Nominalphrasen (NP) und Präpositionalphrase (PP)). Über dieses Bestimmungskriterium

⁷³ Also nicht im zweiten Kapitel, wie die meisten Anapherntypen.

(**Bezugsentität**) hinaus, wird der Begriff **Adverb** auch sehr oft in **der traditionellen deutschen Grammatik** nach seiner **inhaltlichen Bedeutung** bestimmt. Diesem Kriterium nach, werden Wörter, die Angaben über die Art und Weise; den Ort; die Zeit und den Grund eines Geschehens vermitteln, als **Adverbien** betrachtet (Pittner & Berman, 2010: 21). Doch **diese semantische Bestimmungsvorgehensweise**, im Gegensatz zum zuvor obig erwähnten **grammatischen Ansatz**, überschneidet sich mit **der funktionalen Betrachtung des Adverbs**. In der Tat, wird die **Funktion** eines Wortes oder einer Wortgruppe, die eine Handlung bzw. ein Geschehen näherbestimmt, als **Adverbial**⁷⁴ bezeichnet. Diesem Bestimmungskriterium nach, können also **Wortgruppen** nicht nur als Bezugsentität von Adverbien erscheinen, sondern auch selbst als **Adverbien** fungieren. Von dieser Feststellung ausgehend, kann man also den Begriff **Adverb** entweder als **adverbiales Einzelwort** (Ein Wort bezieht sich auf ein Verb) oder als **adverbiale Wortgruppe** (eine Wortgruppe bezieht sich auf ein Verb) bestimmen. Allerdings muss man zu dieser Schlussfolgerung noch etwas nuancieren. Denn diese beiden möglichen Bestimmungen eines Adverbs (**adverbiales Einzelwort & adverbiale Wortgruppe**) berücksichtigen überhaupt nicht den Fakt, dass manche Adverbien gar keine **adverbiale Funktion** in der sprachlichen Kette aufweisen können. In der Tat kommt es manchmal vor, dass Adverbien statt sich auf Verben⁷⁵, zu beziehen, sich eher auf Substantive beziehen und bilden zusammen ein Satzglied, das entweder als Subjekt oder Objekt im Satz fungieren kann. Und im Gegensatz zu Adjektiven stehen sie, aber direkt nach dem Substantiv und bleiben unflektiert. Dies ist meistens der Fall

⁷⁴Ebenso wie die **Wortart** Adverb von **der Funktion** Adverbial (**adverbiale Bestimmung**) zu unterscheiden ist, sind die Wortarten Substantiv und Verb jeweils von **der Funktion** Subjekt/Objekt und Prädikat zu differenzieren. Allerdings können Adverbien, außer der adverbialen Funktion, verschiedene weitere Funktionen im Satz ausüben. Darunter zählen beispielsweise **die attributive, prädikative und ergänzende Funktion** (vgl. Pittner & Berman, 2010: 37).

- Von **der adverbialen Funktion** ist es die Rede, wenn das Adverb **als selbstständiges Satzglied** verwendet wird und sich dabei entweder auf **das Verb** oder auf **den ganzen Satz** bezieht. Als Beispiel gelten die beiden folgenden Sätze: 23) Heute will ich weiterforschen. 24) Gestern war ich unterwegs.
- Von **der attributiven Funktion** ist es die Rede, wenn das Adverb **kein selbstständiges Satzglied mehr** ist. Sondern bildet es, zusammen mit dem Wort bzw. Wortgruppe, das es näherbestimmt, also als **Attribut** gilt, **ein Satzglied**. Dieses Satzglied kann entweder **Subjekt oder Objekt** im Satz sein. Als Beispiel dafür gilt der folgende Satz 25) **Sie fährt sehr langsam**. Philosophiert man nun etwa über die möglichen grammatischen Funktionen des Modaladverbs *sehr*. Dann ergibt es sich, dass dieses Modaladverb in einer **attributiven Funktion** erscheint, vor allem, weil es sich direkt auf das Adjektiv *langsam* bezieht. Allerdings könnte man ihm auch **eine adverbiale Funktion** zuschreiben, denn das zusammen mit dem Adjektiv gebildete Satzglied, sich wiederum selbst auf das Verb bezieht. Zusammengefasst, kann also gesagt werden, dass das Adverb *sehr* im 25 eine attributive Funktion hat, aber das Satzglied *sehr nett* eher sowohl **eine adverbiale als auch eine prädikative Funktion** aufweist.
- Von **der ergänzenden Funktion** ist es die Rede, wenn das Adverb direkt **hinter der Präposition** steht. Als Beispiel dafür gilt der folgende Satz 26) Ich bin in **frühestens** drei Monaten wieder da!

⁷⁵Adverbien können **Verben näher bestimmen** (↑ 1186, 1195); dabei können sie als **Angaben** (Adjunkte) sowie als **Ergänzungen** vorkommen: Angabe: **Dort** wird gefeiert. (Feiern bedarf **keiner Ergänzung**.) Ergänzung: Sie wohnt **dort**. / **Dort** wohnt sie. (Wohnen bedarf **einer lokalen Ergänzung**) (vgl. (Dudengrammatik, 2009: 571/ Rdnr. 844).

von **Orts- und Zeitadverbien** (Artemis Alexiadou, 2013: 459). Nehmen wir zur Illustration dieses Beispiel: 27) Das Buch **dort** gehört der Familie Sofacs. In diesem Beispiel wird also bemerkt, dass das Umstandswort **dort** zusammen mit der NP (**Das Buch**), grammatisch betrachtet, als **Subjekt** des Satzes fungieren. Dabei bleibt das Wort **dort** nichtflektiert und dies, obwohl es sich auf ein Substantiv bezieht. Davon ausgehend, wird dieses Wort als **Adverb** bezeichnet, das in **keiner adverbialen Funktion** auftritt, sondern eher, zusammen mit einer NP, eine Funktion als Subjekt aufweist. Ferner ist auch festgestellt worden, dass Adverbien auch weitere Adverbien⁷⁶ und sogar ganze Sätze⁷⁷ näher beschreiben können.

Im Gegensatz zur traditionellen deutschen Grammatik, in der **das inhaltliche Kriterium** den Vorrang hat, werden in **der gegenwärtigen Darstellung der deutschen Grammatik syntaktische Parameter** zur Bestimmung des Begriffs **Adverb** berücksichtigt. Darunter zählen beispielsweise die **Nichtflektierbarkeit** und das **Satzgliedstatut** (Fähigkeit eines Worts vor dem finiten Verb (**Vorfeldposition**) zu stehen). Nach diesem Kriterium grenzen sich also Adverbien, erstens, von **flektierbaren Wortarten**, wie Adjektive ab und zweitens von **nichtflektierbaren Wörtern**, die nicht für sich allein den Status eines Satzglieds haben können (Zu dieser gehören, vor allem, **Präpositionen** und manche nicht-vorfeldfähigen **Partikeln**. (Dudengrammatik, 2009: 572 / Randnr. 845).

3.2. Klassifikation von Adverbien: Inhaltliche und Funktionale Kriterien

In diesem Abschnitt werden wir Adverbien, je nach ihren inhaltlichen und funktionalen Merkmalen klassifizieren. Das Ziel dieses Abschnitts, ist es die Arten und Rollen von Adverbien innerhalb von Sätzen bzw. Texten einzusortieren und somit den Einstieg in den besonderen Teilbereich unserer Arbeit (**Pronominaladverb**, Gegenstand dieser Arbeit) vorzubereiten.

3.2.1. Der inhaltliche Faktor: Semantische Klassen von Adverbien

⁷⁶Adverbien können andere **Adverbien** modifizieren: Die Notbremse befindet sich **rechts** oben. Der Briefträger kommt **schon** morgens (vgl. Ebd.: 571).

⁷⁷ Manche Adverbien beziehen sich auf **den gesamten Satz**. Dies gilt grundsätzlich für die sogenannten **Kommentaradverbien** (↑ 868): **Natürlich** haben sie gewonnen. Sie kommt **vielleicht**. **Leider** geht er nicht zu der Sitzung (vgl. Ebd.: 571).

Wie gerade oben eingeführt, sagen uns Adverbien, unter welchen **Umständen** etwas geschieht. Dafür **beziehen**⁷⁸ sie **sich** auf sprachliche Einheiten, wodurch das Geschehen versprachlicht worden ist. Daher werden sie auch in manchen Forschungsliteraturen als **Umstandswort**, **Beiwort** oder noch **Eigenschaftswort** bezeichnet. Doch im Gegensatz zu anderen Eigenschaftswörtern wie **Adjektiven**, die Personen oder Sachen näher bestimmen, beschreiben Adverbien, im allgemeinen, eher **die näheren Umstände einer Tätigkeit, eines Ereignisses, eines Vorganges oder eines Zustandes**. Des Weiteren gehören **Adverbien** zusammen mit **Partikeln; Präpositionen** und **Junktionen** zur grammatischen Gruppe der **nicht flektierbaren Wortarten** und können somit an mehreren Stellen im Satz stehen (vgl. Dudengrammatik, 2009: 568). Doch im Gegensatz zu Präpositionen und Junktionen lassen sich Adverbien nicht **nur steigern (Positiv, Komparativ und Superlativ)**, sondern auch nach verschiedenen Kriterien einstufen. Man kann, in der Tat, Adverbien je nach ihrem **lexikal- semantischen Gehalt; ihrer morphologischen Ausprägung** (Bildungsweise); **ihrer funktionalen Rolle und ihrem Geltungsbereich (Skopus)** u.a. **einordnen**. Was den inhaltlichen Gesichtspunkt angeht, werden Adverbien hauptsächlich in vier Gruppen eingeteilt: Lokal-; Temporal-; Kausal- und Modaladverbien. Zu dieser Einteilung sagt die Dudengrammatik, nämlich, dass *es weitaus mehr lokale und temporale Adverbien als modale gibt, wobei die kausalen Adverbien die kleinste Gruppe ausmachen* (vgl. ebd. 572).

1. Lokaladverbien - Adverbien des Ortes

Als Lokaladverbien werden Adverbien bezeichnet, die Auskunft über den Ort eines Geschehens geben. Es wird zwischen **Ortsadverbien** (statisch) und **Richtungsadverbien** (direktional) unterschieden. Die Richtungsadverbien enthalten oft das Element -hin oder -her und antworten auf die Fragen Wohin? und Woher? Zu den lokalen Adverbien im Allgemeinen gehören: **überall, draußen, vorn, hinten, hier, dort, dorthin, zuoberst, links, rechts, da, irgendwo, irgendwohin, irgendwoher, irgendworauf, irgendwodurch, nirgends, nirgendwo, nirgendwohin, nirgendwoher etc...** Weiterhin, werden, wie schon zuvor erwähnt, auch

⁷⁸ Wir werden im Rahmen dieser Arbeit weiterhin zwischen **zwei Verben** unterscheiden, die anscheinend die gleiche Bedeutung haben, die aber unterschiedlich im Rahmen dieser Dissertation eingesetzt werden. Diese sind die Verben, die **sich auf etwas beziehen** und **auf etwas referieren**. Während der Gebrauch des Letzteren **einem phorisch-deiktischen Gebrauch** unterlegen wird, wird der Gebrauch des Ersteren eher nur **grammatisch orientiert**. Denn es handelt sich dabei um Abhängigkeiten bzw. **grammatische Abhängigkeiten zwischen Wörtern einer selben Wortkette**. Nehmen wir zur Illustration des Gebrauchs dieser beiden Verben das folgende Beispiel. 27) Das Buch **dort** gehört der Familie Sofacs. Wie schon bereits erklärt, fungiert das Adverb **dort** zusammen mit der NP (Das Buch), **grammatisch** betrachtet, als Subjekt des Satzes, wobei das **Adverb Dort sich** auf das **Substantiv Buch** **bezieht**. **Doch textlinguistisch bzw. phorisch-deiktisch** betrachtet, fungiert dieses Lokaladverb eher als ein **Deixis** und **referiert** somit **auf** einen Ort, nämlich den, worauf das Buch steht.

Wortgruppen (Meistens NP & PP), die sich auf Verben beziehen, auch als „Adverb“ aufgefasst werden. Denn mit ihrem Bezug zum Verb erfüllen sie **eine adverbiale Funktion** und werden in manchen Literaturen (Dudengrammatik, 2009:1048; Pittner& Berman, 2010: 38-47; Altmann &Hahnemann 2007; Haider, 2010: 191–192) dementsprechend als **Lokaldverbialen** genannt, falls sie einem Ort bezeichnen. Und im Einklang mit den Lokaladverbien werden sie mit den gleichen Fragewörtern abgerufen. Nehmen wir zur Illustration den folgenden Satz **28)** Nelson geht im Park campen. Im Beispielsatz **28)** fungiert das Satzglied „im Park“ als **Lokaladverbialen** und wird durch die Frage „Wo geht Nelson campen?“ abgerufen. Ferner, wurde auch festgestellt, dass bei solchen Fällen, also bei Fällen, in denen **die adverbiale Funktion** von einer Wortgruppe geführt wird, die grammatischen Funktionen **Präpositionalobjekt** und **adverbiale Bestimmung** leicht zu verwechseln sind⁷⁹. Nehmen wir, zur Illustration, den folgenden Beispielsatz **29)** Nelson denkt **in seinem Zimmer** an seine Zukunft. Hier fungiert das präpositionale Satzglied **in seinem Zimmer** als Lokaladverbial, während das andere präpositionale Satzglied des Satzes **an seine Zukunft** eher ein **Präpositionalobjekt** ist. Die gleichen Überlegungen und Illustrationen gelten für die folgenden Paare **Temporaladverbialen⁸⁰ und Temporaladverbien; Modaladverbialen und Modaladverbien⁸¹; Kausaladverbialen und Kausaladverbien⁸².**

2. Temporaladverbien - Adverbien der Zeit

Als Temporaladverbien werden Adverbien bezeichnet, die Auskunft über die Zeit eines Geschehens geben. Somit tragen Temporaladverbien dazu bei, Sachverhalte im zeitlichen Raum zu situieren. Dabei bestimmen sie, so der Dudengrammatik (2009), entweder **den Zeitpunkt; die Dauer oder die Frequenz** eines Geschehens; und werden durch die folgenden Fragen abgerufen Wann? Bis wann? Seit wann? Wie oft? und wie lange? Zu ihnen gehören beispielweise **montags, abends, wieder, heute, jetzt, sofort, stets, nochmals, derzeit, zwischendurch, zunächst, bald, danach, davor, gestern, heute, immer, nie, oft, regelmäßig, täglich, sofort, irgendwann, einmal, mal, nie, niemals etc...**

⁷⁹<https://www.contify.de/glossar/richtig-schreiben/adverbiale-bestimmung/> (Abgerufen am 01.05.2022)

⁸⁰Beispiel von Temporaladverbialen₁ und Temporaladverbien₂: **30) Kunta und Nelson gehen an jeden Freitag₁ zum Sport.** (Während das Satzglied **an Jeden Freitag** als Adverbiale fungiert, übernimmt das Satzglied **Zum Sport** eher die syntaktische Funktion des Präpositionalobjekts) **31) Kunta und Nelson gehen freitags₂ zum Sport.**

⁸¹Beispiele von Modaladverbialen₁ und Modaladverbien₂: **32) Mit großem Glück₁ gab ich dieser Versuchung nicht nach! 33) Glücklicherweise** gab ich dieser Versuchung nicht nach!

⁸² Beispiel von Kausaladverbialen₁ und Kausaladverbien₂: **34) Wegen Autolärm₁ kann ich nicht weiterschlafen. 35) Autos machen vieler Lärm, deswegen₂ kann ich nicht weiterschlafen.**

3. Kausaladverbien - Adverbien des Grundes

Kausaladverbien, wie die Bezeichnung schon verrät, geben Informationen über den Grund bzw. die Ursache eines Geschehens oder einer Handlung an. Der Dudengrammatik nach, enden Kausaladverbien meistens mit *-(et)wegen oder -halber* und antworten auf Fragen wie Weshalb? Warum? und Wozu? Zu ihnen gehören nämlich: **folglich, darum, dennoch, notfalls, anstandshalber, daher, deshalb, also, demnach, somit, trotzdem, irgend wozu, meinetwegen, gesundheitshalber, anstandshalber, deutlichkeithalber etc...** Über diese semantischen und morphologischen Merkmale hinaus, weisen Kausaladverbien, textlinguistisch betrachtet, auch **anaphorische Funktionen** auf. In der Tat, beziehen sich die meisten kausalen Adverbien (wie etwa infolgedessen, folglich, deswegen, deshalb) **auf den vorangehenden Satz** und weisen somit zugleich textverknüpfende Funktionen auf. Aus diesem Grund werden sie auch zur Kategorie der **Konjunkionaladverbien** gezählt (vgl. Dudengrammatik, 2009: 864– 86 7).

4. Modaladverbien - Adverbien der Art und Weise

Als Modaladverbien werden Adverbien bezeichnet, die entweder Auskunft über die **Art und Weise (Qualität)** oder die **Menge und Ausmaß (Quantität)** geben, in der sich eine Handlung nachvollziehen lässt. Davon ausgehend, überschneiden sich **Modaladverbien** und **Satz-oder Kommentaradverbien**; denn sie geben beide Hinweise über die Modalität eines Geschehens. Daher werden sie auch beide zusammen zur übergeordneten Kategorie **der Modalwörter** eingestuft. Modaladverbien antworten auf folgende Fragen Wie? Auf welche Weise? Wie sehr? Wie viel? Und zu ihnen gehören Adverbien wie **blindlings, gratis, einigermaßen, überhaupt, ungefähr, anders, sehr, gern, ziemlich, möglicherweise, barfuß, hoffentlich, irgendwie, genauso, genug, kaum, vergebens, nebenbei, schnellstens, beispielsweise, ebenfalls, einigermaßen, teilweise etc....**

Letztendlich wird also gemerkt, dass diese Klassifikation nach dem inhaltlichen Kriterium exklusiv Adverbien berücksichtigt, **die lexikal-semantisch** betrachtet (man kann an deren semantischen Gehalt rankommen ohne zuvor sich auf ein anderes Wort stützen zu müssen), selbständig sind. Daher werden sie also **Absolute Adverbien** genannt und stehen somit den sogenannten **Pro-Adverbien** gegenüber (Dudengrammatik, 2009: 572; Rdnmr 845).

3.2.2. Nach dem Funktionalen Faktor: Funktionsklasse von Adverbien

Adverbien lassen sich, wie schon oben gerade eingeführt, nicht nur nach ihrer Bedeutung, sondern auch nach ihrer **Funktion innerhalb sprachlicher Ketten** einstufen. Außer deren

Grundfunktion als **Adverbialen** bzw. Adverbialbestimmung, weisen Adverbien auch weitere Funktionen auf. Man unterscheidet, in der Tat, Adverbien **die pronominale Funktion** ausüben (**Pro-Adverbien**). Das ist beispielsweise der Fall von **Pronominaladverbien**; doch zu ihnen kommen wir nochmal im Rahmen eines speziellen Abschnitts zurück (Siehe hierzu Abschnitte 3.2 & 3.3). Doch geben wir im Voraus einige Beispiele von **phorisch-deiktischem** Gebrauch einiger Adverbien. Wie schon in den ersten Abschnitten dieses Kapitels erwähnt, haben wir **mit einem deiktischen Gebrauch** zu tun, wenn sich Adverbien direkt auf außersprachliche Gegenstände referieren, ohne zuvor auf eine sprachliche Einheit verweisen zu müssen. Wenn aber der Sprachverwender sich auf eine zuvor oder nachher stehende sprachliche Einheit stützen muss, um den Referenten des Adverbs zu erschließen, dann handelt es sich entweder um **einen anaphorischen** oder **kataphorischen Gebrauch** von Adverbien. Nehmen wir zur Illustration die folgenden Satz- und Textbeispiele.

36) Dort musst du gehen (**deiktischer Gebrauch**). **37)** Ich werde nach **Tokyo** fliegen. **Dort** werde ich an eine Veranstaltung teilnehmen (**anaphorischer Gebrauch**).

38) Dort werde ich übernachten. **Im Haus meines Vaters** (**Kataphorischer Gebrauch**).

Ferner, bezeichnet Artemis Alexiadou (2013) diese Kategorie von Adverbien, die entweder verweisende oder referierende Funktion aufweisen als **Demonstrativadverbien**. Dazu sagt er nämlich Folgendes:

***Ein Demonstrativadverb** kann auf einen Umstand verweisen, der in der Sprechsituation zu beobachten ist (**Deixis**) oder im Text zuvor erwähnt wurde (**Anaphorik**), zum Beispiel kann auf die Art und Weise eines Geschehens verwiesen werden (**demonstratives Modaladverb**, erstes Beispiel unten) oder auf einen Ort (zweites Beispiel): **So** möchte ich auch tanzen können! **Dort** brüten viele Pinguine (Artemis Alexiadou, 2013: 459).*

Über den deiktisch-phorischen Gebrauch von Adverbien hinaus, unterscheidet man auch, Adverbien, die Relativsätze einleiten; diese werden als **Relativadverbien** genannt und teilen gleiche lexikalische Eigenschaften (w-Wörter) mit denjenigen Adverbien, die Interrogative Sätze einleiten: die sogenannten **Interrogativadverbien**. Doch während Interrogativadverbien entweder an der Satzspitze eines selbstständigen w-Fragehauptsatzes (mit Verbzweitstellung) oder an der Spitze eines w-Fragenebensatzes (mit Verbendstellung) stehen, können **Relativadverbien** eher nur am Anfang eines w-Nebensatzes (mit Verbendstellung). Davon ausgehend, können also **Interrogativadverbien**, wie es im Beispiel (39) gezeigt wird, ausschließlich **deiktisch** gebraucht werden: (39) **Wo** hat sie den Schlüssel vergessen? Man fragt

sie, **wo** sie den Schlüssel vergessen hat). Hier referiert also das Wort „Wo“ direkt zu einem außersprachlichen Gegenstand. Im Gegensatz zu Interrogativadverbien, können **Relativadverbien**, wie es im Beispiel (40) illustriert wird, eher ausschließlich **anaphorisch** gebraucht werden:

(40) Sie weißt nicht mehr den **Ort₁**, **wo₁** sie den Schlüssel vergessen hat.

(41) Isaac hatte noch gute Leistungen gezeigt₁, **was₁** mich sehr erstaunt hat.

So betrachtet, beziehen sich Relativadverbien immer auf eine zuvor genannte **Bezugsgröße** und können daher inhaltlich kategorisiert werden. Als Beispiel für **Relativ-Interrogativadverbien des Ortes** haben wir (wo, woher, wohin, wozwischen, woran, usw...). Zu **Relativ-Interrogativadverbien der Modalität** haben wir (wie) und zu **Relativ-Interrogativadverbien der Gründe** gehören (warum, weshalb, weswegen, wieso). Weiterhin unterscheidet man auch die sogenannten **komplexen Relativ-Interrogativadverbien**; diese bestehen aus einem W-Wort + einer Präposition) und werden meistens als **Präpositionaladverbien bzw. Pronominaladverbien** (Wonach, worauf, wodurch) bezeichnet. Dazu kommen wir gleich im Weiterverlauf unserer Arbeit zurück (vgl. Artemis Alexiadou, 2013: 459).

Des Weiteren unterscheidet man auch Adverbien, **die verknüpfende Funktionen** üben. Noch als Konjunkionaladverbien genannt, werden sie dafür eingesetzt Sachverhalte zu verknüpfen. Zu ihnen gehören die Adverbien wie **deshalb, daher, deswegen, sonst, außerdem, schließlich, danach, zuvor, trotzdem...** Als Beispiele für den Gebrauch von Konjunkionaladverbien haben wir die folgenden Texte:

42) Du hast kein Geld, **deshalb** nimmst du kein Taxi.

43) Gib mir ein Stück Brot, **sonst** werde ich hungern.

Von diesen Satzbeispielen ausgehend, verbinden **Konjunkionaladverbien** also zwei Sätze in eine Relation zueinander und beziehen sich damit auf zwei oder mehrere Sätze (**anaphorischer Aspekt**). Übrigens ist auch festgestellt worden, dass **Konjunkionaladverbien und Kausaladverbien** sich überschneiden. Dennoch gibt es klare Unterschiede; ein Adverb wird als **Konjunkionaladverb** genannt, wenn es lediglich dafür verwendet wird, zwei Satzteile zu verbinden. Dies entgegnend, wird ein Adverb als **ein Kausaladverb** eingestuft, wenn man es ausdrücklich dafür nutzen, einen Grund zu beschreiben (vgl. Dudengrammatik, 2009: 568). Solche Überschneidungen zwischen Wortkategorien tauchen meistens bei **nicht flektierbaren Wortarten** auf und werden als **Homonymien** bezeichnet. Unter **Homonymien** versteht man

Wörter, die **morphologisch** betrachtet, gleich geschrieben sind, die aber **funktional** betrachtet, unterschiedliche Funktionen innerhalb der Wortkette üben. Dabei nehmen sie auch meisten **verschiedene syntaktische Positionen** auf und haben auch verschiedene syntaktische **Werte** und gehören somit zu verschiedenen Klassen. Nehmen wir zur Illustration die aus der Dudengrammatik (2009) stammenden Beispiele.

*Sie arbeitet außerhalb der Stadt (**Präposition**) – Sie arbeitet außerhalb. (**Adverb**);*
*während ihrer Dienstzeit (**Präposition**) – während sie arbeitet (**Subjunktion**);*
*Damit ist sie zufrieden (**Adverb**) – damit sie zufrieden ist (**Subjunktion**);*
*denn er hat keine Zeit (**Konjunktion**) – Hat er denn keine Zeit? (**Abtönungspartikel**);*
*schon am frühen Morgen (**Adverb**) – Was hast du schon zu sagen! (**Abtönungspartikel**);*
*Er geht nicht ins Kino (**Negationspartikel**) – Willst du nicht ins Kino mitkommen?*
*(**Abtönungspartikel**)* (vgl. ebd.: 567).

Letztendlich unterscheidet man Adverbien, die **textgliedernde Funktionen** aufweisen. Zu dieser Kategorie gehören Adverbien deren Geltungsbereich (Skopus) im Gegensatz zu anderen Adverbien, wie etwa lokale, temporale oder modale, **umfangreich** bzw. **Text- oder Abschnittsweit** sind. Dazu zählen beispielsweise Adverbien, die innerhalb des Textes Ordnung markieren, die sogenannten **Textadverbien** (erstens, zweitens, einerseits, andererseits, ferner, weiterhin, etc...). Textadverbien tragen also dazu bei, den Text zu gliedern. Weiterhin wird auch gemerkt, dass manche **Konjunkional- und Präpositionaladverbien** zur Kategorie der Textadverbien gehören, denn sie tragen nämlich auch dazu bei, **den Text inhaltlich zu verbinden bzw. kohärent zu machen und somit dann semantisch zu strukturieren**. Daher erweisen sie ebenso wie Textadverbien **textgliedernde Funktion** auf (vgl. Dudengrammatik, 2009: 1070, Rdnmr:174).

3.3.Pro-Adverbien und Pronominaladverbien (PAn)

Ziel dieses Abschnitts ist es, den Adverbientyp, der im Mittelpunkt unserer Untersuchung steht, so umfangreich⁸³, also morphologisch, syntaktisch semantisch und pragmatisch, zu beschreiben, dass am Ende der Beschreibung Parameter bzw. Regularitäten und Heuristiken zur automatischen Auflösung der **Pronominaladverbien-Anapher** (im Folgenden PA-anapher) aufbereitet werden können. Dazu werden wir zunächst die funktionale Kategorie der

⁸³ Vgl. Die phonetischen Eigenschaften der PAn werden im Rahmen dieser Untersuchung nicht berücksichtigt, vor allem, weil solche Merkmale beim Aufbau unseres Algorithmus nicht miteinbeziehen werden. Jedoch kann die phonetischen Merkmale der PAn in den Arbeiten von Negele (2012b) und Fleischer (2002) aufgerufen werden.

Pro-Adverbien darstellen. Und im Anschluss daran werden wir Einiges über die Morphologie der *Pronominaladverbien* darstellen. Schließlich wird auch die Konkurrenz näher erörtern, die sich zwischen den beiden meisten gebrauchten Begriffen zur Bezeichnung **der da-, hier- und wo-Verbindungen** abspielt.

3.3.1. Pro-Adverbien

Pro-Adverbien und **Pronominaladverbien** gehören zusammen zur oberen Kategorie der **Pro-Wörter bzw. Pro-Formen** und sind deswegen keine Synonyme und dies, obwohl die beiden Begriffe, morphologisch betrachtet, sich ähneln. In der Tat wird der Begriff **Pro-Adverbien** gebraucht, um auf alle Kategorien von Adverbien, die pronominal gebraucht werden können, zu referieren (vgl. Dudengrammatik 2009: 579). So gesehen steht also der Begriff **Pro-Adverb** für alle Adverbien, die an der Stelle entweder **eines Substantivs, einer Präpositionalphrase oder auch eines ganzen Satzes stehen können**. Zu Pro-Adverbien zählen beispielsweise, außer Adverbien, die *in Ihrer Bedeutung nicht autonom sind (dort, dann, da, hier, her, hin, so, sonst, wann, wie, wo* (vgl. Helbig 2004a: 287) auch **Interrogativadverbien, Relativadverbien und Konjunkionaladverbien und manche absoluten Adverbien** (vgl. Dudengrammatik 2009: 579).

Im Gegensatz zu **Pro-Adverbien** wird der Begriff **Pronominaladverbien (PAn)** verwendet, wenn es um die Bezeichnung von Wortzusammensetzungen geht, die **aus einem anaphorischen bzw. deiktischen Element (da(r)-, hier-, wo(r)-des-, dem-, -so- etc.) und einem präpositionalen (relationalen) Element (-an, -nach, -durch etc.) bestehen** (vgl. Knees 2009: 17). Als Beispiel davon zählen PAn wie etwa **darüber; darauf; worüber; worauf; hieran usw...**). Solche PAn werden, funktional betrachtet, fast immer als Stellvertreter entweder für eine sprachliche Entität (**Phorischer Gebrauch**⁸⁴: Anaphorischer/Kataphorischer) oder für eine in der Sprechsituation stehende Entität (**deiktischer Gebrauch**), wobei sie entweder demonstrativ, interrogativ oder relativ verwendet werden (vgl. Jung 1988: 314 zit. nach. Selma AkolGötkas 2020: 54). So gesehen haben **Pro-Adverbien** und **Pronominaladverbien** gemeinsam, dass sie die Pronominalität üben können.

⁸⁴ Bei solchen Fällen wird es die Rede von **Pronominaladverbienanapher** bzw. **Pronominaladverbienkatapher**.

Und gehören daher, wie es **Vater (1975)**⁸⁵ auch behauptet, zur oberen Kategorie der Pro-Wörter/ Pro-Form; Dazu sagt Helbig (2004) nämlich Folgendes:

*Beide -Pronominaladverbien und Pro-Adverbien- sind als Pro-Wörter zu verstehen. Es gibt Prowörter verschiedener Art und verschiedenen Grades. Der Prowort-Charakter (oder: **Pronominalität**) kann somit keinesfalls auf die herkömmliche Klasse der **Pronomina** beschränkt werden (Helbig 2004a: 288).*

Diese Stellungnahme Helbigs wird eben auch in der Dudengrammatik (2007) hervorgehoben. Dabei wird nämlich das PA als *ein Adverb, das sich in seiner Grundfunktion **pronominal auf eine Präpositionalgruppe bezieht*** (Dudengrammatik 2007: 738).

Allerdings herrscht für die Einordnung und Bezeichnung **der da-, hier- und wo-Verbindungen** in der Forschungsliteratur, sowohl **ein Kategorisierungs- als auch ein Bezeichnungskonflikt**. Zum Bezeichnungskonflikt kommen wir noch später in Einzelheiten zurück. Doch was die Einstufung der **da-, hier- und wo-Verbindungen** zu einer bestimmten grammatischen Klasse angeht, lässt sich feststellen, dass diese Wortverbindungen entweder der Wortart **der Pronomina** (vgl. Helbig/Buscha 2001: 236; Kars/Häussermann 1997: 113) oder der Wortart **der Adverbien** (vgl. Weinrich 2007: 568; Duden 2009: 579; Hentschel/Weydt 2013: 244) zugeordnet werden. Der Grund eines solchen **Kategorisierungskonflikts** geht nämlich auf **die Perspektive** zurück, worunter die gestrittenen Wortzusammensetzungen bearbeitet wurden. Ist **der Schwerpunkt** auf die Funktion der **da-, hier- und wo-Verbindungen** gesetzt worden, dann wurden die Wortverbindungen als Angehöriger der Kategorie **Pronomen** eingestuft. Lag der Schwerpunkt aber auf der Form der **da-, hier- und wo-Verbindungen** dann wurden diese zur Kategorie **der Adverbien** gezählt.

Hinsichtlich des **Bezeichnungskonflikts** rezensiert man in der Forschungsliteratur eine dutzende Bezeichnung für die **da-, hier- und wo-Verbindungen**. Und die Motivationen zur Bezeichnungswahl gehen weithinaus über die Parameter **Form** (Präpositionaladverb) oder **Funktion** (Pronominaladverb), sondern stützen sich auch besonders auf **die argumentative Struktur** und somit auf die Zielsetzung der jeweiligen Arbeiten/Autoren. Rüttenauer (1978) zieht es vor, den Terminus **adverbiale Proformen** anstelle des Terminus **Pronominaladverbien** zu benutzen, und zwar, weil dieser Terminus sich mehr auf semantische Funktionen bezieht

⁸⁵Zu den **Proadverbialen** gehören auch die sogenannten „**Pronominaladverbien**“, sie sind nichts Anderes als **adverbiale Pro-Formen**, die eine **Präposition** enthalten (und dadurch etwas spezifischer sind als die bisher erwähnten adverbialen Pro-Formen): 74) In der Ecke steht ein Tisch. **Darauf** liegen Bücher. 75) Hier ist eine Lampe. **Damit** kannst du den Tisch abwischen.“ (Vater 1975: 39)

und wird ihm deswegen dabei helfen den semantischen von dem pragmatischen Aspekt der **da-, hier- und wo-Verbindungen** deutlich voneinander bearbeiten zu können (vgl. Rüttenauer 1978: 2). Oppenrieder (1991), seinerseits, bevorzugt es, um die Spaltungs- und Verdoppelungskonstruktion der **da-, hier- und wo-Verbindungen** hervorzuheben, den Begriff **Proadverb** zu verwenden, (vgl. Oppenrieder 1991: 159.). Rehbein (1995) verwendet den Terminus **zusammengesetzte Verweiswörter** für die Bezeichnung der da-, hier- und wo-Verbindungen. Und als Rechtfertigung dieser Wahl sagt er Folgendes:

Wie Ehlich in einer Reihe von Arbeiten gezeigt hat und auch nach der obigen Diskussion verweisen deiktische Ausdrücke auf nicht-sprachliche Elemente, phorische nur auf sprachliche ‚Pronominaladverb‘ ist deshalb eine prinzipiell unzutreffende Bezeichnung für zusammengesetzte Verweiswörter.“ (Rehbein 1995: 174 zit. nach. Selma AkolGötkas 2020:96)

Rehbein (1995) nach ist also der Terminus **Pronominaladverb** zur Bezeichnung der **da-, hier- und wo-Verbindungen** unzutreffend. Dies hingegen präferiert Günthner (2001) den Gebrauch von dem Terminus **Pronominal- und Relativadverb** und dies vor allem, weil seine Arbeit darauf abzielte, das Vorkommen von **wobei** in der gesprochenen Sprache zu untersuchen (vgl. Günthner 2001: 313). Brinker (2010) seinerseits, teilt aber wiederum die Ansicht von Rehbein (1995) mit und behandelt daher die entsprechenden Wortzusammensetzungen unter dem verallgemeinernden Terminus **Pro-Formen** (vgl. Brinker 2010a: 30). Weiterhin werden Bezeichnungen, wie **Adverbkonnektoren (Adverbialkonjunktionen)** bei Zimmermann (2012) und Breindl (2012) erfasst.

3.3.2. Pronominaladverbien oder Präpositionaladverbien? Formbestand und Wortbezeichnung

3.3.2.1. Allgemeinheiten

Wie schon bereit hingewiesen, versteht die Dudengrammatik (2009) unter **Pronominal- oder Präpositionaladverbien** Wortzusammensetzungen, die aus einem **Adverb** (da-, hier-, oder wo-) als erstem Teil und **einer einfachen Präposition**⁸⁶ als zweitem Teil (**an, auf, aus, bei, durch,**

⁸⁶ Die Dudengrammatik (2009) stellt **eine Klassifikation von Präpositionen nach ihrem Komplexitätsgrad** dar. Dieser Klassifikation nach, werden Morpheme wie (in, auf, mit, nach um usw.) als **einfache bzw. primäre Präpositionen** bezeichnet. Daneben werden Wörter wie (mithilfe, zufolge, anhand, anstelle, wegen, dank, zuliebe, vermöge, während, bezüglich, rechts usw.) als **sekundäre Präpositionen** eingestuft. Und als **tertiäre bzw. mehrteilige Präpositionen** werden Lokutionen wie (im Verlaufe von, in Bezug auf, in Anbetracht, im Gefolge, an Stelle usw.) (vgl. Dudengrammatik 2009: 580).

für, gegen, hinter, in/ein, mit, nach, neben, über, um, unter, von, vor, zu, zwischen) zusammengesetzt sind. Diese Wortzusammensetzungen weisen, morphosyntaktisch betrachtet, weder **deklinierbare** noch **komparierbare Eigenschaften** auf. Sie regieren also keinen Kasus und können zugleich nicht gesteigert werden (vgl. Dudengrammatik 2009: 580). Als Beispiel dafür haben wir darauf/drauf⁸⁷, hierauf, worauf, damit, hiermit womit usw.... Im Gegensatz zu dieser Auffassung, in der **Pronominal- oder Präpositionaladverbien** nur mit den oben zitierten einfachen Präpositionen gebildet werden können, werden auch in manchen Werken Wortzusammensetzungen wie „Darob“ und „Dawieder“ als **Pronominal-bzw. Präpositionaladverbien** gezählt (Dudengrammatik 1998: 373). Aber da der Gebrauch solcher Verbindungen in der Alltagssprache eher selten ist, werden wir im Rahmen unserer Untersuchung nur **die da-, hier-; und wo- Verbindungen, die die zuvor erwähnten 20 einfache Präpositionen als zweites Glied haben**, berücksichtigen. Als weitere veraltete Erscheinungsformen von **Pronominal und Präpositionaladverbien**, die aber heute noch gelegentlich vorkommen, gelten beispielweise „darnach“, „daneben“ (vgl. Dudengrammatik 2009: 580).

Von diesen etablierten veralteten Abweichungen ausgehend, stellen wir fest, dass die Auffassung von was unter **Pronominal-oder Präpositionaladverbien** zu verstehen ist, im Laufe der Forschungsgeschichte, um diesen Gegenstand herum, gewachsen ist. Helbig/Buscha (2001), beispielweise, definieren in ihrem grammatischen Lehrbuch die PAn als „**Verbindungen der Adverbien da-, hier- und wo- (mit vorwiegend lokalen) Präpositionen**“ (Helbig/Buscha 2001: 308). Manche Autoren, wie Götze/Hess-lüttich (1999) und Engel (2004) hingegen weisen aber darauf hin, dass **nur da-Verbindungen** am meistens mit Präposition **mit lokaler Bedeutung (lokale Präposition)** funktionieren, also als PAn verzeichnet werden können. Weiterhin behaupten die beiden Autoren, dass die Wo-Verbindungen nicht mit lokalen

⁸⁷Wenn **die deiktische und die relationale Komponente** eines PAs jeweils mit einem Vokale endet, dann wird bei der Zusammensetzung der beiden Komponente **das Infix „r“** an die Morphem-Grenze hinzugefügt. Dies ist beispielweise der Fall bei da- und wo-Zusammensetzungen mit den Präpositionen wie (–auf, –an, –um, –aus, –in/ein und –unter) (vgl. Carola Strobl 2019). Des Weiteren ordnet Weinrich (2007) die **Kurzformen von da-Zusammensetzungen**, wie etwa (drauf, dran, draus, drunter, drum, drüber, drin/drein) **der Umgangssprache** zu und bezeichnet er sie als **diskontinuierliche Formen der PAn**. Zu diesen diskontinuierlichen Formen der PAn zählt er auch **die aufgesplitterte und die verdoppelte Formen**. Noch unter der Bezeichnung **Spaltungskonstruktion** bekannt, verweist der aufgesplitterte Gebrauch von PAn auf eine Gebrauchsweise von PAn, in der die Präposition sich vom Adverb trennt und im **Mittel- oder Nachfeld im Satz** nach dem Adverb (da, hier, wo) vorkommt. Folgende Beispiele illustrieren die **diskontinuierlichen Formen** der PAn: 45) „Da kann ich nichts **für**. (richtig: **Dafür** kann ich nichts.) 46) Wo brauchst du das **für**? (richtig: **Wofür** brauchst du das?) 47) Hier habe ich gar nichts **zu** sagen. (richtig: **Hierzu** habe ich nicht zu sagen). Als **verdoppelte Formen oder Verdoppelungskonstruktion der PA** zählen Zusammensetzungen wie **dadaran/dadran, dadarauf/dadrauf wodran, wodrauf und hierdran, hierdrauf** usw. (vgl. Dudengrammatik 2005: 587).

Präpositionen (wie **neben**, **hinter** und **zwischen**) möglich sind; stattdessen werden an der Stelle ein **Lokaladverb** eintreten. (Götze/Hess-lüttich 1999: 299& Engel 2004).

Weiterhin werden auch Wortzusammensetzungen, deren Formbestand gar nicht denen der **üblichen bzw. traditionellen PAn** ähnelt, in manchen Literaturen⁸⁸ als PAn verzeichnet. Dazu gehören beispielweise Wörter wie **deshalb**, **deswegen**, **trotzdem**, **demzufolge**, **demgegenüber**, **demgemäß**, **demnach**, **infolgedessen**, **usw...**(vgl. Hentschel/Weydt 2013: 246.). Und im Gegensatz zu traditionellen PA, die aus einem deiktischen Teil da-, hier- und wo- plus einer einfachen Präposition bestehen, bestehen diese Sonderformen von PAn, wie Sandberg (2004: 17)⁸⁹ in seiner Arbeit beschrieben hat, eher aus einem **Pronomen** (*die Dativform Neutrum des Demonstrativpronomens*) und einer **Präposition**, wobei die Präpositionen, im Gegensatz zu traditionellen PA, bei manchen Fällen an erster Stelle stehen (**trotzdem**; **infolgedessen**, **außerdem**, **nachdem**, **indem**, **seitdem**, **währenddessen**). Und als Grund der Auffassung **solcher abweichenden Fälle** als PAn hebt Sandberg hervor, dass die traditionelle Grammatik bei der Bestimmung von welchen Wörtern zu den PAn gehört, vorwiegend den Schwerpunkt auf den **morphologischen Kriterien** gelegt hat. Daher bilden die PA, so Sandberg (2004), **keine homogene Gruppe hinsichtlich ihrer Funktion**. Diese Hinsicht Sandbergs wird noch durch den folgenden Kommentar von Hentschel (2010) unterstützt:

*„Bei einem weiteren Begriffsverständnis werden auch aus einem Pronomen und einer Präposition gebildete Wörter wie **deshalb**, **deswegen**, **trotzdem** usw. zu den **Pronominaladverbien** gerechnet. Auf Grund ihrer Funktion werden diese Wörter jedoch meist den **Konjunktionaladverbien** zugeteilt.“ (Hentschel 2010: 294)*

Wie dieses Zitat von Hentschel (2010) hervorhebt, herrscht bei der Bearbeitung, von was als PA verstanden werden soll, **ein dynamisches Verhältnis** innerhalb der Forschung. Während manche Autoren die Problematik **aus einer morphologischen Perspektive** greifen, gehen die anderen sie eher **aus einer pragmatischen bzw. funktionalen Perspektive** rein. Es ergibt sich also, dass die Auseinandersetzungen um dieses Thema herum, entweder **den Formbestand oder die Funktion der Wörter** als Ausgangspunkt haben. Aus diesen Gründen wird es nämlich

⁸⁸ **Starke (1982)** schließt in seiner Arbeit die Wörter wie **weshalb**, **weswegen**, **wohingegen**, **deswegen** in die **PA** ein, in der er *wo-Verbindungen in der Funktion „Relativadverb“* als weiterführende Nebensätze erforscht (vgl. **Starke 1982: 215**). Wie **Starke (1982)** erläutert **Rehbein**, dass die von ihm bezeichnetem zusammengesetzten Verweiswörter **deswegen**, **deshalb**, **somit**, **indes** usw. **inklusive der PA** umfassen (vgl. Rehbein 1995: 166) [...]. **Schrodtt** zählt die Wörter wie *derart*, *deswegen*, *demnach*, *demzufolge* usw. **zu den PA**, aber untersucht die PA in **Konnektorfunktion** (vgl. Schrodtt 1988: 21 zit. nach Selma AkolGötkas 2020:93)

⁸⁹ Autoren wie Rüttenauer (1978: 83); **Starke (1982: 215)**; Rehbein (1995: 166); Knees (2009: 17) und Zimmermann 2012: 54 sind auch dieser Meinung.

in der Forschungsliteratur, im Allgemeinen, zwei bestimmte Begriffe zur Bezeichnung solcher Wörter anerkannt:

1. **Präpositionaladverb**: Wenn die Form bzw. der Formbestand der Wortgruppe betrachtet wird.
2. **Pronominaladverbien**: Wenn die Funktion der Wortgruppe innerhalb der sprachlichen Kette, also innerhalb des konkreten Sprachgebrauchs (Pragmatischer Aspekt), berücksichtigt wird. Hierbei wird dann hauptsächlich zwischen **der pronominalen** (Prowörter) und **der konjunkionalen** (Konnektor) **Funktion** unterschieden.

Diesen Befund wird durch das folgende Zitat aus der Dudengrammatik (2009) bekräftigt.

*Der Terminus **Präpositionaladverb** bezieht sich auf die Bildungsweise des Adverbs mit einer Präposition als zweitem Bestandteil und mit da(r)-, hier- oder wo(r)- als erstem Bestandteil, z.B. darauf, davon, hierauf, hiervon, worauf, wovon [...]. Der Nebenterminus **Pronominaladverb** bezieht sich auf die Funktion, nämlich für eine Präpositionalphrase oder auch für einen ganzen Satz zu stehe* (vgl. Dudengrammatik 2009: 580).

3.3.2.2. Pronominaladverbien vs Präpositionaladverbien

Wie gerade eben eingeführt, werden **da-, hier-; und wo- Verbindungen** als **Pronominaladverbien** oder als **Präpositionaladverbien** bezeichnet, jeweils, wenn der Fokus auf die Funktion oder den Formbestand gesetzt wird. Dazu wird beispielweise in der Dudengrammatik (1998) folgendes gesagt **Die Bezeichnung Pronominaladverb stellt die pronominale Funktion heraus, während Präpositionaladverb nur auf die Bildungsweise Bezug nimmt** (Duden 1998: 372 zit. nach Selma AkolGötkas 2020: 45). Der Gebrauch des Terminus **Präpositionaladverb** wird beispielweise bei den Autoren, wie Zifonun/Strecker/Hoffmann (1997: 54.); Engel (2004: 419) und Weinrich (2007: 568) erhoben. Hingegen bevorzugen Helbig/Buscha (2001: 236), Hentschel/Weydt (2003: 267), Eisenberg (2006b: 197), König (2012) und Zimmermann (2012) eher den Terminus **Pronominaladverb** und stufen den konkurrierenden Terminus **Präpositionaladverb als Nebenbezeichnung** ein. Daneben gibt es aber auch Autoren, die beiden Begriffe simultan gebraucht haben. Dazu zählen beispielweise Dreyer/Schmitt (2012: 91) und Hentschel (2010: 15; 294).

Weiterhin ließ sich, im Laufe der Forschung feststellen, dass die Wahl der geeigneten Bezeichnung auch vorwiegend von den Zielsetzungen der jeweiligen Arbeiten bedingt war.

Schauen wir uns nun, zum besseren Verständnis von, was damit gemeint ist, die Bearbeitung dieser beiden Termini bei einigen Autoren. Wie gerade eben eingeführt, zählen Helbig/Buscha (2001) zu den Autoren, die in ihren Arbeiten die Bezeichnung **Pronominaladverbien** bevorzugt haben. Diese Wahl lässt sich nämlich von deren argumentativen Vorgehensweise nachvollziehen. In der Tat, bearbeiten sie das Thema „Pronominaladverbien“ innerhalb eines Kapitels betitelt „**Substantivwörter**“. Diese Strukturierung gibt schon Hinweise darauf hin, dass die da-, hier-; und wo- Verbindungen **einige Merkmale von Substantiven erweisen**. Und da Substantive sich von Pronomen ersetzen lassen, schlagen wir beispielweise vor, dass genauso **dieses Merkmal**, mehr oder weniger, den Bezug zwischen den beiden Phänomenen (**da-, hier-; und wo- Verbindungen** und **Substantiv**) ermöglicht hat. Des Weiteren wird diese Hypothese nämlich durch **die Klassifikation** bestätigt, die Helbig/Buscha (2001) den „Pronominaladverbien“ zugeschrieben haben. In der Tat fassen sie dieser Letzte als eine eigenständige Wortart, die zur oberen Kategorie **der Pro-Adverbien** gehört, zusammen (vgl. Helbig/Buscha, 2001: 314). Im Einklang zu Helbig/Buscha ordnen Kats/Häussermann (1997) auch solche Wortzusammensetzungen der Kategorie der Pronomen zu, und präferieren daher eben auch den Begriff **Pronominaladverbien** anstatt des Begriffs **Präpositionaladverbien** (vgl. Kats/Häussermann 1997: 113). Allerdings zählen sie als PAn nur Wortzusammensetzungen, die aus **da- und wo-Verbindungen** bestehen. **Hier-Verbindungen** wurden leider nicht berücksichtigt (vgl. ebd.: 113).

Eisenberg (2006b), seinerseits, verrät in seiner Vorgehensweise schon, dass er den Begriff **Präpositionaladverb** anstatt **Pronominaladverb** gebrauchen wird. In der Tat bearbeitet er die **da-, hier-; und wo- Verbindungen** in einem Kapitel betitelt „**Präpositionen und Konjunktionen**“ dabei werden nur **da- und hier- Verbindungen** als solches, also als Präpositionaladverbien betrachtet. **Wo-Verbindungen**, ihrerseits, werden eher in einem anderen Kapitel betitelt „**Frageadverbien**“ bearbeitet (Eisenberg 2006b: 196). Wie schon von Anfang an gesagt, werden wir im Rahmen unserer Untersuchung die **da-, Hier- und Wo-Verbindungen** als **Pronominaladverbien** bezeichnen. Und zwar, weil uns **die stellvertretende Funktion** (Pragmatische Funktion: Sprachökonomie) und nicht **die verknüpfende Funktion** (syntaktische Funktion) interessiert. Doch bevor wir zu den Gründen unserer Bezeichnungswahl in Einzelheiten eingehen, wollen wir zunächst die Auffassung Weinrichs (2007) bearbeiten. Ebenso wie Eisenberg (2006b) bevorzugt er den Terminus **Präpositionaladverbien** und hebt diesbezüglich noch weitere Spezifikationen. Diese werden am Besten durch das folgende Zitat von **Selma AkolGötkas (2020)** dargestellt:

Weinrich (2007:568f.) kategorisiert die PA als die Bezeichnung „**Präpositional-Adverbien**“ zu den Adverbien und unterscheidet zwei Gruppen wie „**einfache Positions-Adverbien kombiniert mit Präpositionen**“ wie z.B. daran, darauf, hieran, hierauf usw. und „**Direktions-Adverbien kombiniert mit Präpositionen**“ wie z.B. hindurch, hinein, herein, herum usw. Weinrich (2007) hebt hervor, dass die Präpositionaladverbien mit den einfachen Positionsadverbien „da“ und „hier“ und einer Präposition kombiniert werden können, **allerdings nicht mit „dort“** und bei Weinrich (2007) spielen die **wo-Verbindungen gar keine Rolle**. Hinzu wird betont, dass es **mehr Möglichkeiten zur Kombination mit „da“ als mit hier gibt**. Nach Weinrich existiert **gar keine Kombination mit „dort“**, während die Kombinationen der Präpositionen mit dem Positions-Adverb da am häufigsten und mit **dem Positions-Adverb hier seltener sind**. Ferner betont er auch, dass diese Kombinationen als Pro-Formen (Pronominal-Adverbien) häufig sind und in pronominaler Funktion **anstelle der unterschiedlich langen Texte stehen** (vgl. Selma AkiGötkas 2020: 50).

Letztendlich werden wir zu dieser Bezeichnungskonkurrenz die folgende Stellung nehmen. Da unsere Arbeit darauf abzielt, **die Pronominalität von da-, hier- und wo- Verbindungen**, im Rahmen **eines computerlinguistischen Ansatzes**, zu untersuchen. Ist es zur Erreichung unserer Forschungsziele angebracht, dass wir die Bezeichnung **Pronominaladverbien (PA)** als **Stammbezeichnung** verwenden, wobei nur die in der folgenden stehenden Tabelle **60 PA-Erscheinungsformen** als Gegenstand unseres beabsichtigten AR-Systems bzw. Auflösungsalgorithmus fungieren werden .

Primäre einfache Präpositionen	bzw. da(r)-	hier-	wo(r)-
an	daran	hieran	woran
auf	darauf	hierauf	worauf
aus	daraus	hieraus	woraus
bei	dabei	hierbei	wobei

durch	dadurch	hierdurch	wodurch
für	dafür	hierfür	wofür
gegen	dagegen	hiergegen	wogegen
hinter	dahinter	hierhinter	wohinter
in/ein	darin/darein	hierin/hierein	worin/worein
mit	damit	hiermit	womit
nach	danach	hiernach	wonach
neben	daneben	hierneben	woneben
über	darüber	hierüber	worüber
um	darum	hierum	worum
unter	darunter	Hierunter	worunter
von	davon	hiervon	wovon
vor	davor	hiervor	wovor
zu	dazu	hierzu	wozu
zwischen	dazwischen	hierzwischen	wozwischen

3.3.3. Gebrauch und Funktionen von Pronominaladverbien (PAn)

3.3.3.1. Allgemeinheiten zum Gebrauch von Pronominaladverbien

Im vorherigen Abschnitt (3.3.1) haben wir **die morphologischen Eigenschaften der PAn**⁹⁰ erläutern. Dabei wurde insbesondere gemerkt, dass die PAn, je nach der Perspektive bzw. Zielsetzung der wissenschaftlichen Auseinandersetzung, in der sie als Gegenstand fungieren, entweder zur grammatischen Kategorie der **Pronomina** oder **des Adverbs** zugeordnet werden. Daher fällt es einem nämlich schwer **die Funktionen** von PAn so reibungsfrei zu veranschaulichen, wie es etwa bei einem Wort, das einer einzelnen grammatischen Kategorie gehörte, der Fall wäre. Darüber äußern sich Helbig/Buscha (2001) wie folgt:

*[...] Falls die ihnen zugrundeliegende Präpositionalgruppe **Objekt** ist, werden sie als substantivische **Pronomina**, falls sie **Adverbialbestimmung** ist, als **Pro-Adverbien** eingeordnet (vgl. Helbig/Buscha 2001: 315).*

Will man sich zunächst einen Überblick über **die funktionalen Eigenschaften von PAn** schaffen, wenn diese als **Pronomen** fungieren, also, wenn diese als Stellvertreter⁹¹ von sprachlichen oder nicht-sprachlichen Einheiten erscheinen. Dann lässt sich u.a. feststellen, dass **Pronominaladverbien** nur als Ersetzung **für Substantive, Satzglieder oder Teilsätze, die vorher** (anaphorisch) **oder nachher** (kataphorisch) **im Satz erscheinen**, fungieren (vgl. Götze/Hess-Lüttich 1999: 299). Hierbei referiert jede dieser sprachlichen Einheiten (Bezugseinheiten) entweder auf **Objekte oder Sachverhalte**. Dementsprechend können die PAn nur eingesetzt werden, um auf sprachliche Einheiten zu verweisen, die ihrerseits selbst auf **nichtanimierte Lebewesen**⁹² referieren; d. h. PAn können nicht auf sprachliche Einheiten, **die Menschen** oder **Tiere** versprachlichen, verweisen (vgl. Hentschel/Weydt 2003: 268). Dafür, also für den Verweis auf solchen sprachlichen Einheiten wird eher **eine Koppelung von Präpositionen und Personalpronomen** verwendet (vgl. ebd.). Zudem lässt sich einerseits

⁹⁰Neben der Funktion der PA werden semantisch bestimmte Klassen eingeteilt wie: **konsekutiv** (darum, deshalb, deswegen usw.), **adversativ** und **konzessiv** (dabei, dagegen, dafür usw.), **additiv** (außerdem, zudem, überdies usw.), **instrumental** (dadurch damit), **temporal** (danach, davor seitdem usw.), **substitutiv** (stattdessen usw.) und **speziellere Relationen**, für die es jeweils nur einen Ausdruck gibt (angesichts dessen, anhand dessen usw.) (vgl. Grammis-Projektgruppe 2012).

⁹¹ Dieser Funktion entgegen steht **der Nicht-pronominale Gebrauch von PAn**: Dazu gehören beispielweise der Gebrauch von PAn als **Konjunktionen** und als **Verbpartikeln (Verbzusatz)**.

⁹² Zu dieser Kategorie werden auch **Pflanzen** gezählt.

merken, dass **ganze Abschnitte** auch als potentielle Bezugseinheiten wahrzunehmen sind und andererseits, dass **Präpositionalphrase**, wie Selma AholGötkas (2020) es auch hervorhebt, eher als typische **Bezugsentität von PAn** gelten:

[...] Götze und Hess-Lüttich (1999) betonen nebenbei, dass die PA sich nur auf die „Konkrete oder Abstrakta, die als präpositionale Ergänzungen/ Angaben auftreten“, beziehen können. Götze/Hess-Lüttich (1999: 300) (vgl. Selma AholGötkas 2020:88)

Weiterhin lässt sich auch merken, dass der pronominale Gebrauch von PAn noch weiteren Parametern unterlegen ist. In der Tat sind die da-, hier- und wo-Verbindungen⁹³ nicht beliebig auszuwählen. Alle drei können nicht an derselben Stelle beim Sprachgebrauch auftreten. Während **da- und hier- Verbindungen** anstelle von **Personal-und Demonstrativpronomina + Präpositionen (Präpositionalgruppe/Präpositionalfügung) auftreten**, passen lieber die Wo-Verbindungen anstelle von **Interrogativ- und Relativpronomina + Präposition** (vgl. Helbig/Buscha 2001: 237). Betrachten wir nun zur Veranschaulichung, was gerade theorisiert wurde, einige Beispiele.

- **Wo-Verbindungen**

48) Worauf stützt sich sein Plan? (Interrogativer Gebrauch).

49) Sein Plan stützt sich auf etwas, **worüber** er nicht sprechen darf. (Relativer Gebrauch).

50) Es gibt viele Menschen, **an sie** er denken muss (Nicht **woran**, da es um Lebewesen geht).

- **Da- und Hier-Verbindungen**

51) Der junge Mann läuft **daneben/hierüber**. (deiktischer Gebrauch von daneben/hierüber).

52) Es liegen viele Löcher. **Daneben/Hierüber** läuft der junge Mann. (anaphorischer Gebrauch von daneben/hierüber).

53) Der junge Mann ging **davon** aus, dass der Weg nicht Reibungsfrei wäre. (Kataphorischer Gebrauch von davon).

⁹³ Was die Gebrauchsfrequenz von PAn angeht, wurde in manchen linguistischen Forschungsliteraturen festgestellt, dass **die da-Verbindungen häufiger als hier- und wo-Verbindungen sind** (vgl. Negele 2012b: 13).

54) Es gibt viele Menschen. **Neben ihnen** schafft sich der junge Mann einen Weg (nicht **daneben**, weil es sich hier um animierte Lebewesen geht. Etwas, das im vorherigen Beispiel nicht der Fall ist.

Wie bereit illustriert, steht es fest, dass die deutsche Rechtschreibung den Gebrauch von PAn erfordert, wenn es bei der Wiederaufnahme um Sache oder Sachverhalte geht. In der linguistischen Literatur lässt sich jedoch Fälle merken, wobei der Gebrauch von der Koppelung (Präposition+ demonstrativ-Personalpronomen) als gleichrangige Alternative zu den PAn eingestuft wird. Dazu nehmen Heuer/Flückiger/Gallmann (2008) die folgenden Beispiele als Illustration:

*[...] Sie hat keine Zeit dafür/hierfür (**weniger gut**: für das; **unmöglich**: für es). Darauf (**weniger gut**: Auf das) habe ich nicht gedacht. Das ist ein ausgezeichnete Schlagbohrer; damit/hiermit (**neben**: mit dem/ mit ihm) habe ich schon die größten Löcher gebohrt (Heuer/Flückiger/Gallmann 2008: 157).*

Beim letzten angegebenen Beispiel von Heuer/Flückiger/Gallmann (2008) stellen wir also fest, dass der Gebrauch eines PAs bei solchen Fällen nicht systematisch ist, und dies obwohl es sich um **ein innanimiertes Objekt** (*ein ausgezeichnete Schlagbohrer*) handelt. Eine weitere Abweichung der etablierten Sprachregel findet beispielweise statt, wenn PAn zum Wiederaufnahme von sprachlichen Einheiten, die auf Personen⁹⁴ referieren bzw. die Lebewesen, wie Menschen und Tiere versprachlichen. Dazu äußern sich Helbig/Buscha (2001) mit folgenden Wörtern

*Für Personen ist der Ersatz der genannten Pronomina durch **ein Pronominaladverb** nur möglich, wenn es steh um **eine Personengruppe** handelt: In der Klasse stand 24 Kinder, darunter 10 Mädchen. (unter den Kindern) (vgl. HelbtgBuscha 2001: 237 zit. nach Selma AkolGötkas 2020:62).*

Dieser Meinung schließen sich auch Hentschel/Weydt (2003) an und äußern sich eben mit folgenden Wörtern:

*[...]Offenbar werden lokale Adverbialbestimmungen nur **Personen**, wenn es sich dabei um eine Gruppe Unbekannter handelt, in erster Linie als **Ortsangaben** empfunden, so dass die Kategorie „**Belebtheit**“ nicht zum Tragen kommt. 233) Am Bahnhof stand **eine***

⁹⁴ Für weitere Einzelheiten, siehe Zifonun/Strecker/Hoffmann (1991: 55, 557); Götze/Hess-Lüttich (1999: 301) und Dudengrammatik (1998: 375; 2009: 582).

*Reisegruppe. Ich stellte mich **dazu**. Aber 234) Da kommt ja endlich **meine Reisegruppe**. Ich warte schon seit Stunden **auf sie** (vgl. Hentschel/Weydt, 2003 zit. nach Selma AholGötkas 2020: 62).*

Von diesen beiden Zitaten ausgehend, lässt sich also schlussfolgern, dass das relevante Parameter zum Gebrauch von PAn nicht **die Belebtheit oder nicht Nichtbelebtheit eines Referenten**, sondern eher die Art und Weise, wie dieser Referent kategorisiert bzw. konzeptualisiert wurde. Dies ist beispielweise der Fall im Beispiel 233, worin die NP „eine Reisegruppe“ nicht als eine Menge von Personen kategorisiert wurde, sondern eher als eine Gruppe, die sich an einem bestimmten Ort befindet. Deswegen wurde eben das PA **dazu** anstatt der Koppelung (zu + ihnen) eingesetzt.

3.3.3.2. Grammatische Subkategorien von Pronominaladverbien: Nichtpronominaler Gebrauch von PAn?

Sobald ein Überblick über die wiederaufnehmende Funktion von PAn skizziert wurde, schauen wir uns nun einige Kontreverse über **den nichtpronominalen Gebrauch der PAn** an. Ziel dieses Abschnitts ist es, Fälle zu identifizieren, worin die Pronominaladverbien nicht als Stellvertreter auftreten. Es kommt, in der Tat, in manchen Forschungsliteraturen vor, dass einige PAn als Stellvertreter eingestuft werden **Negele (2012b) & (Fleischer 2002)** während sie bei anderen Autoren eher als nichtstellvertretend beschrieben werden (Rüttenauer 1978). Und, da **das Ziel dieser Auseinandersetzung ist es, nur stellvertretende Pronominaladverbien in Betracht zu ziehen**, scheint es uns somit sehr wichtig und auch sehr nötig diesen Unterschied zwischen stellvertretenden und nichtstellvertretenden PAn⁹⁵ **vor dem Aufbau des angestrebten Algorithmus klar zu erläutern.**

3.3.3.2.1.) Pronominaladverbien als Verbzusatz⁹⁶

⁹⁵ Immerhin lässt sich allerdings in der Forschungsliteratur Zustimmung über die Klassifizierung von nichtstellvertretenden PAn feststellen. Im Allgemeinen sind dann die folgenden Kategorien zu unterscheiden: **PA als Verbpartikel; PAn als subordinierende Konjunktion (Subjunktion) und PA als Konjunkionaladverb (KA).**

⁹⁶ In den nachgeschlagenen Werken werden, je nach Zielsetzung, unterschiedliche Begriffe zur Bezeichnung dieses Erstglieds verwendet. Während es beispielweise Begriffe wie **Verbpartikel** (Eichenger 1989; Eisenberg 2006: 255); **trennbares Präfix** (Fleischer 1969); **Halbpräfix** (Wellman 1998); **postponierbares Präverb** (Šimečková 1994), **Nachverb** (Weinrich 1993) in einer morphologischen Perspektive eingesetzt worden sind, tritt, der Dudengrammatik (2009) nach, der Begriff **Verbzusatz** eher in einer syntaktischen Perspektive auf (vgl. Die Dudengrammatik 2009: 696).

Vom **PA als Verbzusatz** ist es die Rede, wenn das PA als ein Teil **eines trennbaren Verbes** vorkommt. Noch unter **Partikelverb** bekannt, verweist der Begriff **trennbares Verb** auf ein Verb, dessen erster Teil von dem Stammverb beim Konjugieren aufgelöst wird (vgl. Dudengrammatik 2009: 696). Hierbei steht dieser abtrennbare Teil am Satzende. Zu der Gruppe abtrennbarer Verben gehören beispielweise Verben, wie **ankommen** (Ich komme in eine Stunde **an.**); **abfahren** (Der Bus fährt in eine Stunde ab.); **zumachen** (Vor gerade zwei Minuten machte ich diese Tür **zu.**). Allerdings können außer den Präpositionen auch weitere Wortkategorien, wie etwa **Substantive, Adjektive, Verben** selbst **und Adverbien** als **Verbpartikel** auftreten (vgl. Eisenberg 2006a: 267).⁹⁷

Wie oben bereits hingewiesen, zählen auch die **Adverbien** zur Wortkategorien, die als Verbzusatz eines Partikelverbs erscheinen können (**hierlaufen, reingehen usw...**). Kommt es doch dazu, dass wir mit **einem Adverb pronominaler Natur bzw. mit einem Pronominaladverb (PA)** als erstes Glied zu tun haben, dann geraten wir genau in den Fall (**PA als Verbzusatz**), der im Fokus dieses Abschnitts steht. Doch bevor wir diesen Aspekt in Einzelheiten bearbeiten, wollen wir zunächst darauf hinweisen, dass nur die **da-Verbindungen**, der Dudengrammatik (2009) nach, **als Verbzusatz** auftreten können. **Hier- und Wo-Verbindungen**, seien hingegen nicht zu diesem Gebrauch beobachtet worden (vgl. Dudengrammatik 2009: 699). Nehmen wir zur Illustration die folgenden Beispiele:

55) Gestern gab es einen Überfall. Anscheinend **war** Azazou auch **dabei**.

56) Warte! Fahre jetzt bitte nicht los! Ein weiterer Passagier **kommt** noch **drin**!

57) *Beispiel 54: Peter hat heute das Fenster unserer Klasse kaputt gemacht. Wahrscheinlich **kommt** er mit einer Warnung **davon*** (vgl. Selma AkiGötkas 2020: 84).

Im zweiten Satz des letzten Beispiels (57) fungiert das Wort „**davonkommen**“ als Verb. Gemerkt wird eben, dass die abtrennbare Partikel (**davon**) sich, wie oben schon hingewiesen, am Ende des Satzes befindet. Bedeutet dies aber, dass dieses PA, weil es als Verbpartikel auftritt, längs **keine pronominale Funktion mehr ausübt**? In der Tat lässt sich in manchen Forschungsliteraturen feststellen, dass der Auftritt des PAs als Verbzusatz keine

⁹⁷ Als Beispiele für Partikelverben mit **Substantiven** als erstes Glied gelten beispielweise Verben wie: Fahrradfahren (er fährt gerade **Fahrrad**); **Fußballspielen** (Gestern spielte er **Fußball**); Schuldhaben (er hat **Schuld**) usw. Für **Adjektive** als Verbzusatz fungieren beispielsweise Verben wie: **loslaufen** (Wir laufen jetzt **los**!); **festhalten** (Zum Glück hielt ich damals mein Handy gut **fest**). Und als Beispiele für Partikelverben mit einem **Verb** als erstes Glied gelten beispielweise **spazierengehen; sitzenbleiben; stehenbleiben; kennenlernen usw.** (vgl. Eisenberg 2006a: 267).

stellvertretende Funktion mehr aufweisen könnte (vgl. Rüttenauer 1978: 85). Allerdings beschränken sich solche Einstellungen bzw. Schlussfolgerungen nur auf die **morphosyntaktischen Eigenschaften** von solchen PA. Würde man aber **den semantischen bzw. referenz-semantischen Aspekt** von solchen PA betrachten, dann würde man dazukommen, dass solche PA eben auch wie andere, die nicht als Verbzusätze auftreten, **pronominale Funktion üben**. Zur Erläuterung dieser Stellungnahme stützen wir uns auf die semantischen Eigenschaften jedes einzelnen Morphems, das ein PA konstituiert. Dabei steht es nämlich fest, dass jedes PA aus zwei Morphemen besteht: **einem relationalen Morphem** (die Präposition) und **einem phorisch-deiktischen Morphem** (da-; hier-; wo-) (vgl. Knees 2009: 17). Und genau dieses phorisch-deiktische Morphem ist für die Erläuterung unserer Stellungnahme von besonderer Relevanz und zwar nämlich, weil egal, ob das PA frei oder als Verbzusatz erscheint, spielt dieser phorisch-deiktische Teil immer seine Rolle. Eine Rolle, die nicht anders ist, entweder auf ein im Text vor/nach erwähntes Element zu verweisen, oder auf ein in der Sprechsituation stehendes Objekt/ abgelaufene Sachverhalte zu referieren. Sowieso sei es bei der phorischen oder bei der deiktischen Wirkung dieses Teils, **wird die stellvertretende Funktion immer lesbar/nachvollziehbar sein**.

Kommen wir nun doch zu unserem Beispiel zurück und versuchen wir den eben theorisierten Ansatz zu illustrieren. Wie gesagt, handelt es sich tatsächlich in diesem Beispiel um das Verb „davonkommen“. Würde der Leser nur den zweiten Satz des Textes lesen, dann würde er nicht in der Lage sein jedem einzelnen Satzelement eine völlig semantische Repräsentation/Interpretation zuzuordnen; und somit auch den ganzen Satz einen Sinn zu geben. Dafür müsste er unbedingt auf den ersten Satz des Textes zurückgreifen, um nämlich zu erschließen, worauf sich das PA (davon) bezieht⁹⁸. In der Tat bezieht sich **davon** hier auf die **Tatsache/Proposition**, dass Peter eine Dummheit/Infraktion begangen hat, und deswegen, sollte man darauf warten, dass er von den zuständigen Behörden gewarnt wird. So illustriert, steht es fest, dass **der phorisch-deiktische Teil des PAs** ihm eine stellvertretende Funktion zuschreibt, und dies, obwohl das PA als **Verbzusatz**⁹⁹ auftritt. Wir werden deswegen nicht

⁹⁸ Etwas das nicht der Fall bei dem PA „damit“ ist, wenn dieses als Subjunktion auftritt. Denn „**damit**“ als **Subjunktion** weist schon eine voretablierte Bedeutung auf. Man benötigt also nicht den Kontext heranzuziehen, um es interpretieren zu können; wie es beispielsweise der Fall bei **damit als PA** wäre.

⁹⁹ An dieser Stelle möchte ich darauf hinweisen, dass dieser gerade erläuterte Ansatz sich nur auf PA als Verbzusatz, die in konjugierter Form auftreten also nicht im Infinitiv. Denn der angestrebte Algorithmus zielt darauf ab, die einzelauf tretenden/alleinerscheinenden **da-; hier- und wo-Verbindungen** automatisch aufzulösen. Allerdings ist die Interpretationsmöglichkeit für Computer wenig anstrengend, wenn ein PA als Verbzusatz im Infinitiv auftritt, denn es handelt sich um ein Wort, das lexikalisch betrachtet, nicht **semantisch leer ist**. Und von daher benötigt der Computer nicht unbedingt einen Stützpunkt im Vor- oder Nachtext zu haben, um das entsprechende

gezwungen, beim Aufbau unseres Algorithmus ein sortierendes Modul zu erstellen, das dabei helfen würde, die PAn von den PAn als Verbzusatz abzutrennen.

Diese gerade eben erläuterte Stellungnahme ist diejenige, die im Einklang mit der Auffassung Fleischers (2002) und Negeles (2012b) **über den stellvertretenden Charakter von PAn als Verbzusatz** steht. Wie es am Anfang dieses Abschnitts hingewiesen wurde, teilen hingegen Autoren wie Rüttenauer (1978) diese Meinung nicht. Ihm nach, handelt es sich beim Auftritt von PAn als Verbzusatz eher um **eine präpositionale Verwendung** von PAn. Dazu äußert sich Selma AkolGötkas (2020) mit folgenden Wörtern:

*Rüttenauer (1978: 49) ist der Meinung, dass es nicht **um pronominale, sondern um präpositionale Verwendung in solchen Fällen** geht, in denen bestimmte Verben über feste Präpositionen verfügen und diese vor einem Satz stehen (vgl. Selma AkolGötkas 2020: 120).*

Mit diesem Zitat steht es klar, dass Verben mit festen Präpositionen wie **warten + auf; denken + an; danken + für; gehören +zu;** usw. als Basis für die Bildung von PAn gelten können. Zudem ist es auch darauf hingewiesen worden, dass solche gebildete PAn, **also PAn, deren relationales Glied eine für ein Verb feststehende Präposition ist, keine Pronominalität ausüben können.** Jedoch teile ich dieser Meinung nicht. Denn, obwohl es beobachtet wird, dass solche PAn mit festen Präpositionen von Verben gebildet worden sind, weisen sie immerhin doch gleiche **syntaktische und morphologische Eigenschaften**, wie die zuvor erläuterten PAn als Verbzusatz, auf. In der Tat lässt sich syntaktisch feststellen, dass solche PAn, eben wie die PAn als Verbzusatz, immer **eine Letztstellung im Satz** aufweisen. Und obwohl solche PAn im Gegensatz zu PAn als Verbzusatz **am Ende eines Hauptsatzes** stehen, welcher seinerseits in fast allen Fällen **einen subordinierenden Satz (Dass-Satz, W-Satz oder Infinitivsatz) einleitet**, beeinträchtigt diese Tatsache nicht den Fakt, dass die aus festen Präpositionen gebildeten PA immerhin am Ende eines beliebigen Satzes stehen. Nehmen wir zur Illustration zunächst einige Beispiele. Diese werden uns dabei helfen, unseren Standpunkt weiter unterstützen zu können.

58) Schon seit letztem Monat **warte** ich **darauf**, dass sie sich bei mir meldet. (Warten+ auf)

59) Ich danke Ihnen **dafür**, dass Sie mir geholfen haben. (danken+ für)

Wort interpretieren zu können. Sondern wird er nur auf seine lexikalische Datenbank zugreifen müssen, um das entsprechende Wort semantisch auflösen zu können.

60) Ich **denke daran**, ihn morgen anzurufen¹⁰⁰ (denken+ an)

Bei der Analyse des Hauptsatzes im ersten Beispiel (58) lässt sich verstehen, dass der Sprecher seit einiger Zeit auf etwas wartet. Und dieses Etwas wird dann im folgenden Nebensatz bekannt gegeben. Somit steht es fest, dass der kohäsive Bezug zwischen den Inhalten beider Sätze (HS & NS) durch das PA **darauf** und die Subjunktion **dass** gewährleistet wird.

In dieser Hinsicht besteht also ein klarer Bezug zwischen dem PA **darauf** und dem subordinierenden Nebensatz: **Das PA darauf vertritt/ steht für die Proposition, welche im Nebensatz auftritt/ versprachlicht worden ist.** Und genau da, bei dieser Eigenschaft solchen PA, steht die Grenze jener Auffassung, die besagt, dass **solche PAn keine stellvertretende Funktion üben können.** Denn es wird eben bemerkt, dass obwohl es bei solchen PAn um keinen **Rückwärtsverweis/Anaphorik** (die stellvertretende Form par Exzellenz) geht, handelt es sich immerhin um einen Verweis, der nach vorne orientiert ist. (**Vorwärtsverweis/Kataphorik**): das PA **darauf** hier bezieht sich auf Elemente des Vortextes. So gesehen haben wir im Gegensatz zu Rüttenauer (1978) doch mit einer pronominalen Verwendung von solchen PAn zu tun, die von Verben mit festen Präpositionen ausgebildet worden sind. Solcher Gebrauch von PA wird in der Forschungsliteratur unter dem Ausdruck **PA als Korrelat** aufgezeichnet (Dončeva 1982: 221; Breindl 1989: 11; Sandberg 2004: 30 ua.). Wir werden im Rahmen dieser Studie solche PA eher unter dem Begriff **PA als Katapher** aufzeichnen, vor allem, weil wir uns im Forschungsdomain der Textlinguistik befinden.

Wie ausführlich erläutert, sei es um **PA als Verbzusatz** oder um **PA als Korrelat**¹⁰¹, ließ sich letztendlich nachweisen, dass beide Gebrauche sich als Stellvertretend nachgewiesen haben. In dieser Hinsicht benötigen wir, wie oben schon hingewiesen, kein besonderes Modul oder

¹⁰⁰ Bei der Näherbetrachtung der Beispiele könnte man auch zur Schlussfolgerung kommen, dass es sich tatsächlich um folgende **trennbare Verben** geht: **daraufwarten (58); dafürdanken (59) und darandenken (60).** Und dies vor allem, weil das PA sich am Satzende (Satzende des Hauptsatzes) findet. Doch dieser Meinung gegenüber stehen andere Auffassungen, die besagen, es handelt sich um **Verben mit festen Präpositionen (warten auf; danken für; denken an).** Letztendlich kommt man sowieso zur Schlussfolgerung, dass während trennbare Verben **PA als Verbzusatz/Partikel** einleiten, leiten **Verben mit festen Präpositionen** eher **PA als Korrelat** ein. Manchmal wird es zwischen diesen beiden Erscheinungsformen von PAn verwechselt. Dabei wird meistens die Rede sein, dass wenn ein **PA als Verbzusatz** auftritt, dann fungiert **es de facto auch als Korrelat.** Etwas das, wie erläutert, nicht der Fall ist. Ein PA kann in der Tat **als Verbpartikel fungieren ohne jedoch dabei als Korrelat aufzutreten** (Siehe Beispiele 55, 56, 57) und umgekehrt, also, **als Korrelat ohne als Teil eines trennbaren Verbs fungieren zu müssen** (Siehe dazu die Beispiele 58, 59, 60).

¹⁰¹ Bitte nicht zu verwechselt. Bei PA als Verbzusatz steht das PA **am Satzende.** Dabei handelt es sich meistens um **einen einzelnen Satz.** Dies hingegen handelt es sich bei PA als Korrelat meistens **um zwei Sätze**, wobei das PA dann **am Ende des Hauptsatzes** steht. Gemeinsam haben beide PA-Formen, die Tatsache, dass sie **eng mit Verben verbunden sind.** Während die eine Form **als Verbpartikel** gilt, fungiert die andere Form eher **als Ableitung von zu Verben gehörende festen Präpositionen.**

Heuristik auf Beine zu stellen, um diese beiden Arten von PAn weder voneinander zu unterscheiden, noch vor anderen erscheinenden Formen von PAn zu identifizieren.

3.3.3.2.2. Pronominaladverbien als Subjunktionen: Fall von „Damit“

Ein weiterer Fall, wo die Kontrovers über den pronominalen und nicht-pronominalen Gebrauch von PAn herrscht, ist beispielweise, wenn das PA als **Konnektor** in Betracht gezogen wird. In der Tat, ließ sich bei der oben dargelegten Skizzierung der pronominalen Verwendung von PAn feststellen, dass PA tatsächlich als **Verknüpfungsmittel** zwischen zwei Sätzen auftreten können¹⁰². Sie funktionieren in dieser Hinsicht also wie **Konjunktionen**; dennoch weisen sie, syntaktisch betrachtet, nicht die gleichen Eigenschaften, wie **anerkannte Konjunktionen** (denn, aber, sondern, und usw.). Denn während diese Letzten **das Subjekt** direkt nach ihnen fordern, fordern die PAn eher **das Verb** direkt nach ihrem Auftauchen im Satz. So gesehen können wir also nicht von **PAn als Konjunktionen** im ursprünglichen Sinn reden, geschweige denn noch von **PAn als Subjunktionen**.

Es gibt, in der Tat, manche PAn (vor allem **wo-Verbindungen**), die Sätze unterschiedlicher Grade miteinander verbinden, wobei das Verb, eben wie beim Gebrauch anerkannte Subjunktionen, **auch am Satzende gestellt wird**. Man würde eben auch solche **PAn als Subjunktionen** subklassifizieren, denn sie funktionieren eben wie Subjunktionen (vor allem, wenn sie in relativem Gebrauch, also nicht in ihren interrogativen Formen)¹⁰³ auftauchen. Trotzdem gehören sie, aber nicht zu den anerkannten **Subjunktionen**, wie es etwa bei „Damit“ der Fall ist. Vielleicht ist es so auch, weil sie ihre stellvertretende Funktion nicht verlieren, wenn sie als Subjunktionen auftreten, **etwas, das bei „Damit“ doch der Fall ist**.

Darüber hinaus lässt sich in den nachgeschlagenen Werken merken, dass der Begriff **Konjunktionaladverb** (KA) meistens dann eingesetzt wird, wenn es darum geht, PAn zu bezeichnen, **deren verknüpfende Funktion im Mittelpunkt der wissenschaftlichen Auseinandersetzung steht** (vgl. Hentschel 2010: 294). Der Begriff Konjunktionaladverb wird also meistens gebraucht, wenn die **Verknüpfungsfunktion** den Vorrang gegeben wird. Doch, wie schon im Abschnitt (3.3.2.) über **die terminologische Annäherung zum Begriff PA**

¹⁰²52) *Es liegen viele Löcher. Daneben/Hierüber läuft der junge Mann.* Die PAn Daneben/ Hierneben sind **Verknüpfungsmittel**. Solche PAn werden von Rüttenauer (1978) und Brinker (2010a) als **Textverknüpfungsmittel** genannt.

¹⁰³49) *Sein Plan stützt sich auf etwas, worüber er nicht sprechen darf.* (Relativer Gebrauch)

erläutert wurde, weist der Begriff Konjunkionaladverb (KA), in sich selbst, einen breiten Umfang auf. In der Tat wird er auch eingesetzt, um nicht traditionelle PAn zu bezeichnen, also um PAn zu bezeichnen, die nicht zur Gruppe von da-, hier- und wo-Verbindungen gehören. Dies ist beispielsweise der Fall von PAn wie **deshalb, deswegen, trotzdem, demzufolge, demgegenüber, demgemäß, demnach, infolgedessen** (vgl. Hentschel/Weydt 2013: 246). Gemerkt ist also, dass obwohl man den Vorrang der verknüpfenden Funktion gibt, **bleibt die stellvertretende Funktion immer existent**. Und von einer Kontreverse ist eben auch die Rede, weil während Autoren einen **solchen zweidimensionalen Charakter** von PAn hinsichtlich deren Funktionen anerkennen, erkennen sie ihn andere Autoren aber nicht. Dies wird durch das folgende Zitat Selmas (2020) unterstützt:

Während Fleischer (2002:260) die als Konjunkionaladverb vorkommenden PA als nichtpronominale Verwendung betrachtet, weist Negele auf solche Verwendungen als pronominale Verwendung hin und zieht als Beleg den Duden (2009: 864) heran, in dem hervorhoben wird, dass die Konjunkionaladverbien über Verweisfunktion verfügen und anaphorisch auf den vorher stehenden Satz verweisen (vgl. Negele 2012b: 18) (vgl. Selmas AkolGötkas 2020: 120).

Im Gegensatz zu Fleischer (2002) und im Einklang mit Negele (2012) bin ich der Meinung, dass alle Pronominaladverbien und ggf. auch Konjunkionaladverbien **dank ihrem phorisch-deiktischen Teil eine stellvertretende Funktion aufweisen können**¹⁰⁴. Jedoch muss hier schon darauf hingewiesen, dass nur ein einziges PA von dieser Regel abweicht. Es handelt sich in der Tat um das PA „damit“. Dieses PA kann in der Tat in **einer nichtpronominalen Verwendung** (Kein Verweis auf vorher oder nachher im Text erwähnte Sachverhalte) auftreten. Dabei fungiert es eher als **eine Subjunktion** und unterscheidet sich von den anderen PAn, dadurch, dass es **syntaktisch** betrachtet, wie andere Junktionen funktioniert. Das Verb wird also am Satzende versetzt und nicht direkt nach dem PA, wie es in normalen Fällen wäre. Also in Fällen, in denen **damit** als **PA** und somit nicht als **Subjunktion** auftritt.

Das Ziel dieses Abschnitts war, wie bereits schon am Anfang dieser sich gerade abschließenden Erläuterung hingewiesen, **Eigenschaften von PAn**, die als **Subjunktionen** fungieren zu

¹⁰⁴ Diese Stellungnahme soll längst aber nicht bedeuten, dass **die stellvertretende Funktion die verknüpfende Funktion** ausklammert, sondern eher, dass PAn **Anaphern** sind und verhalten sich deshalb auch sowohl als **kohäsive Mittel** (Verknüpfungsmittel) als auch **als Kohärenzmittel** (Aufbewahrung des inhaltlichen Zusammenhangs der verknüpfenden Sätze durch das PA). Wie schon erläutert, bin ich der Meinung, dass die beiden Funktionen gleichzeitig auftreten. Nun wird, je nach Zielsetzungen, einer Funktion oder einer anderen den Vorrang gegeben.

verzeichnen bzw. so zu verarbeiten, dass wir am Ende, also bei dem Aufbau unseres Algorithmus, in der Lage versetzt sind, unser System so fähig zu gestalten, dass es **effizient** **Erscheinungen von *damit* als Subjunktionen** von **Erscheinungen von *damit* als PA** unterscheiden kann. Dafür wollen wir die Angaben der folgenden Tabelle auswerten.

DAMIT	Als Subjunktion	Als PA
Wortstellung (Stellung von Damit)	Satzanfang	Alle Satzpositionen
Verbstellung	Satzende	Vor-/Mittelfeld
Subjektstellung	Immer direkt nach der Subjunktion (Damit)	Nimmer direkt nach dem PA (Damit)
BEISPIELE	<i>Putins Strategie sei durchsichtig, so Habeck. »Er versucht, die große Unterstützung für die Ukraine zu schwächen und einen Keil in unsere Gesellschaft zu treiben. Dafür schürt er Unsicherheit und treibt die Preise. Dem setzen wir Geschlossenheit und konzentriertes Handeln entgegen. Wir treffen Vorsorge, damit wir durch den Winter kommen.«</i> https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1 (25/07/2022 um 17Uhr)	<i>Ich bin damit einverstanden, dass mir externe Inhalte angezeigt werden. Damit können personenbezogene Daten an Drittplattformen übermittelt werden.</i> <u>Der russische Staatskonzern Gazprom schränkt die Gaslieferungen über die Pipeline Nord Stream 1 ab 27. Juli noch stärker ein. [...]</u> Die Gaslieferungen würden damit auf 33 Millionen Kubikmeter pro Tag von mehr als 160 Millionen Kubikmeter bei voller Kapazität verringert.

		https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1 (25/07/2022 um 17Uhr)
--	--	---

Tabelle 5: Syntaktische Eigenschaften von Damit als Subjunktion und als PA

Wie es in der Tabelle zu merken ist, lässt sich nur die Kategorie **SUBJEKTSTELLUNG** und vielleicht auch das Merkmal **WORTSTELLUNG** zur Unterscheidung der beiden Damit-Vorfälle gut mechanisieren bzw. automatisieren¹⁰⁵. In der Tat mit Hilfe *eines syntaktischen Parsers* wird unser AR-System in der Lage sein, **Wortarten** und **Wortfunktionen** jedes einzelnen automatisch analysierten Satzes zu identifizieren. Und mit einer Heuristik¹⁰⁶, die beispielweise, wie folgendes formuliert werden könnte (beim Auffinden des Worts **damit**; verifizieren zunächst, ob es Vorfeld steht (Dafür muss ein Punkt direkt vor **damit** stehen); dann verifizieren, ob ein Substantiv/ein Pronomen ihm folgt (Ist es der Fall, dann hat man mit einer Erscheinung **von damit als Konjunktion zu tun**; ist es nicht der Fall dann hat man mit einer Erscheinung **von damit als Pronominaladverbien** zu tun). Zudem könnten auch weitere Heuristiken zur Verstärkung des Algorithmus (Auswahl- und Auflösungsprozess) dienen. Man könnte in der Tat, wie folgendes verfahren: zunächst mit Hilfe eines syntaktischen Parsers herausfinden, ob ein Punkt dem Wort **damit** unmittelbar folgt. Ist es der Fall, dann hat man unbedingt mit **einem damit als PA** zu tun. Denn das Wort **damit** steht hier am Satzende. Steht aber der Punkt vor dem Wort **damit**, und wird das Wort **damit** zugleich von einem Verb gefolgt, dann hat man auch mit **einem damit als PA** zu tun; folgt aber ein Substantiv/Pronomen hat man, eben wie oben schon verarbeitet, mit **einem damit als Subjunktion** zu tun. Und sobald der Unterschied zwischen den beiden Erscheinungen gezogen worden ist, kann dann der Auflösungsprozess weitergehen. Mehr über diesen Auflösungsprozess erfährt man erst im zweiten Kapitel des zweiten Teils dieser Arbeit.

Hiermit wollen wir diesen Abschnitt abschließen. Ziel dieses Abschnitts war, wie bereits erwähnt, nichtpronominale von pronominalen Erscheinungen von PAn voneinander zu

¹⁰⁵ Zu den Einzelheiten über die Auswertung dieser Tabelle siehe das Kapitel über den Aufbau des angestrebten Algorithmus.

¹⁰⁶ Heuristiken sind Regeln bzw. Verfahren, die angefertigt sind, um bestimmte Ziele zu erreichen. Zu diesem Begriff und weiteren anderen Begriffen dieser Kategorie kommen wir gleich im zweiten Teil unserer Arbeit zurück.

unterscheiden. Letztendlich ergab sich, dass wir im Rahmen dieser Forschungsarbeit nur **PAn**, die als Subjunktionen fungieren, als **nichtstellvertretend** verzeichnen werden. Alle weiteren Formen von PAn (PA als Verbzusatz & PA als Korrelat) werden hingegen als **pronominal** betrachtet. Versuchen wir nun weiteren Eigenschaften von PAn in Einzelheiten zu bearbeiten.

3.3.3.3. Textlinguistische Funktionen von Pronominaladverbien: Pronominaler Gebrauch von PAn

Nachdem wir, im Abschnitt (3.3.3.1) über die Allgemeinheiten zum Gebrauch von Pronominaladverbien, **den pronominalen Gebrauch** nur kurz erörtert haben, beabsichtigen wir nun, im Rahmen dieses Abschnitts, dieses Funktionalmerkmal von PAn in einer umfangreichen Perspektive zu bearbeiten. Thematisiert wird hier nämlich **die stellvertretende Funktion** von PAn. Diese Funktion zählt, wie vorher schon hingewiesen, zu den Forschungsschwerpunkten der Textlinguistik und wird dort unter dem Begriff **phorisch-deiktische Funktion** bearbeitet (vgl. Finkbeiner 2018: 192).

Die textlinguistischen bzw. phorisch-deiktischen Funktionen von PAn lassen sich, wie der Begriff schon hinausweist, in zwei Gruppen einteilen. Zum Einen, haben wir die Funktionen, die **textimmanent** sind, die also exklusiv nur auf Textebene nachvollziehbar sind. Für einen solchen stellvertretenden Charakter von PAn wird der Begriff **phorische Funktion** eingesetzt (vgl. ebd.: 192). **Phorische Funktionen** lassen sich, ihrerseits, wiederum in zwei unteren Kategorien einordnen: **Die anaphorische oder rückwärtsverweise Funktion** (Das PA steht für eine sprachliche Entität, welche ihm, in der sprachlichen Kette, vorangeht); **die kataphorische oder vorwärtsverweise Funktion** (Das PA steht für eine sprachliche Entität, die, in der sprachlichen Kette, nach ihm kommt). Im Gegensatz zu phorischen Erscheinungen von PAn, in denen die stellvertretende Funktion **eindimensional** ist (Text zu Text), lassen sich deiktische Erscheinungen von PAn eher als **zweidimensional** einstufen (Text zu Welt). Es geht, in der Tat, bei **einem deiktischen PA**, um ein PA, das an der Stelle von nicht im Text versprachlichen Entitäten (**Referenten**) steht (vgl. ebd.: 194).

Allerdings, werden wir im Rahmen dieses Abschnitts nicht dieser oben wiedergegebenen Klassifikation folgen. Und zwar, vor allem, weil wir darauf abzielen, **nur die anaphorische Funktion von PAn in Betracht zu ziehen bzw. automatisch zu verarbeiten**. Und, da der Unterschied zwischen PAn, die entweder **in anaphorischer oder deiktischer Funktion** auftreten, schwer zu ziehen ist, beabsichtigen wir diese beiden Erscheinungsformen von PAn

integrativ zu verarbeiten. **Die kataphorischen Erscheinungen von PAn** werden hingegen in einem einzelnen Unterabschnitt erörtert, und zwar, vor allem, **weil sie sich anhand leicht und einfach aufgebauten Heuristiken von anaphorischen und deiktischen Erscheinungen** differenzieren lassen. Weiterhin werden wir am Ende dieses Abschnitts in der Lage sein, Heuristiken aufzubauen, deren maximale Zielsetzung ist es, **Pronominaladverbienanaphern von Pronominaladverbienkataphern und Pronominaladverbien als Deixis** zu differenzieren. Anders formuliert, werden diese aufgebauten Heuristiken unserem angestrebten AR-System dabei helfen, **Erscheinungen von PAn als Anapher innerhalb eines beliebigen Textes zu sortieren.**

3.3.3.3.1. Pronominaladverbien als Katapher: Die Pronominaladverbienkataphern

Noch unter der Bezeichnung **PA als Korrelat**¹⁰⁷ (Eisendberg 2006: 328) bekannt, verweist der Begriff **Pronominaladverbienkatapher** auf ein PA, das eine kataphorische Funktion in der sprachlichen Kette aufweist. **Eine Kataphorik** besteht dann vor, wenn die völlige Interpretation eines Wortes (**Katapher**) von einem Bezugsausdruck (**Antezedent**) abhängig ist, welcher **im folgenden Satz-oder Textteil** steht (**Vorwärtsverweis**)¹⁰⁸ (vgl. Finkbeiner 2018: 192). Von **einer Pronominaladverbienkatapher** ist dann die Rede, wenn ein PA als Katapher fungiert bzw. wenn die völlige Interpretation eines PAs nur beim **Vorwärtsverweis** möglich ist. Dabei verweist das kataphorische PA hauptsächlich entweder auf **einen Nebensatz (dass-Satz, ob-Satz, w-Satz)** oder auf **eine Infinitivkonstruktion**. Dazu äußert sich Sandberg (2004) mit folgenden Wörtern:

Das kataphorische Pronominaladverb bezieht sich also auf einen Nebensatz oder eine Infinitivgruppe. Der Nebensatz wird am häufigsten von der Konjunktion dass eingeleitet, andere Konjunktionen kommen selbstverständlich auch vor (vgl. Sandberg 2004: 15).

Wie es schon im Abschnitt (3.3.3.2.1.) über die Erscheinung von PA als Verbzusatz erwähnt wurde, werden manche PAn auf Basis **einer zum Verb feststehenden Präposition** aufgebaut. Solche PAn, also PAn, die von Verben regiert sind, werden deswegen auch als Korrelat genannt, weil ihre völlige Interpretation, eben wie oben gesagt, **von dem nachfolgenden Satz-**

¹⁰⁷ Wie schon hingewiesen, können nur **da-Verbindungen** als Korrelat auftreten. **Hier- und wo-Verbindungen** werden deshalb beim Aufbau der Heuristiken **zur Identifizierung von kataphorischen PAn** außeracht gelassen

¹⁰⁸ Im Gegensatz zur **Anaphorik (Rückwärtsverweis)**.

oder Textteil abhängig ist. Allerdings werden korrelate Pan jedoch **nicht exklusiv von Verben** regiert/eingeleitet. Sie können, in der Tat, auch von anderen sprachlichen Einheiten wie **Substantive** und **Adjektive**, welche **eine Ergänzung mit bestimmten Präpositionen** fordern, abhängig sein. Dazu äußert sich Eisenberg (2006b) mit folgenden Wörtern:

*Auch kann **beim Substantiv** in gleicher Weise **wie beim Verb** ein Pronominaladverb als Korrelat verwendet werden (die Hoffnung **darauf**, dass Inge schläft vs. Wir hoffen **darauf**, dass Inge schläft) (vgl. Eisenberg 2006b: 265).*

In Anlehnung an Eisenberg formulieren wir die folgenden Beispiele:

61) Die Angst davor, nie das Ziel zu erreichen (**Substantiv mit fester Präposition**)

62) Paul hat kein Interesse daran, weiter zu schweigen. (**Substantiv mit fester Präposition**)

63) Der Bitcoin ist dafür nützlich, die Bankenmonopole über die Geldpolitik zu zerbröckeln. (**Adjektiv mit fester Präposition**)

64) Manche humanitäre Organisationen sind stolz darauf, dass der Bitcoin sich weltweit als anerkanntes Zahlungsmittel durchgesetzt hat. (**Adjektiv mit fester Präposition**)

Über **die Morphologie (Wortbildung)** kataphorischer PAN hinausgehend, lässt sich merken, wie schon mehrmals hingewiesen, dass PAN, im allgemeinen, **semantische leere** Entitäten sind und müssen daher in Anlehnung entweder auf vorkommende (Anaphorik) oder nachkommende (Kataphorik) sprachliche Entitäten entschlüsselt bzw. interpretiert werden. Doch **syntaktisch** betrachtet, weisen PAN, wie Preyer (1997) mit dem folgenden Zitat hinausweist, besondere Eigenschaften. Sie erfüllen nämlich die Funktion eines **Präpositionalobjekts** und weisen somit **Satzgliederfunktion** auf.

*„Korrelate im Hauptsatz sind zwar semantisch leer, syntaktisch gesehen aber **Satzglieder im Hauptsatz**; sie können deshalb nur auftreten, wenn der Nebensatz ein Satzglied des Hauptsatzes ist.“* (Peyer 1997: 150 zit. nach S.A. Götkas 2020:75)

Von diesem Zitat ausgehend, steht es also klar, dass der Bezug **zwischen korrelatem (kataphorischem) PA und dessen Antezedent nur syntaktisch nachvollziehbar** ist. Von einem PA in korrelater Funktion kann, dem Zitat Peyers (1997) entsprechend, dann nur die Rede sein, wenn der unmittelbare folgende **Nebensatz/Infinitivsatz auch Teil des Hauptsatzes** sein kann. Versuchen wir nun diese Aussage anhand einiger Beispiele zu erörtern.

65) Ich danke dir dafür, dass du mir dieses Auto geschenkt hast.

66) Ich danke dir für das geschenkte Auto.

Im Beispiel 65) hat man mit zwei Sätzen zu tun: einem Haupt- und einem Nebensatz. Außerdem lässt sich, einerseits, feststellen, dass das PA (*dafür*) im Hauptsatz, syntaktisch betrachtet, als **Präpositionalergänzung** fungiert, und verlangt, andererseits, **zur völligen Interpretation** (semantischer Aspekt) einen Bezug zum Nebensatz. Dieser Bezug erfolgt nämlich, indem man **die durch den Nebensatz ausgedrückte Proposition** mit dem PA in Verbindung setzt. Dabei wird das PA durch die Proposition ersetzt und mit Hilfe **dieser Ersetzung**¹⁰⁹ wird man dann **den völligen semantischen Gehalt des PAs** entschlüsseln können. So gesehen, kann behauptet werden, dass das **PA als Platzhalter/Stellvertreter** für den unmittelbaren folgenden Nebensatz fungiert. Und da der Platzhalter im Hauptsatz steht, und da derjenige, dem dieser Platz zugehört, im Nebensatz steht, können wir im Einklang mit Peyer (1997) schlussfolgern, **dass der Nebensatz (Dass-Satz) ein Satzglied des Hauptsatzes ist**. Diese Schlussfolgerung lässt sich nämlich im Beispiel 66) verifizieren. Dort wurde das PA (*dafür*) durch die im Nebensatz ausgedrückte Proposition ersetzt; und gemerkt wurde nämlich, dass der Satz immer einen Sinn aufweist.

Versuchen wir nun, bevor wir einige Heuristiken aufbauen, welche uns dabei helfen werden, kataphorische PAn innerhalb eines Textes auszusortieren, einige Beispielsätze bzw.-Texte zu analysieren. Ziel der folgenden Analyse ist es, **syntaktische Regularitäten/Merkmale von kataphorischen PAn zu verzeichnen**; und somit alle PAn, die solche Merkmale nicht aufweisen, entweder **als deiktisch oder anaphorisch** aufzuzeichnen. Wie schon mehrmals erwähnt, zielen wir in dieser Arbeit, darauf ab, einen automatischen Auflösungsansatz nur für **anaphorische PAn** zu entwerfen. Und, da innerhalb eines Textes PAn in verschiedener Gebrauchsweise auftreten, scheint es uns zur Erreichung der Forschungszielsetzung unabdingbar, **zunächst kataphorische PAn von deiktischen und anaphorischen PAn zu differenzieren**. Dafür wollen wir zunächst häufige Erscheinungsformen von kataphorischen PAn anhand syntaktischer Merkmale verzeichnen.

67) DIHK beklagt bürokratische Hemmnisse bei Maßnahmen gegen Energiekrise

¹⁰⁹Über den stellvertretenden Charakter von korrelaten/kataphorischen PAn äußert sich Holmlander (1979) mit folgenden Wörtern: „**Der pronominale Bestandteil des PA entspricht, historisch gesehen, dem akkusativischen es und dem genetivischen dessen, die ebenfalls beim jeweiligen Kasus vorausweisend verwendet werden können. Dieser inhaltlich leere Bestandteil bekommt, wie es und dessen, die inhaltliche Ergänzung durch den NS/INF.**“ (Holmlander 1979: 12 Zit. Nach. zit. nach S.A. Götkas 2020:111).

08.20 Uhr: Der Deutsche Industrie- und Handelskammertag (DIHK) drängt die Bundesregierung zur Eile bei den Maßnahmen gegen drohende Gasknappheit und beklagt bürokratische Hemmnisse. So würden seit März die Voraussetzungen für die Umstellung der Energieversorgung von Unternehmen von Gas auf Öl diskutiert, sagte DIHK-Hauptgeschäftsführer Martin Wansleben im ZDF. Der Genehmigungsprozess werde dann voraussichtlich noch einmal acht Monate dauern. Das sei zu lang. »Wir müssen davon ausgehen, dass unser Standort erheblich schlechter geworden ist«, sagte er. Er verweist **darauf, dass Gas derzeit in den USA nur ein Zehntel von dem koste**, was in Deutschland dafür verlangt werde. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

67a) Sie überraschte ihre Schwester **dabei, wie sie Geld abhob** (vgl. Dudengrammatik 2009: 581);

67b) Das Gespräch **darüber, ob die Autobahn bei Nebel gesperrt werden darf, ist noch im Gange** (vgl. Dudengrammatik 1998: 374);

68) Ukraine rechnet mit Beginn der Getreideexporte noch für diese Woche

14.15 Uhr: Trotz der jüngsten russischen Angriffe auf die ukrainische Hafenstadt Odessa rechnet die Ukraine schon für die nächsten Tage mit den ersten Getreideexporten auf dem Seeweg. »Wir gehen **davon aus, dass das Abkommen noch in dieser Woche umgesetzt wird**«, sagte der ukrainische Infrastrukturminister Alexander Kubrakow. Er bezog sich auf die am Freitag in der Türkei unterzeichneten Vereinbarungen zur Wiederaufnahme der durch den Krieg blockierten Lieferungen ukrainischen Getreides. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

69) DIHK beklagt bürokratische Hemmnisse bei Maßnahmen gegen Energiekrise

08.20 Uhr: Der Deutsche Industrie- und Handelskammertag (DIHK) drängt die Bundesregierung zur Eile bei den Maßnahmen gegen drohende Gasknappheit und beklagt bürokratische Hemmnisse. So würden seit März die Voraussetzungen für die Umstellung der Energieversorgung von Unternehmen von Gas auf Öl diskutiert, sagte DIHK-Hauptgeschäftsführer Martin Wansleben im ZDF. Der Genehmigungsprozess werde dann voraussichtlich noch einmal acht Monate dauern. Das sei zu lang. »Wir müssen **davon ausgehen, dass unser Standort erheblich schlechter geworden ist**«, sagte er. Er verweist

darauf, dass Gas derzeit in den USA nur ein Zehntel von dem koste, was in Deutschland dafür verlangt werde. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

70) Der Lehrer hat **davon** gesprochen, **dass wir mit unserer Klasse bald einen Ausflug machen.** 86) Er hatte sich noch nicht **daran** gewöhnt, **dass er keinen mehr hat.** "[T07 SEP.03769 die Tageszeitung. 26.09.2007, S. 6]

71) Ich bin **damit** einverstanden, **dass mir externe Inhalte angezeigt werden.** Damit können personenbezogene Daten an Drittplattformen übermittelt werden. Mehr dazu in unserer Datenschutzerklärung. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

72) Das Vorkommen des Fugen-r bei Pronominaladverbien, die mit da- und wo- gebildet werden, ist **davon** abhängig, **ob die Präposition mit einem Vokal oder einem Konsonanten beginnt. Wenn die Präposition mit einem Konsonanten beginnt, fällt das Fugen-r weg.** <https://deutscheграмmatik20.de/adjektiv/adverb/pronominaladverbien/>

73) Faeser und Heil treffen ukrainischen Ministerpräsidenten
17.22 Uhr: Der ukrainische Ministerpräsident Denys Schmyhal hat sich mit Bundesinnenministerin Nancy Faeser und Bundesarbeitsminister Hubertus Heil getroffen. Beide sind zurzeit in der Hauptstadt Kiew. »Die Einigkeit der Ukraine und der EU ist entscheidend für den Sieg über den Aggressor«, twitterte Schmyhal. Er habe mit Faeser und Heil **darüber** gesprochen, **wie man die Zusammenarbeit noch weiter verstärken könne.** <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr)

74) „Nein, deutsche Rechtsanwälte haben mich im Auftrag ihrer Mandanten gebeten, eine Einschätzung abzugeben. Es handelt sich um Voranfragen. Welche Argumente füllen die potenziellen Kläger ins Feld? Die Argumentationslinie sieht so aus: Wenn die LTG-Bank ihre Kunden **darüber** informiert hätte, **dass ihr 2002 Kundendaten gestohlen worden waren**, hätten

die Anleger die deutsche Steueramnestie des Jahres 2004 in Anspruch nehmen können."
(T08/JUL.03292 die Tageszeitung, 22.07.2008, S. 2 zit. nach S.A. Götkas 2020:78)

75) Wenn sie mir **darüber** informieren wollen, **dass er heute noch kommt**. Dann haben sie Pech, denn ich kümmere mich nicht mehr darum, undisziplinierte Kandidaten, wie ihn, zu helfen.

76) Wenn sie mir darüber informieren wollen, dass er heute noch kommt. Dann haben sie Pech, denn ich kümmere mich nicht mehr **darum, undisziplinierte Kandidaten, wie ihn, zu helfen.**

77) Er wartet **darauf, ein Geschenk zu bekommen.** (Schulz/Griesbach 1972: 208. zit. nach S.A. Götkas 2020:76).

78) Die Demonstranten verzichten (**darauf**), **Gewalt anzuwenden.**" (Dudengrammatik 1998: 374);

79) Ich warte **darauf, ihn morgen zu sehen.** (Fleischer 2002: 22 zit. nach S.A. Götkas 2020:120);

80) Er wird sich (**davor**) hüten, **über den Vorfall zu sprechen.** (vgl. Dudengrammatik 1998: 374);

81) Er ist **damit** beschäftigt, **die Kosten für eine Solarversorgung des Hotels im Kopf zu überschlagen.**" (SOI/MAR.00224 Der Spiegel, 12.03.2001, S. 168 zit. nach S.A. Götkas 2020:79);

82) Während des vorübergehenden Stopps russischer Gaslieferungen durch die Ostseepipeline Nord Stream 1 aufgrund von Wartungsarbeiten wurde in Deutschland zwar Gas eingespeichert, aber auf sehr niedrigem Niveau. Uniper hatte sogar **damit** begonnen, **Gas aus Speichern zu entnehmen.** Um einen Mangel im Winter zu vermeiden, will Deutschland die Speicher so schnell wie möglich auffüllen. Bis zum 1. November sollen sie zu 95 Prozent gefüllt sein.
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr)

83) Die für den Vermieter einfachste Lösung sei laut Energie-Experten allerdings häufig der Einbau einer Stromdirektheizung. Das würde jedoch für die Mieter vergleichsweise teuer und ineffizient werden. Das BMWK schreibt dazu: "Reine Stromdirektheizungen sind viel weniger effizient als Wärmepumpen, sie sollten daher nur in besonders gut gedämmten Häusern mit einem äußerst niedrigen Wärmebedarf eingesetzt werden. In diesen Fällen können sie eine

kostengünstige Investition sein. Auch hier wird unterstellt, dass der Strom über die Nutzungsdauer der Stromheizungen schrittweise vollständig dekarbonisiert wird" – verbindlich sind diese Voraussetzungen nach jetzigem Stand jedoch nicht. Energie-Experten fordert das BMWK deshalb **dazu auf, diese Option in der angedachten Mieterschutzvorschrift ebenfalls zu berücksichtigen.** https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829 (25/07/2022 um 17Uhr)

84) Versuchen wir nun, bevor wir einige Heuristiken aufbauen, welche uns **dabei** helfen werden, **kataphorische PA innerhalb eines Textes auszusortieren**, einige Beispielsätze bzw.-Texte zu analysieren. Ziel der folgenden Analyse ist es, **syntaktische Regularitäten/Merkmale von kataphorischen PAn** zu verzeichnen. Wie es schon mehrmals erwähnt wurde, zielen wir in dieser Arbeit **darauf ab, einen automatischen Auflösungsansatz nur für anaphorische PAn zu entwerfen.**

85) Es bleibt **dabei: Wir reisen morgen ab.** " (vgl. Dudengrammatik 2009: 581);

85a) Ich bin weiterhin **dafür: Das Ehegattensplitting muss abgeschafft werden.** (vgl. ebd.: 581);

Wir sind am Ende der Analyse gelangen, welche darin bestanden hat, **kataphorische PAn und deren Antezedenten zu markieren¹¹⁰**, ergibt sich, dass kataphorische PAn folgende syntaktische Merkmale aufweisen:

Syntaktische Regel N°1 Wenn ein kataphorisches PA einen **Nebensatz** (Dass-/ob-/W-Satz) als Antezedent hat, dann lässt es sich mit folgenden syntaktischen Regeln automatisch auffinden.

PA + Komma+ Dass/ob/Fragewort¹¹¹; (Beispiel 67; 67a; 67b)

PA + Verbpartikel + Komma + Dass/ob/Fragewort; (Beispiel 68)

¹¹⁰ Anders ausgedrückt, handelte es sich bei der Analyse, um eine manuelle bzw. **menschliche Auflösung** von **kataphorischen PAn**. Was, worauf wir am Ende dieser Forschung abzielen, ist es, eine Auflösung von **anaphorischen PAn** durch den Computer (**automatische Auflösung**) zu ermöglichen. Dafür müssen zuvor aber kataphorische PAn **auch automatisch aussortiert werden**. Dies ist dann Aufgabe der entworfenen syntaktischen Regeln. Zu näheren Einzelheiten kommen wir in die folgenden Kapitel zurück.

¹¹¹ Zur Fragewörterkategorie gehören Wörter wie: **Wer (Wen, wem, wessen), wo (wohin, woher), was, wie (Wie lange, wie oft, wie viel); Warum, weshalb, wieso, weswegen, wozu, welche etc.** Quelle: <http://www.mein-deutschbuch.de> (02.08 2022 um 12h). Eine Ausführliche Liste wird demnächst aufgestellt.

PA + Verb (in Infinitiv) + **Komma** + **Dass/ob/Fragewort**; (Beispiel 69)

PA + Verb (Partizip 2) + **Komma** + **Dass/ob/Fragewort**; (Beispiel 70 & 73)

PA + Adjektiv + **Komma** + **Dass/ob/Fragewort**; (Beispiel 71 & 72)

PA + finites Verb (Auxiliar + Partizip 2/Modal+ Infinitiv) + **Komma** + **Dass/ob/Fragewort**; (Beispiel 74 & 75)

Syntaktische Regel N°2 Wenn ein kataphorisches PA einen **Infinitivsatz** als Antezedent haben, dann lässt es sich mit folgenden syntaktischen Regeln automatisch auffinden.

PA+ Komma + **NP/Pronomen**; (Beispiele 76; 77; 78; 79)

PA+ Komma + **PP** (Präposition+ NP/Pronomen); (Beispiele 80)

PA+ Adjektiv+ Komma + **NP/Pronomen**; (Beispiel 81)

PA + Verb (Partizip 2) + **Komma** + **NP/ Pronomen**; (Beispiel 82)

PA + Verbpartikel + **Komma** + **NP/Pronomen**; (Beispiel 83 & 84)

PA + finites Verb (Auxiliar + Partizip 2/Modal+ Infinitiv) + **Komma** + **NP/Pronomen**; (Beispiel 84)

Syntaktische Regel N°3 Wenn kataphorische PAn einen **Hauptsatz** als Antezedent haben, dann lässt es sich mit folgenden syntaktischen Regeln automatisch auffinden.

PA+ Doppelpunkt + **NP/Pronomen**; (Beispiel 85; 85a)

Über diese syntaktischen Merkmale hinaus, welche sich insbesondere auf **nachfeldstehende kataphorische PAn stützt**, weisen manche kataphorische PAn auch **Vorfeldposition** im Satz auf. Jedoch tauchen **vorfeldstehende PAn** im sprachlichen Gebrauch extrem selten¹¹² auf und, lassen sie sich, wenn sie dann vorkommen, nur schwer mit den etablierten syntaktischen Regeln auffinden. Dies soll aber auch heißen, dass manche kataphorische Vorfelderscheinungen von PAn sich doch **mit den aufgebauten Regeln automatisch identifizieren lassen**¹¹³. Wenn dies

¹¹² Es wurde nämlich im Laufe der Forschung gemerkt, dass es sich meistens entweder um **eine Anapher oder eine Deixis** handelt, wenn **ein PA im Vorfeld** steht. Kataphorische PAn treten jedoch, wie gesagt, eher selten im Vorfeld auf.

¹¹³ Fälle von vorfeldstehenden kataphorischen PAn, welche sich nicht durch die etablierten Regeln identifizieren lassen. **86)** „Damit haben wir nicht mehr gerechnet, dass du noch kommst. (Schulz/Griesbach 1972: 209). **87) Dagegen hatte mein Vater nichts einzuwenden, dass ich mir ein neues Auto gekauft habe.** (vgl. ebd: 209). Solche Erscheinungen kataphorischer PAn weisen eher syntaktische Merkmale von anaphorischen/ deiktischen PAn auf, deswegen werden sie dann, wie hingewiesen, als solches eingeordnet. Die folgenden Beispiele sind Fälle, in denen das kataphorische PA Vorfeld steht aber doch sich trotzdem als solches von den etablierten Regeln identifizieren

aber nicht der Fall ist, dann wird unser angestrebte AR-System solche PAn höchstwahrscheinlich als anaphorische PAn verzeichnen, etwas, **das zu möglichen Grenzen unseres AR-Systems zählen wird bzw. negativ auf die Performanz des AR-Systems wirken wird**. Letztendlich werden wir den angestrebten Algorithmus so einstellen, dass PAn, die das Eine oder das Andere des eben etablierten Regelpakets/Merkmalsspaket aufweisen, als **kataphorische PAn verzeichnet bzw. identifiziert werden**. Solche PAn werden deswegen nicht mehr im Weiterverlauf des Algorithmus berücksichtigt, **weil uns nur die automatische Auflösung anaphorischer PAn interessiert**. Versuchen wir nun herauszufinden, wie wir anaphorische PAn innerhalb eines Textes **effizient automatisch** identifizieren und zugleich auflösen können.

3.3.3.2. Pronominaladverbien als Deixis oder als Anapher

Die Wahl **deiktischer und anaphorischer PAn** integrativ zu behandeln, stützt sich, so Brinker (2010a), auf das Faktum, dass diese beiden Erscheinungsformen von PAn, **im Gegensatz zu kataphorischen PAn**, gleiche **Merkmalsbündel hinsichtlich ihrer jeweiligen Syntax** innerhalb einer beliebigen sprachlichen Kette (Satz/Text) aufweisen (Brinker 2010a: 24). Davon ausgehend, ist also ein klarer Unterschied zwischen diesen beiden Erscheinungsformen von PAn (Deixis¹¹⁴ und Anapher), auf Basis syntaktischer Regeln, gleich gut, wie unmöglich zu ziehen. Daher könnte es also zum Stande kommen, dass **der von uns angestrebte automatische Auflösungsprozess für Pronominaladverbienanaphern**, auch beim Auftritt **deiktischer PAn**, ungestört weiterverlaufen wird. Dies wird nämlich dazu führen,

lässt. 88) *Dafür, dass er mir jetzt geholfen hat, bin Ich Ihm für Immer dankbar.* (vgl. ebd: 209). 89) *Darüber, dass der Urlaub vorbei ist, ärgert sich Michael.* (vgl. Aufmkolk 2014: 9).

¹¹⁴90) *Ismael spielt dahinter.* 91) *Die Schule namens Victor-Hugo* steht **davor**. Als Ismael klein war, spielte er immer **dahinter**. Im Beispielsatz (90) steht das PA (*Dahinter*) für ein Element, das zuvor nicht versprachlicht worden ist, sondern eher nur durch eine Geste dem Rezipienten bewusstgemacht wird (**PA als Deixis**). Im Gegensatz zum Beispiel (90) wird im Beispiel (91) gemerkt, dass das PA (*Dahinter*) sich auf die NP (*Die Schule namens Victor-Hugo*) verweist (**PA als Anapher**). Versucht man nun das PA (*davor*) im Beispiel (91) manuell zu analysieren bzw. aufzulösen (**keine durch den Computer geführte automatische Auflösung**), dann lässt sich, anhand semantischer Inferenzen, feststellen, dass dieses PA ein Deixis ist. Und zwar, weil es direkt auf einen Ort referiert, welcher zuvor durch keine andere sprachliche Einheit im Text versprachlicht worden ist. Doch bei der automatischen Auflösung solcher PAn sehen die Sachen ganz anders aus, denn solche PAn weisen, wie schon zuvor erwähnt, gleiche syntaktische Merkmale wie anaphorische PAn auf, und werden deswegen auch als solches während des Auflösungsprozesses identifiziert. Etwas, dass die Performanz des angestrebten Algorithmus bzw. AR-Systems beeinträchtigt wird. Dagegen könnte man eben schon **semantische Inferenzen** auf der Ebene der maschinellen Auflösung von Anaphern miteinbeziehen. Doch solche Aspekte der Computerlinguistik gehören nicht **dem Forschungsfeld der theoretischen Computerlinguistik (CL)**, sondern eher dem der praktischen CL und genauer gesagt, **dem der Wissensrepräsentation nicht sprachlichen Wissens durch Maschinenlearning** (vgl. Kai-Uwe Carstensen 2010: 539-544).

dass **deiktische PAn**, weil sie im Laufe des Identifizierungsprozesses als **anaphorische PAn** durch das AR-System eingestuft worden sind, **Antezedenten zugeordnet werden**, und dies, obwohl sie keinen bedürfen. Allerdings werden wir im Laufe **des Entwurfs eines automatischen Auflösungsprozesses für anaphorische PAn** immer wieder versuchen, Ansätze zu elaborieren, welche uns dann dabei helfen werden, deiktische PAn von anaphorischen PAn differenzieren zu können und somit zugleich auch die Performanz unseres Algorithmus bzw. AR-Systems zu steigern¹¹⁵. In dieser Hinsicht, werden wir in den folgenden Abschnitten zunächst einige Beispiele analysieren (manuelle Auflösung von anaphorischen PAn) und anhand erhobener Regularitäten, **Heuristiken zur automatischen Auflösung von anaphorischen da-; hier- und wo-Verbindungen entwerfen**.

3.3.3.3.2.1. Die Da- und Hier- Pronominaladverbienanaphern

Nach Brinker (2010a) können PAn, textlinguistisch betrachtet, als **Stellvertreter für Substantive, Sätze, Satz- und Textteile im Text erscheinen**.¹¹⁶ Sandberg (2004) ist auch dieser Meinung und fügt hinzu, dass anaphorische PAn sich auch **auf die ganze Proposition**¹¹⁷, beziehen können (vgl. Sandberg 2004: 14). Zusammengezogen, ist also von einem anaphorischen PA die Rede, wenn das PA **auf eine Proposition des vorangehenden Textteils** verweist, welche, ihrerseits, entweder durch den unmittelbaren vorangehenden Satz/Satzteil (ein Substantiv, eine substantivische Wortgruppe, eine Nominalphrase (NP) Präpositionalphrase (PP) oder Substantive der Präpositionalphrase (vgl. Eisenberg 2006b: 198) oder eher durch den unmittelbaren vorangehenden Textteil bzw. Textabschnitt (Mehr als ein Satz) versprachlicht bzw. geäußert worden ist. Dies lässt sich nämlich durch die folgenden Zitate bekräftigen:

¹¹⁵ Einer dieser Ansätze könnte wie folgendes erläutert werden: Wenn ein PA Vorfeld im Satz auftritt, welcher seinerseits auch als erster Satz eines Beispieltexes fungiert, dann handelt es sich unbedingt entweder um **ein deiktisches oder ein kataphorisches PA**. Denn es besteht bei solchen Erscheinungen von PAn keine Möglichkeiten **Rückwärtsbezüge** zu schließen. Dies hingegen sind **Vorwärtsbezüge** und **außersprachliche Bezüge** möglich.

¹¹⁶ In diesem Zusammenhang weisen Pronominaladverbienanaphern gleiche Merkmale **wie komplexe Anaphern** (Siehe Abschnitt 2.3.2.5.). Aus diesem Grund können wir die Pronominaladverbienanaphern als eine weitere Variante **von expliziten Anaphern bzw. als direkten Anaphern** einstufen

¹¹⁷ Gansel/Jürgens (2009) verstehen den Begriff **Proposition** als der Inhalt eines Satzes bzw. als der durch einen Satz ausgedrückte Sachverhalt (vgl. Gansel/Jürgens 2009: 44). Sie bekräftigen ihren Standpunkt nämlich mit der folgenden Haltung „*Die wichtigste Komponente des Satzinhalts ist das Prädikat, wobei der Terminus Prädikat nicht im Sinne der traditionellen Satzgliederung (Subjekt + Prädikat) verstanden werden darf, sondern im Sinne der Prädikatenlogik (als Aussage:präzisieren = etwas aussagen).*“ (vgl. ebd.: 44)

[...] Dies heißt m.a.W., dass **die anaphorische Funktion des Pronominaladverbs** zwei Unterabteilung besitzt: **Erstens** kann das Pronominaladverb sich **auf eine ganze Proposition** beziehen, **zweitens** kann es sich **auf ein einzelnes Substantiv** beziehen, dessen Denotat keine Person darstellt. (vgl. Sandberg 2004: 14)

Beim anaphorischen Rückweisen kann sich das Präpositionaladverb auf verschiedene Größen beziehen: **auf ein Substantiv, auf eine Nominalphrase oder auf einen ganzen Satz**: Sie schoben **die Bücher** beiseite; **dahinter** (hinter den Büchern) kam Staub zum Vorschein. **Er nahm das Geld entgegen**; **dabei** (beim Nehmen des Geldes) hatte er ein schlechtes Gewissen. (vgl. Dudengrammatik 2009: 581)

Aus den im Zitat der Dudengrammatik (2009) erwähnten Beispielen, lässt sich feststellen, dass während das PA **dahinter** das Substantiv **Bücher** als Antezedent hat, hat das PA **dabei** eher die vom vorangehenden Satz **ausgedrückte Proposition** (Satzinhalt) als Antezedent. Jedoch besteht **die Schwierigkeit einer Automatisierung** solcher Auflösungsprozesse darin, dass es anhand rein syntaktischer Parameter, fast nichteinstellbar ist, zu normieren, wann der Antezedent **ein ganzer Satz, Satz- oder Textteil** ist (eine Proposition oder ein Sachverhalt) oder eher, wann **ein Substantiv** (meistens Gegenstände/ Objekte aber auch Handlungen) als Antezedent fungiert. Dazu schließen sich weiterhin weitere Problematiken an. Es wird nämlich beispielweise schwer zu unterscheiden, wann es sich um ein Substantiv als Antezedent handelt, wenn dieses als **Subjekt oder eher als Objekt (Teil eines Präpositionalobjekts)** fungiert. Die gleichen Überlegungen gelten auch für die **Hier- und Wo-Verbindungen** (siehe für die Einzelheiten die folgenden Analysebeispiele, wobei Antezedenten den jeweiligen anaphorischen PAn zugeordnet werden). Ziel dieser manuell aufgeführten Auflösung von Pronominaladverbienanaphern, ist es, wie schon am Anfang dieses Abschnitts erläutert, einige Regularitäten beim Auffinden des Antezedens herauszuschließen. Diese Regularitäten werden uns dann bei dem Entwurf des Algorithmus weiterhelfen.

- **DARAN & HIERAN**

92) Nach der angekündigten weiteren Drosselung russischer Gaslieferungen nach Deutschland hat Wirtschaftsminister Robert Habeck noch einmal den Ernst der Lage betont. »Wir sind in einer ernsten Situation. Es wird auch Zeit, dass das alle verstehen«, sagte er in den ARD-»Tagesthemen«. **Deutschland müsse den Gasverbrauch runterbringen**. »**Daran** arbeiten wir. «Die Maßnahmen müssten konsequent weiter umgesetzt werden. Das Land müsse zusammenstehen und sagen: »Ja, Putin hat das Gas, aber wir haben die Kraft.«

<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr);

93) In der Funktion `countWords` werden durch Leerzeichen voneinander getrennte Wörter aus dem im Funktionsaufruf (Zeile 25) übergebenen Datei-Datenstrom gelesen; in der `mapwords` werden die so segmentierten Wörter als Schlüssel angelegt und der Zähler für ihre Häufigkeit durch den sogenannten Autoinkrement Operator `++` erhöht, wenn das jeweilige Wort erneut gefunden wird. **In der in Zeile 27 beginnenden for-Schleife wird über die Einträge der map iteriert, wobei die Attribute first und second aus dem Template pair stammen.** Hieran sieht man, wie die Struktur aus Schlüssel-/Wertpaaren intern realisiert wird. (vgl. Kai-Uwe Carstensen, 2010: 471);

- **DARAUF & HIERAUF**

93) Wir haben also im Prinzip zwei Arten von Bedeutungen: eine „weltenübergreifende“ und eine bzgl. einer bestimmten Welt. Im Hinblick auf erstere denotiert ein Satz eine Proposition, im Hinblick auf letztere einen Wahrheitswert. **Diese Unterscheidung von Bedeutungsarten ist auch in anderer Hinsicht relevant** und wir werden **darauf** in einem späteren Abschnitt zur Intensionalität zurückkommen. Im Moment genügt es, sich die Satzbedeutung bzgl. einer bestimmten Welt als Wahrheitswert vorzustellen. (vgl. ebd.: 2010: 333);

94) **Einer der Aspekte von NLS, der (Computer-)Linguisten überwiegend bekannt, Laien hingegen überwiegend unverständlich ist, betrifft die Frage, warum es überhaupt so schwierig ist, Systeme zu konstruieren, die auch nur annähernd über eine der menschlichen Sprachfähigkeit ähnliche Sprachkompetenz verfügen.** Eine –wenn auch grobe– Antwort **hierauf** lässt sich anhand dreier grundlegender Problembereiche geben: der Sprachverstehensproblematik, der Wissensverarbeitungsproblematik und der Sprachgenerierungsproblematik (Kai-Uwe Carstensen, 2017: 33);

- **DARAUS & HIERAUS**

95) **In der zweiten Methode überwacht das System die Aktionen des Benutzers einschließlich seiner Antworten auf Fragen** und zieht **daraus** Schlussfolgerungen über ihn. Größtenteils können die so gewonnenen Informationen als die in Benutzeraussagen enthaltenen Präsuppositionen (s. Abschnitt 3.7.3) aufgefasst werden. (vgl. Kai-Uwe Carstensen, 2010:430);

96) In frühen Anwendungssystemen der Künstlichen Intelligenz (KI), den Expertensystemen, führte das Fehlen von Wissen über die Welt zur so genannten „Zerbrechlichkeit“ (brittleness): Schon die geringsten Abweichungen von vorgegebenen Eingabemustern führ(t)en zu Fehlern und Systemabstürzen, die für die Benutzer nicht nachvollziehbar waren. Hieraus entstand das Desiderat allgemein und wiederverwendbarer Wissensressourcen (sog. (common sense) knowledge sharingandreuse), heute allgemein Ontologien genannt. (vgl. ebd.: 2010: 532);

- DABEI & HIERBEI

97) Millionen von Textseiten werden heute jeden Tag mit dem Computer erfasst und bearbeitet: Briefe, E-Mails, Aufsätze, Zeitungen, Bücher usw. Dabei schleichen sich Fehler ein. Schon früh entstand die Idee, zur Erkennung und Korrektur dieser Fehler den Computer zu verwenden und Systeme zu entwickeln, die Fehler in Texten automatisch korrigieren oder wenigstens markieren und dem Benutzer Korrekturvorschläge machen (vgl. ebd. : 555);

98) Mit dieser Darstellung kann die Frage, welche Dokumente am besten zu einer Anfrage passen, unmittelbar beantwortet werden: Es sind die Dokumente, die der Anfrage relativ zu einer gegebenen Metrik (der Gewichtung) am ähnlichsten sind. Das Vektorraummodell bietet daher den Vorteil, dass man eine uniforme Repräsentationsebene für Dokumente und Anfragen besitzt, mit der verschiedene Ähnlichkeitsfunktionen untersucht und verglichen werden können. Grundlage für viele Ähnlichkeitsfunktionen bildet der Cosinus. Hierbei wird die Matrix als m-dimensionaler Vektorraum interpretiert, wobei jedes Dokument einem Punkt entspricht (vgl. ebd.: 2010: 590).

- DADURCH & HIERDURCH

99) Ein etwas anderer Einsatzbereich für Defaultregeln sind Fälle von Neutralisierung. z. B. enden die meisten Formen des Verbparadigmas gleichlautend auf -(e)n. (z. B. klau-en = Infinitiv, Präsens 1.pl. und Präsens 3.pl). Hier könnte man annehmen, dass die Regel, die -(e)n an Verben suffigiert als Default anderen Regeln gegenübersteht, die gegebenenfalls Endungen wie -(e)st, 2.sg. einführen. Dadurch wird es überflüssig, drei verschiedene Regeln anzunehmen, die -(e)n affigieren.

100) Sowohl im Call-Center-Umfeld als auch im Sicherheitsbereich kann es sinnvoll sein, statt einer vollständigen Transkription nur eine Schlüsselwortsuche durchzuführen. Dabei werden szenarioabhängig ca. 50 bis 100 Schlüsselwörter festgelegt, die eine Kategorisierung des Gesprächs erlauben (z.B. „Herzlichen Dank“ als Hinweis auf ein freundliches Gespräch), und nach denen dann im Gespräch gesucht wird. Zwar reduziert sich **hierdurch** die extrahierbare Information deutlich, aber ein „Word Spotter“ ist schneller auf neue Gesprächsthemen übertragbar und mit wenigem Sprachmaterial trainierbar als ein Transkriptionssystem (vgl. ebd.: 2010:618).

- **DAFÜR & HIERFÜR**

101) Putins Strategie sei durchsichtig, so Habeck. »Er versucht, die große Unterstützung für die Ukraine zu schwächen und einen Keil in unsere Gesellschaft zu treiben. Dafür schürt er Unsicherheit und treibt die Preise. Dem setzen wir Geschlossenheit und konzentriertes Handeln entgegen. Wir treffen Vorsorge, damit wir durch den Winter kommen.«
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

102) Maschinelle Übersetzung gesprochener Sprache: Die Übersetzung von gesprochener Sprache hat in den letzten Jahren einen rasanten Aufschwung genommen. Die Gründe **hierfür** liegen zum Einen in großen Fortschritten insbesondere bei der statistischen Übersetzung und zum Anderen in dem stark angewachsenen Interesse an Simultanübersetzungen aufgrund der veränderten politischen und wirtschaftlichen Randbedingungen (vgl. Kai-Uwe Carstensen, 2010:620).

- **DAGEGEN & HIERGEGEN**

103) Solaranlagen-Besitzern droht bald ein Bußgeld: Das müssen Sie jetzt **dagegen** tun
https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829 (25/07/2022 um 17Uhr).

- **DAHINTER & HIERHINTER**

105) Sie schoben die Platten beiseite, **Dahinter** (= hinter den Platten) war eine Geheimtür. (vgl. Dudengrammatik 1998: 374).

- **DARIN, DAREIN & HIERIN, HIEREIN**

107) Die kommunikative Absicht, einem Auskunftsuchenden Kenntnis über den Weg von einem Ort A zu einem Ort B zu verschaffen, bedarf bei nicht-trivialer Wegstrecke zu ihrer Umsetzung einer Folge sorgfältig geplanter und realisierter Äußerungen. Darin werden die Abschnitte, aus denen sich ein solcher Weg zusammensetzt, in geordneter Weise und geeigneter Ausführlichkeit angeführt, je nach situativen Gegebenheiten und Informationsbedürfnis des Adressaten, wie etwa in folgendem Beispiel: [...] (vgl. Kai-Uwe Carstensen, 2010: 436).

- **DAMIT & HIERMIT**

108) Der russische Staatskonzern Gazprom schränkt die Gaslieferungen über die Pipeline Nord Stream 1 ab 27. Juli noch stärker ein. Eine weitere Turbine müsse wegen Wartung außer Betrieb gesetzt werden, teilte das Unternehmen mit. Die Gaslieferungen würden **damit** auf 33 Millionen Kubikmeter pro Tag von mehr als 160 Millionen Kubikmeter bei voller Kapazität verringert. Das würde nur noch gut 20 Prozent entsprechen.
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

109) Direkte MÜS gehören zu den ältesten Systemen für maschinelle Übersetzung (z. B. war das System der Georgetown University, aus dem später Systran hervorging, von diesem Typ) entsprechend sind sie recht einfach strukturiert. Meist handelt es sich um Wort-für-Wort-Übersetzungen, bei der eine morphologische Analyse vorgenommen wird. Hiermit werden immerhin Konstruktionen wie It's oder Ein- und Ausgang in die zugrundeliegenden Vollformen (It is und Eingang und Ausgang) abgebildet, so dass ihre ökonomische Übersetzung möglich wird (vgl. Kai-Uwe Carstensen, 2017: 189).

- **DANACH¹¹⁸ & HIERNACH**

110) „Zuerst die Suppenbrühe zum Kochen bringen, den Reis zugeben und ausquellen lassen. Paprika entkernen, waschen und die weißen Trennwände herausschneiden. Tomaten kurz in kochendes Wasser tauchen, kalt abbrausen, enthäuten und den grünen Stengelansatz

¹¹⁸ Darüber hinaus lässt sich auch in der Forschungsliteratur feststellen, dass, wenn die Proposition, worauf man sich stützt, um das PA völlig interpretieren zu können, einen Zeit- oder Situationspunkt ist, dann redet man von **temporalen Anaphern** (vgl. Knees 2009) und dabei ist es meistens der Fall von PAN wie „**danach, darauf und davor**“ (vgl. Eisenberg 2006b: 213). Über den Fall von dem PA **danach** äußert sich Knees (2009) mit den folgenden Wörtern: „Syntactically, the antecedent of **danach** is either an S-based phrase (a sentence, a clause, a VP etc.) referring to a situation or an NP-based phrase (an NP or PP) referring to a situation, a time-entity or an Object (coerced into a situation). Considering the ontological status of the referents of the antecedent and of the anaphor. We can distinguish between **situational reference** (the referent is a situation) and **temporal reference** (the referent is a time-entity).“ (Knees 2006: 58)

entfernen. *Danach* die Zwiebel schälen, Zucchini und Petersilie waschen. (Mannheimer Morgen, 05.03.1989, Lokales; 1. Tag Vitaminreiches Zit. nach. Selma AkolGötkas 2020).

111) *Die Sprecherin des Instituts für Plasmaphysik, Isabella Milch, bestätigte gestern einen entsprechenden Bericht des Magazins „Stern“, Danach hat die Firma NTG (Neue Technologien GmbH) gegen die wegen Verstoßes gegen das Außenwirtschaftsgesetz ermittelt wird, von einem Mitarbeiter des Instituts eine Lizenz eine Entseuchungsanlage hochradioaktiven Tritums erworben. (Mannheimer Morgen, 05.01.1989, Polotik; Wissenschaftler in illegale Atomexporte verwickelt).“* (vgl. Sandberg 2004: 30; nach. Selma AkolGötkas 2020).

- DANEKEN & HIERNEKEN

112) Knees erklärt, dass das Antezedens von danach¹¹⁹ syntaktisch sowohl eine Satz-basierte Phrase (ein Satz, Satzteil, VP usw.) ist, die sich auf eine Situation bezieht, als auch eine NP-basierte Phrase (ein NP oder PP), die sich auf eine Situation, eine Zeiteinheit oder ein Objekt bezieht, sein könnte. *Daneben* betont sie, dass zwischen der situativen (der Referent ist eine Situation) und der zeitlichen Bezugnahme (der Referent ist eine Zeiteinheit) in Anbetracht des ontologischen Status der Referenten des Antezedens und der Anapher zu unterscheiden ist (vgl. Selma Akolötkas, 2020).

- DARÜBER & HIERÜBER

113) Der Rasen des Spielfelds muss in Zukunft besser gepflegt werden. *Darüber* müssen wir noch einmal sprechen. Sie fuhr den Wagen nicht in die Garage, sondern stellte ihn davor ab. (vgl. Dudengrammatik 1998: 374)

113a) Das Thema ist noch nicht erschöpft; darüber/hierüber müssen wir noch einmal sprechen. (vgl. ebd.: 374)

- DARUNTER & HIERUNTER

114) *Wegen Verbrechen gegen die Zivilbevölkerung im Donbass seien bereits mehr als 1300 Strafverfahren gegen mehr als 400 Personen eingeleitet worden. In den Vorermittlungen seien*

¹¹⁹ Hier haben wir beispielweise mit einem PA, das keinen relationalen Charakter aufweist – „**Relationale Adverbien nehmen Bezug auf den Sprecher/Schreiber, den Ort und die Zeit des Sprechers/Schreibens. Sie haben zeigende (deiktische) Funktion oder dienen im Text der Rückverweisung oder Vorausweisung (anaphorische und kataphorische Funktion)**“ (Duden 1998: 365)–, das also weder anaphorisch, kataphorisch und vielleicht auch nicht deiktisch gebraucht worden ist. Beim Auftreten solcher PAn wird unser angestrebtes AR-System diesem Wort höchstwahrscheinlich ein Antezedent zuordnen und dies, obwohl dieses keinen bedarf.

gut 220 Personen der Verbrechen gegen die Menschlichkeit und den Frieden überführt worden – Anklagepunkte, die nicht verjährten. **»92 Kommandeure und ihre Untergebenen wurden angeklagt, 96 weitere, darunter 51 ukrainische Offiziere, zur Fahndung ausgeschrieben«**, sagte Bastrykin. Er berichtete zudem von Ermittlungen gegen Briten, US-Amerikaner, Kanadier, Niederländer und Georgier wegen Söldnertums. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

115) **Die Sprachtechnologie befasst sich mit der Entwicklung marktreifer Anwendungen der maschinellen Sprachverarbeitung (Computerlinguistik, CL).** Ein paar reale (!) Szenarien sollen andeuten, was **hierunter** zu verstehen ist: [...] (vgl. Kai-Uwe Carstensen, 2017: 7).

- **DAVON & HIERVON**

116) Seit dem Start der russischen Invasion am 24. Februar wurden in Deutschland schon mehr als **915.000 Kriegsflüchtlinge** im Ausländerzentralregister erfasst. Das teilte das Bundesinnenministerium auf Anfrage der Deutschen Presse-Agentur mit. **Davon** seien 890.605 ukrainische Staatsangehörige (Stichtag 19. Juli). Wie viele der Personen sich derzeit noch in Deutschland aufhalten, ist aber unklar. Eine erhebliche Zahl könne bereits in andere Staaten weitergereist oder in die Ukraine zurückgekehrt sein. <https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr).

117) Der Wert für SYNSEM ist wieder eine Merkmalsstruktur vom Typ SYNSEM, die die beiden Attribute LOCAL und Non-LOCAL enthält. Non-LOCAL enthält nichtlokale Information, also Information, die z. B. bei der Analyse von Abhängigkeiten weit voneinander entfernter Teilkonstituenten (long distance dependencies) Verwendung findet und hier nicht näher betrachtet werden soll. **Innerhalb des Wertes von LOCAL gibt es eine weitere Unterteilung in das Attribut CAT (von engl. category), das syntaktische Informationen enthält, das Attribut CONT (von engl. content), das semantische Informationen enthält und das Attribut CTXT (von engl. context), das pragmatische Informationen über den situativen Kontext (wie Informationen über Sprecher und Hörer, Zeitpunkt der Äußerung, etc.) bereitstellt.** Hiervon interessieren uns wieder nur die syntaktischen Angaben in CAT. (vgl. Kai-Uwe Carstensen, 2010: 296).

- **DAVOR & HIERVOR**

118) Sie fuhr den Wagen nicht in die Garage, sondern stellte ihn **davor** ab. (vgl. Dudengrammatik 2009).

- DAZU & HIERZU

119) Die für den Vermieter einfachste Lösung sei laut Energie-Experten allerdings häufig der Einbau einer Stromdirektheizung. Das würde jedoch für die Mieter vergleichsweise teuer und ineffizient werden. Das BMWK schreibt **dazu**: "Reine Stromdirektheizungen sind viel weniger effizient als Wärmepumpen, sie sollten daher nur in besonders gut gedämmten Häusern mit einem äußerst niedrigen Wärmebedarf eingesetzt werden. In diesen Fällen können sie eine kostengünstige Investition sein. Auch hier wird unterstellt, dass der Strom über die Nutzungsdauer der Stromheizungen schrittweise vollständig dekarbonisiert wird https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829(25/07/2022 um 17Uhr).

120) Desweiteren gingen wir davon aus, dass die Informationsanfrage in Form einer vollständigen natürlichsprachlichen Frage formuliert ist und dass Timple dies schon entsprechend interpretieren kann. Hierzu muss Timple jedoch „verstehen“, dass mit der Frage das Auffüllen der Tabelle Umsatzmeldungen „gemeint“ ist (vgl. Kai-Uwe Carstensen, 2010: 583).

- DAZWISCHEN & HIERZWISCHEN

Siehe hierzu den empirischen Teil der Arbeit.

3.3.3.3.2. Die Wo- Pronominaladverbienanaphern

Ebenso wie die Da- und Hier- Verbindungen, können die **Wo-Verbindungen als Deixis und/oder als Katapher** in einer sprachlichen Kette erscheinen. Dies ist dann der Fall, wenn sie **interrogativ** gebraucht werden. Dabei stehen sie meistens Satzanfangs und weisen daher keine Möglichkeiten auf, Rückwärtsbezüge zu etablieren/schaffen. Im Gegensatz zum interrogativen **Gebrauch von Wo-Verbindungen**, welche als deiktischer/ kataphorischer Gebrauch aufgefasst werden können, weist **der relative Gebrauch** von Wo-Verbindungen **eher anaphorische Merkmale auf**. Es ist in der Tat durchaus möglich einem in relativen Gebrauch auftretenden WO-Pronominaladverb ein Antezedens zuzuschreiben (vgl.

Dudengrammatik 2009: 583). Die folgenden Beispiele sollen zur Bekräftigung dieser Erläuterung dienen.

Interrogativer	Gebrauch:	DEIXIS/ Relativer Gebrauch : ANAPHER KATAPHER
<i>Wovor hast du Angst?</i>		<u>Etwas</u> , <i>wovor</i> du Angst hast.
<i>Worüber spricht sie?</i>		<u>Etwas</u> , <i>worüber</i> sie spricht.
<i>Worauf warten sie?</i>		<u>Etwas</u> , <i>worauf</i> sie warten.
<i>Woran denkt er?</i> (vgl. Dudengrammatik 2009: 583).		<u>Etwas</u> , <i>woran</i> er denkt. Der Brief, <i>wovor</i> ... / <i>worüber</i> ... / <i>worauf</i> ... / <i>woran</i> ...; (vgl. Dudengrammatik 2009: 583).

Tabelle 6: Deiktische/kataphorische und anaphorische Erscheinungen von Wo-Verbindungen

Sobald dieser Unterschied zwischen Wo-Verbindungen in ihren deiktischen (ggf. Kataphorischen) und ihren anaphorischen Erscheinungen gezogen wird, wollen wir nun das Funktionieren **der WO-Pronominaladverbienanaphern** anhand einiger Beispieltex te darstellen. Dabei soll in jedem Beispieltex t die Anapher und das entsprechende Antezedens markiert werden. Ziel dieser Manöver ist es, Regularitäten/ Heuristiken für eine Automatisierung dieses Prozesses (Auflösungsprozess) zu erarbeiten.

- **WORAUS**

121)Die Anforderungen, die in der TREC an die teilnehmenden Systeme gestellt werden, werden ständig hinterfragt und geändert. **Insgesamt lässt sich ein erheblicher Anstieg der Anforderungen über die Jahre feststellen,** *woraus* gleichzeitig ein theoretischer und praktischer Fortschritt in der Technologie resultiert. Der Effekt der TREC zeigt sich nicht zuletzt am stattfindenden Technologietransfer. Beispielsweise wurzelt das LCC-System in einem TREC-QA-Sieger (vgl. Kai-Uwe Carstensen, 2017: 110).

122)Auch die Optimalitätstheorie lässt sich mit regulären Modellen modellieren. **Die Grundidee der Optimalitätstheorie ist, dass die Abbildung von der phonologischen auf die phonetische Ebene nicht deterministisch vorgegeben ist, sondern dass ausgehend von einer phonologischen (lexikalischen) Repräsentation alle phonetischen Repräsentationsmöglichkeiten in einer Generatorkomponente GEN frei generiert werden, die dann durch eine geordnete (ranked) Menge von universellen Constraints CON in einer**

Evaluationskomponente EVAL gewichtet werden, die die Anzahl der Constraint-Verletzungen registriert, woraus schließlich die Übersetzung mit den wenigsten Constraint-Verletzungen als die optimale Übersetzung gewählt wird (vgl. Kai-Uwe Carstensen, 2010: 209).

- WOBEI

123) Konzepte können über eine Konjunktion von Konzeptbeschreibungen definiert werden, wobei Konzeptbeschreibungen u. a. Konzeptnamen (wie bei der Definition von Mann) oder Rollenbeschreibungen (wie bei der Definition von Essen) sein können. Die angegebenen Rollenbeschreibungen formalisieren die Wertebeschränkungen für die jeweilige Rolle (vgl. ebd.: 539).

124) Tabelle 2 fasst die Ergebnisse dieser Analyse zusammen, wobei die Häufigkeiten als prozentualer Anteil der Gesamtzahl für die jeweilige PA (dafür, dazu, davon, damit) im jeweiligen Korpus (L2, L1) wiedergegeben wird (vgl. Selma AkolGötkas, 2020).

125) In den darauffolgenden sechs Spalten sind die Prozentzahlen für kataphorische PA eingetragen, wobei zwischen dem Vorverweis auf Infinitivsätze, auf mit “dass” eingeleitete Ergänzungssätze, und auf indirekte Fragesätze (FS) differenziert wird (vgl. Selma AkolGötkas, 2020).

- WODURCH

126) Im zweiten Schritt wird dem jeweiligen Token aufgrund im Lexikon abgelegter Häufigkeiten über das Auftreten eines Wortes mit einer bestimmten Wortart der wahrscheinlichste Wortarttag zugewiesen, wodurch sich der Parameterraum extrem reduziert. Daraufhin findet eine Klassifikation durch den Lernalgorithmus statt: Zur Disambiguierung kommen sowohl neuronale Netze als auch Entscheidungsbäume zum Einsatz, wobei beide Verfahren ähnlich gute Ergebnisse liefern (vgl. Kai-Uwe Carstensen, 2010 : 218).

127) In silbenbasierten Erkennen ist also die Silbenfolge und bei lautbasierten Erkennen zusätzlich die Lautfolge spezifiziert, während im Fall von Ganzworteinheiten der Schritt der lexikalischen Dekodierung komplett entfällt, wodurch sich die Struktur des Erkennens vereinfacht. Tabelle 3.2 zeigt die maximale Struktur eines solchen Lexikoneintrags (vgl. auch Abbildung 3.19 auf Seite 221) (vgl. ebd.: 226).

- WOFÜR

128) Eine aktuell sehr im Fokus stehende Anwendung ist **die Personensuche im Web**, wofür es auch bereits erste spezialisierte Suchmaschinen gibt, z. B. <http://www.spock.com/> oder <http://www.zoominfo.com/>. Meistens werden die zugrundeliegenden persönlichen Einträge und Profile zunächst von der jeweiligen Person selbst zur Verfügung gestellt (vgl. Kai-Uwe Carstensen, 2010: 598).

129) Grundsätzlich versteht sich dieses Buch als eine Art **Meta-Überblick** (bildlich gesprochen: die Spitze der Spitze des Eisbergs) über den Bereich der Natürlichsprachlichen Systeme, da oft Überblicksartikel über einzelne Anwendungsgebiete die wesentliche Grundlage bilden. **Ein solches Vorgehen ist notwendigerweise selektiv und bei dieser Auswahl subjektiv**, wofür ich beim Leser um Verständnis bitte. Diese Selektion ist aber in jedem Fall Prinzip: da an jeder Stelle beliebig detaillierter und ausladender dargestellt werden könnte, würde sich schnell eine Informationsmenge ergeben, die eher hinderlich ist als dass sie nützt (vgl. Kai-Uwe Carstensen, 2017: 5).

- **WOGEGEN**

130) Wir wollen daher als relevante Information den Anteil an der potentiellen Information bezeichnen, der von einem Benutzer als „nützlich“ betrachtet wird. **Man beachte, dass die Bewertung, ob Information relevant ist oder nicht, benutzer- und situationsabhängig ist, wogegen** „Information“ dies per se nicht ist, d. h., Information kann auch irrelevant sein (vgl. Kai-Uwe Carstensen, 2010: 577).

131) Methoden zum Clustering können grob aufgeteilt werden in hierarchische und flache Verfahren. Beim hierarchischen Clustering ergibt der Clusteringprozess eine Baumstruktur. **Jeder Clusterknoten enthält Kindercluster**, wogegen Geschwistercluster die Punkte des Elternclusters partitionieren (vgl. Berkhin 2006). Methoden zum hierarchischen Clustering können weiter unterteilt werden in bottom-up (agglomerative) und top-down (divisive) (vgl. ebd.: 593).

- **WOMIT**

132) **In Forschungsprototypen von QA-Systemen wird Clustering oft zur Erkennung von Paraphrasen eingesetzt und zur Gruppierung von semantisch ähnlichen Textpassagen**, womit eine gezielte Antwortextraktion möglich wird (vgl. Abschnitt 5.3.4). Aktuell werden Clustering Verfahren verstärkt im Bereich der IE exploriert, insbesondere im Bereich der

Behandlung mehrdeutiger Eigennamen (vgl. Abschnitt 5.3.3) und der webbasierten, domänenoffenen IE (vgl. Abschnitt 5.3.3) (vgl. ebd.: 593).

133) Zeile 8. ist das Resultat der Anwendung von Regel WV auf alle Zweige aus Zeile 7. (siehe Punkt 2. oben). Damit wurden schon drei Zweige geschlossen und müssen nicht mehr weiter betrachtet werden. **Eine weitere Expansion mittels $F \wedge$ verzweigt das Tableau erneut und liefert auch wiederum geschlossene Zweige,** womit das ganze Tableau geschlossen ist und somit die Gültigkeit der Folgerung bewiesen ist (vgl. ebd.: 44).

- **WONACH**

134) **Die Implikatur des Satzes (3.178b), wonach** die besagte Tankstelle geöffnet ist, ergibt sich beispielsweise aus der Annahme, dass der Sprecher das Kooperationsprinzip befolgt (dem Autofahrer also helfen will), und dass er die Relevanzmaxime befolgt (der Verweis auf eine geschlossene Tankstelle wäre irrelevant) (vgl. ebd.: 412).

135) **Das Frame-Paradigma betont den objekt-orientierten und schematischen Aspekt der Wissensrepräsentation,** wonach das relevante Wissen über eine Entität direkt an ihrem Stellvertreter verfügbar ist. Zentrales Konstrukt dieses Paradigmas ist das des Frames („Rahmen“, s. Minsky 1975) als Repräsentation schematischen Wissens über Entitäten der Welt (Stühle, Kindergeburtstage etc.) (vgl. ebd.: 417).

- **WORÜBER**

136) Und so bleibt es dabei: Scholz erinnert sich an nichts. Er hat **keine Ahnung, worüber** er seinerzeit wohl mit Christian Olearius, dem damaligen Miteigentümer der Warburg-Bank, bei drei Treffen gesprochen hat. Was er aber weiß: Mit der Entscheidung, der Bank die Rückzahlung von unrechtmäßig erstatteten Steuern in Höhe von 47 Millionen Euro zu erlassen, damit hat er nichts zu tun. »Da war nichts.« <https://www.spiegel.de/politik/deutschland/news-ukraine-krieg-gas-krise-ampel-olaf-scholz-cum-ex-donald-trump-joerg-meuthen-a-186f7cb5-9200-46c9-a5ac-bcfd3b281d96> (19.08.2022).

- **WORUNTER**

137) Eine breite oder wohl-definierte syntaktische Abdeckung. Erfahrungen zeigen, dass Benutzer generell Probleme mit intransparenten Beschränkungen der sprachlichen Eingabe haben, dass sie sich andererseits aber schnell an klar definierte Beschränkungen gewöhnen. **Sinnvoll kann daher die Verwendung sog. kontrollierter Sprache sein,** worunter man die

Verwendung eines Ausschnitts des prinzipiell möglichen Sprachinventars (Syntax, Lexik...) in einem NLS versteht (s. den Exkurs auf der nächsten Seite) (vgl. Kai-Uwe Carstensen, 2017: 85).

- WOVON

138) Die Schallschwingungen, die durch die artikulatorischen Quellen und Filter erzeugt wurden, werden durch die Luft und andere Medien als komplexe Druckwellen übertragen, die mit handelsüblichen Aufnahmegeräten aufgenommen werden. Das Sprachsignal wird mit den Methoden der akustischen Phonetik auf unterschiedliche Weise analysiert und dargestellt, wovon Folgende die drei wichtigsten sind: [...] (vgl. Kai-Uwe Carstensen, 2010: 176).

3.3.3.3. Prototypischer/ vorübergehender Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern

Im Anschluss an die gerade eben abgeschlossene Analyse wollen wir nun, auf Basis einiger während der Analyse erhobenen Details & Merkmale, die folgenden **Heuristiken** zur automatischen Auflösung von **da-hier- und wo- Pronominaladverbienanaphern** formulieren. Im Allgemeinen konnte festgestellt werden, dass Pronominaladverbienanaphern aller Arten meistens **Propositionen als Antezedenzien** haben. Jedoch werden diese Propositionen, je nach der Art der entsprechenden Anapher, jeweils entweder **durch einen ganzen Satz** (Dies ist der Fall bei den Da- und Hier-Pronominaladverbienanaphern) oder eher **durch einen Satzteil** (dies ist der Fall bei den Wo-Pronominaladverbienanaphern). Weiterhin wurde auch festgestellt, dass **Substantive und NPs auch als Antezedenzien** fungieren können, aber solche Fälle kommen nur selten vor. Beschränkt man sich aber nun darauf, nur die häufig auftretenden Fälle auszuwerten, dann lassen sich die beiden folgenden Aussagen formulieren: a) Was das Antezeden von Da- und Hier-Pronominaladverbienanaphern angeht, handelt es sich meistens **um den unmittelbaren Vorgängersatz**; b) Was das Antezeden von Wo-Pronominaladverbienanaphern betrifft, handelt es sich auch in fast allen Fällen **um den Hauptsatz des Satzes, in dem die Wo-Pronominaladverbienanapher als Relativpronomen auftritt**.

So gesehen, sind wir schon in der Lage, **einen automatischen Auflösungsprozess** für Pronominaladverbienanaphern aufzubauen. Somit gelangen wir letztendlich am Ende dieses umfassenden Kapitels, dessen Hauptziel darin bestand, maximale Kenntnisse über die Wortart

PRONOMINALADVERB zu sammeln und anhand der gesammelten Kenntnisse, Wege zum Aufbau des angestrebten Algorithmus bzw. AR-System zu schaffen. Aus den geführten Überlegungen ergibt sich, dass Pronominaladverbien nicht nur anaphorisch in einer sprachlichen Kette (Satz/Text) auftauchen können, sondern auch deiktisch und kataphorisch gebraucht werden können. Nach diesem Befund musste man, in der Tat, nochmal darüber überlegen, wie sich diese nichtrelevanten Erscheinungen von PAn (Deixis & Katapher) von der für uns relevanten Erscheinung (Anapher) automatisch differenzieren lassen. In dieser Hinsicht wurde untersucht, welche syntaktischen Merkmale ein PA aufweisen muss, um als Katapher (Siehe hierzu **Abschnitt 3.3.3.1**) oder als Deixis¹²⁰ aufgezeichnet zu werden. Und anhand der entworfenen Verfahren und Ansätze wird unser Algorithmus in der Lage sein, **anaphorische PAn von anderen PAn auszusortieren**.

Sobald dieser Sortierungsschritt erledigt war, blieb es uns nun weitere Ansätze zu formulieren, wonach eine automatische Auflösung von Pronominaladverbienanaphern möglich wäre. Dies wurde letztendlich auch geschafft, und zwar indem man auf Basis einiger, während der manuellen Auflösung, erhobenen Regularitäten fungierende syntaktische Regel zur Auflösung von Pronominaladverbienanaphern etabliert hat. So gesehen, kann zusammengefasst, unser vorübergehender *Algorithmus*¹²¹ wie folgt aussehen:

¹²⁰ Wie vorher erwähnt, wird ein PA als Deixis aufgezeichnet, wenn es Vorfeld im Satz auftritt, welches seinerseits auch als erster Satz des entsprechenden Beispieltexes fungiert. Denn es besteht, in der Tat, bei solchen Erscheinungen von PAn keine Möglichkeiten **Rückwärtsbezüge** zu schließen. Dies hingegen sind **außersprachliche Bezüge** und **Vorwärtsbezüge** möglich.

¹²¹ Im Moment handelt es sich nur um einen Entwurf, der nur sprachwissenschaftliche Aspekte enthält. Doch am Ende des zweiten Arbeitsteils, betitelt **Algorithmen und AR-Systeme**, werden wir in der Lage sein, einen vollständigen Algorithmus auf Beine zu stellen, welcher dann auch die methodischen Prinzipien des Forschungsgebiets COMPUTERLINGUISTIK erachten wird. Wir werden, in der Tat, im Laufe dieses zweiten Teils, maximale Kenntnisse über den Aufbau von bereits existierenden, und somit anerkannten, Algorithmen sammeln. Und in Anlehnung an die dabei verwendeten verschiedenen Prinzipien und Methoden werden wir eine, für die Erreichung unserer Zielsetzung, **angemessene Vorgehensweise strukturieren können**.

<Taucht ein PA während der Aufarbeitung eines Textes auf, dann:

<Prüfen Sie, ob es sich um eine Da- oder Hier-Verbindung handelt.

<Ist dies der Fall dann Prüfen Sie, ob es sich entweder um einen kataphorischen oder deiktischen¹ Gebrauch handelt.

<Ist dies der Fall, dann weglassen und auf ein anderes PA überspringen.>

<Ist dies aber nicht der Fall, dann handelt es sich entweder um eine Da- oder Hier- Pronominaladverbienanapher, die den unmittelbaren vorangehenden Satz als Antezedens hat.>>>

<Prüfen Sie, ob es sich um eine Wo-Verbindung handelt.

<Ist dies der Fall dann Prüfen Sie, ob es sich um einen deiktischen¹ Gebrauch handelt.

<Ist dies der Fall, dann weglassen und auf ein anderes PA überspringen.>

<Ist dies aber nicht der Fall, dann handelt es sich um eine Wo-Pronominaladverbienanapher, die den unmittelbaren Hauptsatz (Teil des komplexen Satzes, in dem die Anapher eingebettet ist), als Antezedens hat.>>>>

Abbildung 6: Vorübergehender Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern

KAPITEL 4: ANSÄTZE UND THEORIEN ZUR AUTOMATISCHEN ANAPHERNAUFLÖSUNG: Algorithmen, AR-Ansätze und AR-Systeme

Das vorliegende Kapitel setzt sich als Ziel, dem Leser einen breiten Überblick über **die bestehenden computerlinguistischen Ansätze zur automatischen Resolution von Anaphern** bereitzustellen. Dafür wird sowohl historisch als auch thematisch vorgegangen, wobei jedes Mal **die Architektur und Stammsäule jedes einzelnen Systems** hervorgehoben wird. In dieser Hinsicht werden Gemeinsamkeiten und Unterschiede hinsichtlich **der Sprache, der Strukturierung und der Performanz der berücksichtigten Algorithmen; AR-Ansätze und AR-Systeme** verarbeitet. Ziel dieser kontrastiven Verarbeitung ist es, maximale **Spezifikationskenntnisse** zu sammeln, welche beim Entwurf eines passenden Ansatzes zur automatischen Resolution von **Pronominaladverbienanaphern** von besonderer Relevanz sein werden. Zu merken ist, dass der Algorithmus, den wir zuvor in Kapitel 3 entworfen haben, sich zurzeit nur von Menschen gebrauchen lässt. Nun müssen **Wege erforscht** werden bzw. **Mittel eingesetzt** werden, damit Computer auch von diesem Algorithmus Gebrauch machen können. Die maximale Zielsetzung dieses Abschnitts ist es also unser eigenes AR-System auf Beine zu stellen.

4.1. Historische und thematische Darstellung von bestehenden AR-Ansätzen; - Algorithmen und -Systemen (Von 1960 bis 1990)

4.1.1. Das System STUDENT von Bobrow (1964); SHERDLU von Winograd (1972); LUNAR von Woods & Co (1977) und der Algorithmus von Hobbs (1977)

Die maschinelle Verarbeitung der Sprache fing wirklich, wie bereits zuvor im Kapitel 1 ausführlich dargestellt, mit der Vularisierung der Computer (etwa gegen die 80er Jahren) an. (Vgl. A. V. Abyzov & A. G. Rys 2017: 148). Doch bezüglich der automatischen Resolution von Anaphern lassen sich die ersten Systeme schon in den 60er und Anfangs 70er aufspüren. Dies ist Beispielsweise der Fall von dem System STUDENT von Bobrow (1964) und das

Dialogsystem SHERDLU von Winograd (1972), welche beide auf Techniken der *Heuristiken* und *des Mustervergleichs* basieren¹²².

Das System STUDENT von Bobrow (1964) ist ein System, das dafür konzipiert worden ist, **mathematische bzw. algebraische Probleme aufzulösen** (G. Hirst 1981: 33). Doch, da der Input eines beliebigen Problems ins System durch natürliche Sprache realisiert wird, benötigt das System den eingeschriebenen Inputtext zuvor zu verstehen, bevor nachher Ergebnisse zu dem gestellten Problem widergegeben bzw. abgerechnet werden. (Vgl. Ebd.: 33) In dieser Hinsicht muss das System so gerüstet sein, dass es in der Lage sein wird, anaphorische Ausdrücke im Inputtext zu identifizieren und zugleich auch sie aufzulösen¹²³. Was das System STUDENT angeht, steht es fest, dass nur **einige (einfache) Heuristiken** zur Anaphernaufflösung zum Einsatz kommen. Eine dieser Heuristik ist beispielweise wie folgendes formuliert:

- Beim Auftauchen eines Demonstrativpronomens am Satzanfang eines zweiten (S+1) oder dritten Satzes (S+2) eines Inputtextes; dann den ganzen vorherigen Satz als Antezedent für das (anaphorische) Demonstrativpronomen . (Vgl. Ebd.: 34)

Dazu äussert sich Graeme Hirst (1981) mit folgenden Wörtern

*[...] STUDENT does not actually resolve the pronouns at all. Phrases containing **this** are usually taken to refer to the consequence of the immediately **preceding item** without looking at the rest of the phrase. Thus, in (3-2):*

*(3-2) An umber is multiplied by 6. **This** product is increased by 44.*

The word product could be changed to result or sasquatch without changing the assumed referent of this. (Vgl. G. Hirst 1981: 34)

¹²² Während **Heuristiken** auf Lösungsverfahren verweisen, *die auf Hypothesen, Analogien oder Erfahrungen aufgebaut sind*. (Balzert 1999:560). Verweist **die Technik des Mustervergleichs (Pattern Matching)** eher auf ein Lösungsverfahren, in dem *Zeichenketten auf ihre Ähnlichkeit miteinander verglichen werden* (Renata Sevcikova 2005: 7). Ziel dieses Vergleichs ist es nämlich gleichaufgebaute und bei ggf. gleichbedeutende Zeichenkette herauszufinden.

¹²³ Genauer betrachtet, ist das System **STUDENT** ein Mechanismus zur Identifizierung von Korreferenzen (Aufbau von Ketten koreferenter sprachlicher Ausdrücke) und zur Auflösung von **koreferenten Anaphern** während der automatischen Sprachverarbeitung. In der Tat wurde dieses System im Rahmen eines umfangreichen Programms zum Sprachverstehen implementiert (Vgl. Bobrow 1964: 20). Somit steht es fest, dass das System STUDENT sowie alle andere Systeme zur Korreferenz- und oder Anapherauflösung meistens als Module/Teilkomponente von größeren Einheiten/ Programmen implementiert sind.

Wie bereits schon eingeführt, wurde auch **die Technik des Mustervergleichs** eingesetzt um anaphorische Konstruktionen in Inputtexte aufzulösen. Ein Problem könnte beispielsweise mit dem folgenden Text ins System STUDENT eingegeben werden.

(3-1) *The number of soldiers **the Russians** have is half the number of guns **they** have. The number of guns is 7000. What is the number of soldiers **they** have?* (Vgl. Ebd.: 34)

Hier geht es beispielweise um die Auflösung des anaphorischen Personalpronomens „**They**“. Dafür wird die Maschine, also das System STUDENT, so G. Hirst (1981), zunächst Sätze Wort zu Wort vergleichen. Und erst beim Feststellen, dass einige Sätze sich nur bei einem einzelnen Wort unterscheiden, zu der Schlussfolgerung kommen, dass die beiden verschiedenen Wörter zur gleichen anaphorischen Kette gehören. Dies wurde zum Beispiel mit dem ersten und dritten Satz des obig stehenden Zitats durchgeführt. Wie die folgende Tabelle illustriert, steht es nämlich fest das das NP „The Russians“ durch das Personalpronomen „They“ ersetzt wurde.

Satz 1	<i>The</i>	<i>number</i>	<i>of</i>	<i>soldiers</i>	<i>the</i>	<i>Russians</i>	<i>have</i>
Satz 3	<i>the</i>	<i>number</i>	<i>of</i>	<i>soldiers</i>	<i>they</i>		<i>have?</i>

Tabelle 7: Die Technik des Mustervergleichs im AR-System STUDENT von Bobrow (1964)

Im Gegensatz zum System STUDENT, das nur mit Hilfe von zusätzlichen Informationen aus KI-Datenbank brauchbar ist, (deswegen ist es auch nur zugänglich im Rahmen eines größeren Programms zum automatischen Sprachverstehen) läuft das System SHRDLU von Winograd (1972), *auf Basis komplexeren Heuristiken* und kann daher unabhängig von weiteren umfangreichen Sprachverarbeitungssystemen hochgefahren werden (Ludger Paschen 2013: 58). Gemerkt wurde nämlich, dass das System SHRDLU (1972) beim Auflösen von Anaphern schon den möglichen Antezedentskandidaten (NPs) Werte zuschrieb. Und der Kandidat mit der höheren Punktzahl wird dann als Antezedent für die bearbeitete Anapher automatisch ausgewählt (Vgl. Ebd.:58). Jedoch lässt sich merken, dass der Aufbau von Punktkategorie komplexeren Heuristik unterlegt ist. Und zwar vor allem, weil die angewiesenen/angegebenen Punktzahlen zusammen mit syntaktischen Positionen und somit syntaktischen Funktionen von Satzelementen korrelieren¹²⁴. In dieser Hinsicht wird beispielweise, und dies je nach

¹²⁴ Beim Betrachten des Systems von Lappin &Co. (1994) werden wir diesen Punktmechanismus ausführlich darstellen.

Sprache, das Subjekt des unmittelbaren vorangehenden Satzes mit der höchsten Punktzahl auskommen, während die NP, die als Präpositionalobjekt mit der niedrigsten Punktzahl aufgezeichnet wird¹²⁵.

Somit steht es fest, dass die Komplexität des Systems SHERDLU (1972) der des Systems STUDENT (1964) überlegen ist. Und zwar vor allem, weil das erste es fordert, wie schon zuvor eingeführt, die Formulierung von mehreren Heuristiken, welche die Aufgabe der Punktverteilung zwischen den möglichen Antezedentskandidaten regieren (Vgl. G. Hirst 1981: 34). Daher wird in der Literatur hervorgehoben, dass SHERDLU (1972) auf **komplexeren Heuristiken** basiert. Allerdings werden auch einige einfache Heuristik innerhalb des Algorithmus SHERDLU (1972) beobachtet. Eine davon wird beispielweise wie Folgendes formuliert:

*Gross heuristics cover some simpler cases: If **it** or **they** occurs twice in the same sentence, or in two adjacent sentences, the occurrences are assumed to be coreferential.* (Vgl. G. Hirst 1981: 34)

Ebenso wie das System **SHRDLU** (1972), basiert das System **LUNAR** von Woods & Co (1977) auf komplexe Heuristiken. Das Besondere an die Realisierung von solchen Heuristiken liegt in der Tatsache, dass sie Zwischenmodule zur syntaktischen Analyse des Satzaufbaus erfordern. In dieser Hinsicht, findet man im System LUNAR (Woods 1977) außer einem **ATN Parser** (Woods 1970) zur syntaktischen Analyse, auch und einen auf Basis der prozeduralen Semantik aufgebauten **Semanticinterpretator** (Woods 1968). Noch unter dem Akronym **LSNLIS (The Lunar Sciences Natural Language Information System)** bekannt, erfolgt die Auflösung von Anaphern in LUNAR zunächst, indem die Anapher und das entsprechende Antezedens durch den Parser identifiziert werden, worauf schließend die völlige Interpretation der Anapher erst durch den *Semanticinterpretator* verraten wird. Allerdings werden zwischen dem System STUDENT (Bobrow 1964) und LUNAR (Woods & Co 1977) einige Ähnlichkeiten bezüglich des Gebrauchs der Methode **Patternmatching** zur Auflösung von Anaphern hervorgehoben. Jedoch basiert sich das System LUNAR bei Patternmatching mehr auf die tiefere Struktur der Sprache, als auf die lexikalische Eben, wie es bei STUDENT der Fall ist. Was den bearbeitenden Anapherntyp angeht, werden insbesondere intersentiale pronominale Anaphern berücksichtigt.

¹²⁵ Zu den Wertanzahlen ins Deutsche kommen wir noch in ganzer Breite zurück, vor allem, wenn wir über die Ansätze von Lappin & Less (1994) reden werden.

Ebenso wie Woods & Co (1977) entwirft Jerry Hoobs (1977) auch einen Algorithmus zur automatischen Auflösung von pronominalen Anaphern, welcher eben auch auf die Technik der Analyse der Satzaufbaustruktur (*Parse tree of Sentence*) berührt. Der Algorithmus basiert somit auch auf verschiedene syntaktische Einschränkungen (Heuristiken) und insbesondere auf diejenigen, die eine Pronominalisierung im Satz voraussetzt (Hoobs 1977: 2-7). Der Algorithmus wurde auf Texte aus einem archäologischen Buch; auf eine Novelle und auf eine Wochenzeitung getestet (Vgl. Ebd.: 6). Insgesamt wurden 100 anaphorische Konstruktionen manuell annotiert (Goldstandard¹²⁶). Und danach wurde der Algorithmus bei jeder einzelnen Anapher eingesetzt. Es ergibt sich dann, dass der Algorithmus eine Rate von 88% beim Auffinden von richtigen Antezedenten aufweist. Zu merken ist jedoch, dass dieser Durchschnitt sinkt, wenn der Algorithmus entweder auf Anaphern zutrifft, die mehr als einen möglichen Antezedenten haben (82%) oder wenn die syntaktischen Einschränkungen, worauf der Algorithmus basiert ist, vereinfacht wird (92%). Zum genaueren Funktionieren des Algorithmus von Hobbs (1978) drückt sich Ludger Paschen (2013) mit folgenden Wörtern aus:

Der filterbasierte Ansatz für die Pronomenauflösung im Englischen von Hobbs (1978) enthält einen „naive[n]“ (ibid.: 339) Algorithmus, der die syntaktische Struktur eines zuvor geparsen Satzes nach möglichen Antezedenten durchsucht. Dazu werden vom Pronomen ausgehend bei allen NP- und SKnoten auf dem Weg zum initialen S-Knoten die NPen links des Weges abhängig von ihrer Einbettung auf Kongruenz und selektionale Beschränkungen hin abgeglichen. Komplementiert wird der syntaktische Algorithmus durch einen semantischen Ansatz, der hauptsächlich auf lexikalische Inferenzen und Kausalketten zurückgreift. (Ludger Paschen 2013: 59)

Ebenso wie der Algorithmus von Hoobs (1978) zielt das System von Taylor & Rosenberg (1975) darauf ab, pronominale Anaphern automatisch aufzulösen. Doch das Besondere an diesem System liegt darin, dass die Aufgabe durch grammatische Einschränkungen gemeistert¹²⁷ wird (Vgl. Fillmore 1977).

¹²⁶ Menschliche Auflösung von Anaphern (Assoziation von Antezedenzen und Anaphern), womit die automatische Auflösung geprüft wird)

¹²⁷ Es wird, in der Tat, nur anaphorische Pronomen berücksichtigt, welche zusammen, in der sprachlichen Kette, mit besonderen Verben stehen. Somit könnte es also dazu kommen, dass einige anaphorische Pronomen nicht als solche anerkannt werden, geschweige denn aufgelöst werden. Eben, weil sie nicht von im Programm verzeichneten Verben zusammenstehen. (Vgl. G. Hirst 1981: 38)

4.1.2. Der BFP Algorithmus von Brennan, Friedman & Pollard (1987): Eine Implementation der Centeringtheorie

Ein weiterer Ansatz, welcher eben auch, wie diejenigen, die oben zuvor erläutert wurden, auf *komplexe Heuristiken* basiert, ist der Ansatz von Brennan & Co. (1987). Dieser Ansatz beabsichtigt, mit Hilfe der Centerintheorie (CT)¹²⁸, einen Algorithmus zur Auflösung anaphorischer Pronomen zu entwickeln. Wie schon in meinen früheren Arbeiten (Sofac 2021 & 2022a) ausführlich erläutert, ist die Centerintheorie (CT) eine Theorie zur Formalisierung der Auflösung von anaphorischen Pronomen (Sofac 2021a: 44-49). Brennan & Co (1987) haben sich diese linguistische Theorie bedient, weil diese, in sich, bereits schon, einige automatisierbare Fertigkeiten aufweist. Grob gesehen, sind, innerhalb der CT, Mechanismen entwickelt worden, um das bestmögliche Antezedens einer Anapher identifizieren zu können (Vgl. Ludger Paschen 2013: 59-60). Diese Mechanismen wurden eben auch von den dreien Wissenschaftlern, Brennan; Friedman & Pollard (1987), als Heuristiken umgewandelt, welche, ihrerseits, dann später die Automatisierung des Verfahrens ermöglicht haben. Im Folgenden stellen wir zunächst dar, wie die CT funktioniert und anschließend darauf, erläutern wir den BFP Algorithmus von Brennan & Co. (1987).

In der Tat, werden, in der CT, die bestehenden Verhältnisse zwischen Bezugsausdrücken (Antezedenzen) und anaphorischen Ausdrücken (Anaphern) durch mehrere Prinzipien aufgefasst. In der Tat, geht die CT davon aus, dass die Bestimmung des Bezugs einer anaphorischen Einheit von mehreren Parametern abhängt (Vgl. M.Krifka WS 2007: 9). Als **Hauptparameter** gilt die **syntaktische Position** des potentiellen Bezugsworts im Satz. Man redet auch, in diesem Zusammenhang, von der **Oblikheits-Hierarchie**: *Subjekt > direktes Objekt > indirektes Objekt > Adjunkte*¹²⁹. Es geht nämlich bei diesem ersten Parameter um die

¹²⁸ **Die Centering-Theorie (CT)** ist eine Theorie, die von Barbara Grosz, Candy Sidner und Aravind Joshi (1986) entwickelt worden ist. Sie ist im Laufe der Zeit von Autoren wie Walker et al. (1998) und Poesio e.al. (2004) weiteraktualisiert worden (Meyer 2007: 85). Sie lässt sich im Allgemeinen als *eine Theorie, die sich auf die Analyse der Kohärenz und Salienzstruktur in Texten fokussiert* (T. Jung 2007: 5). Noch spitzer aufgefasst, lässt sich die CT als *eine Theorie zur Bestimmung des Bezugs von anaphorischen Elementen* verstehen (M.Krifka WS 2007: 8). So betrachtet, beschränkt sich die CT nämlich auf die Beschreibung der Verhältnisse, die zwischen Bezugsausdrücken und Wiederaufnahmeausdrücken (Anaphern) bestehen (Meyer 2007: 85). Diese Auffassung steht auch im Einklang mit der Urauffassung von Grosz, Joshi und Weinstein (1995), die den Begriff „**Center einer Aussage**“ benutzen, um auf die Einheiten zu referieren, welche dazu dienen, diese Aussage mit anderen Aussagen im Diskurssegment, welches sie enthält, zu verknüpfen. (vgl. Jung 2007:5)

¹²⁹ Zu merken ist jedoch die Tatsache, dass diese Ranking-Hierarchie im Rahmen von wissenschaftlichen Auseinandersetzungen, die die englische Sprache als Material hatten, empirisch gewonnen worden war. Doch für den Gebrauch dieses Begriffs auf das Deutsche muss Einiges modifiziert werden, nämlich, dass der Konstituent, der satzinitial ist, also im Vorfeld steht, hochsalient ist, als die weiteren folgenden Konstituenten des Satzes. (Vgl. M.Krifka WS 2007: 9)

Bestimmung der **Zugänglichkeit oder Salienz** eines Bezugsausdrucks. Ein sprachlicher Ausdruck bzw. eine sprachliche Realisierung **eines Diskursreferenten**¹³⁰ des vorangehenden Textes wird als zugänglicher oder mehr Salienz als einen anderen, wenn er eine höhere Position in der Obliks-Hierarchie besitzt (Strube & Hahn 1999: 5). Die CT bezeichnet alle Diskursreferenten des Vortextes also alle potentiellen Bezugsausdrücke mit dem Begriff „**Forward-Looking-Center**“(CF) und zwischen allen diesen potentiellen Bezugsausdrücken wird dann derjenige, der Hochsalienz also zugänglicher als die andere ist als „**Preferred Center**“(CP) genannt. (Vgl. M.Krifka WS 2007: 9)

Als weiterer wichtiger Parameter, den die CT theorisiert, gilt **die Eigenschaft der Anapher**. Die CT berücksichtigt hauptsächlich zwei Merkmale des wiederaufnehmenden/anaphorischen Ausdrucks (Anapher). Einerseits **die grammatikalische Kategorie (Natur)** und andererseits **die syntaktische Position (Ranking)**. Was die Bedingung Natur der Anapher angeht, geht die CT davon aus, dass der Gebrauch von einfachen sprachlichen Ausdrücken (wie etwa **Personalpronomen**) sich am meistens/ bestens für die Wiederaufnahme von Bezugsausdrücken, die Hochsalienz sind (**Erstrangiger Antezedent**). Dies Hingegen, eignet sich der Gebrauch von komplexen sprachlichen Ausdrücken (wie etwa **Artikelpronomen** oder noch **NP**) eher für die Wiederaufnahme von Bezugsausdrücken, die weniger Salienz sind (**Zweitrangiger Adezedent**) (Meyer 2007: 85). Was die Bedingung Ranking/Position des anaphorischen Ausdrucks angeht, wird diese Eigenschaft insbesondere gebraucht, falls der anaphorische Ausdruck für einen Ausdruck des weiteren Textes als Bezugsausdruck fungiert (Die Anapher wird zum Antezedent einer neuen anaphorischen Relation). Dabei wird sie also von Relevanz sein, nämlich insofern als sie dabei helfen wird, diese neue anaphorische Kette aufzulösen.

Wie oben schon angedeutet, will die CT die Zugänglichkeit von DRen im Diskurs anhand des Prinzips der **Obliks-Hierarchie** erfassen (Vgl. M.Krifka WS 2007: 9). Dafür wird, in der CT, als Grundannahme die Tatsache genommen, dass ein Text aus einer Sequenz von Sätzen ($S_1+S_2+S_i+S_{i+1}+...S_n$) besteht. Weiterhin geht die CT davon aus, dass jeder beliebige Satz S_i eines Textes aus einer begrenzten Zahl von DRen besteht. Diese Liste von DRen wird, wie oben schon hingedeutet, **als vorwärtsbezogenes Zentrum (Forward-Looking Center)** genannt

¹³⁰Unter dem Begriff **Diskursreferent (DR)** versteht man konzeptuelle Einheiten, die Personen, Dinge aus der konkreten Welt abbilden. DRen werden beim Sprachgebrauch (schriftlich/mündlich) durch sprachliche Einheiten, wie etwa NP oder Pronomina versprachlicht bzw. realisiert. (Vgl. ebd: 7) In manchen Literaturen über die CT wird auch den Terminus **Diskurseinheit** eingesetzt, um auf die gleiche Realität zu referieren.

und wird wie folgendes notiert **CF (S_i, D)**. Wobei das hochrangierte Element dieses Zentrums/dieser Liste wird **als präferiertes Zentrum** genannt und wird wie folgendes notiert **CP (S_i, D)** (Vgl. Ebd.: 10). Zu merken ist jedoch, dass die nicht initialen Sätze sowohl von ihren DRen (also das eigene CF (S_i, D)) als auch von denen des unmittelbar vorangehenden Satzes (also von CF (S_{i-1}, D)) bestimmt/ charakterisiert sind. Bei solchen Fällen wird in der Liste von DRen des vorangehenden Satzes CF (S_{i-1}, D) nur das präferierte Zentrum CP (S_{i-1}, D) berücksichtigt, denn dieses Zentrum wird zugleich **als rückwärtsbezogenes Zentrum (Bakward-Looking Center) von S_i** genannt und wird wie folgt notiert CB (S_i, ,D) (Vgl. Strube & Hahn 1999: 6) Nehmen wir zur Illustration den Beispieltex (5) aus dem zweiten Kapitel.

(5) Eine Frau₁ kommt in die Botschaft. **Sie₁** befindet sich im Warteraum.

Würde man mit klassischen Prinzipien der Anaphernresolution¹³¹, wie etwa die morphologischen Kongruenzen, vorgehen, dann würde man merken, dass **das grammatische Genus** bei der Auflösung dieser pronominalen Anapher nicht hilft. Trotzdem könnte man ganz spontan behaupten, „Sie“ würde sich, mit höher Wahrscheinlichkeit, auf „Frau“ und nicht auf „Botschaft“ beziehen. Doch genauso bei solchen Fällen wird die CT von Nutzen sein, denn sie wird mit dem Begriff der Zugänglichkeit oder Salienz die Auflösung solcher Ambiguitäten ermöglichen. Konkret betrachtet, sieht eine Anaphernresolution mit Hilfe der CT wie Folgendes aus:

	CB	CF
Eine Frau₁ kommt in die Botschaft.		DR Frau; DR Botschaft
Sie₁ befindet sich im Warteraum.	DR Frau	DR Frau; DR Warteraum

¹³¹Als weiteres Grundprinzip, das bei der Auflösung bzw. Findung des geeigneten Bezugsausdrucks eines anaphorischen Ausdrucks hilft, zählt man auch **die Proximitäts-Leseart von Demonstrativa**. (Vgl. Finkbeiner 2018: 193). Noch unter der Bezeichnung **Obviation** (Vgl. M.Krifka WS 2007: 9) oder noch **Die Nähe/Ferne Regel** (Vgl. Kubczak 2016: 231) bekannt, besagt dieses Prinzip, dass der potentielle Antezedent-ausdruck, der dem anaphorischen Demonstrativpronomen in der linearen Kette am nächsten steht, wird bei Ambiguitätsfall präferiert, als derjenige, der von der anaphorischen Wiederaufnahme (das Demonstrativpronomen) entfernter steht.

Tabelle 8: Beispiel einer Anapherresolution durch die CT ohne Rücksicht auf (CB_{Si} & CB_{Si+1})- Übergänge

Anhand der Abarbeitung der Informationen, die sich in dieser Tabelle befinden, lässt sich nachvollziehen, dass die konzeptuelle Entität FRAU durch zwei unterschiedliche linguistische Einheiten sprachlich realisiert ist. Somit könnte man behaupten, der anaphorische Ausdruck „sie“ hat die NP „Eine Frau“ als Bezugsausdruck. Aber bevor man zu dieser Schlussfolgerung abrupt gelangt, gibt es zu merken, dass zwei DRen in dem ersten Satz sprachlich realisiert sind (DR_{Frau} und DR_{Botschaft}). Zwischen diesen beiden ist ein Salienter als der andere. Und wie es in der Tabelle festzustellen ist, wurde dieser saliente DR als CB des nächsten Satzes ausgewählt. Und weil es in der CT üblich ist, dass die sprachliche Einheit, wodurch der CB eines Satzes realisiert ist, immer als Bezugsausdruck für ein sprachliches Element im CF (meistens dieses, das den hochsalienten DRen sprachlich realisiert hat) des gleichen Satzes fungiert. So könnte man, im Fall des Texts (5), die obig schon erwähnte Schlussfolgerung bekräftigen¹³².

Jedoch, ist es den CT-Theoretikern, aufgefallen, dass es unterschiedliche Arten von Übergängen geben könnte. Diese werden in der unteren stehenden Tabelle in Anlehnung an T. Jung (2007) vorgestellt; und gleich nachher anhand eines Beispieltexes aus dem zweiten Kapitel illustriert.

	CB(S _i) = CB(S _{i-1}) oder CB(S _i) = nicht vorhanden	CB(S _i) ≠ CB(S _{i-1})
CB(S _i) = CP(S _i)	CONTINUE	SMOOTH SHIFT
CB(S _i) ≠ CP(S _i)	RETAIN	ROUGH-SHIFT

Tabelle 9: Häufig auftretende (CB_{Si} & CB_{Si+1})- Übergänge bei anaphorischen Beziehungen

- **CONTINUE** ist der unauffälligste Übergang: Das rückwärtsbezogene Zentrum bleibt gleich und ist das präferierte Zentrum.
- **RETAIN**: Das rückwärtsbezogene Zentrum ist nicht das präferierte Zentrum.

¹³² Genauso dieses Verfahren beabsichtigen Brennan, Friedman & Pollard (1987) durch ihren Algorithmus zu automatisieren. Schwerpunkt dieses BFP Algorithmus (1987) ist es, alle, bis hier, erwähnten Grundprinzipien und Parametern der CT in automatisierbare Heuristiken umzuwandeln.

- **SMOOTH SHIFT:** Das rückwärtsbezogene Zentrum verändert sich.
- **ROUGH SHIFT:** Das rückwärtsbezogene Zentrum verändert sich und ist auch nicht das präferierte Zentrum. (T. Jung 2007: 7).

Betrachtet man nun den folgenden Text mit der CT. Es ergibt sich in der darauffolgenden stehenden Tabelle folgende Beschreibungen.

(7) **Die Priester₁** gehen in den Essraum der katholischen Gemeinde. **Einer₁** ruft den **Kellner₂** an. Doch **dieser₂** kommt nicht.

SÄTZE	CB	CF	ÜBERGÄNGE
Die Priester₁ gehen in den Essraum der katholischen Gemeinde.		DR _{Priester} ; DR _{Essraum} ; DR _{Gemeinde}	
Einer₁ ruft den Kellner₂ an.	DR _{Priester}	DR _{Priester} ; DR _{Kellner}	CONTINUE
Doch dieser₂ kommt nicht.	DR _{Kellner}	DR _{Kellner}	SMOOTH SHIFT

Tabelle 10: Beispiel einer Anaphernresolution durch die CT mit Rücksicht auf (CB_{Si} & CB_{Si+1})- Übergänge

Die Analyse dieser Tabelle anhand des Prinzips der Oblickheits-Hierarchie ist nur bis zum zweiten Satz nachvollziehbar. Denn das Rückwärtsbezogene Zentrum des zweiten Satzes ist, wie das Prinzip der Oblickheits-Hierarchie es will, das präferierte Zentrum des vorwärtsbezogenen Zentrums CF (S₁, D) des ersten Satzes. So gesehen, bezieht sich also das indefinite Pronomen „Einer“ auf die NP „Die Priester“. Doch diese Regel wird bei dem Übergang zum dritten Satz nicht respektiert. Man merkt in der Tat, dass das CB des dritten Satzes nicht das präferierte Zentrum des vorwärtsbezogenen Zentrums des zweiten Satzes CF (S₂, D) ist. Die Beschreibungsregeln für solche Fälle sind in der obig stehenden Tabelle 8 sortiert.

In einem Artikel erschienen im Jahre 1996 und wieder im Rahmen der Jahrestagung für Computerlinguist im 1999 aktualisiert und veröffentlicht, stellten Michael Strube und Hahn den BFP Algorithmus von Brennan; Friedman und Pollard (1987) unter einer funktionalen Perspektive dar. Doch dafür beschrieben sie zunächst im Voraus, den Aufbau und das Funktionieren des **BFP-Algorithmus**. Wie oben schon eingeführt, verweist dieser BFP-Algorithmus auf einen Auflösungsmechanismus für pronominale Anaphern, der sich hauptsächlich in drei Etappen gliedert. **In der ersten Phase sollen alle möglichen CB-CF-Kombinationen generiert werden.** Hierfür werden insbesondere alle möglichen und unmöglichen Verbindungen zwischen Pronomen von CF_{Si+1} ¹³³ mit Elementen von CF_{Si} des vorangehenden Satzes etabliert (Vgl. Strube & Hahn 1999: 9). Nach dieser Phase, kommt **die der Ausfilterung**, wie ihr Name schon verrät, werden, in dieser Phase, alle entworfenen Regeln und Heuristiken eingesetzt um das bestmögliche Antezedent bzw. die wahrhaftige Kombination zwischen Pronomen (Also Elemente von CF_{Si+1}) und Elementen von CF_{Si} aufzufangen. Dafür wurden beispielweise die folgende Heuristik¹³⁴ entworfen: „Wenn ein Diskursreferent aus CF (S_{i+1} , D) in S_i bereits durch ein Pronomen realisiert ist, dann ist auch in CB (S_i , D) durch ein Pronomen realisiert“ (Walker e.a. 1998 zit. nach Ebd.: 13). In dem letzten Schritt dieses Algorithmus werden dann die nach dem zweiten Schritt verbliebenen bzw. restlichen Kombinationen nach dem Prinzip der **Übergangshierarchie** weiteraussortiert. Dabei wird insbesondere der Übergang CONTINUE hochrangiert als die anderen Übergänge. Dies bedeutet vor allem, dass, wenn es am Ende des zweiten Schritts dazu käme, dass zwei, drei oder mehrerer Kombinationsmöglichkeiten zur Auflösung einer pronominalen Anapher stehen, dann wird der Algorithmus das Antezedens auswählen, das zusammen mit dem aufzulösenden anaphorischen Pronomen einen CONTINUE-Übergang¹³⁵ beschreibt. Zu dieser Rankinghierarchie gelten die folgenden Vorschriften **CONTINUE > RETAIN > SMOOTHSHIFT > ROUGH-SHIFT**. (Vgl. Strube & Hahn 1999: 13).

- **Fazit**

¹³³ Hier werden also nur Elemente des Vorwärtszentrums, die durch Pronomen realisiert sind, berücksichtigt.

¹³⁴ *Filter by constraints, e.g., contra-indexing, sortal predicates, centering rules and constraints. This way, possible antecedents are filtered out because of morphosyntactic, binding, and semantic criteria. Also the realization of noun phrases in the current utterance (e.g., realization as a pronoun vs. realization as a definite noun phrase or proper name) comes into play.* (Vgl. Strube & Hahn 1999: 13).

¹³⁵ *Rank by transition orderings. This is the step, where the pragmatic constraints of centering apply. Basically, CONTINUE transitions are preferred, i.e., the antecedent of a pronoun is more likely to turn up as the Cb of the previous utterance than any other element of the Cf. In certain configurations, the algorithm includes a preference for parallelism in linguistic constructions.* (Vgl. Strube & Hahn 1999: 14).

Wie bereits erläutert, beschränken sich Algorithmen und Ansätze der 60er meistens auf Verarbeitung von einzelnen Typen von Anaphern. Sie werden daher als traditionell¹³⁶ aufgezeichnet und zwar vor allem, weil sie nur die Satzebene im Kauf nehmen. Solche Ansätze behandeln somit nur explizit im Satz/ Text genannten Antezedenten. Dies entgegenend fokussieren sich, wie schon erläutert, die Systeme der 70er und 80er Jahren eher auf die Auflösung von vielen Anapherotypen. Sie sind daher theorielästig¹³⁷ und zwar, vor allem, weil sie große Menge von linguistischen und nicht linguistischen Daten während der Auflösung im Kauf nehmen. Somit weisen sie komplexere Grundstruktur als ihre Vorgänger auf. Dies alles führte nämlich dazu, dass solche theorielästigen Ansätze **nicht implementierbar waren bzw. sich nur eingeschränkt implementieren lassen**¹³⁸. Zu den wichtigen Arbeiten dieses Zeitraumes zählen die Ansätzen von Walker (1989), Carter (1986), Rich und LuperFoy (1988) und Carbonell und Brown (1988). (Renata Sevcikova 2005: 7)

Weiterhin führte die **Nichtimplementierbarkeit** von theorielästigen bzw. wissensreichen Ansätze aus den 70er & 80er Jahren dazu bei, dass Wege zur Implementierung von Ansätzen erforscht werden. In dieser Hinsicht tauchen ab den 90er Jahren neue Ansätze, die dank ihrer wissenarmen Charakter sich leicht implementieren lassen. Gemerkt ist nämlich, dass im Gegensatz zu Ansätze der 80er, welche meistens darauf abzielte sowohl mehrere **Textsorten** als auch mehrere **Anapherotypen** innerhalb eines gegebenen Ansatzes aufzulösen, beschränken sich die Ansätze in den 90er Jahren vorwiegend auf die Auflösung von bestimmten Anaphern in bestimmten Textsorten. Diese Praxiswende wurde nämlich durch Fortschritte im Bereich der Informatik verursacht und unterstützt. Dazu äußert sich **Renata Sevcikova (2005)** mit folgenden Wörtern:

¹³⁶ Dazu äussert sich G. Hirst (1991) mit folgenden Wörtern: “[...]*Their resolution methods tend to work at the sentence level, and may bring to bear **world knowledge and low-level linguistic knowledge**. Antecedents not explicit in the text are not handled. This characterization is of course a generalization; not all approaches classified as traditional fit this description in every detail. On the other hand, modern methods recognize the importance of focus and **discourse-level knowledge for resolution**. Implicit antecedents may also be handled*“.

(Graeme Hirst 1981: 33)

¹³⁷ Syntaktische und semantische Beschränkungen, teilweise aber auch Wissen über das Diskursthema und den lokalen Fokus, bilden das Fundament für diese Ansätze. (Mitkov et al. 2002c:88 in Renata Sevcikova 2005: 8)

¹³⁸ Es handelt sich in der Tat um Ansätze, welche den ganzen Text im Kauf nehmen. Sie müssen dafür sich sowohl um die oberflächliche Struktur des Textes als auch um die tiefere Struktur kümmern. In dieser Hinsicht sind sie auch in der Lage implizite Antezedenten aufzulösen, denn diese lassen sich nur exklusiv durch semantische Einschränkungen/Heuristik/Parametern identifizieren. Aus diesen Gründen lassen sich solche Ansätze auch schwer implementieren also als automatische System umwandeln.

Zudem wurden diese wissensarmen Ansätze auch durch das Aufkommen von verlässlichen computerlinguistischen Werkzeugen wie **POS-Taggern**¹³⁹ und **Parsern**¹⁴⁰ sowie Ressourcen wie **Korpora**¹⁴¹, **Ontologien**¹⁴² und Evaluationsmethoden begünstigt. Ausserdem hat man begonnen, **Methoden des Maschinenlernens**¹⁴³, im Folgenden mit ML abgekürzt, **probabilistische Ansätze**¹⁴⁴ und mehrsprachige Systeme zur Behandlung von Anaphern einzusetzen. (Vgl. Renata Sevcikova 2005: 8f)

Diese Fortschritte in der Automatisierung der Sprachverarbeitung spitzte sich im Mitte der 90er weiter zu. Gemerkt wurde nämlich, dass Aufgaben **wie automatische Textzusammenfassungen oder Informationssuche in Texten** schon realisierbar waren. Doch eine effiziente bzw. richtige automatische Zusammenfassung eines Textes erfolgt nur, wenn die im Text bestehenden **anaphorischen Bezüge** zuvor gut identifiziert bzw. richtig aufgelöst sind. In dieser Hinsicht nahm diese Problematik der automatischen Auflösung von Anaphern innerhalb von jährlichen computerlinguistischen Konferenzen und Tagungen¹⁴⁵ weiter an Bedeutung zu. (Vgl. S. McCarthy/ Lehnert 1995, Steinberger Et Al. 2005). Im Folgenden werden weitere AR-Ansätze näher beschrieben. Dabei wird der Fokus, je nach bearbeiteter Sprache (Englisch, Deutsch, Französisch oder noch Portugiesisch) und Textsorte (Zeitungstexte; Gebrauchstexte oder noch Romane) insbesondere daraufgesetzt, zu beschreiben, wie jedes dieser jeweiligen AR-Systeme zuerst **bei der Identifikation der Anapher; dann beim Finden von möglichen Antezedenten und schließlich bei der Auswahl des für die Auflösung geeigneten Antezedens vorgeht**. Wobei im Anschluss die Performanz jedes bearbeiteten Ansatzes besprochen wird.

¹³⁹Ein **POS-Tagger (Part-of-Speech-Tagger)** ordnet jedem Wort die Wortart zu. (Vgl. Renata Sevcikova 2005: 8f)

¹⁴⁰Mit **Parsern** werden die Syntaxstrukturen eines Satzes ermittelt. Dabei wird die Eingabe anhand der Grammatik des Programms auf Korrektheit geprüft. (Vgl. Ebd.: 8f)

¹⁴¹Ein **Korpus** ist eine umfassende Textsammlung. (Vgl. Ebd.: 8f)

¹⁴²**Ontologien** sind Wissensrepräsentationen. Es handelt sich hierbei um hierarchische Ordnungen von Begriffen. (Vgl. Ebd.: 8f)

¹⁴³Mit **Methoden des Maschinenlernens** wird versucht, die Akquisition von Wissen zu automatisieren. Ein System kann aus Beispielen lernen und Verallgemeinerungen aufstellen, d.h. es werden Gesetzmässigkeiten in den Trainingsdaten erkannt. Nach der Lernphase kann das System auch unbekannte Daten beurteilen. (Vgl. Ebd.: 8f)

¹⁴⁴**Wahrscheinlichkeitsbasierte Ansätze**. (Vgl. Ebd.: 8f)

¹⁴⁵Die regelmäßige Gewinnung an Aufmerksamkeit wurde seit der sechsten und siebten Message Understanding Conference (MUC-6154, MUC-7155) beobachtet <http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html> (04.04.08). http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html (04.04.07). (Vgl. Ebd.: 8f)

4.2.Computerlinguistische Ansätze, Algorithmen & Systeme zur Auflösung von koreferenten Pronominalanaphern (Personal- Possessiv- und Reflexivpronomina)

Die Aufgabe der Automatisierung des Auflösungsverfahrens für pronominale Anaphern lässt sich insbesondere durch **zwei Ansätze bzw. Techniken** realisieren. Zum einen, haben wir Systeme, die vorwiegend **auf linguistische Heuristiken** bauen, um *das Antezedens aus einer Liste von möglichen Kandidaten auszuwählen*. (Holger Wunsch 2010: 105). Und zum anderen, haben wir Systeme, die eher **auf Technik des Maschinenlearning (ML) bzw. Deeplearning setzen**, um den anaphorischen Ausdruck aufzulösen. Diese letzteren benutzen insbesondere statistische und probabilistische Verfahren, um das richtige Antezedens auszuwählen (Vgl. Renata Sevcikova 2005: 32). Zu Einzelheiten darüber kommen wir noch, in einem umfassenderen Überblick, später zurück. Nun beabsichtigen wir zunächst Systeme der ersten Kategorie darzustellen.

4.2.1. Auf Heuristiken basierende AR-Algorithmen und oder AR-Systeme¹⁴⁶ für Pronomen im Deutschen und im Englischen.

4.2.1.1. Das AR-System von Shalom Lappin und Herbert J. Leass (1994)

Der Algorithmus von **Shalom Lappin und Herbert J. Leass (1994)** namens **RAP** (Resolution of Anaphora Procedure) wurde insbesondere zur Automatisierung des Auflösungsverfahrens von anaphorischen Personalpronomen **der dritten Person Singular** entworfen (Lappin & Leass 1994:538). Ursprünglich für das Englische entworfen, wurde der RAP-Algorithmus später der deutschen Sprache angepasst. Zunächst werden wir die englische Variante beschreiben und erst im Anschluss daran die entsprechende deutsche Anpassung erläutern. **Das System RAP** (Lappin & Leass 1994) besteht aus **mehreren Komponenten und 2 Anbindungsalgorithmen** (Der Anbindungsalgorithmus **zur Auflösung von Reflexiv- und Possessivpronomen** und der Anbindungsalgorithmus **zur Identifizierung von pleonastischen**

¹⁴⁶ Ein AR-Algorithmus ist ein Auflösungsverfahren, das **in einer natürlichen Sprache (Deutsch/Englisch...)** aufgefasst ist. Dies hingegen, verweist **ein AR-System** eher auf ein Auflösungsverfahren, das **in einer informatischen Sprache** (Prolog; Java; C++...) geschrieben ist. Bei diesem letzteren Fall redet man von der **Programmierung des Auflösungsverfahrens**. Der Algorithmus wurde also für den Computer verständlich und brauchbar gemacht. Es gibt, in der Tat, manche Autoren, die nur AR-Algorithmen und kein AR-Systeme entworfen haben. Dies ist beispielweise der Fall von **Hobbs (1978)**. Allerdings wird in der Literatur öfters **der Begriff AR-Algorithmus** um auf die beiden **Konzepte/Realitäten** zu referieren.

Pronomen). (Vgl. Holger Wunsch 2010: 105.) Jede dieser Komponenten erfüllt eine bestimmte Aufgabe bei der Sprachverarbeitung bzw. bei den verschiedenen Etappen, die zur Auflösung der Anapher führen.

Die erste eintretende Komponente ist **der syntaktische Parser** (Vgl. Ludger Paschen 2013: 60). Dieser funktioniert auf Basis der **McCord's SLOT-Grammatik** und ist in einem umfassenderen Überblick in Lappin and McCord (1990a) dargestellt worden. Doch im Rahmen dieser Arbeit werden wir nur einen begrenzten und zielgerichteten Überblick über **seine Natur, seine Eigenschaften, sein Funktionieren und seine Ziele** wiedergeben. Wie der Name schon hinausweist, verweist **ein syntaktischer Parser** auf einen automatisierten Prozess zur Zerlegung von Sätzen in kleinen grammatischen Einheiten (**Nomen, Artikel, Adverb, Verb, Pronomen**) (Vgl. Holger Wunsch 2010: 100). Ein syntaktischer Parser ist also ein Algorithmus dessen Ziel darin besteht, alle bestehenden Einheiten eines Satzes als solches zu verzeichnen. Und da diese Aktion (das Parsen/ die Zerlegung) eben nicht von Menschen, sondern eher von Computer durchgeführt wird, soll dieser Zerlegungsprozess in einer Programmiersprache aufgefasst werden. Anders gesagt, man muss **den Zerlegungsprozess bzw. den Zerlegungsalgorithmus mit Hilfe einer informatischen Sprache schreiben bzw. programmieren**¹⁴⁷. Und die Programmiersprache, mit der der von Lappin & Leass (1994) eingesetzte syntaktische Parser programmiert worden ist, heißt **PROLOG**¹⁴⁸.

Die Besonderheiten von Programmiersprachen liegt in der Tatsache, dass sie nur in einem bestimmten Arbeitsfeld (Computerfenster) editiert und genutzt werden können. Dies hat nämlich für Vorteil, dass Informationen **über vorherigen Parser- Ergebnisse** abrufbar sind. Es wird in der Tat Angaben über das Parsen von vorherigen Sätzen **in einem Zwischenlager** gespeichert (Vgl. Lappin & Leass 1994: 535). Außer Informationen über die grammatische Natur und syntaktische Beziehungen (Kopf-Argument/ Kopf-Adjunkt) einer einzelnen sprachlichen Einheit werden auch weitere Informationen für den guten Weiterverlauf des Auflösungsprozesses bereitgestellt. Dazu äußern sich Lappin & Leass (1994) mit folgenden Wörtern aus:

***RAP** fonctionne principalement sur une représentation clausale de l'analyse Slot Grammar de la phrase courante dans un texte (McCord et al. 1992). La représentation clausale consiste en un ensemble de clauses d'unité Prolog qui fournissent des*

¹⁴⁷ Und da es verschiedene Programmiersprache gibt, wird man auch zwischen verschiedenen Parsers-Typen unterscheiden können.

¹⁴⁸ Auch mit dieser Programmiersprache (PROLOG) ist das RAP-System (Lappin & Leass (1994) hergestellt worden.

informations sur les relations **tête-argument et tête-ajout** de la structure de phrase que **la Slot Grammar** attribue à une phrase (phrase). Les représentations clauses des quatre phrases précédentes dans le texte sont conservées dans **l'espace de travail Prolog**. **La représentation du discours** utilisée par notre algorithme est constituée de ces représentations clausales, ainsi que de **clauses unitaires supplémentaires** déclarant les **référents** du discours évoqués par les SN dans le texte et spécifiant **les liens anaphoriques** entre les référents du discours ? Toutes **les informations** relatives à un référent de discours ou à son NP évocateur sont accessibles via **un identifiant (ID)**, un terme Prolog contenant deux entiers. Le premier entier identifie la phrase dans laquelle se produit le SN évoquant, les phrases d'un texte étant numérotées consécutivement. Le deuxième entier indique la position du mot principal du NP dans la phrase. (Vgl. Lappin & Leass 1994 : 535 ; Traduit de Anglais vers Français via www.onlinedoctranslator.com le 28.01.2023 à 16h)

Wie Jean-Paul Bronckart & Ecaterina Bulea Branckart (2011) in einem Artikel zum Thema, **Clause, Phrase, Proposition? des Enjeux d'un débat conceptuel** feststellen, existiert tatsächlich in der Syntax verschiedene Ansätze zur Beschreibung der syntaktischen Einheit „Satz“. Je nachdem welches Merkmal im Fokus steht, werden verschiedene Begriffe eingesetzt um auf den „Satz“ zu referieren. Dies ist Beispielweise der Fall vom Begriff **„Clause (Nomen) oder clausal (Adjektiv)“**. Dieser kommt insbesondere zum Einsatz, wenn es sich darum handelt, den Fokus auf **die Fähigkeit eines Verbs Objektergänzungen zu fordern**; man redet auch von **der „Syntaxe de rection“** (J.P. Bronckart & E.B Bronckart 2011: 209). Genau in diesem Paradigma ist **die Solt-Grammatik (Lappin and McCord 1990a)** geankert worden. Diese Grammatik zielt nämlich darauf ab, die Clause (also Sätze) auf Basis von in Clausen bestehenden **Kopf-Argument & Kopf-Adjunkt** zu analysieren.

Aus dem obig stehenden Zitat ergibt es sich, dass der syntaktische Parser, in sich selbst, ein Algorithmus zur Satzanalyse auf Basis der Solt-Grammatik (Lappin and McCord 1990a) ist. Der Algorithmus ist, wie schon oben erwähnt, mit Hilfe einer Programmiersprache namens **PROLOG** implementiert worden. Davon ausgehend, lässt sich also der von Lappin & Leass (1994) eingesetzte Parser als **ein Analysator (Tool)**¹⁴⁹, der darauf abzielt, Sätze eines Textes **in Prolog-Clause** zu verzeichnen (Vgl. Lappin & Leass 1994 : 535). Diese Prolog-Clause/-

¹⁴⁹ Man redet auch von Natural Language Processing resources (NLP-Ressource). Es gibt in der Tat verschiedene Ressourcen solcher Arten, die schon zum Einsatz bereits sind. Man brauche sie dennoch nur an dem angestrebten Ziel/ AR-System (In unserem Fall) anzupassen.

Einheiten werden für jedes Element eines Satzes gebildet, wobei nur Angaben über die vierletzten verarbeiteten Sätze im Zwischenlager archiviert bleiben. (Vgl. Ebd.: 535). Und wie schon zuvor hingewiesen, werden außer Informationen über Kopf-Argument-Beziehungen und Kopf-Adjunkt- Beziehungen auch Angaben über im Text erwähnten Referenten sowie deren jeweiligen NP¹⁵⁰ zum Abruf bereitgestellt. Nun wird auf Basis der Repräsentation des Textes in dieser **Prolog-Clause/ Prolog-Unit** der RAP-Algorithmus (Lappin & Leass 1994) entwickelt und später auch mit Hilfe der PROLOG-sprache implementiert.

Zusammengefasst, lässt sich also feststellen, dass Lappin & Leass (1994) mit dem Einsatz **des syntaktischen Parsers** (Lappin and McCord (1990a) darauf abzielen, nicht nur alle NPs und Pronomen (P) eines Textes sowie auch deren jeweiligen Referenten zu identifizieren und als solche für den Weiterverlauf der Prozedur, **in Prolog-clausen**, zu verzeichnen, sondern auch alle im Inputtext bestehenden grammatischen (Kopf-Argument und Kopf-Adjunkt) zu verzeichnen. Außerdem wird u.a. im Parser auch **zwischen bestimmten und unbestimmten NPs** einerseits und zwischen sprachlichen Einheiten, die neue Referenten einführen und solchen, die schon erwähnten Referenten wiederaufnehmen, unterschieden. (Vgl. Renata Sevcikova 2005: 33). Und wie in dem obig stehenden Zitat deutlich erklärt, werden all diesen Informationen über NPs und Pronomen dank **Prolog-Einheiten/ Prolog-clausen** jeder Zeit abrufbar sein; zumindest für die 3 Sätzen, die den aktuellen Satz vorangehen (Vgl. Lappin & Leass 1994: 535)

Im Anschluss an der Identifizierung von im Input bestehenden NPs und Pronomen (also nach dem Einsatz des syntaktischen Parsers) wird, in einer nächsten Etappe, erstens eine Prozedur zur Identifikation **von pleonastischen Pronomen eingeschaltet** und zweitens, nach Übereinstimmungen in Kongruenzmerkmalen zwischen dem Reflexivpronomen (Anapher) und den möglichen Antezedenzien (NPs) , **einen Anbindungsalgorithmus** (Lappin et McCord 1990b) **zur Identifizierung und zugleich Auflösung von Reflexivpronomen**¹⁵¹ losgefahren.

¹⁵⁰ *Toutes les informations relatives à un référent de discours ou à son NP évocateur sont accessibles via un identifiant (ID), un terme Prolog contenant deux entiers. Le premier entier identifie la phrase dans laquelle se produit le SN évoquant, les phrases d'un texte étant numérotées consécutivement. Le deuxième entier indique la position du mot principal du NP dans la phrase.* (Vgl. Lappin & Leass 1994 : 535)

¹⁵¹ **Die linguistischen Heuristiken**, die den Anbindungsalgorithmus **zur Auflösung von Reflexivpronomen** zugrunde stehen, sind bei Lappin & Leass (1994) wie folgendes formuliert worden: **A noun phrase N is a possible antecedent for a reflexive or reciprocal A if one of the five rules below can be applied: 1. A is in the argument domain of N, and N fills a higher argument than A. 2. A is in the adjunct domain of N. 3. A is in the NP domain of N. 4. N is an argument of a verb V, there is an NP Q in the argument domain or the adjunct domain of N such that Q has no noun determiner, and (i) A is an argument of Q, or (ii) A is an argument of a preposition PREP and PREP is an adjunct of Q. 5. A is a determiner of a noun Q, and (i) Q is in the argument domain of N and N fills a**

(Vgl. Holger Wunsch 2010: 107). Die im Satz verbliebenen Pronomina werden als potentielle anaphorische Pronomina verzeichnet. Und für jedes dieser Pronomina werden auch alle möglichen Antezedens-kandidaten (NPs) intra- oder intersententialer Natur auf eine Liste eingetragen. Dabei wird jedem Kandidaten (NP) einen **Salienz-Wert/ eine Salienz-Gewichtung** zugeschrieben, welche ihrerseits die Summe von mehreren **Salienz-Faktoren** ist: Man redet auch vom **Gewichtungs-Mechanismus**¹⁵² (Vgl. Ludger Paschen 2013: 60)

Zur Berechnung dieses **Salienz-Werts/ dieser Salienz-Gewichtung** und der daran anschließenden Relevanz von **Äquivalenzklassen** (Elemente einer koreferentiellen Kette/Gruppe von koreferentiellen Ausdrücken) kommen wir gleich zurück. Doch zuvor möchten wir darauf hinweisen, dass Kandidaten, die nach dem Pronomen auftreten, schwer benachteiligt werden. (Rückzug von etwa 100 Punkten). Man redet auch vom **CATAPHORA PENALTY** (Vgl. L.Paschen 2013: 60). Dies hingegen, werden Kandidaten, die (in der sprachlichen Kette) vor dem anaphorischen Pronomen auftreten begünstigt (Zuschreibung von positiven Werten, je nach der syntaktischen Funktion der NP). Davon ausgehend, wird **eine Salienz-Werteschwelle** (etwa der Durchschnitt des Salienz-Werts von allen in der Liste vorhandenen Antezedens-kandidaten) berechnet. Somit werden alle Antezedens-kandidaten, die unter der Schwelle stehen, automatisch vom Auswahlprozess ausgeschlossen (Vgl. Lappin & Leass 1994: 537). Die verbliebenen Kandidaten werden dann zunächst einen Numerus-Genusstest (Einsatz eines auf **Basis Genus-Numerus-prinzipien aufgebauten morphologischen Filters**) unterzogen und danach, falls nötig, auch einen syntaktischen Test durchführen müssen. Ziel des auf Basis **Bindungsprinzipien von Noam-Chomsky (1981)**¹⁵³

higher argument slot than Q, or (ii) Q is in the adjunct domain of N. (Vgl. Lappin & Leass 1994:538-539 zit. nach Holger Wunsch 2010: 107)

¹⁵² Noch als **Scoringssystem** (L. Paschen 2013) bezeichnet, schreibt **der Gewichtungsmechanismus** jeder NP aufgrund von bestimmten **Faktoren** einen Wert zu. Als Gewichtungsfaktoren **gelten Satznähe, Subjekte, existentielle Konstruktionen** (Konstruktionen mit *there is/there are*), **Objekte, Hauptnomen und nicht-adverbiale NPs**. Diese Gewichtungsfaktoren haben **einen bestimmten anfänglichen Wert**. Der Anfangswert für **Satznähe** ist beispielsweise 100, der Anfangswert für **Subjekte** 80. Die Anfangswerte sind in [Lappin et al. 1994:541] aufgeführt **und werden im Verlauf des Prozesses vermindert**. Mit jedem neuen Satz wird der aktuelle Wert für **jeden Faktor halbiert**. Wenn ein Faktor **einen Wert nahe 0** erreicht, wird er entfernt. Auch jede koreferentielle Kette erhält eine Gewichtung. Dieser Gewichtungswert ist die Summe der aktuellen Gewichtung von allen Faktoren, in deren Reichweite mindestens ein Element der koreferentiellen Kette liegt. (Vgl. Renata Sevcikova 2005: 33).

¹⁵³ Wie gerade schon angesagt, haben Lappin et McCord (1990a) **einen syntaktischen Filter** auf Beine gestellt, um das System in die Lage zu versetzen, alle Paare zu finden, die nicht anaphorisch sein können. (Vgl. Holger Wunsch 2010: 105). Dieser syntaktische Filter basiert auf folgende **Bindungsprinzipien von Noam-Chomsky (1981)**:

1) Argument domain: *P is in the argument domain of a phrase N if and only if P and N are both arguments of the same head.* **2) Adjunct domain :** *P is in the adjunct domain of a phrase N if and only if N is an argument of a head H, P is the object of a preposition PREP and PREP is an adjunct of H.* **3) NP domain:** *P is in the NP domain of N if and only if N is the determiner of a noun Q and P is an argument of Q or P is the object of a preposition PREP and PREP is an adjunct of Q.* **4) Containment A:** *P is contained in a phrase Q if P is either immediately contained*

hergestellten **syntaktischen Filters**¹⁵⁴ ist es, zu prüfen, ob sich in den verblieben Kandidaten keiner befindet, der syntaktisch betrachtet, niemals in einer anaphorischen Situation mit dem Pronomen stehen könnte (Vgl. Holger Wunsch 2010: 105). Blicke es nach all dieser Etappen mehr als einen Kandidaten, dann wird derjenige mit der höchsten **Salienz-Gewichtung** ausgewählt. Haben die verbliebenen Kandidaten die gleiche Salienz-Gewichtung, **dann wird derjenige Kandidat vorgezogen, welcher dem anaphorischen Pronomen am nächsten steht.**

Wie schon zuvor erwähnt, sieht **die Berechnung der Salienz** eines Ausdrucks **bei Computern** etwa anderes **als bei Menschen** aus. In der Tat, wird bei menschlicher bzw. bei manueller Auflösung von Anaphern direkt einem Antezedenskandidaten (NP), **je nach seiner syntaktischen Funktion im Satz**, eine Werte zugeschrieben (Sofac 2021: 45). Diese Werte bleibt bis zum Prozessende (Auswahl des richtigen Antezedens) fest und unverändert. Anders ausgedrückt, man wird bei **Ambiguitäten-fällen**¹⁵⁵ immer den hochrangierten Antezedenskandidaten auswählen. In diesem Sinne werden beispielweise immer Subjekte als Objekte präferiert (**Oblikheits-Hierarchie: Subjekt > direktes Objekt > indirektes Objekt > Adjunkte**. (Vgl. M.Krifka WS 2007: 9). Neben der **Oblikheits-Hierarchie** als Desambiguierungsmittel treten noch weitere Mittel ins Spiel. Dazu zählen **die Fern-Nähe-Regel** (Vgl. Kubczak 2016: 231) und **der syntaktische Parallelismus**¹⁵⁶ (Vgl. Renata

in Q, i.e. an argument or an adjunct of Q; or P is immediately contained in some phrase R, and R is contained in Q. (Vgl. Lappin & Leass 1994:538-539 zit. nach Holger Wunsch 2010: 107)

¹⁵⁴ Auf Basis dieser obenstehenden Einstellungen haben Lappin & Leass (1994) **die folgenden Heuristiken zur Identifizierung und Verzeichnung in Prolog-Unit von nicht-koreferenten Ausdrücken eines gegebenen Satzes.** (Vgl. Ludger Paschen 2013: 60)

RESTRIKTIONEN	BEISPIELE
P und N stimmen in den Kongruenzmerkmalen nicht überein	*The woman _i said that he _i is funny.
P ist im Argumentbereich von N	*She _i likes her _i .
P ist im Adjunktbereich von N	* She _i sat near her _i .
P ist ein Argument des Kopfes H, N ist kein Pronomen und in H enthalten	*He _i believes that the man _i is amusing.
P ist im NP-Bereich von N	* John _i 's portrait of him _i is interesting.
P ist ein Artikel des Nomens Q und N ist in Q enthalten	*His _i portrait of John _i is interesting.

Abb. 16: Syntaktischen Einschränkungen zur Ausschliessung von Koreferenz, basierend auf [Lappin et al. 1994:538] (Vgl. Renata Sevcikova 2005: 33).

¹⁵⁵ Eine Ambiguität während der Auflösung einer Anapher steht vor, wenn es immer **noch mehr als einen Kandidaten** gibt; und dies obwohl **morphologische Kongruenzen** zwischen Anapher und potentiellen Antezedenzen schon durchgezogen worden sind.

¹⁵⁶ Man redet noch von **dem Prinzip des Parallelismus in Bezug auf grammatikalische Rollen** (Vgl. Renata Sevcikova 2005: 33).

Sevcikova 2005: 33). Während das erste Prinzip es besagt, dass der Kandidat, **der der Anapher am nächsten steht**, präferiert wird, weist das zweite darauf hin, dass der Antezedens-kandidat, der zusammen mit der Anapher **die gleichen syntaktischen Funktionen** teilt, besser bewertet wird als einer, der sie nicht teilt.

Doch im Gegensatz zu Menschen, die es leicht zwischen Subjekte und Objekte in Sätzen unterscheiden können, ist dies bei Computern leider nicht der Fall. Computer können **dank syntaktischen Parsens** lediglich nur **die grammatische Natur von Satzelementen** identifizieren. Sie müssen also mit diesen wenigen Angaben weitervorankommen können; und zwar bis sie, grob betrachtet, **den Antezedenskandidaten** gefunden haben. Der Weg dorthin lässt sich, wie schon seit dem Anfang dieses Teils illustriert, **in mehreren Etappen strukturieren**. Die erste Etappe besteht darin, den Satz zu parsen, also den Satz so zu zerlegen bis jede seiner Konstituenten erkennbar wird und anhand ihrer jeweiligen grammatischen Natur in **einem Zwischenlager (Arbeitsplatz) im PROLOG-Arbeitsfenster** gelagert wird (Vgl. Lappin & Leass 1994: 543-546). Die Speicherung im Zwischenlager erfolgt wie Folgendes:

- Für jedes Satzelement wird **ein Prolog-ID (Identifiant Prolog)** eröffnet.
- Der Prolog-ID besteht aus zwei Nummern. Während die erste Nummer auf die Position des Satzes im Text verweist, referiert die zweite Nummer eher auf die Position des gespeicherten Elements im Satz. (Vgl. Ebd.: 543-546)
- In jedem Prolog-ID befinden sich die folgenden Merkmale: **Informationen** über **die grammatische Natur** des entsprechenden Satzelements; **Angaben** über Kopf-Argument & Kopf-Adjunkt Beziehungen; **Angeben** über die Bestimmtheit/Unbestimmtheit einer NP; **Angaben**, darüber, ob es sich um definite, indefiniten oder reflexive Pronomen handelt; **Angaben** darüber, ob das Element zum ersten Mal erscheint usw....(Vgl. Ebd.: 543-546)
- Weiterhin werden auch, nach dem Einsatz von weiteren Komponenten des AR-System, wie etwa **der syntaktische Filter**, weitere Angaben zu jedem Prolog-ID automatisch hinzugefügt. Dazu zählt beispielweise **Angaben über die Möglichkeiten von Satzelementen in einer anaphorischen Beziehung zu stehen**. (Vgl. Ebd.: 543-546)

Wie gesagt, **RAP** (Lappin & Leass 1994) ist ein Algorithmus zur Auflösung von anaphorischen Pronomen der dritten Person Singular. Doch eben wie andere Algorithmen lässt sich auch RAP (Lappin & Leass 1994) mit Hilfe einer informatischen Sprache automatisieren, also in einem vom Computer verständlichen Verfahren umwandeln. Die informatische Sprache **PROLOG** wurde für den Automatisierungsprozess ausgewählt. **RAP** (Lappin & Leass 1994) lässt sich

daher nur auf **eine clausele Repräsentation von Texten** einsetzen. Und eine clausele Repräsentation eines Textes ist nicht anders als eine Repräsentation von Texten mit Hilfe von **PROLOG-ID** (Vgl. Ebd.: 543-546). Und wie schon oben eingeführt, wird **RAP** (Lappin & Leas 1994) nach dem Einsatz des Parsers (Lappin and McCord (1990a) in der Lage sein, **Pronomen von NP** zu unterscheiden. Und nach dem jeweiligen Einsatz von **zwischen Modulen** (zwischen Algorithmus) zur Auffindung und Auflösung **von pleonastischen und reflexiven Pronomen**, werden dann Antezedenten für die verbliebenen Pronomen gesucht.

Dafür wird, wie oben schon hingewiesen, **jeder NP eine Prolog-ID** gegeben. Dieser besteht aus zwei Nummern, die erste Nummer verweist auf den Satz, in dem sich die NP befindet, und die zweite Nummer eher auf die Position des Kopfs der NP im Satz. **Diese Verzeichnung von NP anhand Prolog-ID wird jeweils für jede 4 Sätze durchgezogen.** (Vgl. Lappin & Leass 1994: 543-546). Nach dem vierten Satz werden die vorherigen Daten gelöscht und eine neue Identifizierung fängt wieder an. Das Besondere an **dieser Methode der ID-Identifizierung** liegt dann in der Tatsache, dass sie dem Computer ermöglicht auf vorherige etablierte Merkmale zuzugreifen, um dann Entscheidungen zu treffen. Es wäre in der Tat für den Computer nicht möglich, ohne diesen Zugriff auf gespeicherte Informationen zu wissen, ob eine NP neue im Text erscheint oder schon erwähnt worden ist; bestimmt oder unbestimmt ist; zu einer koreferenten Kette gehört oder nicht und welche syntaktische Funktion sie besitzt (Subjekt oder Objekt). Und mit all diesen Informationen und **anhand eines Gewichtungsmechanismus**, der gleich erklärt wird, wird der Computer bzw. das RAP-System in die Lage versetzt, pronominale Anapher aufzulösen.

Was **den Gewichtungsmechanismus** angeht, wird jeder NP, je nachdem, ob sie sich **in einer Einflusszone eines Salienz-Faktors**¹⁵⁷ befindet, **eine bestimmte Anfangswerte** zugeschrieben. **Diese Anfangswerte** wird dann halbiert jedes Mal, wenn man von einem Satz zu einem anderen geht. Am Ende des Prozesses könnte es also dazu kommen, dass **ein Salienz-Faktor** bezüglich einer bestimmten NP **eine Werte 0 (null)** erhält, und dies obwohl er zuvor auf 100 war. Konkret gesagt, wird eine Salienz-Werte für jede NP eines Satzes kalkuliert; und zwar egal, ob die NP zu den möglichen Antezedenten einer beliebigen Anapher gehört oder nicht. Und **diese Salienz-Werte** ist, wie schon oben gesagt, die Summe von allen Salienz-Faktoren, die einen Einfluss auf die entsprechende NP üben (Vgl. Lappin & Leass 1994: 543-

¹⁵⁷ Der Begriff **Einflusszone eines Salienz-faktors** lässt sich durch die folgende Situation illustrieren: Wenn ein sprachliches Element (hier eine NP) sich beispielweise vor oder nach **dem Verb** befindet, dann wird es vom RAP (Lappin & Leass 1994) verzeichnet als ein Element, das **in der Einflusszone des SUBJEKT und des OBJEKT** steht und wird deswegen die beiden entsprechenden Gewichtungswerte (80+50) zugeschrieben.

546). Es kann also dazu kommen, dass eine NP etwa **210 Punkte als Salienzwerte** im ersten Satz erweist, aber befindet sich trotzdem 3 weitere Sätzen später mit einer **Salienzwerte von etwa 50 Punkte**. Anders ausgedrückt, es wird **eine Salienzwerte** (Summe aller an einem Moment in Kraft auftretende Salienzfaktoren) für jede NP des Textes kalkuliert, und nur NP mit einer Salienzwerte gleich/über die **Salienzschwelle**¹⁵⁸ werden berücksichtigt. Stellen wir zunächst die im RAP-System (Lappin & Leass 1994) eingesetzten Salienzfaktoren und deren jeweiligen Gewichtung dar.

Salience factor	Value
Positional salience factors	
Sentence recency	100
Salience factors based on grammatical role	
Subject emphasis	80
Existential emphasis	70
Accusative emphasis	50
Indirect object and oblique complement emphasis	40
Head noun emphasis	80
Non-adverbial emphasis	50

Tabelle 11: Salienzfaktoren und Gewichtungen bei RAP (Vgl. Holger Wunsch 2010:106)

Salienzfaktoren	Gewichtung
<i>RECENCY</i> ¹⁵⁹	100
Subjekt (SUBJ)	80
<i>HEAD NOUN</i> ¹⁶⁰	80
Existentielle Konstruktionen (EXIST) Wie z.B.: „Es gibt einen Hund“	70
Direktes Objekt (DO)	50
<i>NON-ADVERBIAL</i> ¹⁶¹	50
Indirektes Objekt (IO)	40

¹⁵⁸ Durchschnitt der **Salienzwerte aller NPs**, die in der Antezedenskandidatenliste vorhanden sind.

¹⁵⁹ ***RECENCY*** belohnt das Auftreten im gerade verarbeiteten Satz und wird *de facto* auf jeden Referenten angewandt.

¹⁶⁰ ***HEAD NOUN*** emphasis ist gegeben, wenn sich die NP des Referenten nicht in einer anderen NP befindet. (Vgl. L.Paschen 2013: 60)

¹⁶¹ ***NON-ADVERBIAL*** emphasis liegt vor, wenn sich die NP nicht in einer adverbialen PP befindet. (Vgl. Ebd.: 60)

PARALLELISMUS ¹⁶²	35
Präpositionalobjekt (PO)	30
CATAPHORA PENALTY ¹⁶³	-100

Tabelle 12: Erweiterung von Salienzfaktoren und deren jeweiligen Gewichtung bei Lappin & Leass 1994 (Vgl. L.Paschen 2013: 60)

- **Fazit & Grenzen**

Die Auflösung von Anaphern in RAP (Lappin & Leass 1994) ist **ein zweistufiger Prozess**. Und da der Suchraum potenzieller Antezedenzen eines Pronomens zunächst alle Nominalphrasen enthält, die dem Pronomen vorangehen oder folgen (was sehr viel sein können), verwendet RAP (Lappin & Leass 1994) **im ersten Schritt** einen morphologischen und einen syntaktischen Filter, die dazu beitragen, den Umfang des Suchraumes zu reduzieren. Während der morphologische Filter dazu eingesetzt wird, alle potenziellen Antezedenzen, die nicht mit dem Pronomen in Person, Numerus oder Geschlecht übereinstimmen, zu entfernen; entfernt der syntaktische Filter alle möglichen Antezedenzen, die **Bindungsprinzipien** (Noam-Chomsky 1981) verletzen würden, wenn sie mit einem Reflexiv- oder Personalpronomen gepaart würden. **Der zweite Schritt** ist der tatsächliche Lösungsschritt. RAP (Lappin & Leass 1994) verwendet eine **Salienzhierarchie**, die den syntaktischen Funktionen und positionellen Merkmalen bestimmte Gewichtung zuschreibt (siehe Tabelle 9&10). Danach summiert RAP alle Gewichtungen, die einem potentiellen Antezedent zugeschrieben wurde und berechnet somit seinen Salienz-wert. RAP (Lappin & Leass 1994) wählt schließlich den Kandidaten mit dem höchsten Salienz-wert als Antezedens des Pronomens aus.

Das System RAP (Lappin & Leass 1994) weist zwar einige Grenzen auf, denen wir im Rahmen dieser Arbeit nicht in Einzelheiten nachgehen werden. Es steht beispielweise fest, dass die Unmöglichkeit des Algorithmus **zwischen Haupt- und Nebensätze zu unterscheiden**, dazu führt, dass falsche NPs als möglichen Antezedens-Kandidaten verzeichnet werden¹⁶⁴ (Vgl. Holger Wunsch 2010: 107). Weiterhin kommt es auch zur Wahl falschen Antezedens, wenn

¹⁶² **DER PARALLELISMUS** steht vor, wenn Anapher und Antezedent in gleicher syntaktischer Funktion auftreten (Vgl. Ebd.: 60)

¹⁶³ **CATAPHORA PENALTY** steht vor, wenn die Anapher vor dem Antezedenten steht. Zu merken ist allerdings, Die letzten zwei Faktoren, sind nicht-permanent, d.h. ihre Werte werden nicht im DM gespeichert, sondern nur für einzelne Pronomen addiert. (Vgl. Ebd.: 60)

¹⁶⁴ Und dieser Fehler wird Auswirkungen auf den ganzen Auflösungsprozess haben. Vor allem, wenn wir bedenken, dass **die Salienzschwelle** höchst von **der Salienzwerte** jeder einzelnen NP abhängig ist.

der richtige Kandidat beispielsweise **in einer anderen NP eingebettet ist oder sich in einer anderen syntaktischen, für die Gewichtung ungünstigen, Position** befindet. (Vgl. Lappin & Leass 1994: 550-551). Aus all diesen Gründen wurde ein auf statistisches Verfahren basierendes **Verbesserungsmodul, namens RAPSTAT** eingebaut. Allerdings liegt die Erfolgsrate von RAP bei 86%; und mit dem Einbau **dieses statistisch gemessenen lexikalischen Präferenzen-Moduls, das mit der Technik von Kollokationsmuster** funktioniert, wurde eine Verbesserung um etwa 3% beobachtet, also 89% für RAP nach dem Einbau von RAPSTAT¹⁶⁵ (Vgl. Ebd.:551).

4.2.1.2. Der wissensarme Ansatz von Kennedy & Bougarrev (1996): Die sogenannte KP-RAP

Ebenso wie RAP von Lappin & Leass (1994) ist der KP-RAP „**knowledge-poor RAP**“ von Kennedy und Bougarrev (1996) auch eine auf Heuristik basierende Prozedur zur Resolution von Anaphern. Doch im Gegensatz zu Lappin & Leass (1994) zielte Kennedy und Bougarrev (1996) darauf ab, die linguistischen (Syntaktischen & morphologischen) Einschränkungen, die bei RAP (1994), wie eben oben erläutert, sehr umfangreich sind, zu reduzieren (Vgl. Don Tuggener 2016:59). Die beiden Forscher gingen nämlich von der Feststellung aus, dass RAP (1994) nur funktionieren kann, wenn diese umfangreichen bzw. **tiefen linguistischen Kenntnisse** (Einschränkungen) schon vorhanden sind bzw. schon mit Hilfe eines Parsers generiert worden sind. Und da **Parser** oder syntaktische **Analysator** meistens für bestimmte Zwecke bzw. zur Analyse bestimmter Texttypen konzipiert sind, fehlt es einem nämlich schwer irgendeinen Parser auszuwählen und ihn auf einen beliebigen Text einzuwenden. Und zwar, weil man nicht sicher sein wird, ob gewünschte linguistische Einzelheiten, die für den Start der Auflösungsprozedur (des Algorithmus) nötig ist, tatsächlich erzeugt werden können. Bei solchen Fällen redet man von **Reimlosigkeit** zwischen der Auflösungsprozedur und dem Parser. (Vgl. Don Tuggener 2016:65).

Von dieser Feststellung ausgehend, haben Kennedy und Bougarrev (1996) eine Resolutionsprozedur entwickelt, welche **wenige linguistische Informationen** aus dem

¹⁶⁵ *Des expériences ont été menées avec une amélioration de l'algorithme qui apporte des informations modélisées statistiquement concernant les relations sémantiques et réelles à la procédure de décision de l'algorithme. Fait intéressant, cette amélioration n'améliore que marginalement les performances de l'algorithme (de 3 %).* (Vgl. Lappin & Leass 1994 : 551 Traduit de Anglais vers Français via www.onlinedoctranslator.com le 28.01.2023 à 16h)

Inputtext benötigt, um funktionsfähig zu sein, also um Anaphern findig zu machen und sie gleich auflösen zu können. Die beiden Forscher konnten, in der Tat, einen Algorithmus entwickeln, der, im Gegensatz zu RAP (1994), nur **mit einfachen bzw. oberflächlichen morphosyntaktischen Annotationen**¹⁶⁶ funktionieren kann. Noch als **KP-RAP « knowledge-poor RAP »** bekannt, funktioniert der Algorithmus von Kennedy und Bougarrev (1996) auch in zwei Schritten. Beim ersten Schritt werden **Daten für die konkrete Auflösungsprozedur** bereitet (Analyse der oberflächlichen Struktur des Inputs). Und im Anschluss darauf wird dann **die konkrete Auflösungsprozedur** losgefahren (Vgl. Ebd.: 63). Die erste Phase, also die Annotation und die darauffolgende Extraktion nötiger linguistischen Kenntnisse aus dem Inputtext, lässt sich mit Hilfe eines morphosyntaktischen Tagger namens **LINGSOFT** durchführen. Im Anschluss an der Annotation werden dann Pronomen der 3ten Person Singular aufgelöst. Das Auflösungsverfahren bei Kennedy und Bougarrev (1996) ähnelt in fast allen Punkten (**Einsatz eines morphologischen Filters, eines syntaktischen Filters und bei der Berechnung der Salienz von einzelnen potentiellen Antezedenten**) dem von Lappin und Leass (1994)¹⁶⁷. Sogar die Auswertungsprozedur ihres Algorithmus ist das gleiche, das zur Auswertung von RAP (1994) gebraucht wurde. Doch mit einer Leistung von etwa 75% liegt KP-RAP (1996) weit hinter der Leistung von RAP (1994), welche sich um 86% erhöht (Vgl. Ludger Paschen 2013: 60). Gründe dieser Weithinterlegung ist nämlich die Tatsache, dass KP-RAP (1996) auf Texte aus verschiedenen Domänen getestet worden ist, während RAP (1994) nur für Texte eines einzigen Bereichs (Informatik) konzipiert sind. Ein weiterer Grund zur Erklärung dieses Unterschieds ist nämlich die Quantität und Qualität von linguistischen Informationen, die beim Auflösungsprozess berücksichtigt werden. Während tiefe Kenntnisse bei RAP (1994) berücksichtigt werden, werden bei KP-RAP (1996) eher oberflächliche Sprachmerkmale aufgehalten und im Prozess mitberechnet. (Vgl. Holger Wunsch 2010:114)

4.2.1.3. Die Übertragung von RAP ins Deutsche (G-RAP)¹⁶⁸

4.2.1.3.1. Der Ansatz von Hinrichs & Co. (2005)

¹⁶⁶ Kenntnisse, die nach Anwendung eines Parsers erzeugt werden.

¹⁶⁷ Aus diesem Grund werden wir diesem Verfahren, in Rahmen dieser Erläuterung, nicht in Einzelheiten nachgehen. Außerdem ist die Grundabsicht dieses Kapitels, genüge Kenntnisse über bestehende AR-Systeme zu sammeln, um dann, im Anschluss darauf, ein eigenes automatisches Auflösungsverfahren auf Beine stellen zu können.

¹⁶⁸ German-RAP

Das System RAP (1994) ist von Autoren, wie **Schiehlen (2004)**, **Hinrichs et al. (2005)**, **Holger Wunsch (2010)** und **Paschen 2013)** zu der deutschen Sprache angepasst worden. Im Gegensatz zur englischen Variante, die auf die Textgattung *technische Handbücher* (Gebrauchtexte aus dem informatischen Bereich) getestet wurde, sind die deutschen Varianten der RAP also die G-RAP eher auf den *TüBa-D/Z-Korpus*¹⁶⁹ (eine aus deutschen **Zeitungstexten** bestehende Datenbank) evaluiert worden (Vgl. Don Tuggener 2016: 71). Doch eben wie KP-RAP (1996) verfolgen die deutschen Varianten immer das ursprüngliche Muster, das bei RAP (1994) ersichtlich ist. Allerdings merkt man einige Abweichungen von den ursprünglichen Prinzipien, vor allem, wenn es darangeht, **Gewichtungswerte** an bestimmten **Salienzfaktoren** zuzuschreiben. Nach etlichen empirischen Versuchen gelingt es den deutschen Forschern Eigenartigkeiten des Deutschen bei der **Modellierung der Salienz** (Mechanisierung der Berechnung der Salienz eines Diskursreferents (DR) / Ausdrucks) zu berücksichtigen (Vgl. Ebd.: 70). Es wurde in der Tat neue **Gewichtungswerte** für das Deutsche entwickelt (Siehe dazu die folgende Tabelle). Und diese Anpassung der **Gewichtungswerte** zu den deutschen Sprachprinzipien ist von großer Relevanz, denn die im RAP (1994) bestehenden **Gewichtungswerte** für die **Modellierung der Salienz** im Englisch entworfen worden sind und lassen sich deshalb nur schwer ins Deutsche Wert zu Wert übertragen. (Vgl. Hinrichs & Co. 2005). Was das Berechnungsverfahren in sich selbst aber angeht, bleibt dieses jedoch dasselbe.

Feature	Weight
<i>Subject emphasis</i>	170
<i>Accusative object emphasis</i>	70
<i>Dative object emphasis</i>	50
<i>Genitive object emphasis</i>	50
<i>Head noun emphasis</i>	80
<i>Short distance cataphora penalty</i>	-80
<i>Long distance cataphora penalty</i>	-175
<i>Parallelism reward</i>	35
<i>Current sentence reward</i>	20

¹⁶⁹ **Die Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z)** wurde von der Sektion Allgemeine Sprachwissenschaft & **Computerlinguistik** der Universität Tübingen entwickelt und beinhaltet in dem hier verwendeten Release 8 (Beck 2013) etwa **75.000 morphosyntaktisch sowie text- und diskurssemantisch** annotierte Sätze aus der deutschen Zeitung „die tageszeitung“, was ca. 1,4 Mio. Tokens bzw. 3.000 Texten entspricht. Mit mehr als **100.000 Anaphorik- und Koreferenzrelationen** bietet die TüBa-D/Z das zurzeit umfangreichste deutschsprachige Korpus mit derartigen **Annotationsebenen** (Vgl. Ludger Paschen 2013: 58).

Tabelle 13: Features and their weights in RAP-G, Hinrichs et al. (2005)'s adaption of Lappin and Leass (1994)'s approach to English third person pronoun resolution. (Vgl.

Hinrichs et al. (2005) zit. nach Don Tuggener 2016: 70)

4.2.1.3.2. G-RAP nach Hollger Wunsch (2010)

Auch H. Wunsch (2010) hat RAP (1994) ins Deutsche überträgt. Ebenso wie Heinrich & Co. (2005) nimmt er als Inputtext Ausschnitte des **TuBa-D/Z** und bewahrt zugleich auch die gleichen für das Deutsche erfundenen **Salienzwerte**. Der Inputtext, auf dem er seine Prozedur evaluiert, besteht aus etwa 1 188 Artikel, dies entspricht etwa 25 312 Sätze und genau **13 278 Pronomen (Personal, Reflexiv und possessiv)** der dritten Person. Doch im Gegensatz zu der Lappin & Leass (1994) Prozedur, in der einen Zwischenalgorithmus zur Ausfilterung von **pleonastischen Pronomen** eingesetzt wurde, durchführte H. Wunsch (2010) diese Aktion mit Hilfe der ursprünglichen Annotation des Korpus, also mit Hilfe des **Goldstandards**. Das folgende Schema illustriert die Menge der verarbeiteten Pronomen.

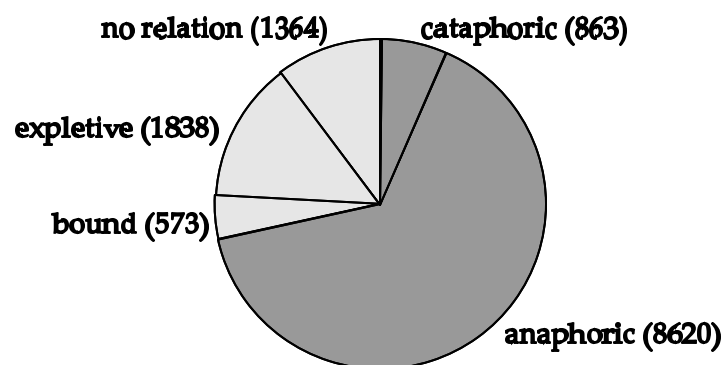


Abbildung 7: Distribution of referential relations of third person pronouns in the first 188 articles of the TuBa-D/Z treebank. The relations in dark gray are handled by G-RAP.

(Vgl. Holger Wunsch 2010: 114)

Aus dem obigen Schema steht es nämlich fest, dass der Ansatz von Holger Wunsch (2010), also die **G-RAP (2010)** eine **Art Verlängerung** und vor allem auch, wie wir es gleich anhand einiger Merkmale sehen werden, eine **Art Verbesserung** der **G-RAP (2005)** ist. **Die G-RAP (2005)** wurde nämlich für Inputtexte entwickelt, die nicht mehr als 350 Pronomina enthalten.

Überschreitet ein Text diese Grenze, dann *buggiert* das System. Im Gegensatz zum System von Hinrichs & Co. (2005) kann G-RAP (2010) genau 13 278 in einem Inputtext bearbeiten. In den beiden folgenden Schemen wird der Ansatz von Holger Wunsch (2010) zusammengefasst. Während die erste Abbildung alle Schritten des Algorithmus bzw. **alle Module des Systems** zeigen, fokussiert die zweite Abbildung eher **auf die Beschreibung der Berechnung der Salienz eines potentiellen Antezedens** im System.

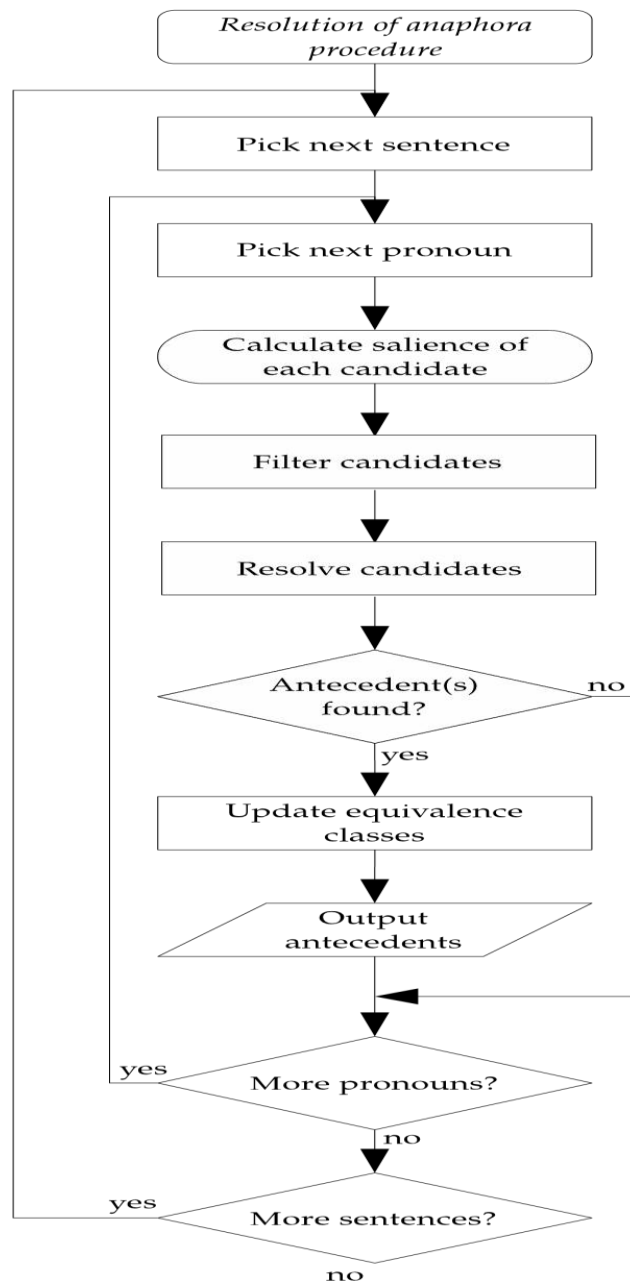


Abbildung 8: Flow chart of German RAP (Vgl. Holger Wunsch 2010: 122)

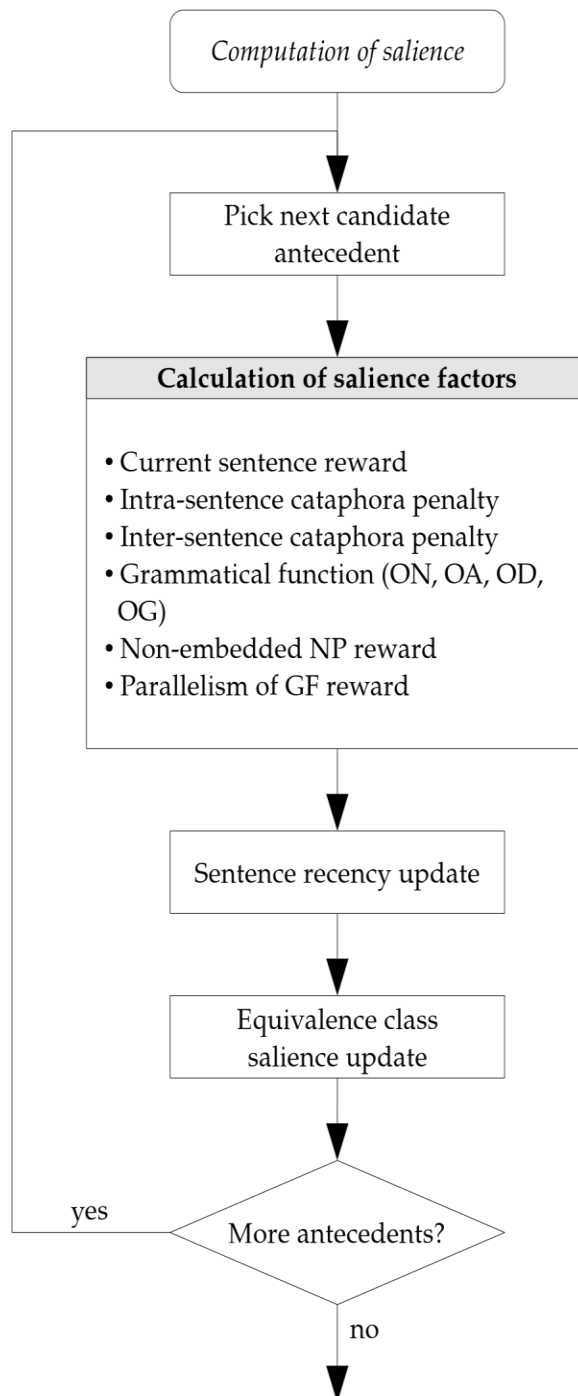


Abbildung 9: Computation of salience value. (Vgl. Ebd. : 123)

4.2.1.3.3. Der Ansatz von Ludger Paschen (2013)

Ein weiterer Autor, der RAP (Lappin & Leass 1994) zu der deutschen Sprache angepasst hat, ist **Ludger Paschen (2013)**. Im Gegensatz zu seinen Vorgängern, die weiterhin mit der **PROLOG-Sprache** gearbeitet haben, implementiert Ludger Paschen (2013) seinen Algorithmus mit **der Version 2.7. der informatischen Sprache/ Programms namens PYTHON** (Vgl. Ludger Paschen 2013: 61). Ebenso, wie seine Vorgänger benutzt er Fragmente des TüBa-D/Z- Korpus als Input seines Programms. Grob betrachtet, wird das Programm nach dem Empfang des Inputs darüber abreiten **Äquivalenzklassen**¹⁷⁰ für alle referierenden Pronomen des Inputtextes zu erstellen bzw. zu berechnen. Und im Anschluss daran werden die vom Algorithmus gebildeten Äquivalenzklassen (Ergebnisse des Auflösungsprozesses) mit den sogenannten **Gold-Äquivalenzklassen**¹⁷¹ (Manuel annotierte korreferente Kette) verglichen, woraus die Performanz des Algorithmus fixiert/ etabliert werden kann. Diese erhöht sich nämlich auf etwa 70%. (Vgl. Ebd.: 64). Die G-RAP (2013) gliedert sich in folgenden Schritten.

¹⁷⁰ Gehören zu **einer Äquivalenzklasse** alle sprachlichen Einheiten, die zu demselben Referenten (außersprachlicher Realität) referieren. Somit handelt es sich bei der Erstellung von **Äquivalenzklassen** um die Auflösung von Pronomina

¹⁷¹ Das besondere bei **der Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z)** ist, dass alle koreferenten Ausdrücke schon zuvor manuell annotiert worden sind. Im Allgemeinen werden manuelle Annotation von Korpora als **Goldstandard** bezeichnet. Also als eine Art Model wogegen die Ergebnisse eines AR-Systems konfrontiert werden können. Und da es hier um die Auflösung **von korreferenten pronominalen Anaphern** geht, ist **die Gold-Äquivalenzklassen** von besonderer Relevanz, um **die Ergebnisse des Systems** bewerten zu können

- i. Definiere Score-Werte, Markables und **Pronomentypen**, erstelle leere Äquivalenzregister.
- ii. Erstelle am Anfang eines Textes jeweils **ein leeres DM**.
- iii. Gehe nun satzweise vor. Sammle alle Anaphern (die oben definierten Pronomen mit Ausnahme jeglicher Expletiva) und alle potentiellen Antezedenten (**Markables**) und merke ihre morphosyntaktischen Werte. Merke gleichzeitig die **Anaphorikannotationen** der TüBa-D/Z.
- iv. Aufbau des DM für referierende Markables und Pronomen: Bestimme die **grammatische Funktion** und weise den Markables die zuvor definierten Score-Werte (s. Tab 1, letzte Spalte) zu und speise sie in das DM ein. Ein OBJA wird als EXIST-Konstruktion gewertet, wenn folgende Bedingungen erfüllt sind:
 - ON ist der Wert von func im node des „es“
 - Im Satz, in dem das OBJA steht, gibt es mind. eine expletive-Relation und eine Wortform „es“
 - Es steht kein SIMPX-Knoten zwischen dem „es“ und dem OBJA (s.u.)
- v. Gleiche alle Pronomen der Reihe nach mit allen Referenten im DM ab. Wende als **harte Constraints** an: Genuskongruenz, Numeruskongruenz (nicht bei PPOSAT) und Bindungsprinzipien. Wende als **weiche Constraints** an: Kataphernrelation und Parallelismuskonstruktion (für die Score-Werte s. 1).
- vi. Ermittle Referenten **mit höchstem Score** (bei Gleichstand wähle den dem Pronomen nächsten) und füge das koreferente Paar dem Äquivalenzregister hinzu.
- vii. Lösche Referenten, die vier oder mehr Sätze zurückliegen und teile alle Werte der verbleibenden Referenten **durch 2**. (Vgl. Ludger Paschen 2013: 63).

Abbildung 10: Übertragung von RAP (Lappin & Leass 1994) ins Deutsche nach Ludger Pacher 2013 (Vgl. Ludger Paschen 2013: 63).

• **Fazit**

Letztendlich möchten wir noch darauf hinweisen, dass es verschiedene AR-Systeme gibt, die auf heuristische Basis entwickelt worden sind¹⁷². Dies ist beispielweis der Fall von dem renommierten System von Mitkov (2002a) namens **MARS** (**M**itkov's **A**naphora **R**esolution **S**ystem). Doch im Allgemeinen weisen alle diese Systeme trotz Unterschiedlichkeiten bezüglich des Wesens und der Struktur mancher Module, eine globale Architektur auf, welche wir anhand der folgenden Abbildung illustrieren möchten.

¹⁷² Dazu gehören die Systeme von Baldwin (1997); von (Denber 1998) und von (Siddharthan 2003)

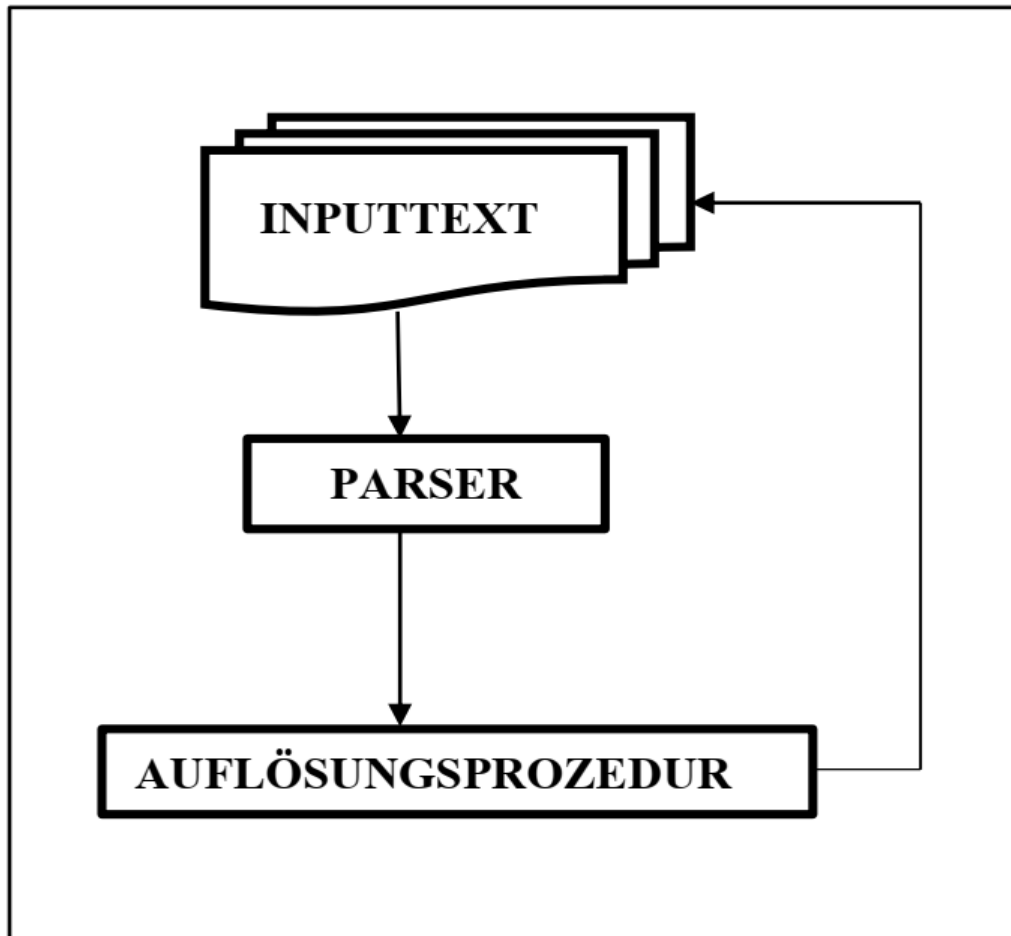


Abbildung 11: Allgemeine Architektur von AR-Systemen (Heuristische Systeme)

4.2.2. Auf Technik des Maschinenslernens basierende AR-Systeme: Einsatz der Künstlichen Intelligenz bei der automatischen Auflösung von Anaphern.

Das **Maschinelle Lernen (ML)** ist eine Technik, die im **Forschungsbereich der Künstlichen Intelligenz (KI)** entwickelt worden ist (*I. Döbel 2018: 26*). Von dieser Betrachtungsweise ausgehend, beabsichtigen wir, zur Erläuterung dieser Technik, zunächst den Begriff **KI** in einem breiten und umfassenderen Überblick zu erörtern. Und erst im Anschluss daran werden wir anhand einiger bestehenden AR-Systeme den Einsatz von KI bzw. ML bei der Resolution von Anaphern illustrieren.

4.2.2.1. Künstliche Intelligenz (KI) und Maschinenslernen (ML)

4.2.2.1.1. Intelligenz und Intelligenzarten

Das Wort **Intelligenz** ist vom lateinischen Wort **intelligere** abgeleitet worden. Dieses lateinische Wort seinerseits ist ein **Determinativkompositum**¹⁷³, das aus **inter** (zwischen) und **legere** (Lesen/ wählen) besteht (Detlef H. Rost 2013). Was wortwörtlich bedeuten würde entweder (zwischen lesen) oder (zwischen wählen). Ersichtlich bei dieser etymologischen Betrachtungsweise des Begriffs **Intelligenz** ist, dass der Begriff zum Einsatz kommt, wenn es vor allem darum geht, die Fähigkeit eines (biologischen/künstlichen) Wesens entweder **zwischen verschiedenen Dingen** zu wählen oder noch **zwischen den Zeilen zu lesen** (Also die Fähigkeit zu entschlüsseln, was nicht explizit dargelegt worden ist). Hierbei sieht man schon eine gewisse **geistige bzw. kognitive Aktivität des** (biologischen/künstlichen) **Wesens**.

Weiterhin lässt sich diese aus der etymologischen Betrachtungsweise abgeleitete Vermutung noch mit der reinen Übersetzung des lateinischen Worts ins Deutsche bestätigen. Das Wort **intelligere** wird ins Deutsche mit „erkennen“, „einsehen“, „verstehen“ übersetzt (Vgl. Detlef H. Rost 2013). Und um etwas zu erkennen, zu prüfen oder noch zu verstehen, benötigt man unbedingt Zugriff auf ein kognitives bzw. Neuronales Netz. In dieser Hinsicht verweist der Begriff **Intelligenz** auf die Fähigkeit eines Wesens Lösungen zur Überwindung bestimmter Hürde zu finden¹⁷⁴. Und die Art der Intelligenz lässt sich dann nicht nur nach **der Art der Herausforderung** (Logik; Sprache; Mathematik usw...) kategorisieren, sondern auch nach **der Art des Wesens**, das die Herausforderung überwindet. Diesem zweiten Kategorisierungsprinzip nach, unterscheidet man insbesondere zwei Arten von Intelligenzen: **die natürliche oder biologische Intelligenz** (für Lebewesen, wie Menschen: Tiere oder noch Pflanzen) und **die künstliche oder machinelle Intelligenz** (für nichtanimierte Wesen wie Maschinen; Computer, Handys, Autos usw....) (Vgl. Jann Raveling 2022). Doch wie können dann **Computer, Autos oder noch Handys intelligent sein?** Anders ausgedrückt, wie können Autos autonom fahren bzw. Selbst entscheiden, wann biegen, wann anhalten; wann

¹⁷³ Wortzusammensetzung, bei welcher der erste Teil den zweiten in seiner Bedeutung näher bestimmt. (Detlef H. Rost 2013)

¹⁷⁴ **Intelligenz** ist ein vielschichtiges Konzept, das **die Fähigkeit eines Individuums** beschreibt, komplexe Informationen **zu verstehen, zu verarbeiten und daraus angemessene Schlussfolgerungen zu ziehen**. Diese Fähigkeit umfasst **kognitive Prozesse** wie Wahrnehmung, Lernen, Erinnern, Problemlösung, kritisches Denken und Entscheidungsfindung. **Intelligenz** wird durch eine Kombination von **genetischen** und **Umweltfaktoren** beeinflusst und kann durch **Bildung, Erfahrung und Training** verbessert werden. Es gibt verschiedene Theorien und Modelle, die versuchen, Intelligenz zu messen oder zu beschreiben, darunter der IQ-Test und das Multiple-Intelligenzen-Modell. <https://blog.neuronation.com/de/definition-der-intelligenz-was-ist-das-eigentlich/> (10.04.2023 um 07h)

beschleunigen bzw. bremsen; oder noch wann parken? Wie kommt es dazu, dass Computer selbst Sprachen und Bilder erkennen können; Text generieren und Gespräche führen? Ganz allgemein ausgedrückt, wie kommt es dazu, dass nichtanimierte Wesen wie Menschen handeln können?

Die Verleihung von menschlichen geistigen Fähigkeiten (hören, sehen, denken und antworten) an Maschinen erfolgt insbesondere durch ein einziges Mittel: **Die Programmierung**. Maschinen werden also zur Durchführung einer bestimmten Aufgabe fähig gemacht, indem sie vom Menschen dafür programmiert werden. Ihnen, den Maschinen wird also, je nach der Aufgabe, ein Bündel an Anweisungen (**Der Code/ der Algorithmus**) vom Programmierer gegeben, dem sie sorgfältig folgen müssen, um erfolgreich die Aufgabe lösen zu können. Verfolgt eine Maschine eine Vorschrift nicht, dann bugiert das System und die Handlung muss wieder noch von vorn angefangen werden. Man redet bei solcher Vorgehensweise zur **Intelligenzmachung bzw. Intelligenzverleihung** von/an Maschinen von **heuristischen Ansätzen (regelbasierten Systemen/Programme/Software)**¹⁷⁵. Hier lässt sich ganz schnell merken, dass es sich um eine Art Robotisierung der Maschinen handelt. Zwar erledigt die Maschine eine Aufgabe, welche Menschen auch erledigen können, aber sie entscheidet sich weder noch vom Selbst, diese Aufgabe zu erledigen, noch denkt sie eigenständig, wie sie verfahren muss, um die Aufgabe zu lösen, also welche Wege sie nehmen muss, um die Aufgabe zu meistern. All dies bleibt **bei solchen heuristischen Ansätzen** noch **Handlungsfeld des Programmierers**. In dieser Hinsicht fehlt es einem doch etwas schwer von Intelligenz der Maschine zu reden. Denn intelligent zu sein berührt vor allem auf das eigenständige Denken und nicht bloß Schritt für Schritt wiederholen, was schon von einem anderen Wesen angefertigt worden ist. Diese Art der Verleihung von Kompetenzen (*Intelligenz*) durch **vom Menschen ausgebaute Algorithmen** unterscheidet sich ganz gründlich von einer anderen Art der Intelligenzverleihung, welcher **eher auf das Prinzip des Selbstlernens von Maschinen** setzt.

In der Tat, handelt es sich, bei dieser zweiten Methode, um den **Entwurf eines Grundalgorithmus**, der sich im Laufe der Zeit, je nach der Menge der inzwischen verarbeiteten Daten, selbst verbessert (**R. Rey 2017**). Hier baut also die Maschine selbst den Algorithmus aus, und dies anhand ihrer eigenen Erfahrung, welcher sowohl während der Lernphase als auch während der Einsatzphase (Mehr dazu berichten wir in der Rubrik Maschinenlearning) gesammelt worden ist. Diese Art vom **Selbstlernen** kann man dann als

¹⁷⁵ Siehe mehr dazu im ersten Teil dieses Kapitels.

eine Ausprägung von **Intelligenz** (eine künstliche) einstufen, denn das Wesen (die Maschine) lernt und perfektioniert sich selber im Laufe der Zeit. Es findet also selbst seinen Weg und bekommt nicht alles vorbereitet und angefertigt vom Programmierer, wie es **bei heuristischen Ansätzen** der Fall war (Vgl. Jann Raveling 2022).

Doch wieso überhaupt soll man die Intelligenz den Maschinen verleihen?¹⁷⁶ Wieso sich die Mühe geben damit Maschine eben auch wie Menschen denken und handeln können? Antworten auf diese Fragen werden uns ermöglichen, überhaupt zu wissen, warum der Mensch sich entschieden hat mit einer Maschine einige geistige Fähigkeiten, wie etwa das Nachdenken¹⁷⁷ zu teilen (Vgl. Jann Raveling 2022). Der Forschungsliteratur nach, liegt der Hauptgrund der Entwicklung solcher Systeme/Programme/Softwares, an zwei Sachen. Die erste ist die Tatsache, dass es für Menschen immer unangenehm ist, **wiederholende Aufgabe** ständig zu lösen. Dafür wurde beispielweise heuristische Ansätze gebraucht um Programme/Systeme/Software zu entwerfen, welche dem Menschen von mühselige Aufgabe entlasten sollen (*Vgl. I. Döbel 2018: 26*). Dies hingegen stehen die Programme, die auf das Prinzip des Selbstlernen basieren. Ihrer Entwicklung hat nicht **die Mühseligkeit der Aufgabe** zu bedanken, sondern eher **die Unfähigkeit des Programmierers**, einen Code zur Realisierung bestimmter Aufgabe zu schreiben. Zu solchen Aufgaben gehören beispielsweise die automatische Bilderkennung, Spracherkennung, Textgenerieren, Textzusammenfassen usw..... Jann Raveling (2022) äußert sich dazu in einer online erschienenen Zeitung zur Digitalisierung mit folgenden Wörtern:

[...] Keine Programmiererin auf der Welt kann einen Satz an Anweisungen schreiben, der immer erkennt wie ich aussehe, ganz gleich ob das Foto nachts, am Strand oder im Auto aufgenommen wurde – in einem regelbasierten System wäre das vollkommen unmöglich, denn dazu müsste die Programmiererin alle möglichen Bilder im Vornherein kennen und beschreiben können. Eine Programmiererin bringt einer KI nun bei, wie sie Menschen erkennen kann, aber nicht, wie sie mich erkennen kann. Auch die KI kennt nicht jedes Bild von mir, aber sie kann aus einer Anzahl an vorhandenen Bildern lernen, wie ich aussehe und diese Regel dann auf neue Bilder übertragen und mich erkennen. Und das nicht nur mit mir, sondern mit Milliarden Gesichtern in Bruchteilen von

¹⁷⁶ Warum ist das wichtig? Weil bestimmte Probleme so komplex sind, dass es unmöglich ist, dafür einen Code von Hand zu schreiben.

¹⁷⁷ Zwar schreibt der Programmierer immer die maximale Zielsetzung aber die Wege dahin baut sich die Maschine selbst aus, deswegen redet man vom Selbstnachdenken und somit auch von Künstlicher Intelligenz (KI) bzw. vom Maschinellen Lernen (ML).

Sekunden. Eine KI ist also in der Lage, **mit bisher unbekannten Daten umzugehen, Muster zu finden oder Handlungen daraus abzuleiten**. Sie lernt **eigenständig** aus den ihr vorliegenden Daten – was sie lernt, wird dabei aber vom Menschen vorherbestimmt, indem dieser die KI designt. **Der Mensch programmiert die KI, die KI lernt jedoch eigenständig, wie sie die ihr einprogrammierte Aufgabe ausführt**. KI sind damit weitaus mächtiger als **regelbasierte Systeme**, da sie auf – im gewissen Rahmen – bisher unbekannte Situationen reagieren können und aus Erfahrung lernen. (Vgl. J. Raveling, 2022 <https://www.wfb-bremen.de/de/page/stories/digitalisierung-industrie40/was-ist-kuenstliche-intelligenz-definition-ki> am 05.04.2023)

4.2.2.1.2. **Zum Begriff Künstliche Intelligenz (KI)**

Der Begriff KI wurde zum ersten Mal vom US-Wissenschaftler **John McCarthy** im Jahre 1956 im Rahmen einer Konferenz erwähnt, welche in der Geschichte als *Geburtsstunde der Disziplin Künstlichen Intelligenz* bekannt ist (Mag. Dunja Darer, 2020: 11). Doch als Impuls zur Entwicklung dieser neuen Forschungsdisziplin galten sensible wissenschaftliche Fortschritte, die zuvor in den 1940er geleistet wurden. Darunter zählen beispielsweise, einerseits, die Entwicklung **der ersten mathematischen Modelle für neuronale Netze** (McCulloch, Pitts und Hebb, 1943) und andererseits, **der technische Sprung, der die damaligen Rechenmaschinen zur programmierbaren Entität geführt haben** (Vgl. Ulrich Eberl, 2018: 10). Der Sprung vom Rechnermaschinen zum Computer wurde nämlich dadurch charakterisiert, dass Rechnermaschine nicht mehr nur Zahlen als Input bearbeiten konnten, sondern auch andere Symbolen, wie etwa Buchstaben. Und an diesen bahnbrechenden Erkenntnissen schlossen sich natürlich auch weitere Herausforderungen an, wie etwa die Entwicklung von neuen Programmiersprachen, die diese neue Eigenschaft von Rechnermaschinen im Kauf nimmt. In dieser Hinsicht entwarf John McCarthy (1956) die erste **Programmiersprache, die speziell für die Verarbeitung von symbolischen Strukturen geschaffen wurde**. (Vgl. Mag. Dunja Darer, 2020: 12). Ihm nach, verfolgt die KI das Ziel, *Maschinen zu entwickeln, die sich verhalten, als verfügten sie über Intelligenz*. (John McCarthy 1956 zit. nach Ebd.: 10).

Doch als sozialer Anstoß dieser Forschungsdisziplin galt die Debatte über die Erweiterung des Einsatzbereichs vom Computer im **Alltagsleben**. Die an dieser Konferenz in Dartmouth anwesenden Forscher gingen nämlich von der Feststellung aus, dass Computer nur geschaffen

wurden, um Menschen **von langen und komplexen mathematischen Aufgaben** zu entlasten (Vgl. Mag. Dunja Darer, 2020: 13). Im Alltagsleben wird jedoch nicht ständig kalkuliert, sondern auch gelesen; gesprochen; gehört; gefühlt und vor allem auch gespielt. Zu diesem letzten erwähnten Aspekt des menschlichen Alltagslebens entschied sich der Elektroingenieur namens **Arthur Samuel (1959)** ein Programm für einen IBM-Großrechner zu schreiben (Vgl. Ulrich Eberl, 2018: 10). Er entwarf nämlich ein Programm für das Brettspiel **Dame**.

Die Regeln, mit denen der Algorithmus/das Programm geschrieben wurde, entsprechen nur die erlaubten Züge des Damenspiels. Während das entworfene Programm nur was im Spiel erlaubt oder nicht erlaubt ist, wusste, kannte **Arthur Samuel (1959)** mehrere Strategien und gewann somit bei jeder Spielrunde den künstlichen Gegner. **Arthur Samuel (1959)** wollte aber die Performanz seinen künstlichen Mitspieler erhöhen und er kam schließlich auf die Idee, ein neues Programm zu schreiben, das im Gegensatz zum ersten nicht nur die elementaren Regeln des Spiels kannte, sondern auch die Strategien von **Arthur Samuel (1959)** vorprogrammiert bekam. Doch das Besondere bei diesem neuen Programm lag nicht an der Tatsache, dass es mehrere Strategien als das erste kannte, sondern vielmehr, an dem Fakt, dass in ihm ein **Zwischenmodul** inkorporiert war, dass bei jedem Zug **die Gewinnwahrscheinlichkeit** für die aktuelle Aufstellung auf dem Brett **bewertete**. (Vgl. Ulrich Eberl, 2018: 10). Und da es während ein Spiel tausende unterschiedliche Aufstellungen von Spielfiguren¹⁷⁸ geben könnte, sollte das Programm bzw. das Zwischenmodul zur **Gewinnwahrscheinlichkeit** mehrere Runde spielen, um all diese Vorfälle mit ihren entsprechenden **Gewinnwahrscheinlichkeitswert** verzeichnen zu können. Und da dies ungeheuerlich viel Zeit in Anspruch nehmen würde, falls **Arthur Samuel (1959)** sich entschieden hätte selbst gegen den Computer zu spielen, wurde eher die Variante vorgezogen, in der der Computer gegen sich selbst spielt. So konnte der Computer, Spiel für Spiel, die Genauigkeit seiner Vorhersagen so verbessern, dass nach einer gewissen Schulungszeit **Arthur Samuel (1959)** keine Chance mehr gegen ihn hatte.

Dies war das erste Mal in der menschlichen Geschichte, dass ein Mensch einer Maschine etwas beigebracht hat *bei dem sie durch **stetiges Lernen** besser wurde als ihr eigener Lehrer*¹⁷⁹. (Vgl.

¹⁷⁸ Hier merken wir, es handelt sich immer noch um eine Art regelbasierte Systemen, aber im Gegensatz von klassischen Fällen, in denen der Start-Algorithmus schon ausgebaut ist. Baut sich hier der Start-/Grundalgorithmus von selbst aus.

¹⁷⁹ Dieses Ergebnis steht im Einklang mit der Bedeutung, die sich die Stanford University (2016) von der KI macht. Sie beschreibt nämlich die KI *als eine **Wissenschaft und Computer-Technologie**, die davon inspiriert ist, wie Menschen ihr Nervensystem und ihren Körper nutzen, um wahrzunehmen, zu lernen, zu schlussfolgern und Handlungen auszuführen*. (Stanford University, 2016). Eine Erweiterung dieser Bestimmung lässt sich in der

Ulrich Eberl, 2018: 10) Man redet auch bei solchen KI-Formen von **Expertensystemen**, denn die Maschine nimmt Rückgriff auf die Wissensbasis (Regeln), die zuvor vom Lehrer also dem Experten vorprogrammiert wurde, um daraus selbst Schlüsse zu ziehen. Noch unter der Kategorie der **symbolischen KI** (Vgl. J. Raveling, 2022) eingeordnet, bestehen solche Ansätze darin Symbole nach vorprogrammierten Regel zu kombinieren, um einem nachvollziehbaren Weg zu Ergebnissen durch zu öffnen.¹⁸⁰ Außer **der Damensoftware** von Arthur Samuel (1959), die auf dieser Technologie der **KI (Expertensystem)** basiert ist, unterscheidet man auch weitere Spielsoftware, wie die **Schachsoftware** namens „**Deep Blue**“ einerseits, die 1997 den amtierenden Schachweltmeister Gary Kasparov besiegte und die Software **AlphaGo Zero** von der **Google-Tochter DeepMind**, die beim Spiel „Go“ (eine komplexere Version von Schach) eben auch den amtierenden Weltmeister Lee Sedol mit 4:1 im 2016 gewonnen hatte (Vgl. J. Raveling, 2022).

Wie schon mehrmals erwähnt, merken wir, dass die Gemeinsamkeit an all diesen Systemen, ist die Tatsache, dass beim Programmieren der Software, Regeln des entsprechenden Spiels miteinbezogen werden. Somit wird die Maschine (der Computer) sowohl aus Regeln als auch aus Erfahrungen¹⁸¹ in der Lage versetzt, den besten Zug vorzuspielen und somit das beste Ergebnisse zu erreichen. Doch da diese Vorgehensweise sich nicht komplett **von reinen heuristischen Ansätzen distanzierte**, erzielte sie ebenso, wie heuristische Ansätze nur beschränkte Ergebnisse. Aus diesem Grund wurde gegen den 70er Jahren die Finanzierung für die KI-Forschung zunächst gekürzt und später auch eingestellt. In der Forschungsliteratur wird diese Phase als **Winter der Künstlichen Intelligenz** gekannt. Dazu äußert sich Ulrich Eberl (2018) mit folgenden Wörtern

*Über Jahrzehnte versuchten Forscher daher, mit sogenannten Expertensystemen alle Probleme zu bewältigen, etwa durch die Anwendung von **Wenn-Dann-Regeln** im Sinne*

Betrachtungsweise von Kreutzer/Sirenberg, (2019) nachvollziehen. „Der Kern der Künstlichen Intelligenz besteht darin, selbstständig große Datenmengen zu verarbeiten, darin **eigenständig Muster** zu erkennen und auf deren Grundlage **autonom Entscheidungen** und/oder Vorhersagen zu treffen. KI-Anwendungen sind bei solchen **Aufgabenstellungen** häufig schneller und – je nach System – auch kostengünstiger“. (Kreutzer/Sirenberg, Künstliche Intelligenz, 9. Zit. nach Mag. Dunja Darer, 2020: 12)

¹⁸⁰ **Eine symbolische KI** kommt auf einem nachvollziehbaren Weg zu Ergebnissen, in demes Symbole (also Wörter, Buchstaben, Ziffern o.ä.) nach vorprogrammierten Regeln kombiniert, um einen Schluss zu ziehen. Ein Beispiel dafür wäre die klassische Logik: **Symbol 1**: "Alle Menschen sind sterblich" **Symbol 2**: "Sokrates ist ein Mensch" **Schluss**: "Sokrates ist sterblich" (Vgl. J. Raveling, 2022)

¹⁸¹ Die Gewinnwahrscheinlichkeiten, die von den verschiedenen Spielfigurenaufstellungen während der Schulung des Programms abgeleitet worden sind, werden als Erfahrung für zukünftige Spiele gespeichert.

von: „Wenn die Nase läuft und der Patient Halsweh hat und hustet, aber nur geringes Fieber hat, dann ist die Wahrscheinlichkeit größer, dass es sich um eine simple Erkältung handelt und nicht um eine gefährliche Virusgrippe.“ Doch Rückschläge blieben nicht aus, denn oft sind solche regelbasierten Systeme gar nicht einsetzbar – vor allem bei Alltagsaufgaben, die wir Menschen scheinbar mühelos beherrschen: etwa beim Verstehen von Sprache oder beim Erkennen von Bildinhalten. Wenn man einem Computer beibringen will, einen Baum zu erkennen, genügt es nicht, ihm zu beschreiben, wie ein Stamm oder Äste aussehen. Denn auch ein Strommast hat so etwas wie einen Stamm und Äste – und im Winter verlieren viele Bäume ihr Laub, sodass sich auch Blätter nicht als Unterscheidungsmerkmal eignen. Es lassen sich zahllose solcher Fälle finden, **in denen man mit vorgegebenen Regeln nicht weiterkommt**. In den 1970er Jahren wandten sich daher viele Forscher frustriert vom Gebiet der Künstlichen Intelligenz wieder ab. Finanzierungsprogramme wurden gekürzt oder ganz gestrichen – im Rückblick wird dies als „**Winter der Künstlichen Intelligenz**“ bezeichnet. (Vgl. Ulrich Eberl, 2018: 10)

Zusammengefasst, ergibt sich aus diesem Zitat, dass man (Forscher und vor allem Förderer) sich insbesondere darüber Gedanken gemacht hat, wie Computer den Menschen bei anderen alltäglichen Aufgaben außer beim Spielen behilflich sein könnten. Zu diesen anderen alltäglichen Aufgaben gehören etwa **Textanalyse; Sprachübersetzung oder Sprach- und Bilderkennung** (Vgl. Ulrich Eberl, 2018: 10). Und um Computer so zu programmieren, dass sie im Nachhinein solche Aufgaben erledigen können, müssen Programme/Software/Systeme entwickelt worden, die auf Basis einer anderen Technik der Künstlichen Intelligenz (Maschinelles Lernen (ML)) funktionieren. In der Tat fungiert diese Technologie der KI als die weitverbreitete kommerzialisierte Technik und wird mit Hilfe von Künstlichen Neuronalen Netzen (KNN) und Big Data (Deep learning) realisiert. (Mehr dazu im folgenden Abschnitt). Angesichts der verarbeiteten Informationen wollen wir letztendlich die folgende Auffassung des KI-Begriffs bewahren: **Die Künstliche Intelligenz (KI)** ist ein Teilgebiet der Informatik (Computerwissenschaft), in dem ständig versucht wird, menschliche kognitive Fähigkeiten nachzuahmen. Diese Nachahmung kann entweder auf programmierte Vorgehensweise (heuristische Ansätze) basieren oder eher durch maschinelles Lernen¹⁸² erzeugt werden. Im

¹⁸² Dabei handelt es sich insbesondere um spezielle Algorithmen, die im Gegensatz zu den aus heuristischen Ansätzen entwickelten Algorithmen nicht aus voretablierten Anweisungen (Regeln) bestehen, sondern eher aus Regeln, die während der Schulung des Algorithmus dank mathematischer Modelle und statistischer Verfahren abgeleitet worden sind. (Vgl. Ulrich Eberl, 2018: 13) In der Tat wird der Algorithmus durch ständige Expositionen

Allgemeinen kann also die KI als eine Art Technologie verstanden, die Computer ermöglicht, menschliche Intelligenz nachzuahmen.

4.2.2.1.3. Das maschinelle Lernen (Machine Learning)

Abgeleitet von der obenstehenden Auffassung des KI-Begriffs ist **das Machinelearning (ML)** eine der beiden **Techniken**, die zur Verleihung von Intelligenz an Maschinen gebraucht werden. Und Angesicht der riesigen Aufgaben, die darunter stecken, wurde sie mittlerweile in der aktuellen Forschungsliteratur als **ein eigenständiges Teilgeiet der KI** aufgezeichnet (Vgl. Mag. Dunja Darer, 2020: 12). Doch ursprünglich betrachtet, wurde sie von Arthur Samuel (1959) als eine **Methode** entwickelt, welche ihm dabei geholfen hat, eine Spielsoftware zu entwickeln, die anhand unzähliger **Spielerfahrungen** von selbst lernen konnte, wann ein Spielzug am bestens zu spielen ist. Das Computersystem bzw. die Spielsoftware lernte also nicht nur die Regel des Spiels selbst zu kennen, sondern auch eigenständig einen guten Spieler zu werden. Damit ist gemeint, dass der gute Spieler nicht nur derjenige ist, der die Spielregeln kennt, sondern auch derjenige, der weist am welchen Zeitpunkt es angebracht ist, einen gegebenen Zug zu spielen. Und diese Lernfähigkeit von Computersystemen sind **spezieller Algorithmen**¹⁸³ zu danken.

Bevor wir zur historischen Entwicklung des Begriffs Machinelearning zurückkommen, wollen wir zunächst, darauf hinweisen, dass die Entwicklung dieser zweiten Methoden zur **Intelligenzverleihung an Maschinen** kam zum Stande als die erste an ihrer Grenze gestoßen war (Vgl. Ulrich Eberl, 2018: 10). Beim ersten Mittel zur Intelligenzverleihung an Maschinen, handelt es sich um **heuristische Ansätze**; also um eine Art Bündel an Anweisungen, die der Mensch (der Programmierer) für die Realisierung einer Aufgabe schon zuvor bereitet hat. Dieses Bündel an Anweisungen wird **Programm (Software)** genannt und wird letztendlich in

auf Datensätze (Schulung des Algorithmus) mathematische Modelle entwickeln können, welche die Merkmale, Gesetzmäßigkeit und Muster wiedergibt bzw. formalisiert, die sich aus dem Phänomen/Datensatz ableiten lassen. Weiterhin wird sich der Algorithmus selbst ausbauen indem er aus diesen formalisierten Gesetzmäßigkeiten Schlüsse eigenständig zieht bzw. Regel eigenständig entwirft, welche letztendlich zur Analyse und Interpretation von neuen ähnlich strukturierten Datensätzen angewendet werden (Vgl. Ebd. 14). Mehr dazu im folgenden Abschnitt

¹⁸³ [...] Die Grundlage für diese automatisierten Lernprozesse **bilden ausgefeilte Algorithmen**. Darunter kann man sich **eine Art Bauanleitung als Abfolge von Schritten und Regeln** vorstellen, mithilfe derer eine Aufgabe gelöst wird. **Diese Algorithmen** werden auf vorhandene Datensets angewendet und erkennen **selbstständig Muster und Gesetzmäßigkeiten**, um im Anschluss passende **Lösungsansätze abzuleiten** (Vgl. Gethmann & Co 2022).

einer gegebenen **Maschine (Hardware)**, sei es Computer, Handys usw... implementiert, also hochgefahren. Im Gegensatz zu diesen Softwares unterscheidet man Programme, die von Computer selbst und nicht vom Programmierer ausgebaut werden. **Diese Kompetenzverleihung** an Maschinen wurde dem revolutionären **Prinzip des Selbstlernens** zu danken. Immerhin werden diese Software, sobald sie anfertigt sind, also ausprogrammiert sind, eben wie die ersten in einer Hardware hochgefahren.

Als erster Wissenschaftler, der dieses Prinzip des Selbstlernens experimentiert hat, gilt der amerikanische Forscher **Arthur Samuel (1959)**. Er führte nämlich ein Experiment durch, in dem einem Computer beigebracht werden sollte, **das Damenspiel eigenständig zu lernen**. (Vgl. Thomas Beznoskov 2019: 9). Am Ende des Experiments ergab sich tatsächlich, dass der Computer in der Lage versetzt worden, in der er selbstspielen konnte. Dafür musste aber zuvor Einiges getan werden, zum einen, wurde **ein Grundalgorithmus auf Basis der Spielregel** entworfen. Und zum zweiten wurde ein Dispositiv aufgebaut, welches dem Grundalgorithmus ermöglichte, mehrere Spielrunden zu machen. Und aus der gewonnenen Erfahrungen¹⁸⁴ konnte also der Grundalgorithmus sich selbst ausbauen (Vgl. Mag. Dunja Darer, 2020: 12). Allerdings lässt sich auch aus dem Experiment feststellen, dass damit Maschine dieses eigenständige Lernniveau erreichen konnten, mussten **die Grundalgorithmen zuvor mit ein paar Heuristiken vom Programmierer ausgestattet** werden (Aufbau des Grundalgorithmus auf Basis von Spielregel). Und von dieser Schwäche ausgehend, stellte man nämlich fest, dass diese erste Variante des ML¹⁸⁵ nicht genug geeignet war, um weitere Aufgaben wie **Spracherkennung, Textgenerierung oder noch Bildererkennung** automatisch zu lernen. Und zwar, weil die Herausforderungen, die der Entwurf dieser nötigen Heuristiken mit sich bringt, einfach zu hoch sind und auch in gewissermaßen unrealisierbar sind, wenn es sich darum handelt, Maschinen beizubringen, wie sie entweder Texte eigenständig generieren können oder noch, wie sie Bilder oder Sprache eigenständig erkennen. (Siehe dazu das Zitat im Abschnitt 4.2.2.1.1.)

¹⁸⁴ Und dies wurde nämlich durch Statistische Verfahren ermöglicht, welche **die Gewinnwahrscheinlichkeit** von jedem Zug berechnete

¹⁸⁵ Auf **einer künstlichen Intelligenzskala** würden wir Maschinen bzw. Systeme/Softwares, die auf Basis heuristische Ansätze konzipiert sind als **KI der ersten Generation** einstufen. Im Gegensatz dazu würden wir KI-Systeme, die mit der Technik des maschinellen Lernens entworfen sind, sei es die erste oder zweite Version, **als KI der zweiten Generation** einstufen. Und zwar weil es bei diesen Letzten nicht mehr darum geht, dass Computer die von Menschen selbst entwickelte Parameter (**das rein fertige Produkt der menschlichen Intelligenz**) exekutieren, sondern eher, dass Computer selbst eigene Heuristiken (Regel/ Wege) entwickelt, um eine bestimmte Aufgabe zu erledigen. (**Eigenständige Ausprägung der Intelligenz einer Maschine**)

In dieser Hinsicht, wurde **eine zweite Variante des ML** entwickelt, welche nämlich auf ein anderes Prinzip des ML basiert ist: **Die Erkennung von Mustern und Gesetzmäßigkeiten während der Trainingsphase**¹⁸⁶. Dazu äußern sich *Russell und Norvig (2012)* mit folgenden Wörtern

Die Forschungsrichtung des maschinellen Lernens **verfolgt das Ziel** bestimmte Muster oder Regeln in vorgegebenen Datenmengen automatisiert zu finden. **Die gewonnene Information kann dann auf weiteren Daten mit ähnlicher Struktur und Aufbau angewendet werden um neue oder effizientere Lösungen zu erreichen.** (Vgl. *Russell und Norvig (2012)* in Thomas Beznoskov 2019: 5).

Gemerkt ist nämlich, dass der Grundalgorithmus unter Berücksichtigung der während der Trainingsphase erfindenen Muster und Gesetzmäßigkeiten selbst ausbauen können wird. Doch konkret betrachtet, bedeutet dies folgendes, wenn beispielsweise ein System zur **Bildererkennung** (Katzenerkennung) realisiert werden soll, dann muss dafür ein Grundalgorithmus entwickelt werden, der während der Trainingsphase, auf 1000de Katzenbilder aufgesetzt wird. Der Grundalgorithmus wird anhand mathematischer Modelle (Aufbau von künstlichen neuronalen Netzen) sich selbst ausbauen¹⁸⁷; er wird also dank dieser künstlichen Neuronen selbst Regel entwickeln, die ihm ermöglichen werden, bei der Einsatzphase in der Lage zu sein, Katzen auf einem Haufen von neuen Bildern jedes Mal erkennen zu können und dies, obwohl er nicht auf diese Bilder trainiert worden sind. Hier merkt man schon, dass es sich um eine Art der Intelligenz, die fast der des Menschen ähnlich ist, handelt. In dieser Hinsicht besteht **die maximale Zielsetzung von maschinellen Ansätzen** darin, aus Input-Daten sinnvolle Zusammenhänge zu erkennen und daraus Regeln (Rückschlüsse) abzuleiten, welche dann beispielsweise bei Aktivitäten/Aufgaben wie Prognosen, Trends erkennen, oder noch bei der Gruppierung von Daten nach bestimmten Kriterien, angewendet werden. (Vgl. Gethmann & Co 2022). Das ML ist also eine Methode zur Erkennung von Zusammenhängen und da die Erkennung von Zusammenhängen nur mit der Erfahrung wächst, wollen wir in Anlehnung an **Tom Mitchell (1997)**¹⁸⁸ und Ulrich Eberl

¹⁸⁶ Und dies gelang dank **verschiedener Techniken des ML** (Entscheidungsbäumen; Künstliche neuronale Netze (KNN)).

¹⁸⁷ Dieser Ausbau kann durch mehrere Techniken/ Lernverfahren des MLs durchgezogen werden. Darunter unterscheidet man Techniken wie **Entscheidungsbaum-Lernen, Lernen mit Hilfe Künstlicher Neuronaler Netzwerke, Bayessches Lernen, Random Forests, Clusteranalyse** usw....(Vgl. Ulrich Eberl, 2018: 12)

¹⁸⁸ Um das Prinzip des Lernens aus Erfahrung dem Vorrang zu geben, betrachtete **Tom Mitchell (1997)**, einer der Vorläufer dieser Disziplin, den Begriff Machinelearning aus einer formalen Perspektive. Dazu sagte er nämlich

(2018)¹⁸⁹, die folgende Auffassung des ML-Begriffs aufbewahren: **Das Machinelearning ist also eine Ausprägung der KI, bei der Algorithmen lernen, ohne zuvor explizit programmiert zu werden.** Beim Lernprozess werden Informationen aus vorhandenen Daten automatisch extrahiert, und anhand dieser Schlüsse gezogen bzw. Regel abgeleitet, welche einerseits zum Ausbau des Algorithmus beitragen werden. (Vgl. Ulrich Eberl, 2018: 12)

Von dieser zweiten Auffassung des Maschinellernens ausgehend, unterscheidet man drei Arten: **Das überwachte ML** (supervised learning), das **nichtüberwachte ML** (unsupervised Learning) und **das bestärkende Lernen** (Reinforcing Learning). Bei überwachtem ML handelt es sich darum, dem Grundalgorithmus eine Aufgabe zu geben und sobald er die Aufgabe gelöst hat, ihm dann die richtige Lösung zur Verfügung stellen, damit er vergleicht, ob die von ihm gegebene Lösung mit **der Goldlösung** übereinstimmt. Ist dies der Fall, dann geht der Algorithmus schlicht zur nächsten Aufgabe. (Vgl. Thomas Beznoskov 2019: 9). Treffen die beiden Lösungen (Antwort des Grundalgorithmus und Goldlösung) nicht miteinander, dann verbessert sich der Grundalgorithmus sodass, er bei der Wiederholung der Aktion/Aufgabe eine Antwort gibt, die mit der Goldlösung übereinstimmt. Dafür gestaltet sich also der Grundalgorithmus neu und zwar indem er Schlüsse¹⁹⁰ von seinem Fehlerschlag zieht, anhand welcher er sich neu strukturieren wird (Vgl. Mag. Dunja Darer, 2020: 10). Der Grundalgorithmus mutiert also zu einem anderen Algorithmus, welchem im Gegensatz zum ihm in der Lage ist, die Goldlösung für die entsprechende Aufgabe zu finden. Und diese vom Grundalgorithmus selbst geleistete Mutation wird sich so ständig während der Lernphase des Algorithmus wiederholen, dass der Grundalgorithmus am Ende dieser Phase schon einen gewissen Entwicklungsstand erweist, welcher ihm erlauben wird, eine gewisse akzeptable Performanz bei der Einsatzphase zu erreichen¹⁹¹. Weiterhin wird sich der Grundalgorithmus bei

folgendes: *Eine Maschine lernt aus der Erfahrung E hinsichtlich einer Klasse von Aufgaben T und dem Performance-Maß P, falls die Performance P hinsichtlich T mit E sich verbessert.* (Tom Mitchell, Machine Learning, 1997)

¹⁸⁹ Bei Maschinellernem erkennen **Computersysteme** Muster in Beispielen und können ihre „Erkenntnisse“ auf andere Beispiele übertragen. So lernen sie, aus Daten immer genauere Schlüsse zu ziehen und Entscheidungen abzuleiten. (Vgl. Ulrich Eberl, 2018: 12)

¹⁹⁰ Durch Vergleich mit was getan werden muss, um die Goldlösung zu erreichen

¹⁹¹ Diese Erläuterung lässt sich nämlich durch die folgende Aussage von Alpaydm E (2019) bekräftigen:

Beim überwachten Lernen (supervised learning) liegt für jeden Datensatz der Input-Daten eine Bewertung vor, d.h. die Daten sind schon in Kategorien bzw. Klassen unterteilt.

Der Ablauf ist wie folgt: Die Input-Daten werden in **Trainingsdaten** und **Testdaten** unterteilt. Auf Basis der Trainingsdaten wird mittels eines passenden Algorithmus ein **Vorhersagemodell**, z.B. ein **Entscheidungsbaum** oder **Neuronales Netzwerk** erstellt, dessen Güte mit Hilfe von Performance-Kennzahlen (Kreuzvalidierung; Vertrauenswahrscheinlichkeit, Genauigkeit, Trefferquote) ermittelt wird. Der **Trainingsprozess** wird solange wiederholt, bis das Modell eine gewünschte Performance erreicht. Danach kann es für die Vorhersage / Klassifikation auf neuen Datensätzen verwendet werden. (Vgl. Alpaydm E 2019: 20)

der Einsatzphase immer weiter verbessern können. Und dies, obwohl er in dieser Phase keine vom Programmierer jedes Mal eingegebenen Ergebnisse (Goldlösung) mehr besitzt, wie es etwa bei der Trainingsphase der Fall war. Bei solchen Fällen von Selbstverbesserung (Ohne Eingriff vom Programmierer) redet man vom nichtüberwachten ML¹⁹²(Vgl. Döbel u. a. 2018 :26).

Ebenso wie beim überwachten Lernen, besteht beim nichtüberwachten Lernen immer auch eine Trainingsphase (Vgl. Thomas Beznoskov 2019: 9). Dem Algorithmus wird eine Aufgabe gegeben, die er auf Inputdaten realisieren soll. Doch im Gegensatz zum überwachten Lernen bekommt er keine Rückmeldung, wenn er die Aufgabe richtig oder falsch gelöst hat. Der Algorithmus baut sich selbst aus und ist am Ende der Trainingsphase in der Lage bestimmte Eigenschaften zu erkennen, also Inputdaten nach besonderen Kriterien auszusortieren. Das besondere bei nichtüberwachten Lernen ML ist, dass es meistens für den Algorithmus darum geht, kleine Auffälligkeiten in Dataset zu erkennen und dann Anhand dieser Auffälligkeiten Daten nach bestimmten Kriterien zu kategorisieren. Also der Algorithmus entwickelt sich selbst, indem er eigenständig Unterscheidungsmerkmale zwischen Daten merkt und benötigt somit keine Hilfe vom Goldmuster, wie es bei dem überwachten Lernen der Fall ist. Dazu äußert sich Alpaydm E (2019) mit folgenden Wörtern:

Beim unüberwachten Lernen (unsupervised learning) liegen für die Input-Daten keine Bewertungen vor, d.h. das Lernverfahren muss ein Modell ohne Zielvorgaben erstellen, indem es Muster in den Daten erkennt. Ein anschauliches Beispiel sind Clusteranalysen, ein Verfahren, mit dessen Hilfe man in einem Datensatz Gruppen ähnlicher Beobachtungen ("Cluster") identifizieren kann. Die Clusterbildung kann mit unterschiedlichen Algorithmen umgesetzt werden. Zwei der verbreitetsten Verfahren sind K-Means-Clustering und hierarchisches Clustering. (Vgl. Alpaydm E 2019: 23)

Ein weiterer Unterschied zwischen den beiden Ansätzen liegt auch an der Art der an ihnen gestellte Aufgabe. Während es bei dem ersten im Allgemeinen darum geht, mathematische Kalküle zu realisieren, handelt es sich beim zweiten darum, Daten zu klassifizieren, also auszusortieren. Ebenso wie **das nichtüberwachte Lernen** bekommt der Algorithmus beim Reinforcing Learning keine Goldmuster während der Trainingsphase. Noch merkwürdiger bekommt er auch keine Aufgabe. Er wird einfach eingeschaltet und erledigt bestimmte

¹⁹² Im Gegensatz zum überwachten Lernen, bei dem der Algorithmus lernt, indem ihm Beispiele mit bekannten Ergebnissen gezeigt werden, lernt der Algorithmus **beim unbewachten Lernen**, indem er selbstständig Zusammenhänge erkennt. (Vgl. Döbel u. a. 2018 :26).

Aktionen innerhalb eines Umfelds¹⁹³ (Das neue Element/ neue Eigenschaft, also etwas, das nicht bei den zwei ersten ML-Arten vorhanden ist). Und wenn die Aktion gut ist, bekommt er einen Anreiz als Belohnung, wenn sie nicht richtig realisiert ist, bekommt er keinen Anreiz. Und somit baut sich der Algorithmus von selbst weiter aus bis er das von seinem Ersteller vordefinierte Ziel erreicht.

*Beim bestärkenden Lernen (reinforcement learning) werden intelligente Agenten eingesetzt, die mit ihrer Umgebung interagieren und Strategien erlernen, um erhaltene Belohnungen zu maximieren. **Bestärkendes Lernen** benötigt keine Bewertung der Input-Daten wie das überwachte Lernen. Die mathematische Grundlage des Reinforcement Learning sind Markovsche Entscheidungsprobleme, die schon länger in anderen Disziplinen (Kontrolltheorie, Operations Research) erforscht und mit Algorithmen der Dynamischen Programmierung gelöst werden. **Bestärkendes Lernen** wird dann angewendet, wenn ein System nur durch **Interaktion** mit seiner Umgebung lernen kann, z.B. in der Prozessautomatisierung bei der Optimierung der Parametrisierung von Controllern. (Vgl. Alpaydın E 2019: 23)*

• FAZIT

Wie wir es herausgestellt haben, wird die Nachahmung der menschlichen Intelligenz bei Maschinen durch verschiedene Ansätze nachvollgezogen. Bei heuristischen Ansätzen baut der Informatiker eigenständig ein Programm (einen Algorithmus) aus, das er bei Maschinen (Computer/ Handy/Roboter) hochfahren wird. Die Maschine wird sich also exekutieren, indem er allen Anweisungen folgt, die zur Realisierung der gegebenen Handlung vom Programmierer niedergeschrieben worden sind. Beim zweiten Ansatz hingegen, also bei maschinellern Lernen, ähnelt der Erwerbprozess eher dessen der Menschen. Computer lernen, in der Tat, eine Handlung zu durchführen, eben auch wie Menschen es tun, das heißt durch ständige Wiederholung, also durch Erfahrung. Je öfter eine Maschine eine Aufgabe erledigt, desto mehr Erfahrung sammelt sie; und aus dieser Erfahrung, entwirft sie selbst ihre Vorgehensweise (Techniken) um eine gegebene Handlung durchzuführen. Wie schon im vorherigen Abschnitt eingeführt, sind viele Techniken solcher Art (also Techniken zum Selbstlernen) in der Literatur verzeichnet. Im Folgenden werden wir den Fokus auf eine AR-Systeme, die mit der Technik des maschinellen Lernens entworfen worden sind.

¹⁹³ Der Entwickler (Programmierer) bestimmt das Umfeld.

4.2.2.2.Einsatz des maschinellen Lernens (ML) bei der Auflösung von Anaphern

Kommen wir nun zu AR-Systemen, die das Prinzip des maschinellen Lernens (ML) verwenden, um das passende Antezedens einer gegebenen Anapher findig zu machen. Von vornherein, wollen wir darauf hinweisen, dass wir, im Gegensatz zu heuristischen Ansätzen, wo wir in Einzelheiten vorgegangen sind, nur einen Grobüberblick liefern werden. Und zwar weil das AR-System, das wir am Ende zu realisieren beabsichtigen, nicht aufs Prinzip des ML berührt, sondern eher auf dessen der Heuristik. Allerdings werden wir, wie hingewiesen, einiges über die Strukturierung sowie über das Funktionieren und die Performanz von ein paar AR-Systemen, die nach ML Prinzipien aufgebaut worden sind, beschreiben.

Noch unter der Bezeichnung ***Datengesteuerte Ansätze (Data-driven Approach) zur Anaphernaufflösung*** bekannt, verweist solche AR-Systeme auf einen Lernalgorithmus, der sich zunächst während der Trainingsphase ausbaut¹⁹⁴ und erst während der Einsatzphase konkret verwendet wird. (Vgl. Holger Wünsche 2010 :129). In der Tat, wird dieser Lernalgorithmus in der Lage sein, dank der aus Trainingsdaten gewonnenen Kenntnisse, ein Auflösungsmodell aufzubauen, das so allgemein ist, dass es bei neuen Datensätzen (Testphase) erfolgreich eingesetzt wird.¹⁹⁵ Allerdings muss darauf hingewiesen werden, dass damit das AR-System bei der Einsatzphase erfolgreiche Ergebnisse wiedergibt, muss zuvor die Schulungsphase gut gestaltet werden. Und die Gutgestaltung einer Trainingsphase bedeutet nicht anders als die Trainingsdaten gut zu annotieren, also die Inputdaten fehlerfrei zu annotieren (Vgl. Ebd.: 129). Damit ist also gemeint, dass wenn ein Lernalgorithmus zur Anaphernresolution auf einen Text trainiert wird, auf dem die Annotationen (Antezedens-Anapher) falsch/fehlerhaft zuvor von einer menschlichen Instanz (einem Linguisten) realisiert worden ist, dann wird das AR-System auch beim Einsatz nicht erfolgreich sein. Denn es hat sich all diese von der menschlichen Instanz durchgeführten Fehler als mäßig bzw. sinnvoll gemerkt. In dieser Hinsicht lässt sich also sagen, dass der größte und mühselige Teil der Arbeit, bei solchen Ansätzen, denjenigen gehört, die dafür verantwortlich sind, Trainingsdaten rechtmäßig/zielmäßig zu annotieren. Und

¹⁹⁴ Dazu äusserst sich Holger Wunsch (2010) mit folgenden Wörtern: *Wie ausführlich in Kapitel 3 besprochen, erfordern Ansätze zur Anaphernaufflösung, die auf **Strategien des maschinellen Lernens** basieren, kein ausgeklügeltes System linguistischer Regeln. Stattdessen extrahieren sie autonom ihr eigenes linguistisches Modell aus Trainingsdaten.* (Holger Wünsche 2010 :129)

¹⁹⁵ Im Gegensatz zu AR-Systemen, die wir am Anfang dieses Kapitels beschrieben haben, sind die auf Strategien des maschinellen Lernens basierende AR-Systeme kein vom Programmierer entworfenes Bündel aus linguistischen Regeln, sondern eher ein Ergebnis eines Extraktionsprozesses von Regeln (gezogene Schlüsse aus ständig vorkommenden Merkmalen) aus Trainingsdaten, der von einem lernenden Algorithmus (***maschineller Lernalgorithmus***) während der Trainingsphase des AR-Systems durchgeführt worden ist. (Vgl. Ebd.: 43)

solche AR-Systeme werden der oberen Kategorie des überwachten ML zugeordnet¹⁹⁶ (Vgl. Holger Wünsche 2010 :43). Im Folgenden besprechen wir einige Beispiele davon.

Ebenso wie heuristische Ansätze zur Anaphernresolution lassen sich Maschinlernansätze fast gleichmäßig unterkategorisieren. Als übliche Unterkategorisierungsparameter zählen wir beispielweise die SPRACHE¹⁹⁷ (Englisch, Deutsch, Französisch, Portugiesisch, Finnisch usw....); der TEXTTYP¹⁹⁸ (Zeitungstexte; Gebrauch- oder Handlungstexte; Wissenschaftliche Texte usw....); der **ANAPHERNTYP**¹⁹⁹ (Personal-Possessiv-Reflexiv- Demonstrativ Relativpronomen, Pronominaladverbien und Nominalphrasen/ NP) oder noch die ERFOLGSRATE (Performanz jedes einzelnen Systems). Aber das besondere bei AR-Systemen aus dem ML-Paradigma ist, dass sie sich nach einem neuen Parameter unterscheiden lassen: DIE NATUR DES LERNALGORITHMUS²⁰⁰

Man unterscheidet in der Tat, je nach der Art und Weise, mit der eigenständig gelernt wird, verschiedene Arten von Selbstlernenden Algorithmen: **Entscheidungsbaumalgorithmen; Algorithmen, die mit der Technik der künstlichen neuronalen Netze lernen; Gedächtnisbasierte Algorithmen** usw..... Soon & Co (2001) bauten beispielweise ein System, das auf **einen Entscheidungsbaumalgorithmus** basiert ist, und für die Auflösung von Pronomen der dritten Person (Personal, Reflexiv, Possessiv) im Englischen konzipiert worden ist. Dieser Lernalgorithmus war zunächst während der ersten Phase auf manuell annotierte Korpora MUC-6 (1995) und MUC-7 (1997) trainiert worden, und ausschließlich während der zweiten Phase auf konkrete Testdaten bewertet worden, wobei er eine Performanz von etwa 65% aufwies. (Vgl. Holger Wünsche 2010 :129-147). Zu dem konkreten Funktionieren von **Entscheidungsbäumen** äußert sich Holger Wünsche (2010) mit folgenden Wörtern:

*Anhand der Trainingsdaten konstruiert der Klassifikator [Lernalgorithmus] einen Entscheidungsbaum. **Entscheidungsbäume** sind Darstellungen eines Klassifikationsprozesses, der aus einer Folge von atomaren Entscheidungen besteht, die schließlich zur Zuweisung **eines Klassenlabels** führen. Jeder Knoten im Baum entspricht einer atomaren Entscheidung. Die Knoten sind mit den Merkmalen beschriftet. Somit*

¹⁹⁶ Ansätze für maschinelles Lernen, die manuell vorbeschriftete Daten verwenden, werden **als überwachte Ansätze** bezeichnet. (Vgl. Ebd.: 43)

¹⁹⁷ Hier handelt es sich um die natürliche Sprache wofür das AR-System entwickelt worden ist.

¹⁹⁸ Hier handelt es sich um den Typ des Textes, wofür das System entworfen worden ist.

¹⁹⁹ Hier handelt es sich um die Natur des verarbeitenden anaphorischen Ausdrucks

²⁰⁰ Bei AR-Systeme der zweiten Generation (ML-Ansätze) wird ein weiteres Unterscheidungsmerkmal berücksichtigt, nämlich die informatische Technologie, mit der der selbstlernende Algorithmus aufgebaut worden ist.

läuft jede Entscheidung darauf hinaus, ein Merkmal zu bewerten und, abhängig vom Wert dieses Merkmals, zum nächsten Knoten überzugehen. Die Konstruktion des Entscheidungsbaums ist die Kernaufgabe des C5-Klassifikators, bei der die meiste Arbeit zu leisten ist. Das Klassifizieren neuer Instanzen beinhaltet nur eine ziemlich einfache Funktionsprüfung. (Vgl. Holger Wünsche 2010 :130)

Allerdings waren Soon & Co (2001) nicht die ersten die diese Technik des maschinellen Lernens durch Entscheidungsbaum eingesetzt haben, **McCarthy und Lehnert, (1995)** hatten ihnen zuvor den Weg bereitet. Und zwar indem sie ein System zur Auflösung von Koreferenz namens RESOLVE entwarfen. RESOLVE (1995) wurde auf eine ältere Version von MUC, MUC-5 trainiert und hatte den C4.5 Entscheidungsbaum von Quinlan (1993) als Lernalgorithmus. Das Koreferenz-Auflösungssystem erreicht zwar eine Performanz von 86,5% aber die Herausstellung seiner Performanz bezüglich der Auflösung von Pronomen ist nicht extra aufgezeichnet worden²⁰¹ (Vgl. **Beata Kouchnir 2004** 31-37) Immer mit der gleichen Technologie, also Lernalgorithmen, die auf Entscheidungsbaum basieren, entwickelten Ng und Cardie, (2002) ein System, das ein globales Ergebnis von 76% erreichte und somit als eine Art Verlängerung der Arbeit von Soon & Co (2001) bezeichnet wurde. Morton (2000) hingegen arbeitet zwar immer noch auf die englische Sprache, aber er benutzt kein **Entscheidungsbaum** als Machinelearning-Technik. Stattdessen benutzte er eine Technik namens **Entropie-Modell**. Das System von Morton (2000) wurde auf Texten vom Wall Street Journal trainiert, und erzielte beim Einsatz eine Performanz von 81,5% bezüglich der Auflösung von pronominalen Anaphern.

Im Gegensatz zu den vorerwähnten Ansätzen, in denen Systeme nur für das Englische entworfen worden sind, entwarfen Strube & Co (2002) einen ML-Ansatz zur Auflösung von Personal- und Possessivpronomen in der deutschen Sprache. Das System wurde auf den Heidelberger Textkorpus trainiert und erreicht während der Einsatzphase eine Performanz von etwa 82,79% (Vgl. Holger Wunsch 2007:115-124)). Ein weiteres System, das für das Deutsche entworfen worden ist, ist das System CAR (Computational Anaphora Resolution) von Erhard W. Hinrichs, Katja Filippova und Holger Wunsch (2007). CAR (2007) lässt sich als ein

²⁰¹ Zu merken ist nämlich, dass Koreferenzsysteme nicht nur anaphorische Beziehungen in Betracht ziehen, sondern eher vielmehr alle Arten von Koreferenzbeziehungen. Dazu gehören sowohl kataphorische als auch iterative Beziehungen. Davon ausgehend, können wir dann vermuten, dass dies höchstwahrscheinlich der Grund war, warum Soon & Co (2001) sich entschieden haben Systeme zu entwerfen, die sich nur auf die Auflösung von pronominalen Anaphern beschränken und daher nicht alle Arten von Koreferenzbeziehungen bearbeiten, wie es etwa der Fall bei McCarthy und Lehnert (1995) war.

hybrides System beschreiben. Und zwar, weil es sowohl aus einer regelbasierten Komponente für die Vorfilterung und einer Maschinenringkomponente für die konkrete Auflösung besteht. Außer dieser Besonderheit der Kombination von zwei Ansätzen zu Einem, wird bei CAR (2007) kein Entscheidungsbaum als Lernalgorithmus verwendet, sondern eher ein **gedächtnisbasierter Algorithmus**.²⁰² Noch im Englischen unter der Bezeichnung *Memory Based Learner* (MBL) bekannt, ist diese Technologie des maschinellen Lernens bei CAR (2007) durch das Modul namens *Tilburg Memory Based Learner – TiMBL* implementiert worden (Telljohann & Co 2004). Außerdem wurde CAR (2007) auf deutsche Zeitungstexte (Die TüBa-D/Z Baumbank) trainiert und distanziert sich somit vom zuvor erwähnten Ansatz von Strube & Co (2002), welcher eher auf Touristeninformationstexte (Heidelberger Korpus) trainiert worden war. (Vgl. Kouchnir 2003).

- **Fazit und Ausblicke**

Zusammenfassend können wir sagen, dass maschinelle Lernansätze zur Auflösung von Anaphern die Zukunft der Forschung um diesen Forschungsgegenstand (Automatische Auflösung von Anaphern) herum ist. Daher beabsichtigen wir während unseren zukünftigen Auseinandersetzungen mit dieser Thematik, das Wissen um das Wesen, das Funktionieren und die Verwendung dieser revolutionären Technologie (Das maschinelle Lernen) zu vertiefen. Wir würden beispielweise untersuchen, wie ein AR-System zur Auflösung von Pronominaladverbienanaphern im Deutschen mit dieser Technik des ML möglich zu entwerfen wäre. Dabei würden wir also Neues erfahren, Neues lernen und uns somit immer tiefer in diesem Forschungsgebiet einwurzeln können.

Doch zuvor, beabsichtigen wir im Rahmen dieser Dissertationsarbeit, einen Ansatz bzw. ein AR-System zu entwerfen, das noch auf Heuristik basiert ist. Und wie schon am Anfang dieses Kapitels 4 erwähnt, war die maximale Zielsetzung dieses sich abschließenden Kapitels, genüge **computerlinguistische Kenntnisse** zu sammeln, um zu sehen, wie solche Systeme konzipiert werden könnten. Und am Ende der Exploration gelangen, können wir dann sagen, wir sind in der Lage, unseren im Kapitel 3 (Abschnitt 3.3.3.3.3.) zuvor entworfenen vorübergehenden Algorithmus, welcher vorwiegend **aus einer rein linguistischen Perspektive** entworfen worden ist, an computerlinguistischen bzw. informatischen Prinzipien anzupassen. Diese Anpassung lässt sich dann nämlich durch das folgende Schema illustrieren.

²⁰² Einbruch eines neuen Paradigmas (Das gedächtnisbasiertes Lernen) beim maschinellen Ansätzen zur Anapherauflösung

4.3. Entwerfener Algorithmus zur automatischen Auflösung von Pronominaladverbienanaphern

In Anlehnung an die allgemeine Architektur eines auf Heuristik basierenden AR-Systems (Siehe Abbildung 6) entwerfen wir die folgende Architektur (Algorithmus), worauf das später (im empirischen Teil der Arbeit) auftretende AR-System zur Auflösung von Pronominaladverbienanaphern basierend wird.

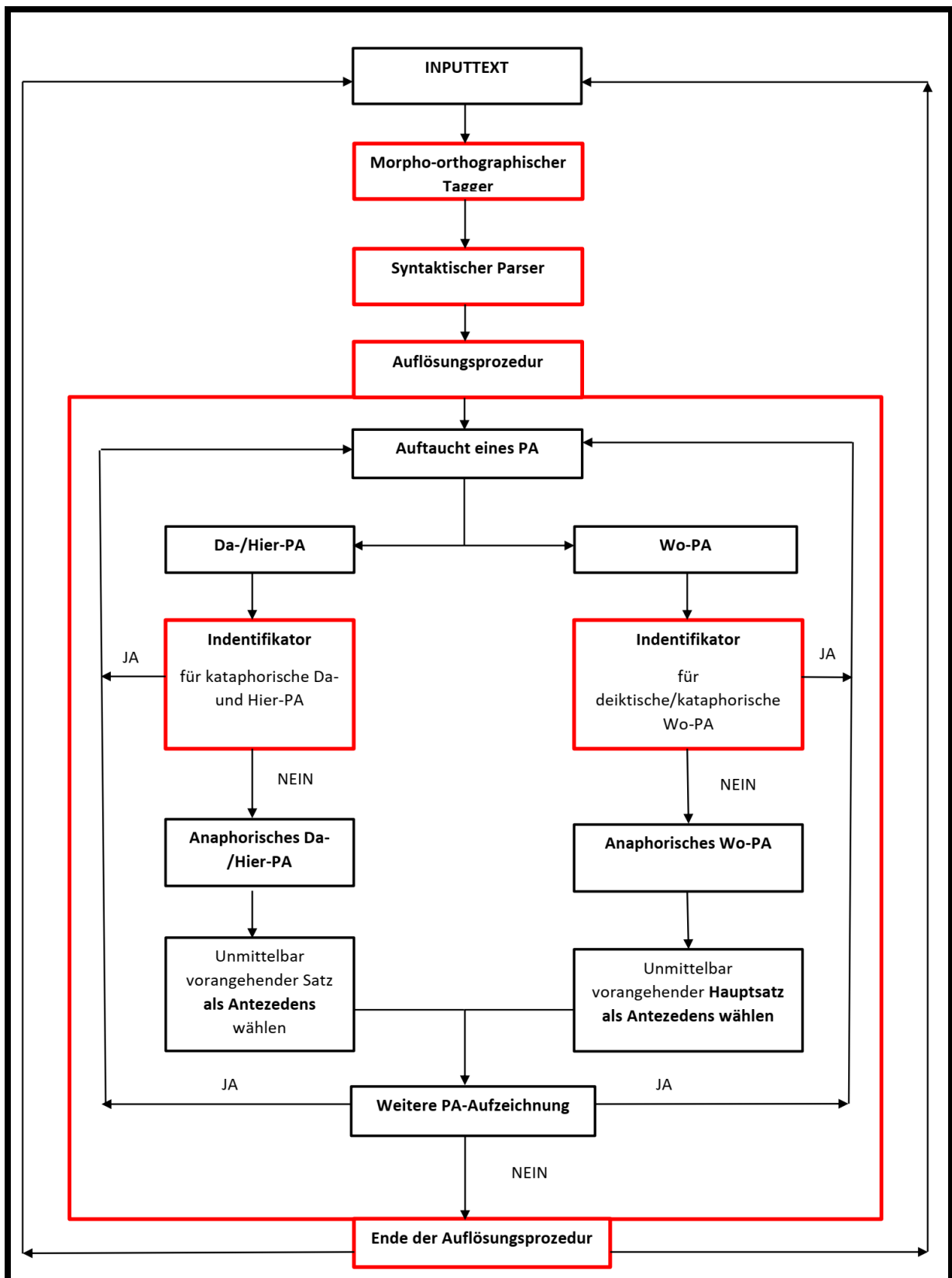


Abbildung 12: Endgültiger Algorithmus zur Auflösung von Pronominaladverbienanaphern

Im Folgenden stehen einige Informationen über das Funktionieren von im rot verzeichneten Zwischenmodulen, wie etwa dem morpho-orthographischen Tagger; dem syntaktischen Parser; dem Filter für kataphorische da- und hier-PA und dem Filter für deiktische und kataphorische wo-PA.

Morpho-orthographischer Tagger –Aktionen ²⁰³
• Abschnittsanfänge taggen (=Kein Token vor dem ersten Token) • Satzanfänge taggen (= Erscheinung eines Tokens nach einem Punkt) • Satzende markieren (= Erscheinung eines Punkts gefolgt durch nichts) • Subjunktionen (Dass, ob &Wie) annotieren • Fragewörter annotieren • Pronominaladverbien (PA) annotieren • Orthographiezeichen (wie Punkte; Doppelpunkte und Komma) taggen

Tabelle 14: Morpho-orthographischer Tagger

Syntaktischer Parser-Aktionen
• Hauptsätze markieren • Nebensätzen markieren

Tabelle 15: Syntaktischer Parser

Identifikator²⁰⁴/Filter für kataphorische Da- und Hier-PA

²⁰³ Informationen, die nach dem Taggen des Inputtextes für die konkrete Auflösungsprozedur zur Verfügung gestellt werden.

²⁰⁴ In Anlehnung an die im Abschnitt 3.3.3.3.1. entworfenen syntaktischen Regeln haben wir diesen Filter zur Sortierung von kataphorischen da- und hier-PA entwickeln.

- *PA+ Maximum 3Token +Komma+ Dass-/ob-/W-Satz*
[Wer (Wen, wem, wessen), wo (wohin, woher), was, wie
(Wie lange, wie oft, wie viel); Warum, weshalb, wieso,
weswegen, wozu, welche.../
- *PA+ Maximum 1Token+ Komma*
- *PA+ Maximum 1Token+ Doppelpunkt*

Tabelle 16: Filter zur Identifizierung von kataphorischen Da- und Hier-PA

Identifikator ²⁰⁵ /Filter für deiktische/kataphorische Wo-PA
• Abschnittsanfang+ PA • Satzanfang + PA

Tabelle 17: Filter zur Identifizierung von deiktischen/kataphorischen Wo-PA

²⁰⁵ In Anlehnung an die im Abschnitt 3.3.3.2.2. berichteten Informationen haben wir diesen Filter zur Sortierung von deiktischen und kataphorischen wo-PA entwickeln.

**DRITTER TEIL: EMPIRISCHER TEIL (HERSTELLUNG; EINSETZUNG UND
AUSWERTUNG DES PAAR/TOOLS)**

KAPITEL 5 Umwandlung des AR-Algorithmus ins AR-System: Entstehung des Pronominaladverbienanaphernreolvers (PAAR)

Die maximale Zielsetzung dieses Kapitels ist es, den Umwandlungsprozess näher wie möglich zu beschreiben. Doch zuvor werden wir, in einem ersten Schritt, weitere Teilmodule des AR-Algorithmus ausbauen müssen. Wir werden in der Tat Algorithmen für die beiden Kataphernfiltern erstellen und im Anschluss daran ein paar praktische Auskünfte über die Auswahl von dem benötigten syntaktischen Parser sowie über den Aufbau von dem gewünschten morpho-orthographischen Tagger geben. Weiterhin werden wir, in einem zweiten Schritt, unsere Kollaboration mit dem Informatiker (Programmierer) ausführlich beschreiben, der für die Umwandlung des Algorithmus ins System (durch den Gebrauch einer Programmiersprache) zuständig war.

5.1. Algorithmen zur Ausschließung von kataphorischen/deiktischen PA

5.1.1. Der Filter für kataphorische hier- und da-PA

Von den Angaben der Tabelle 14 ausgehend, welche sich ihrerseits auf Sprachregularitäten stützen, die während der im Kapitel 3 durchgeführte Betrachtung gewonnen worden sind, wollen wir den folgenden Algorithmus zur Sortierung von kataphorischen hier- und da-PA entwerfen.

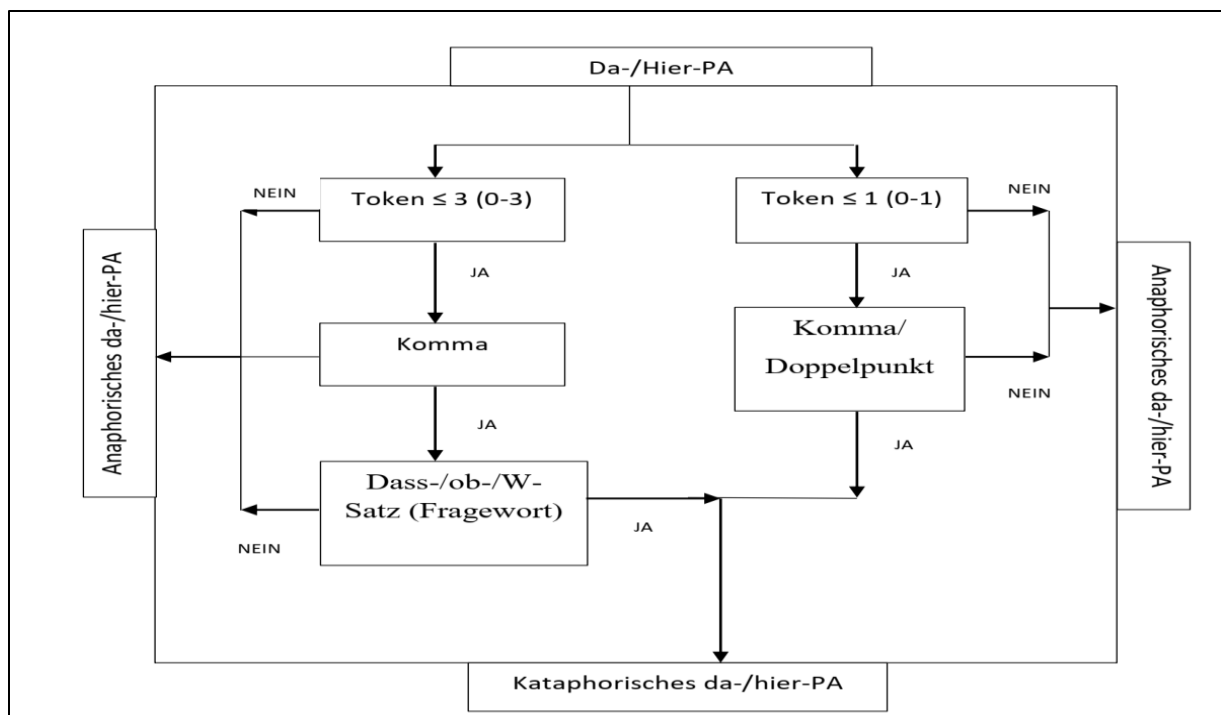


Abbildung 13: Algorithmen zur Filterung von kataphorischen da-/hier-PA

5.1.2. Der Filter für kataphorische/deiktische wo-PA

In Anlehnung an die in der Tabelle 15 enthaltenen Informationen werden wir zwei Prototypen von Algorithmen zur Ausschließung von kataphorischen/deiktischen wo-PA entwerfen. Und erst während des Umwandlungsprozess werden wir, in Berücksichtigung der ganzen Architektur des AR-System, den bestgeeigneten Prototyp auswählen.

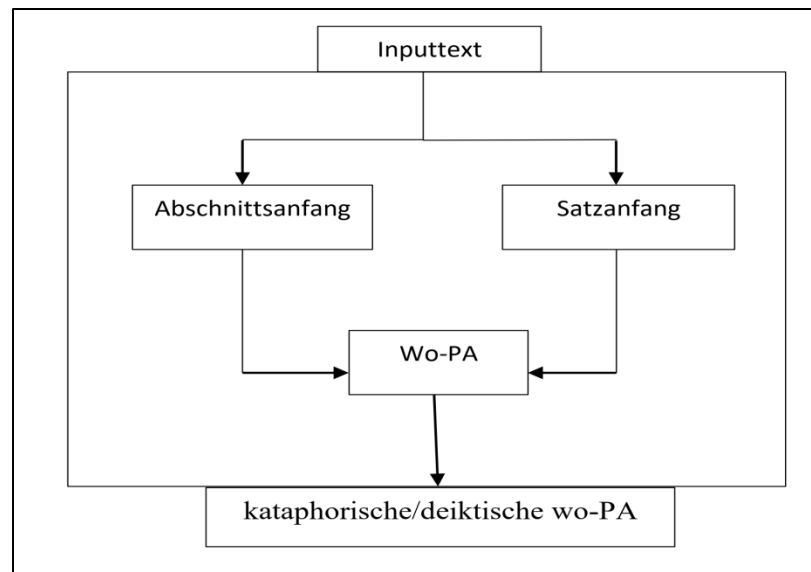


Abbildung 14: Algorithmen zur Filterung von deiktischen und kataphorischen wo-PA

(Prototyp 1)

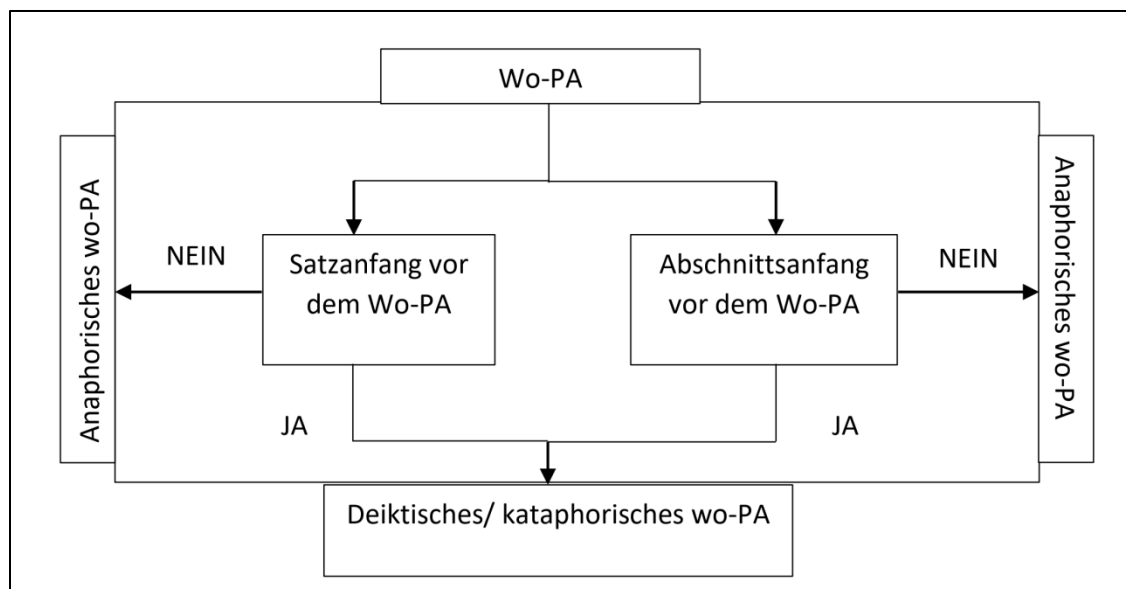


Abbildung 15: Algorithmen zur Filterung von deiktischen und kataphorischen wo-PA

(Prototyp 2)

5.2. Angaben über die Auswahl des Parsers

Wie schon in den vorherigen Kapiteln zuvor hingewiesen, wird mit dem Begriff *Parser*, ein informatisches Programm bezeichnet, das die syntaktische Struktur einer Zeichenkette analysiert²⁰⁶. Berücksichtigt wird nämlich bei der Analyse sowohl die Konstituente eines Satzes (Subjekt, Objekte und Ergänzungen) als auch die Hauptstrukturen eines Satzes (Haupt- & Nebensatz). Genau dieser letztere Aspekt wird den Vorrang bei der Wahl des für uns benötigten Parsers haben. Es gibt in der Tat in der Forschungsliteratur verschiedene computerlinguistischen Ressourcen, darunter zählen wir sowohl (syntaktische) Parser; (morphologische) Tagger; Compiler als auch Textkorpora²⁰⁷. Diese Ressourcen sind im Web abrufbar und können nach Registrierung von einer Plattform heruntergeladen werden.

In dieser Hinsicht bleibt es uns nur die für uns am bestens passende tools (Parser) von den bestehenden auszuwählen. Allerdings können wir zurzeit diese Wahl noch nicht freilich machen und zwar, weil diese Tools meistens mit einer Programmiersprache (JAVA; Python; C++....) aufgefasst und programmiert worden sind. Und da wir zurzeit noch nicht wissen mit welcher dieser Sprachen der Informatiker die Anwendung (unser AR-System) entwickeln wird, wollen wir uns hier mit der Beschreibung des Parsers aufhören und erst später weitergehen. Also erst nachdem das AR-System im Ganzen komplett aufgebaut wird.

5.3. Angaben über den Ausbau eines geeigneten morpho-orthographischen Taggers

Die gleichen im vorangehenden Abschnitt geführten Überlegungen gelten auch hier. In der Forschungsliteratur bzw. in den computerlinguistischen Ressourcendatenbanken sind verschiedene Tagger zu finden. Außer dem Prinzip der Kompatibilität zwischen dem ganzen Modul (Das AR-System) und dem Teil-Modul (Tagger) bezüglich der für ihre jeweilige Entwicklung benutzten Programmiersprache, wird bei unserer Wahl eines geeigneten Taggers auch ein weiteres Prinzip berücksichtigt. Für unser AR-System benötigen wir in der Tat keinen „klassischen/normalen“ Tagger (Annotation von Wörtern nach ihrer grammatischen Natur), sondern einen der sowohl die grammatische Natur von Wörtern als auch die für den Rest der Auflösungsprozedur benötigten Informationen annotiert/markiert. Dies ist Beispielweise der

²⁰⁶*In der Computertechnik ist ein Parser ein Programm, das Teil des Compilers ist, und das Parsing ist Teil des Compiler-Prozesses. Das Parsen erfolgt während der Analysephase der Kompilierung.*
<https://www.computerweekly.com/de/definition/Parser#:~:text=Ein%20Parser%20ist%20ein%20Programm,und ere%20Software%20ihn%20verstehen%20kann> (27.05.2023 um 21h)

²⁰⁷Vgl.: Ebd.

Fall von der Markierung von Satz- oder Abschnittsanfängen, die während dem Einsatz des Moduls TAGGER erfolgen und erst während dem Einsatz des Moduls IDENTIFIKATOR für Wo-PA abgerufen/benötigt werden. Und da solche besondere Tagger in den vorhandenen computerlinguistischen Ressourcendatenbanken noch nicht vorhanden sind, werden wir im Rahmen unserer Kollaboration mit dem Informatiker einen entsprechenden Tagger entwickeln müssen (Entstehung eines weiteren computerlinguistischen Tools)²⁰⁸

5.4. Angaben über die Kollaboration mit dem Programmierer: Entstehung des PAARS

Der Informatiker, der sich um die Umwandlung des AR-Algorithmus ins AR-System gekümmert hat, heißt Alex Hermann SOCKENG. Er ist ein Absolvent der polytechnischen Hochschule von der Universität von Jaunde 1. Er absolvierte nämlich eine Ausbildung **als Computeringenieur ("computational engineer")**. Dabei spezialisierte er sich insbesondere für die Modellierung von mathematischen Gegebenheiten. Mehr über seine Biographie und Bibliographie im Anhang. Die Kollaboration mit ihm findet hauptsächlich in einem Internetcafé namens 4G statt. Dieses Café befindet sich an der Cité-U (Rue des pavés- Carrefour Orly) und ist Campusweit für ihre Fähigkeit bekannt, den Studenten einen Internetzugang von guter und preiswerter Qualität bereitzustellen. Dort trafen wir uns 2 bis 3 Mal pro Woche, wobei die Arbeitssitzungen zwischen 3 bis 4 Viertelstunden dauerten.

Die erste Sitzung fand nämlich am 30.05.2023 statt. Dabei ging es nämlich um eine grobe Darstellung des Forschungsprojekts. Wir stellten dem Computeringenieur nämlich klar vor, was wir von ihm, erwarteten. Am Ende der Sitzung sagte er, er habe das Projekt verstanden, aber er müsse dennoch nachsehen, ob er uns dabei helfen könnte. Drei Tagen später trafen wir uns wieder und er bestätigte uns, dass er in der entsprechenden Forschungsliteratur nachgesehen habe, und dass er tatsächlich im Projekt mitmachen könne. Entscheiden war nämlich, ihm nach, die Tatsache, dass unser Forschungsprojekt im Forschungsbereich **der NLP (Natural Language Processing)** zu verorten ist, und dass es für Arbeiten in diesem Forschungsgebiet schon vorhandene **Methode**, also vordefinierte **Tools**²⁰⁹ gibt, welche dabei helfen könnten, das angestrebte AR-System zu entwerfen.

²⁰⁸ Außer dem AR-System (Tool), das wir zu entwerfen beabsichtigen, werden wir auch einen eigenen morpho-orthographischen Tagger (tool) entwickeln.

²⁰⁹ Hier referiert er nämlich auf vorhandene Parser und Tagger

In einer weiteren Arbeitssitzung fassen wir zusammen anhand **einem TDR²¹⁰ (Termes de références)** weitere Modalitäten der Kollaboration (wie die Dauer und die benötigten Ressourcen des Projekts). Siehe dazu das TDR im Anhang. Weiterhin klärten wir noch einiges über das Funktionieren des AR-Algorithmus. Besprochen wurde nämlich die Rolle von jedem Zwischenmodul, wie etwa dem Parser und dem Tagger, wobei anschließend über die Implementierung der beiden Kataphernfiltern (weitere Zwischenmodule) diskutiert wurde. Und sobald alle zwischen Modulen erfolgreich umgewandelt worden waren, fingen wir an das ganze System auf Beine zu stellen. Dafür entschied er sich nämlich für die Programmiersprache namens **PYTHON**. Denn diese Programmiersprache eignet sich, ihm nach, am bestens für Arbeiten im Bereich der NLP (Natural Language Processing).

Wie alle Programmiersprache, wird **PYTHON** nur in einer Bestimmten Entwicklungsumgebung (**IDE: integrierte Open-Source-Entwicklungsumgebung**) verwendet/manipuliert. Das heißt, das Schreiben eines informatischen Codes zur Automatisierung einer gegebenen Handlung kann nur in der dafür entwickelten Programmierungsumgebung stattfinden. Diese Programmierungsumgebung wird, je nach Programmiersprache, durch entsprechende Software aufgerufen. Und für den Fall der Programmiersprache PYTHON existieren 14 Softwares, die als Entwicklungsumgebung dienen können. Allerdings weist jede dieser Softwares besondere Spezifitäten auf. Während einige sich für heuristische Ansätze am bestens einsetzen lassen, lassen sich andere eher im Rahmen von Machinelearning-Projekten aufrufen. Allerdings lassen sich all diese 14 Software in einer Übersoftware namens **Anaconda-Navigator²¹¹** auffinden. Das heißt, sobald man die Software

²¹⁰L'expression **Termes de références (Tdr)** fait référence à un cadre qui permet de décrire l'objet, la portée et le processus de réalisation (Gestion et aspects techniques) d'un projet. Le TDR a donc pour objectif de présenter de manière claire et brève ce qui est attendu à la fin d'un projet. Dans notre cas par exemple, il aura pour tâche de présenter ce qui est attendu à la fin de notre collaboration avec le « computational engineer ». Comme Rubriques principales d'un TDR, on distingue : *Le Contexte de la recherche/Projet ; les acteurs y participant ; les outils et méthodes réquisitionnés ; la nature et sources des données exploitées* <https://www.primetraining.global/termes-de-reference-lecons-sur-leur-structure-et-linterpretation-des-resultats/> (15.06.2023 um 9h16)

²¹¹ **Anaconda** ist tatsächlich eine Software, die zur Kategorie der **Distribution** gehört. Mehr dazu auf dem folgenden Link ([https://de.m.wikipedia.org/wiki/Anaconda_\(Python-Distribution\)](https://de.m.wikipedia.org/wiki/Anaconda_(Python-Distribution))) (13.06.2023 um 01h30)). Sie wurde nämlich für die Programmiersprachen **Python** und **R** entwickelt und enthält die folgenden Softwares:

- **DataSpell** : *It is an IDE for exploratory data analysis and prototyping machine learning models. It combines the interactivity of Jupyter notebooks with the intelligent Python and R coding assistance of PyCharm in one user-friendly environment.* (Vgl. Die Startseite vom Anaconda Navigator)
- **Jupyter Notebook**: *It is a web based, interactive computing notebook environment. It edits and runs human readable docs while describing the data analysis.* (Vgl. Ebd.)
- **JupyterLab**: *It is an extensible environment for interactive and reproducible computing, based on the Jupyter Notebook and Architecture.* (Vgl. Ebd.)

Anaconda-Navigator heruntergeladen hat, hat man Zugang zu den 14 Python Programmierungsumgebungen.

Der Computeringenieur entschied sich für die Entwicklungsumgebung namens **SPYDER** und fing an, das ganze System zu erstellen also, den dafür geeigneten Code zu schreiben. Und sobald dies gemacht wurde, testeten wir anhand konkreter Beispiele, ob der Code fehlerfrei funktioniert, also ob der Code beim Hochfahren/ bei der Exekution tatsächlich realisiert, was von uns erwartet ist. Zwar mussten wir Einiges jedes Mal verbessern aber nach mehrmaligen Probetesten und Verbesserungen gelang uns letztendlich was auf Beine zu stellen, das tatsächlich exekutiert, was wir uns zuvor erwünscht hatten. Der Programmierer schickte uns den Code per E-Mail. Und nach der Installation auf unserem Rechner von **ANACONDA** und **SPYDER** konnten wir tatsächlich den Code (Das AR-System) benutzen, um die Daten auszuwerten. Das folgende Schema gibt weitere und nähere Angaben über die gesamte Architektur des PAAR

-
- **Spyder : Scientific Python Development environnement** is a Powerfull Python IDE with advanced editing, interactive testing, debugging and introspection futures (Vgl. Ebd.) **Spyder** ist eine plattformübergreifende, **integrierte Open-Source-Entwicklungsumgebung (IDE)** für die wissenschaftliche Programmierung in der Programmiersprache Python. Spyder lässt sich mit einer Reihe von Paketen des wissenschaftlichen Python-Stacks integrieren, darunter NumPy, SciPy, Matplotlib, Pandas, IPython, SymPy und Cython, sowie mit anderer Open-Source-Software. Es wird unter der MIT-Lizenz veröffentlicht und ist Teil der Anaconda-Distribution www.spyder-ide.org [https://de.m.wikipedia.org/wiki/Spyder \(Software\)](https://de.m.wikipedia.org/wiki/Spyder_(Software)) (13.06.2023 um 01h30)
 - **Datalore:** Kick-start your data science projects in a pre-configured environment. Enjoy coding assistance for Python, SQL and R in Jupyter Notebook and benefit from no-code automations. Use Datalore online for free. (Vgl. Ebd.)
 - **IBM Watson Studio cloud:** It provides you the tools to analyze and visualize data to cleanse and shape data, to create and train machine learning models. Prepare data and build models using open source data science tools or visual modeling. (Vgl. Ebd.)
 - **Oracle data science service:** It offers a machine learning platform to build, train, manage and deploy your machine learning models on the cloud with your favorite open-source tools (Vgl. Ebd.)
 - **Glueviz:** It is a multidimensional data visualization across files. Explore relationships within and among related datasets (Vgl. Ebd.)
 - **Orange 3:** Component based data mining framework. Data visualization and data analysis for novice and expert. Interactive workflows with a large toolbox. (Vgl. Ebd.)
 - **PyCharm Professionnel:** it is a full-fledged IDE by JetBrains for both Scientific and web Python Development. Support HTML, JS and SQL (Vgl. Ebd.)
 - **RStudio:** a set of integrated tools designed to help you be more productive with R includes R essentials and notebooks (Vgl. Ebd.)
 - **CMD.exe Prompt:** It runs a cmd terminal with your current environnement from Navigator active (Vgl. Ebd.)
 - **Powershell Prompt :** It runs a Powershell terminal with your current environnement from Navigator active (Vgl. Ebd.)
 - **Qt Console :** It support inline figures, propre multiline editing with syntax highlighting, graphical, calltips and more (Vgl. Ebd.)

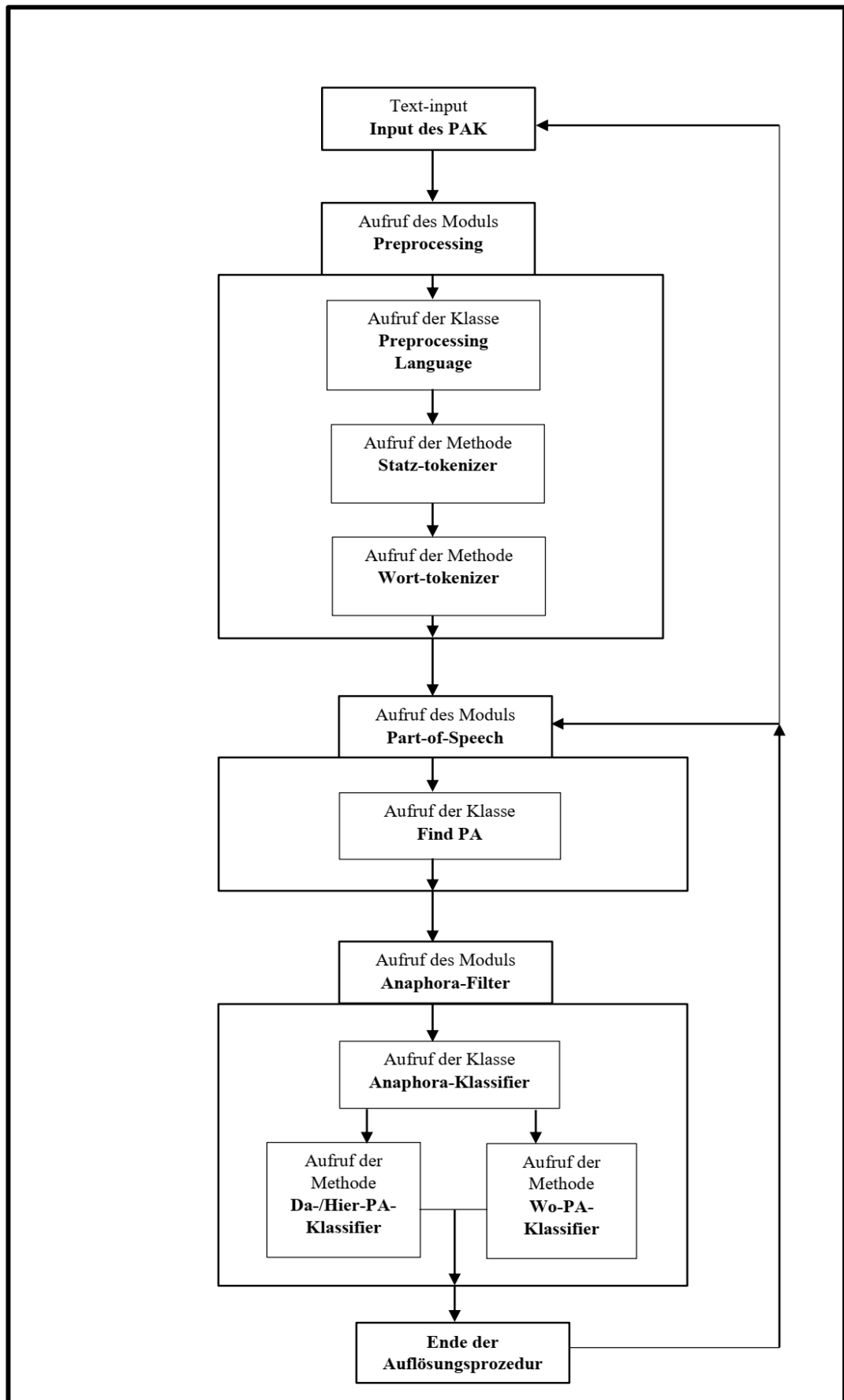


Abbildung 16: Globale Architektur des PAAR

In folgenden Abschnitten und im darauffolgenden Kapitel stellen wir in Einzelheiten den Einsatz des **PAAR** auf ein von uns konstituiertes Textkorpus, namens **PAK** (Pronominaladverbien Korpus) vor. Aber zuvor geben wir einige Angaben über den Aufbau des für die Untersuchung gebrauchten Textkorpus.

5.5. Aufbereitung der Datensätze und Einsatz des PAAR

Dieser Abschnitt gibt praktische Auskünfte über den Aufbau des für die Untersuchung (Einsatz des PAAR) benötigten Textkorpus. In der Tat werden wir aus verschiedenen Textdatenbanken (computerlinguistischen Ressourcen) Texte auswählen und sie zusammenstellen, um unser Textkorpus zu konstituieren.

5.5.1. Aufbau eines Textkorpus

Mit dem Begriff **Textkorpus** (plr. Textkorpora) wird in der Korpuslinguistik eine Sammlung von Texten einer bestimmten Sprache bezeichnet. Diese Sammlung kann sowohl aus Texten unterschiedlicher Stämmen (Gattungen, Autoren...) bestehen als auch *von textlich aufgezeichneten mündlichen Äußerungen* konstituiert werden²¹². (Hagen Hirschmann 2019). Textkorpora werden meistens konstituiert, um einer Forschungsfrage nachzugehen. In dieser Hinsicht unterscheidet man in der Forschungsliteratur zu deutschen Korpora verschiedene Arten von Textkorpora-Datenbanken. Diese Letzte werden dann eben durch Forschungseinrichtungen (wie universitäre Instituten); Forschungsgesellschaften oder -gemeinschaften (wie die Gemeinschaft von Computerlinguisten) konstituiert. In der deutschen Forschungswelt beispielsweise unterscheidet man einen Haufen solcher Datenbanken. Man könnte sie, wenn es sich exklusiv um linguistische Forschungsfrage handelte, entweder als synchronische oder diachronische Datenbanken klassifizieren. Wenn es sich aber um Untersuchungen ganz anderer Phänomene, deren Betrachtung aber textlich aufgefasste

²¹² Dies ist beispielweise der Fall für die Textkorpora, die sich aus der **Datenbank für Gesprochenes Deutsch (DGD)** herunterladen lassen. Die DGD ist eine der Datenbank, die **vom Leibniz-Institut für Deutsche Sprache (IDS)** entwickelt worden ist (<https://dgd.ids-mannheim.de>). Der Zugang zu dieser Datenbank ist zwar kostenlos aber muss über eine Registrierungsanfrage ermöglicht werden. Diese Letzte wird gründlich von einem DGD-Team geprüft und erst genehmigt, wenn die vom Antragsteller angegebenen Ziele tatsächlich eine Relevanz zu den Forschungsschwerpunkten der IDS finden. Im Anhang findet man das E-Mail-Gespräch, das wir mit dem DGD-Team geführt haben und am dessen Ende uns den Zugang zu dem DGD freigeschaltet worden ist.

Datensätze benötigt, werden sie, die Textkorpora-Datenbanken, nach thematischen Kriterien klassifiziert.

Über diesen Klassifikationsaspekt hinausgehend, lässt sich auch merken, dass die Datenbanken in elektronischer Form vom Web abzurufen sind. Allerdings werden, je nach gewünschten Korpora, Zugangseinschränkungen beobachtet. In der Regel genügt eine kostenlose Registrierung aber manchmal ist der Zugang zu gegebenen Textkorpora-Datenbanken von einer Zugehörigkeit zu der Institution (oder zu einer anderen, der schon eine Gebrauchslizenz abgekauft hat) bedingt, der die Datenbank geschöpft hat. Dies ist beispielweise der Fall von **dem Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z)**²¹³, worauf wir gern die Performanz unseres AR-Systems geprüft hätten. Im Folgenden beschreiben wir ein paar Datenbanken, die sich für synchronische Studie eignen und im Anschluss daran werden wir eine davon auswählen, aus der wir Texte zum Aufbau unseres eigenen Textkorpus entnehmen werden.

5.5.1.1. Bestehende Textkorpora für korpuslinguistische Untersuchungen im Deutschen

Das digitale Wörterbuch der deutschen Sprache (DWDS) ist eine Datenbank des Deutschen, die von **der Berlin-Brandenburgischen Akademie der Wissenschaften** konstituiert worden ist. Diese elektronische Datenbank erfüllt sowohl die Funktion eines Wörterbuchs als auch die einer Quelle für Textkorpora. Man wird in der Tat durch den Zugriff auf diese Datenbank nicht nur in der Lage sein, gesuchte Wörter/Begriffe semantisch zu klären, sondern auch einen Zugang zu einer großen Sammlung verschiedener elektronischer Textkorpora haben, die nach linguistischen *Erfassungskriterien* annotiert sind.²¹⁴ Eine weitere Textkorpora-Datenbank, die auch die Funktion eines Wörterbuchs erfüllt, ist das **Projekt Deutscher Wortschatz (PDW)**. PDW ist eine elektronische Datenbank, die **von der Universität Leipzig** aufgebaut worden ist und somit zur **Leipzig Corpus Collection** gehört. Sie umfasst 35 Millionen Sätze mit 500

²¹³ **Die Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z)** wurde von der Sektion Allgemeine Sprachwissenschaft & Computerlinguistik der Universität Tübingen entwickelt und beinhaltet in dem hier verwendeten Release 8 (Beck 2013) etwa **75.000 morphosyntaktisch sowie text- und diskurssemantisch** annotierte Sätze aus der deutschen Zeitung „die tageszeitung“, was ca. 1,4 Mio. Tokens bzw. 3.000 Texten entspricht. Mit mehr als **100.000 Anaphorik- und Koreferenzrelationen** bietet die TüBa-D/Z das zurzeit umfangreichste deutschsprachige Korpus mit derartigen **Annotationsebenen** (Vgl. Ludger Paschen 2013: 58).

²¹⁴ Das **DWDS-korpus** ist durch den folgenden Link abrufbar: <http://www.dwds.de/resource/kerncorpus/> (23.05.2023).

Millionen Wörtern und ist durch den folgenden Link abrufbar <http://wortschatz.uni-leipzig.de/> (23.05.2023). Ebenso wie Korpora der DWDS-Datenbank, ist die **Hamburg Dependency Treebank** auch ein Korpus das nach linguistischen Merkmalen annotiert ist. Sie ist in der Tat das größte Deutschkorpus, das mit **Satzstrukturanalysen** annotiert worden ist²¹⁵. Eine weitere Datenbank, die, ebenso wie die DWDS-Datenbank, den Zugang zu einer großen Sammlung von Textkorpora ermöglicht, ist die **IDS-Textkorpora-Datenbank**. Wie bereits oben eingeführt, ist die IDS-Korpora vom **Leibniz-Institut für Deutsche Sprache** (IDS) in Mannheim entwickelt worden und enthält fast alle in elektronischer Form bestehenden Textkorpora der deutschen Sprache. Einige Korpora, die zu dieser Liste von IDS-Korpora gehören sind: **Das Deutsche Referenzkorpus (DeReKo)**²¹⁶; **das DeWaC-Corpus**²¹⁷ und **das LIMAS-Korpus**²¹⁸. Die Tatsache, dass die IDS-Datenbank aus verschiedenen Textkorpora besteht, ermöglicht einem ein repräsentatives Textkorpus zur Eruiierung einer Forschungsfrage aufzubauen. IDS-Korpora sind durch den folgenden Link herunterzuladen <http://www.ids-mannheim.de/kt/corpora.html> (23.05.2023). Aber da es sich um eine Datenbank handelt, die über 100 Gb groß ist, wurde ein Navigationswerkzeug entwickelt, das ermöglicht, an alle Texte ranzukommen, ohne sie zuvor in ihrer Ganzheit herunterladen zu müssen. Das Navigationstools heißt **COSMAS II**²¹⁹ und ist nach einer Registrierung kostenlos verwendbar.

Die gerade oben zitierten Textdatenbanken eignen sich am bestens für synchronische Studien. Handelt es sich aber um diachronische Forschungsfragen, dann werden eher historische Texte vorwiegend als gegenwärtige Texte präferiert. **Zu diachronischen Textkorpora-Datenbanken**, zählen wir beispielweise *das Kali Korpus*²²⁰; das Textkorpus von **Thomas**

²¹⁵ Die Texte des Korpus stammen aus der deutschen Nachrichtenseite heise.de (1996 – 2001) und sind unter dem folgenden Link erhältlich <http://hdl.handle.net/11022/0000-0000-7FC7-2> (23.05.2023).

²¹⁶ **DeReKo: Deutsches Referenzkorpus:** *Das Deutsche Referenzkorpus ist die weltweit größte Sammlung elektronisch verfügbarer, deutschsprachiger Texte*

²¹⁷ **Das DeWaC-Corpus** ist ein Textkorpus, das aus deutschsprachigen Web-Texten besteht und ist unter dem folgenden Link in seiner Ganzheit herunterzuladen. <https://www.sketchengine.eu/dewac-german-corpus/> (23.05.2023).

²¹⁸ **Das LIMAS-Korpus** ist ein repräsentatives Zeitschnittkorpus der deutschen Gegenwartssprache, das aus 500 Texten verschiedener Textsorten besteht. <http://www.korpora.org/Limas/> (23.05.2023).

²¹⁹ **COSMAS II** steht für *Corpus Search, Management and Analysis System* und ist das Nachfolgesystem von **COSMAS I** (1991-2003) am IDS. Sie ist eine am IDS konzipierte Volltextdatenbank für das linguistisch motivierte Recherchieren in den Textsammlungen (= Korpora) des IDS. <https://www2.ids-mannheim.de/cosmas2/projekt/einsteiger/thema-druck.html?template=/cosmas2/template/print.tpl#was> (09.06.2023 um 06h)

²²⁰ **Das Kali-Korpus** (Kali: "Korpusarbeit Linguistik") ist ein teilannotiertes diachrones Korpus des Deutschen für Forschung und Lehre, das seit Ende 2003 am Deutschen Seminar der Universität Hannover unter der Leitung von Prof. Dr. Gabriele Diewald aufgebaut wird. www.kali.uni-hannover.de (25.05.2023).

Gloning²²¹; das **Bochumer Mittelhochdeutsch Korpus**²²²; die **mittelhochdeutsche Begriffsdatenbank** ²²³(**MHDBDB**); das **Referenzkorpus Altdeutsch : Das Referenzkorpus Altdeutsch; Das Korpus Deutsch Diachron Digital**²²⁴ **DTA: Deutsches Textarchiv**²²⁵.

Ein weiteres Differenzierungsmerkmal, auf welchem Basis wir die bestehenden Textkorpora-Datenbanken differenzieren können, ist die ANNOTATION bzw. das linguistische Phänomen, das in einem gegebenen Text annotiert worden ist. Handelt es sich beispielsweise um eine Untersuchung von **morphologischen** (Wortart) **syntaktischen** (Satzteile) oder noch **textlinguistischen** (Koreferenz/ Anaphorik) Gegebenheiten dann wird der Text entsprechend annotiert. <https://blogs.uni-paderborn.de/lingutools/oeffentliche-korpora/> (23.05.2023). In dieser Hinsicht muss man bei der Auswahl eines Textkorpus für die Führung einer gegebenen Untersuchung nicht nur die Art und Weise, mit der das Textkorpus annotiert worden ist, im Kauf nehmen, sondern auch das Tool (Annotationsprogramme /computerlinguistische Ressourcen) berücksichtigen, womit die Annotation durchgeführt worden ist²²⁶.

Aus all diesen bereits erwähnten Einschränkungen entschieden wir uns zur Erreichung der Zielsetzung unserer Forschung, ein passendes Textkorpus aufzubauen. Und da wir eine synchronische Studie durchführen, wählten wir eine Datenbank, die sich für entsprechende Studie eignet. Wir entschieden uns fürs **Deutsche Referenzkorpus (DeReKo)** (die weltweit größte Sammlung elektronisch verfügbarer, deutschsprachiger Texte). Dieses Korpus befindet sich nämlich in der **IDS-Korpora-Datenbank** vom **Leibniz-Institut für Deutsche Sprache** (IDS) und ist mit Hilfe des **Corpus-Browser COSMAS II** abrufbar. Die folgende Abbildung gibt ergänzende Informationen über die wechselseitigen Beziehungen, die zwischen der IDS-

²²¹ Das Textkorpus **Thomas Gloning** ist eine Schöpfung der Universität von Gießen und ist unter dem folgenden Link kostenlos herunterzuladen <http://www.uni-giessen.de/gloning/etexte.htm> (25.05.2023).

²²² **Das Bochumer Mittelhochdeutsch Korpus** ist eine Schöpfung der Ruhr-Universität Bochum und ist unter dem folgenden Link abrufbar http://www.ruhr-uni-bochum.de/wegera/archiv_1.htm (25.05.2023).

²²³ Sie enthält 4.7 Millionen Wörter und ist vom folgenden Link abzurufen <http://mhdbdb.sbg.ac.at> (25.05.2023).

²²⁴ **Korpus Deutsch Diachron Digital** enthält Korpustexte aus verschiedenen Sprachepochen des Deutschen, darunter dem Althochdeutschen (bis 1050), dem Mittelhochdeutschen (1050-1350) und dem Frühneuhochdeutschen (1350-1650). <https://www.sprache-spiel-natur.de/2020/06/14/textsammlungen-und-datenbanken-zur-verwendung-des-deutschen/> (Abgerufen am 30.05.2023 um 13h)

²²⁵ **Das Deutsche Textarchiv** stellt ein Referenzkorpus deutschsprachiger Texte aus dem Zeitraum zwischen 1600 und 1900 zur Verfügung. Es eignet sich besonders gut zur Untersuchung ganz unterschiedlicher **Sprachwandelprozesse** und bietet viele nützliche Funktionen zur statistischen Auswertung. <https://www.sprache-spiel-natur.de/2020/06/14/textsammlungen-und-datenbanken-zur-verwendung-des-deutschen/> (Abgerufen am 30.05.2023 um 13h)

²²⁶ In der Tat werden Textkorpora von ihren jeweiligen Gründern, je nach dem gewünschten untersuchten Phänomen, annotiert. Und wie schon in der Vorphase dieser Forschungsarbeit hingewiesen, besteht in der Forschungsliteratur noch kein Textkorpus, das die Beziehungen (PAA-Antezedens) im Deutschen annotiert. Dieser Mangel erklärt sich nämlich durch die Tatsache, dass es kein AR-System zur Auflösung von Pronominaladverbienanaphern in der Forschungsliteratur gibt.

Textkorpora und dem **Deutschen Referenzkorpus (DeReKo)** bestehen, dem Textkorpus, aus dem wir Texten zur Konstituierung unseres eigenen Textkorpus auswählen werden.

© **Leibniz-Institut für Deutsche Sprache, Mannheim**

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Mittwoch, den 31. Mai 2023, 9:21:41
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Daran*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.763.793 Treffer

Abbildung 17: Archivierungsnomenklatur (Wechselseitige Beziehungen zwischen IDS-Korpora; DeReKo und COSMAS II)

Vom **DeReKo** wählten wir in der Tat 154 Texte und wir sicherten uns bei der Auswahl, dass die 60-PA, die als Gegenstand unserer Untersuchung gelten, sich tatsächlich auch in der von uns konstituierten Textsammlung befinden. Außer logisch methodischen Prinzipien, die diese Vorgehensweise untermauern, lässt sich diese Herangehensweise auch durch das folgende Zitat bekräftigen.

Die Auswahl von Texten und ihre Zusammenstellung in einem Textkorpus ist ebenso wie die theoretische Verankerung des Analysevorhabens ein methodischer Vorgang, der in hohem Maße Einfluß auf die Ergebnisse hat. <https://www.wortbedeutung.info/Textkorpus/>
(Abgerufen am 23.5.2023)

Konkret berichtet, sind wir so verfahren: Nach einer Registrierung auf den COSMAS II-Server enthielten wir einen Login-ID, der uns den Zugang zu der IDS-Korpora-Datenbank öffnete. Einmal in der Datenbank stellten wir fest, dass es unglaubliche viele Korpora gab, die in verschiedenen Archivrubriken klassifiziert worden sind. Wir entschieden uns für die **W4-Arschiv der geschriebenen Sprache**. Weiterhin stellten wir fest, dass sich viele weitere Textkorpora innerhalb dieser ausgewählten Archivrubrik befinden. Darunter entschieden wir

uns für die *D-Korpora- Korpora aus Deutschland*²²⁷ und fingen gleich danach an, Textabschnitte auszuwählen²²⁸ und diese in ein anderes Dokument aufzulagern. So sind wir verfahren, um das Textkorpus zu konstituieren, dessen Hauptziel ist es, den Rahmen zur Auswertung unseres AR-System namens PAAR zu ermöglichen. Dieses konstituierte Textkorpus wird als PAK (Pronominaladverbien-Korpus/ PA-Korpus) bezeichnet und im Folgenden beschreiben wir den Einsatz des PAAR auf das PAK.

PAK	ANZAHL
Textabschnitte	154
PA	261

Tabelle 18: Konstituenten des PAK

PA	ANZAHL
Daran	4
Hieran	1
Woran	3
Darauf	9
Hierauf	2
Worauf	6
Daraus	1
Hieraus	2
Woraus	3
Dabei	11
Hierbei	3
Wobei	13
Dadurch	6
Hierdurch	4
Wodurch	4
Dafür	6
Hierfür	3
Wofür	3

²²⁷ Wir stellten nämlich fest, dass Elemente aus dem D-Korpora eigentlich Elemente aus dem Deutschen Referenzkorpus (DeReKo) stammen.

²²⁸ Mehr über die Wahl der Textabschnitte in Anhang

Dagegen	6
Hiergegen	1
Wogegen	2
Dahinter	4
Hierhinter	1
Wohinter	3
Darin	4
Hierin	2
Worin	3
Darein	3
Hierein	3
Worein	5
Damit	20
Hiermit	1
Womit	6
Danach	7
Hiernach	2
Wonach	3
Daneben	1
Hierneben	2
Woneben	0
Darüber	3
Hierüber	1
Worüber	6
Darum	2
Hierum	1
Worum	6
Darunter	5
Hierunter	3
Worunter	5
Davon	16
Hiervon	1
Wovon	3

Davor	0
Hiervor	0
Wovor	1
Dazu	18
Hierzu	8
Wozu	7
Dazwischen	6
Hierzwischen	1
Wozwischen	5
TOTAL	261

Tabelle 19: Distribution der PA im PAK

5.5.2. Einsatz des PAAR aufs PAK (Pronominaladverbien-Korpus)

Den Einsatz des PAAR aufs PAK wird man auf zweierlei Weise realisieren. Für den ersten Text gilt eine gründliche Vorgehensweise. Das heißt der Einsatz des PAAR auf PAK (Text N°1) wird Modul für Modul dargestellt, wobei die jeweiligen Ergebnisse anschließend analysiert und interpretiert werden. Und nach dem Text N°1 werden wir den PAAR-Einsatz nicht mehr Modul für Modul darstellen, sondern eher direkt die nach dem Textinput bekommenden Ergebnisse analysieren und interpretieren.

- **EINSATZ DES PAAR AUF DEN TEXT N°1**

Die folgenden Abbildungen beschreiben das Hochfahren der Python-Entwicklungsumgebung **SPYDER**. Dafür muss man aber, wie gesagt, den **Anaconda Navigator** zuerst hochfahren.

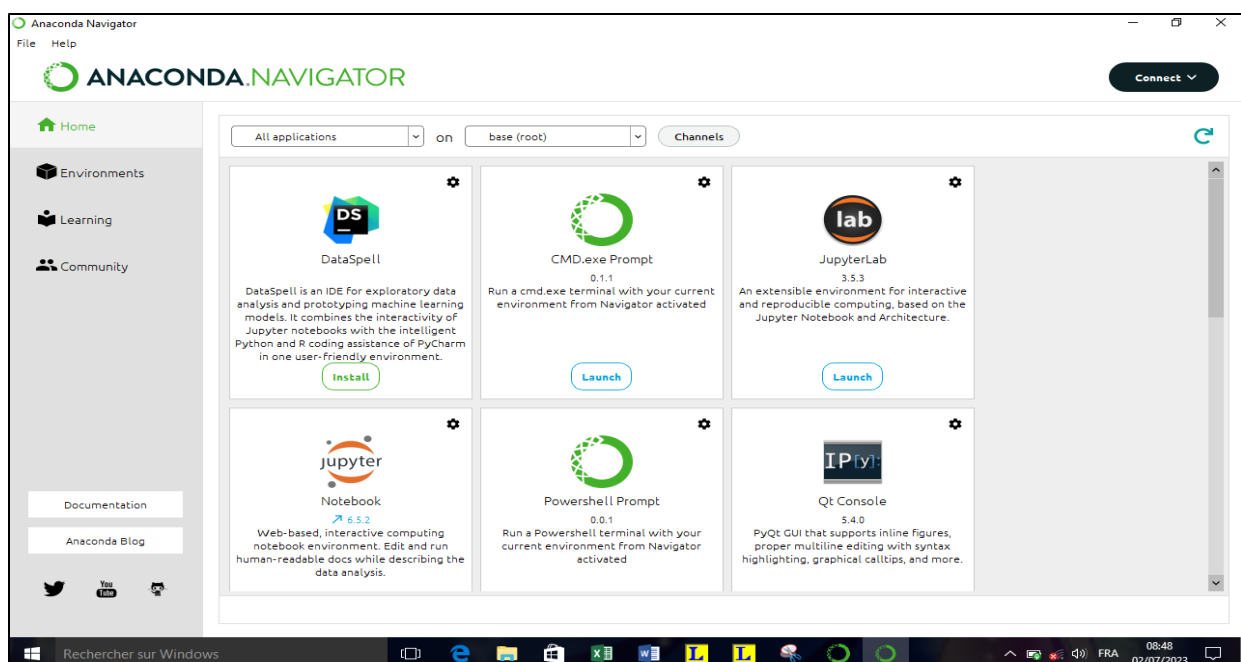


Abbildung 18: Startseite der Software ANACONDA NAVIGATOR (1/3)

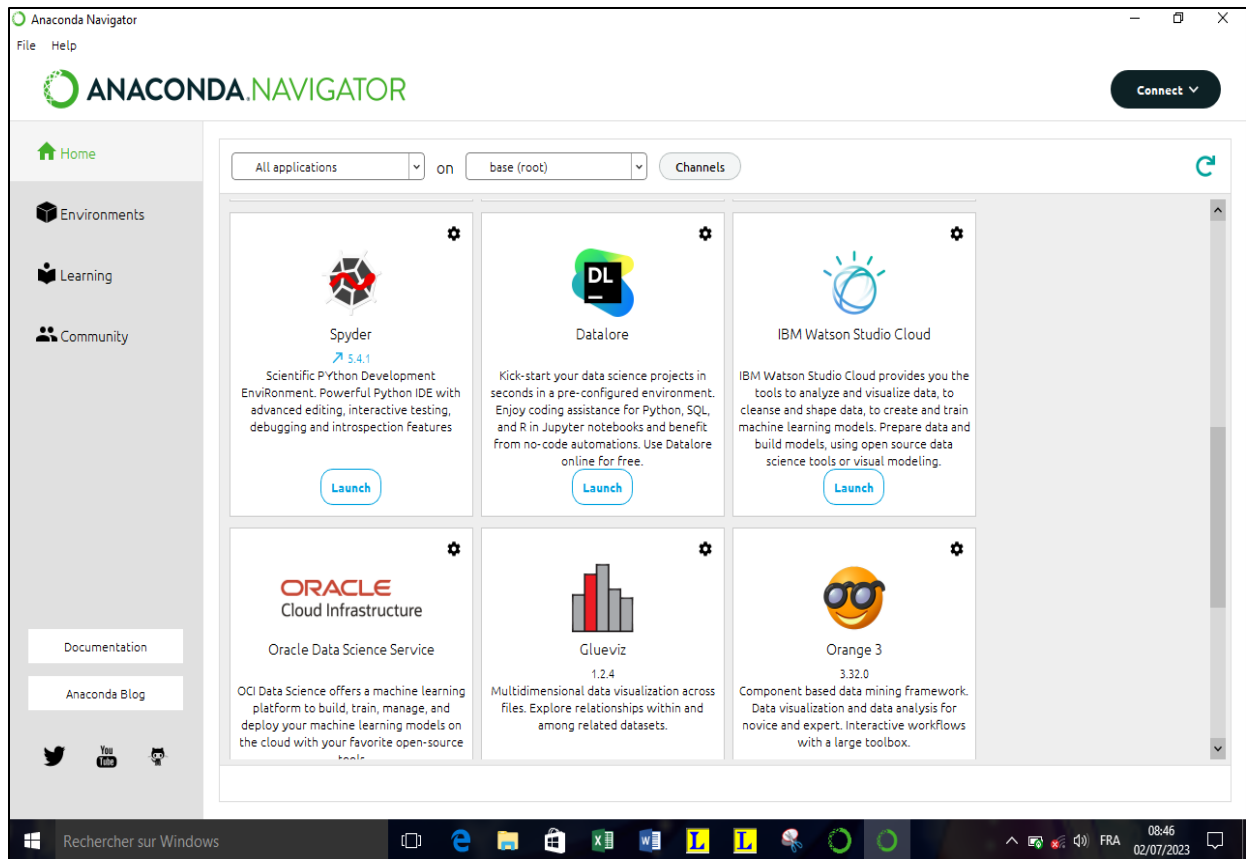


Abbildung 19: Startseite der Software ANACONDA NAVIGATOR (2/3)

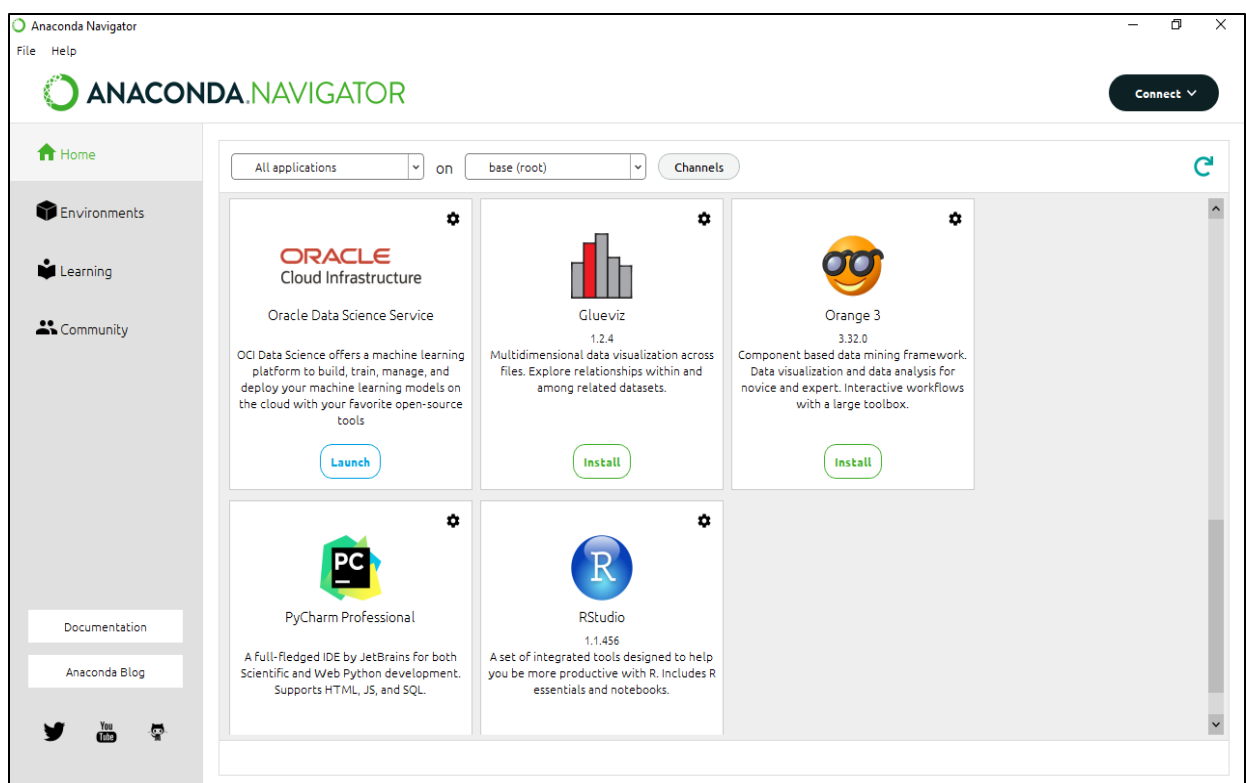


Abbildung 20: Startseite der Software ANACONDA NAVIGATOR (3/3)

In den Abbildungen 13 bis 15 merkt man nämlich die Ikone der zuvor zitierten 14 Python-Entwicklungsumgebungen. Fahren wir nun eine (**SPYDER**) davon hoch.

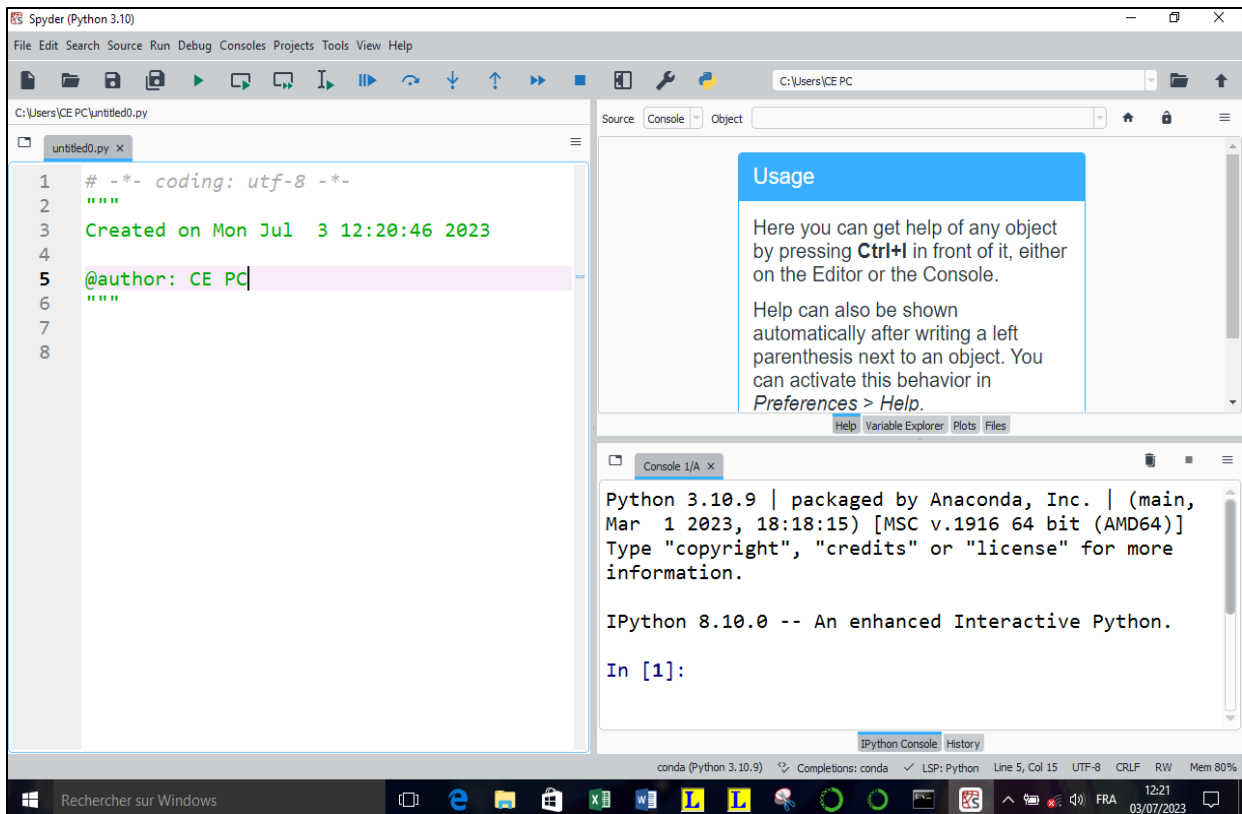


Abbildung 21: Startseite der Python-Entwicklungsumgebung SPYDER

Wie die Abbildung 16 illustriert, besteht das Interface von **SPYDER** aus drei Konsolen/Fenstern. Das erste Fenster links ist der Ort, wo **der informatische Code** vom Programmierer geschrieben wird; also der Ort, wo die gewünschte Handlung codiert wird. Von den beiden anderen Fenstern wird eine **als Hilfsstelle** benutzt (das obig stehende), während das andere eher als Compiler bzw. **als Exekutionsstelle** gebraucht wird. Versuchen wir nun das System PAAR auf den Text N°1 des PAK einzusetzen. Wie zuvor hingewiesen, werden wir den Einsatz des PAAR zunächst Modul für Modul realisieren. In dieser Hinsicht gilt der Schritt **Text-Input**, gemäß der globalen Architektur vom PAAR (Siehe Abbildung 11), als erste Etappe.

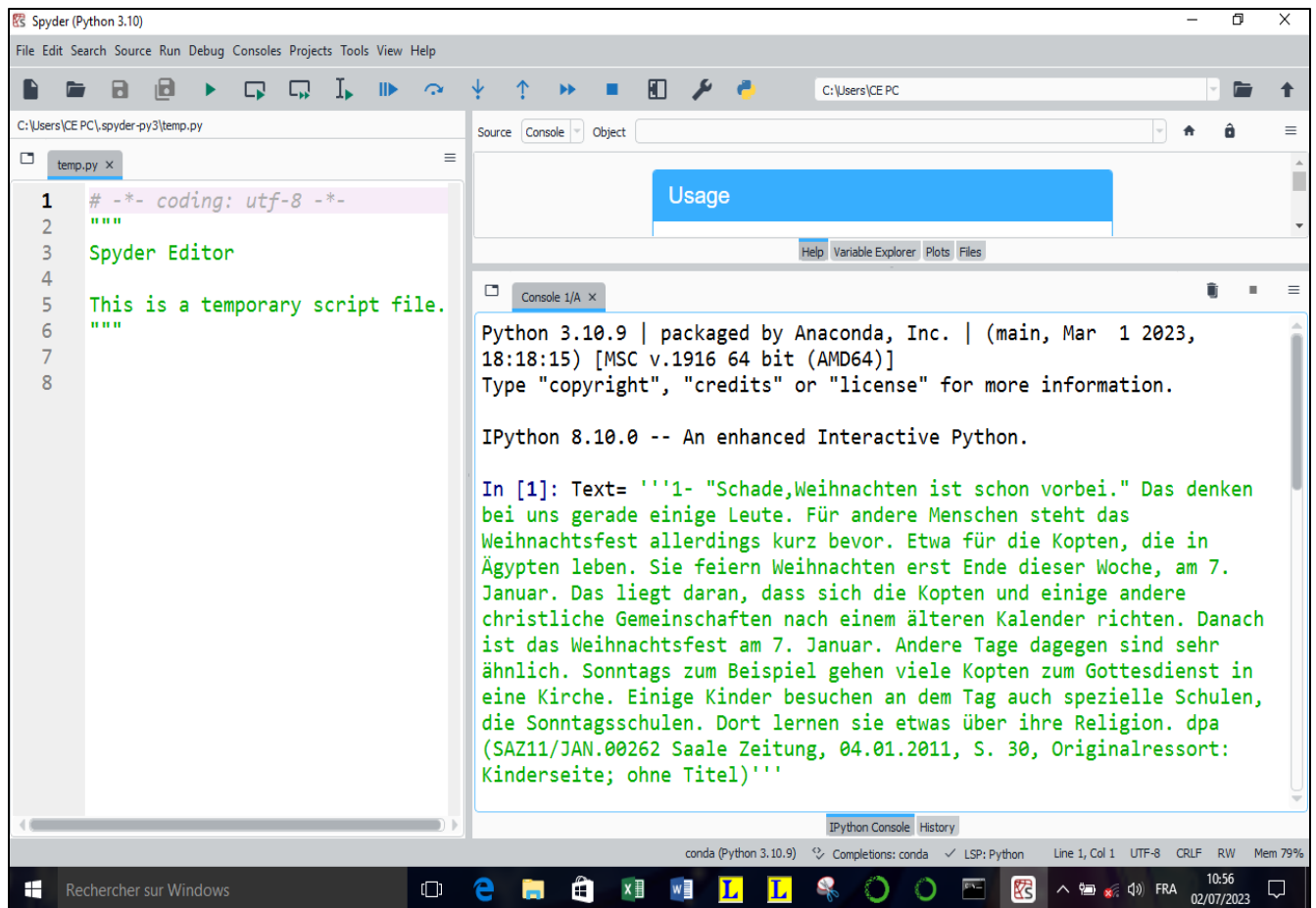


Abbildung 22: Input des Textes N°1 in der Exekutionsstelle von SPYDER (1/2)

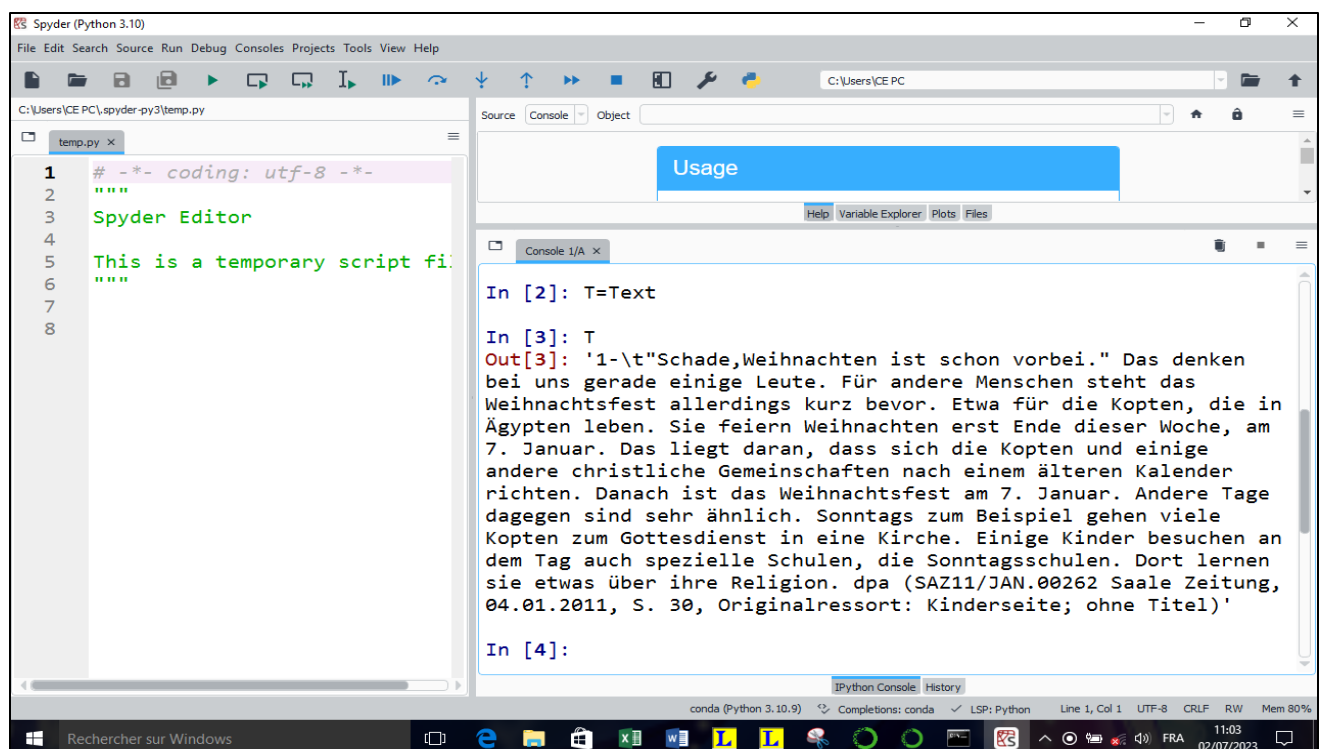


Abbildung 23: Input des Textes N°1 in der Exekutionsstelle von SPYDER (2/2)

Wie es aus den Abbildungen 17 & 18 abzulesen ist, haben wir im **Input [1]** die Variable (**Text**) bestimmt, also ihr einen Wert gegeben und im **Input [2]** haben wir diese Variable als (**T**) umbenannt. Im **Input [3]** verifizieren wir tatsächlich, ob beim Aufruf der **Variable (T)** der zuvor definierte Wert (Also der Text N°1 des PAK) als Output tatsächlich kommt. Etwas, das tatsächlich der Fall ist. Importieren wir nun das Modul **Preprocessing** (Abbildungen 19- 22) und aktivieren wir zugleich die Klasse **Language_Preprocessing** (Abbildung 23), wobei anschließend die Methode **Satz-Tokenizer** (Abbildungen 23&24) und **Wort-Tokenizer** (Abbildungen 25-27) jeweils aufgerufen werden.

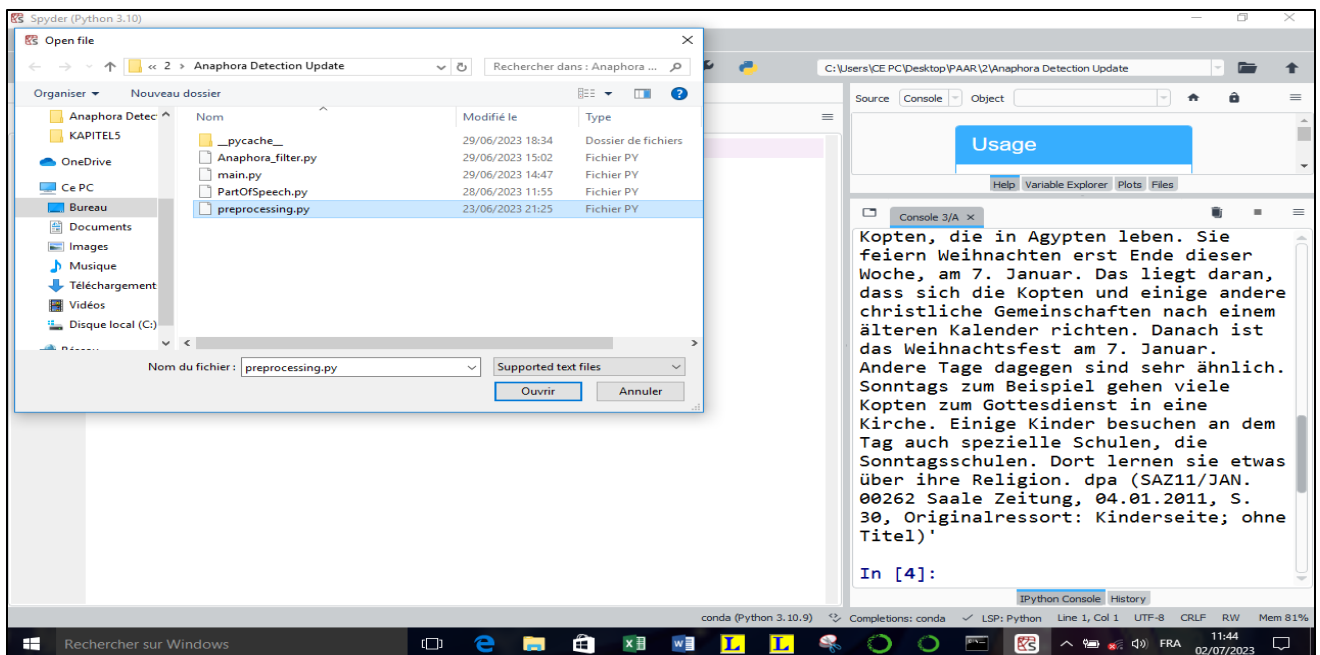


Abbildung 24: Aufruf des Moduls Preprocessing vom PAAR (1/4)

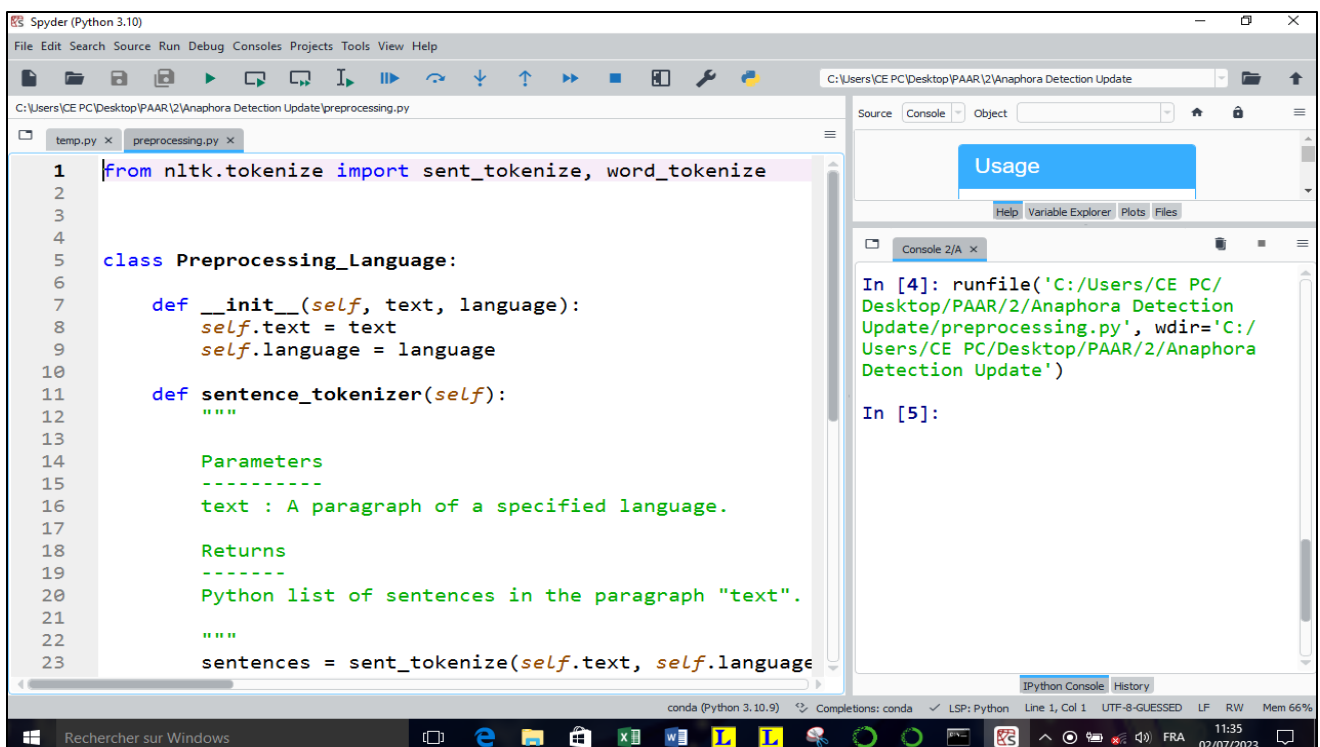


Abbildung 25: Aufruf des Moduls Preprocessing vom PAAR (2/4)

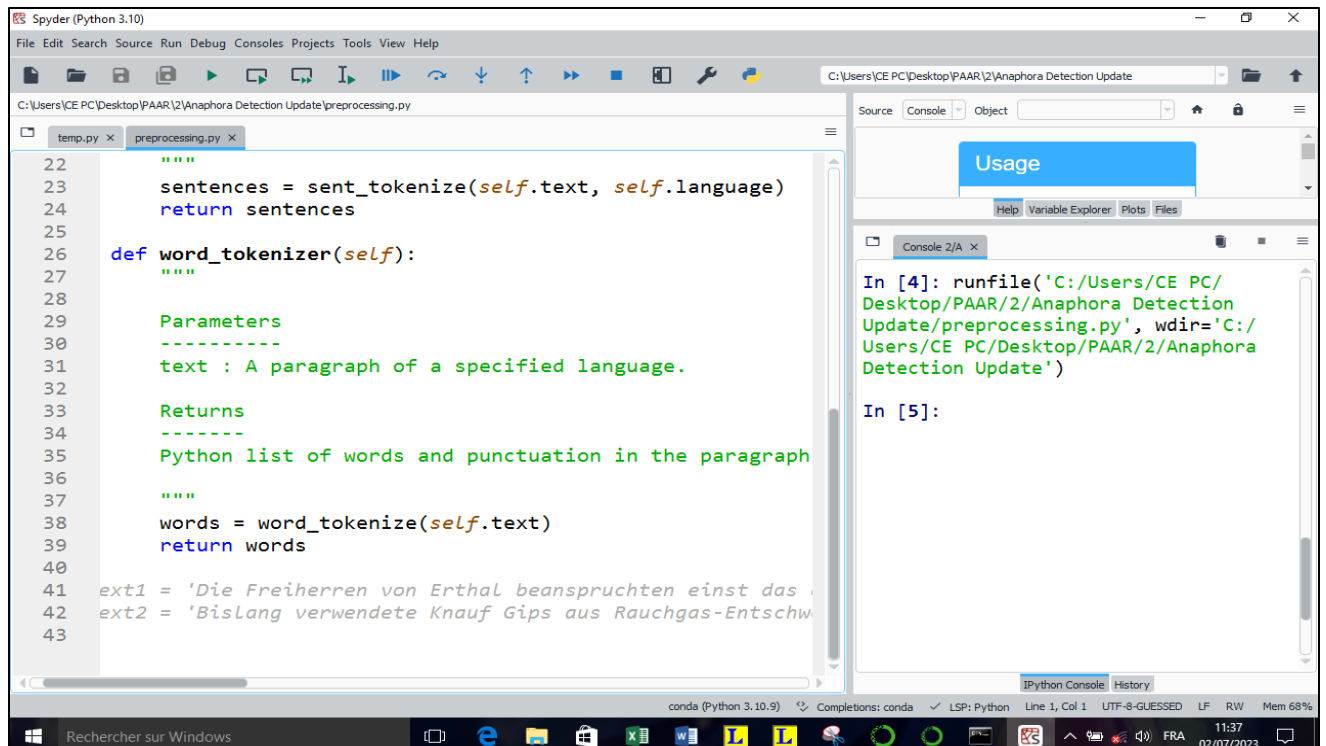


Abbildung 26: Aufruf des Moduls Preprocessing vom PAAR (3/4)

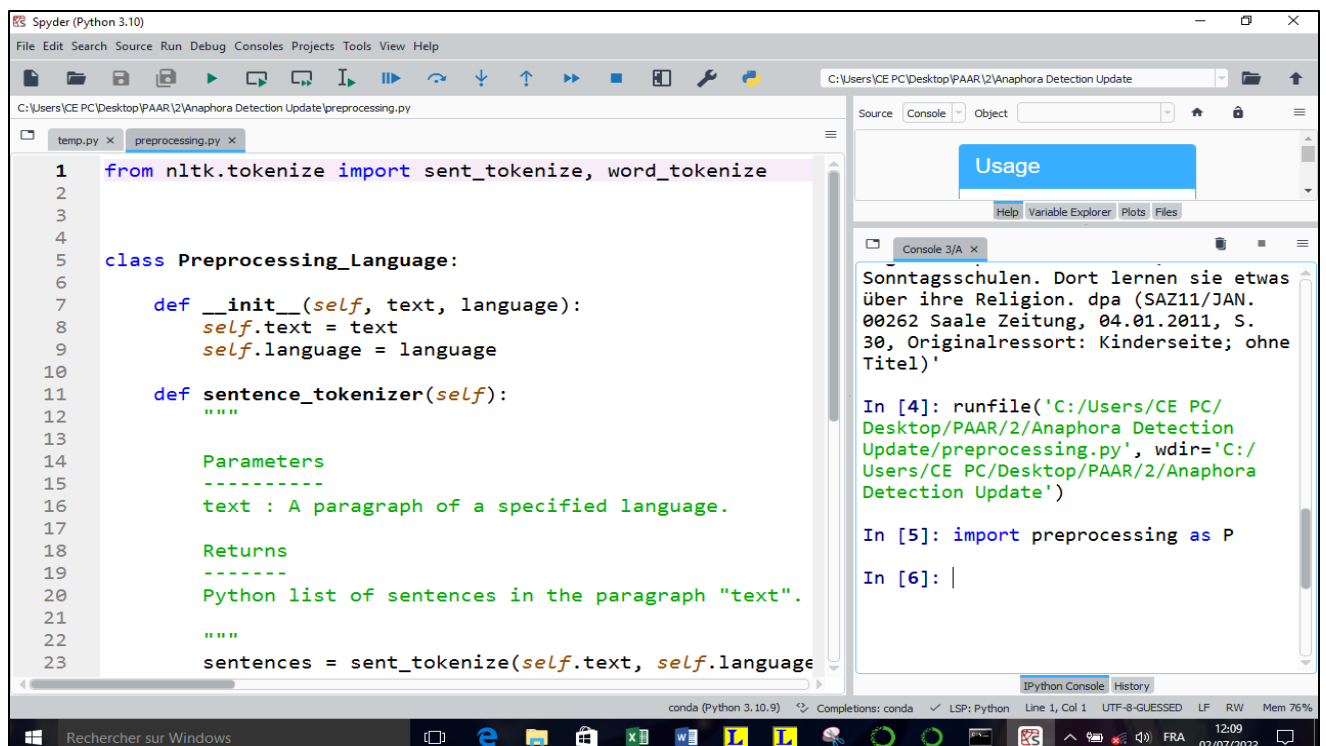


Abbildung 27: Aufruf des Moduls Preprocessing vom PAAR (4/4)

Wie es aus den Abbildungen (19-22) abzulesen ist, haben wir im **Input [4]** das Modul **Preprocessing** vom PAAR exekutiert und im Anschluss daran ihm den Namen (**P**) gegeben (Siehe dazu **Input [5]**). Im **Input [6]** greifen wir auf die Klasse **Preprocessing _ Language** zu. Und dieser Zugriff wird mit einem (**p**) bezeichnet. In einem weiteren Schritt greifen wir auf die Komponente/Methode „**sentence_tokenizer**“ von (**p**) zu (Siehe hierzu **Input [7]**). Dieser

The screenshot shows the Spyder Python IDE with the file `preprocessing.py` open. The code in the editor defines a class `Preprocessing_Lang` with an `__init__` method that takes `text` and `language` as arguments, and a `sentence_tokenizer` method. The IPython console shows the following interactions:

```
In [5]: import preprocessing as P
In [6]: p=P.Preprocessing_Language(T,'german')
In [7]: p.sentence_tokenizer()
Out[7]:
['1-\t"Schade,Weihnachten ist schon vorbei."',
'Das denken bei uns gerade einige Leute.',
'Für andere Menschen steht das Weihnachtsfest allerdings kurz bevor.',
'Etwa für die Kopten, die in Ägypten leben.',
'Sie feiern Weihnachten erst Ende dieser Woche, am 7. Januar.',
'Das liegt daran, dass sich die Kopten und einige andere christliche
Gemeinschaften nach einem älteren Kalender richten.',
'Danach ist das Weihnachtsfest am 7. Januar.',
'Andere Tage dagegen sind sehr ähnlich.',
'Sonntags zum Beispiel gehen viele Kopten zum Gottesdienst in eine
Kirche.',
'Einige Kinder besuchen an dem Tag auch spezielle Schulen, die
```

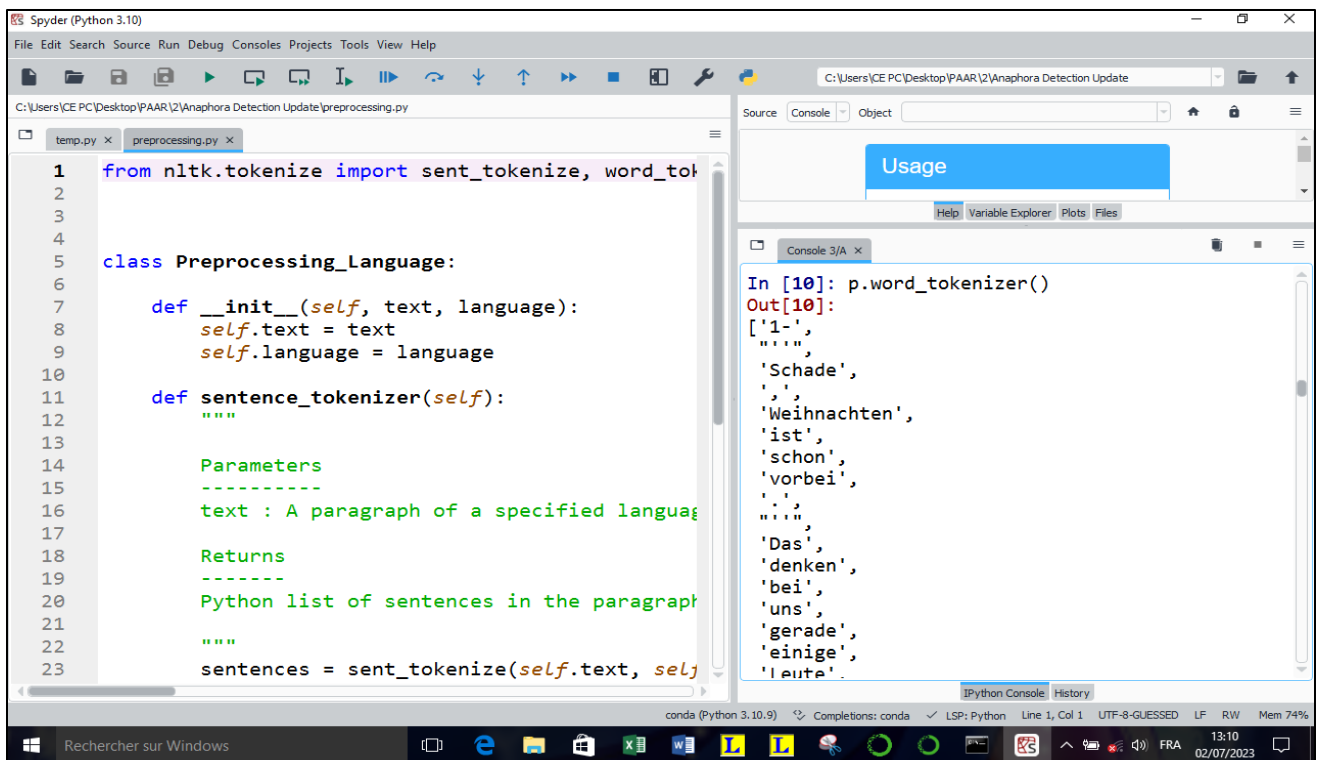
Abbildung 28: Aufruf der Klasse „Preprocessing Language“ und der Methode „sentence tokenizer“ (1/2)

This screenshot continues the previous one, showing the next steps in the IPython console:

```
In [8]: TSatz= p.sentence_tokenizer()
In [9]: TSatz
Out[9]:
['1-\t"Schade,Weihnachten ist schon vorbei."',
'Das denken bei uns gerade einige Leute.',
'Für andere Menschen steht das Weihnachtsfest allerdings kurz bevor.',
'Etwa für die Kopten, die in Ägypten leben.',
'Sie feiern Weihnachten erst Ende dieser Woche, am 7. Januar.',
'Das liegt daran, dass sich die Kopten und einige andere christliche
Gemeinschaften nach einem älteren Kalender richten.',
'Danach ist das Weihnachtsfest am 7. Januar.',
'Andere Tage dagegen sind sehr ähnlich.',
'Sonntags zum Beispiel gehen viele Kopten zum Gottesdienst in eine
Kirche.',
'Einige Kinder besuchen an dem Tag auch spezielle Schulen, die
Sonntagsschulen.',
'Dort lernen sie etwas über ihre Religion.',
'dna (SA711/1AN 00262 Saale Zeitung 04.01.2011 S. 30 Originalressort:
```


Abbildung 29: Aufruf der Klasse „Preprocessing Language“ und der Methode „sentence tokenizer“ (2/2)

Sobald der Text schon in Sätzen zerstückelt (durch den Einsatz des Satz-Parsers von SPYDER) worden ist, wollen wir nun den Text in Wörtern zerstückeln. Dafür greifen wir auf die Methode „word_tokenizer“ von (p) zu. (Siehe hierfür **Input [10]**). Wir werden diesen Zugriff gleich danach (**im Input [11]**) als (**TWort**) nennen. Diese Abkürzung steht nämlich für (Text in Wörtern). Und im **Input [12]** verifizieren wir tatsächlich, ob beim Aufruf der Variable (TWort) der erwartete Wert rauskommt (also der in Wörtern zerstückelte Text).

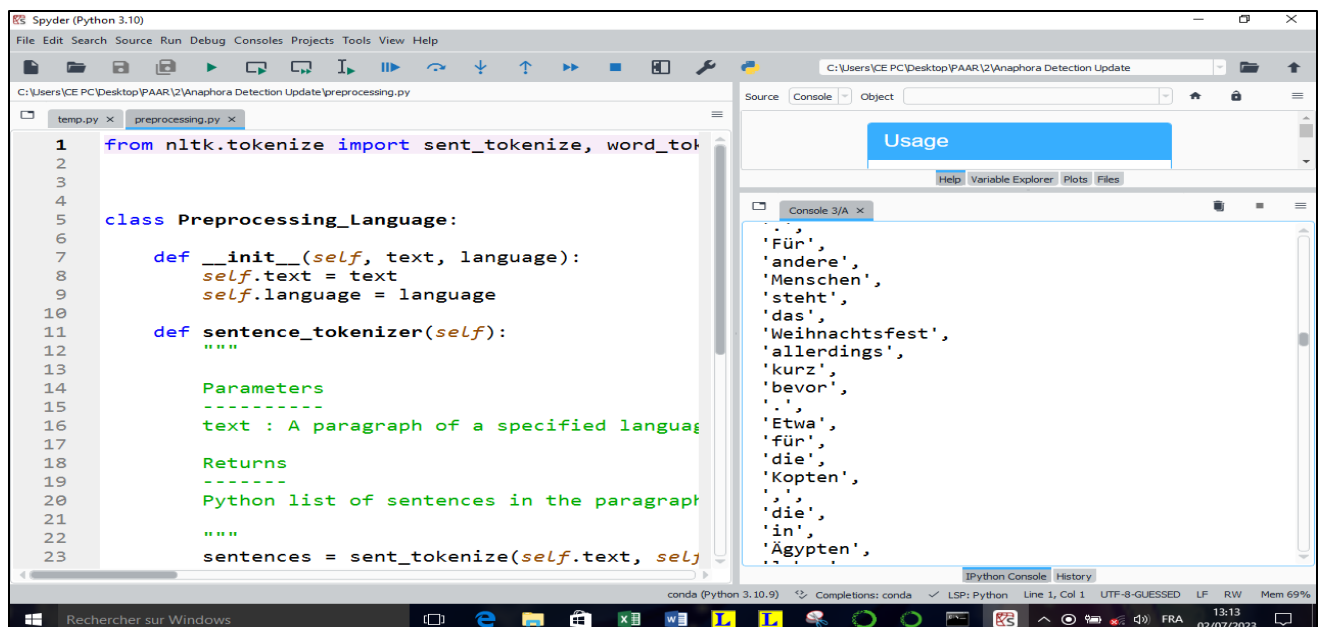


```
1 from nltk.tokenize import sent_tokenize, word_tokenize
2
3
4
5 class Preprocessing_Language:
6
7     def __init__(self, text, language):
8         self.text = text
9         self.language = language
10
11     def sentence_tokenizer(self):
12         """
13         Parameters
14         -----
15         text : A paragraph of a specified language
16
17         Returns
18         -----
19         Python list of sentences in the paragraph
20
21         """
22         sentences = sent_tokenize(self.text, self.language)
23         return sentences
```

Console 3/A x

```
In [10]: p.word_tokenizer()
Out[10]:
['1-',
 '...',
 'Schade',
 '...',
 'Weihnachten',
 'ist',
 'schon',
 'vorbei',
 '...',
 '...',
 'Das',
 'denken',
 'bei',
 'uns',
 'gerade',
 'einige',
 'leute'...
```

Abbildung 30: Aufruf der Methode „word tokenizer“ (1/3)



```
1 from nltk.tokenize import sent_tokenize, word_tokenize
2
3
4
5 class Preprocessing_Language:
6
7     def __init__(self, text, language):
8         self.text = text
9         self.language = language
10
11     def word_tokenizer(self):
12         """
13         Parameters
14         -----
15         text : A paragraph of a specified language
16
17         Returns
18         -----
19         Python list of words in the paragraph
20
21         """
22         words = word_tokenize(self.text, self.language)
23         return words
```

Console 3/A x

```
['Für',
 'andere',
 'Menschen',
 'steht',
 'das',
 'Weihnachtsfest',
 'allerdings',
 'kurz',
 'bevor',
 '...',
 'Etwas',
 'für',
 'die',
 'Kopten',
 'die',
 'in',
 'Ägypten',
 '...'...
```


Abbildung 31: Aufruf der Methode „word tokenizer“ (2/3)

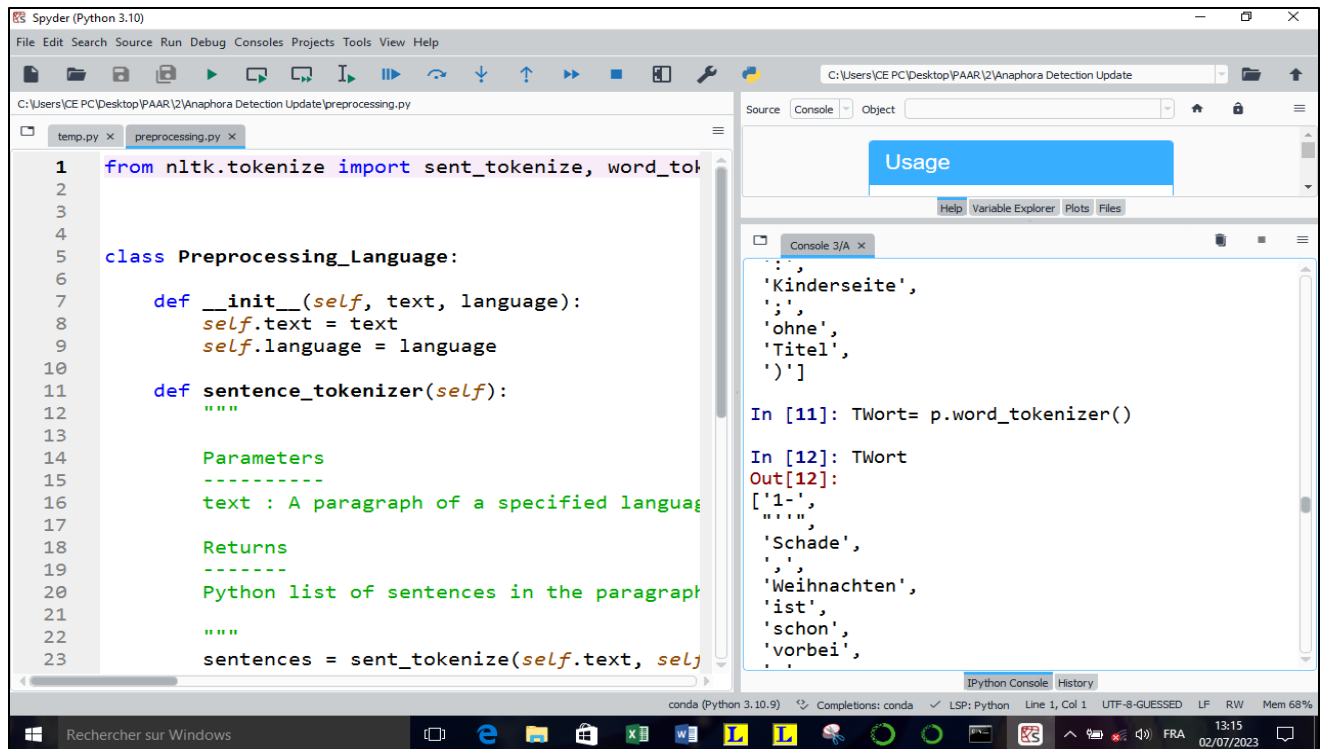


Abbildung 32: Aufruf der Methode „word tokenizer“ (3/3)

Importieren wir nun das Modul „Part-of-Speech“ und nennen wir es als (POS) (Siehe hierfür **Input [13] & [14]** in den Abbildungen 28 & 29). Im **Input [15]** nehmen wir Zugriff auf die Klasse „find_pa“ und wir nennen zugleich diesen Zugriff als (pos) (**Abbildung 30**). In einem weiteren Schritt verifizieren wir tatsächlich, ob beim Aufruf der Variable (pos) tatsächlich alle Pronominaladverbien (PA) des Textes (T) aufgefunden werden (Siehe hierzu **In- & Output [16]**) in der **Abbildung 31**)

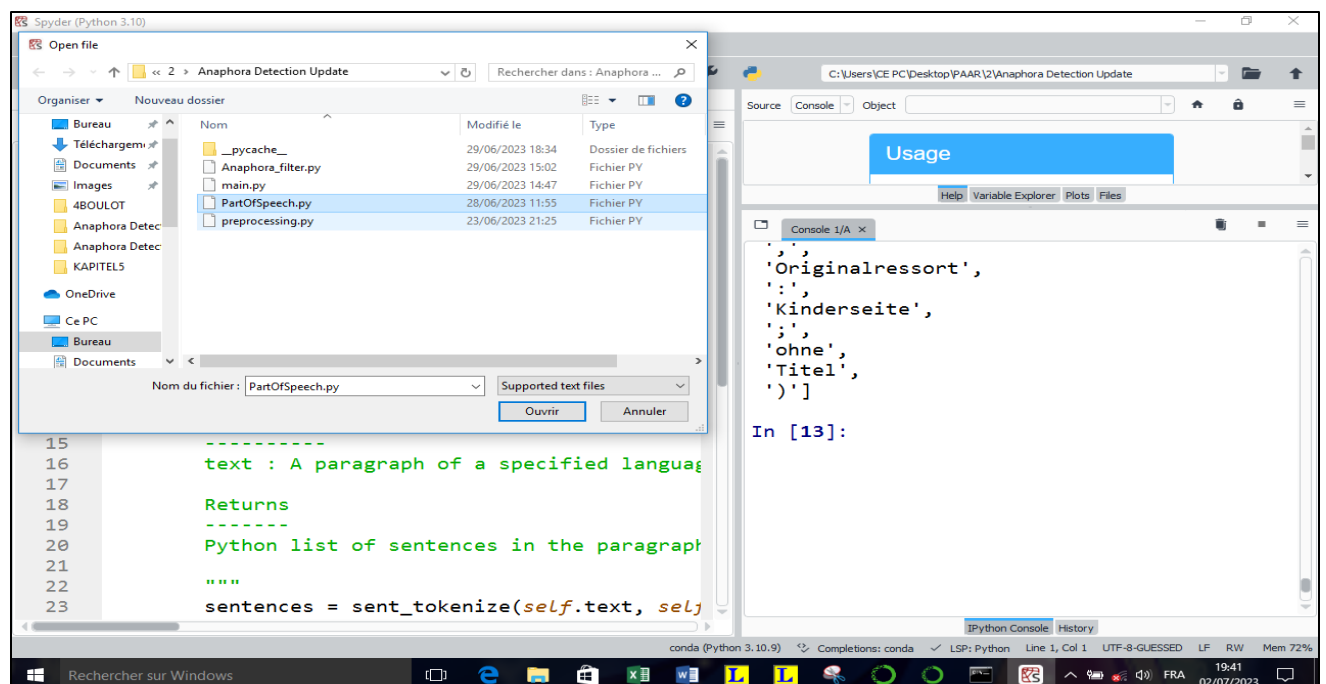


Abbildung 33: Aufruf des Moduls „Part-of-Speech“ (1/3)

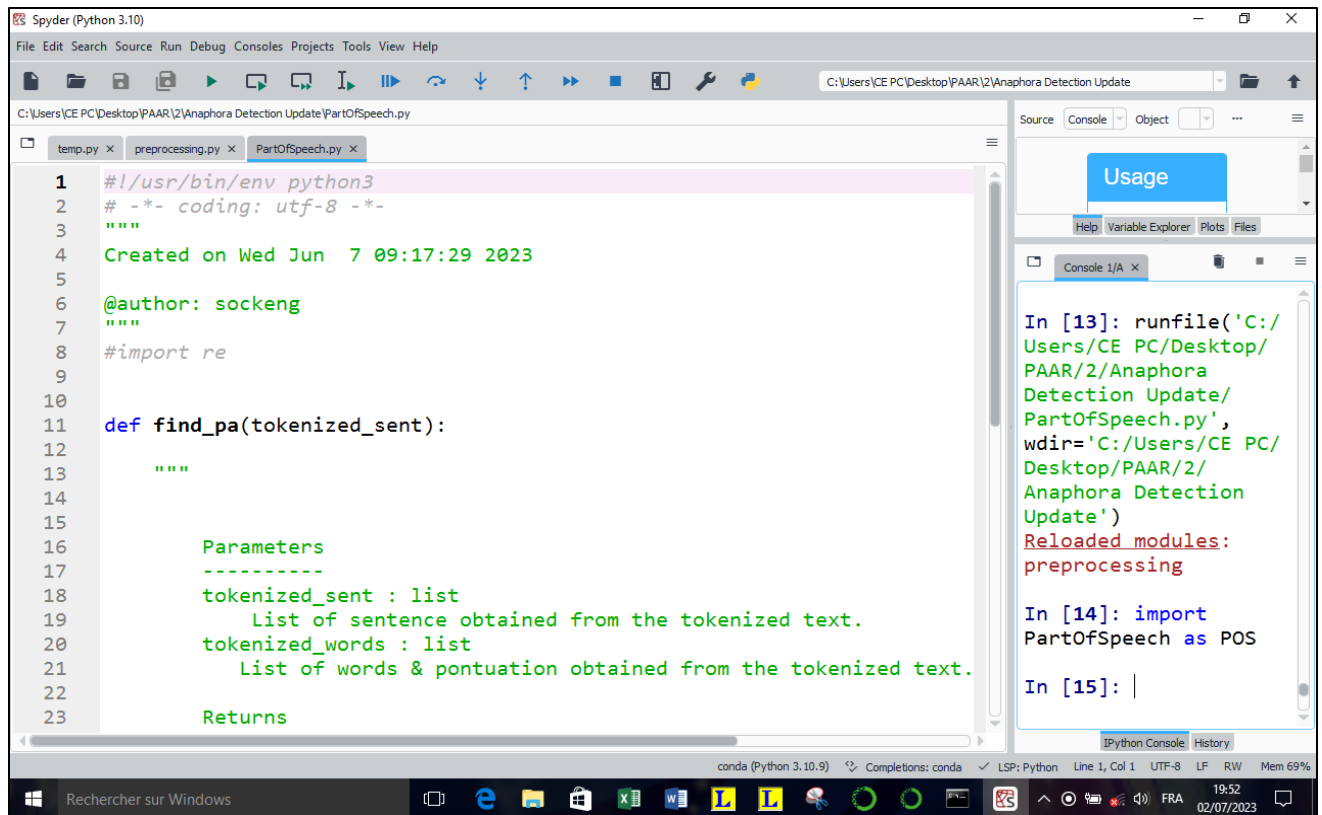


Abbildung 34: Aufruf des Moduls „Part-of-Speech“ (2/3)

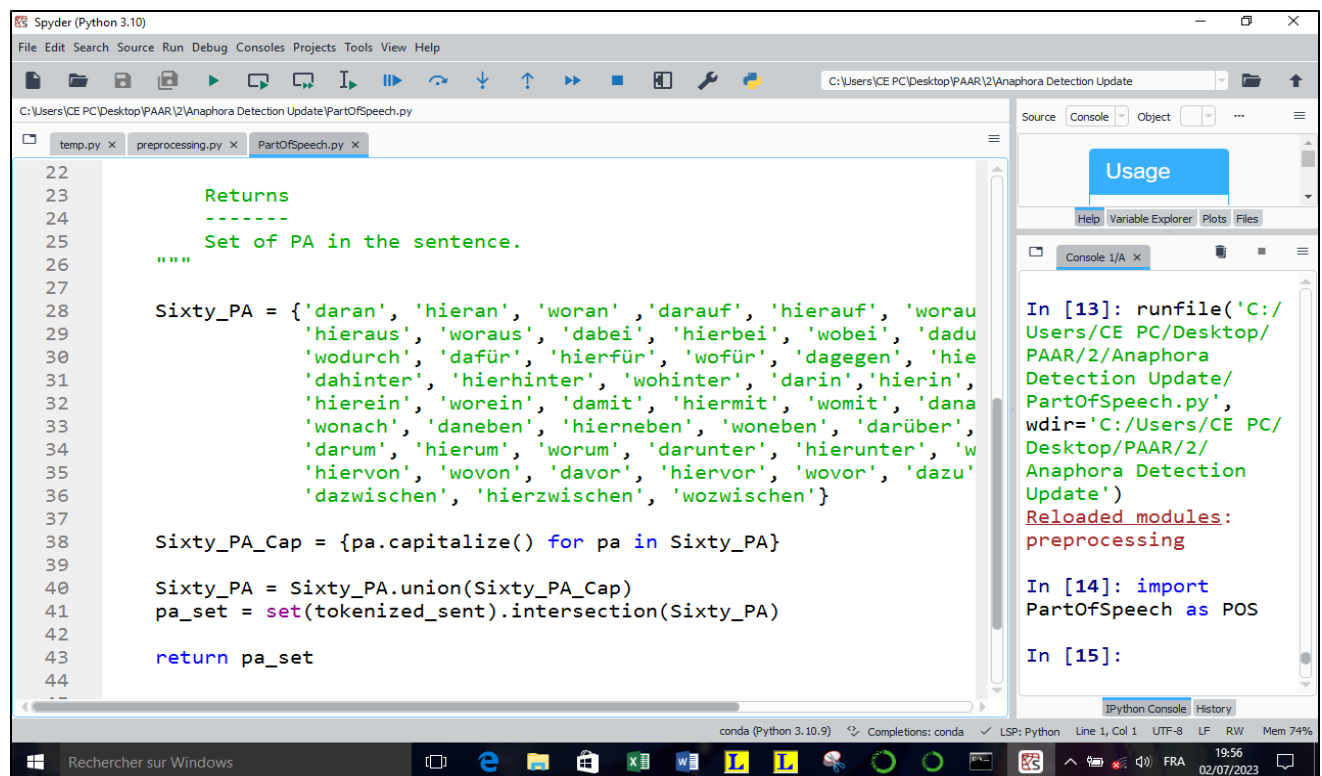
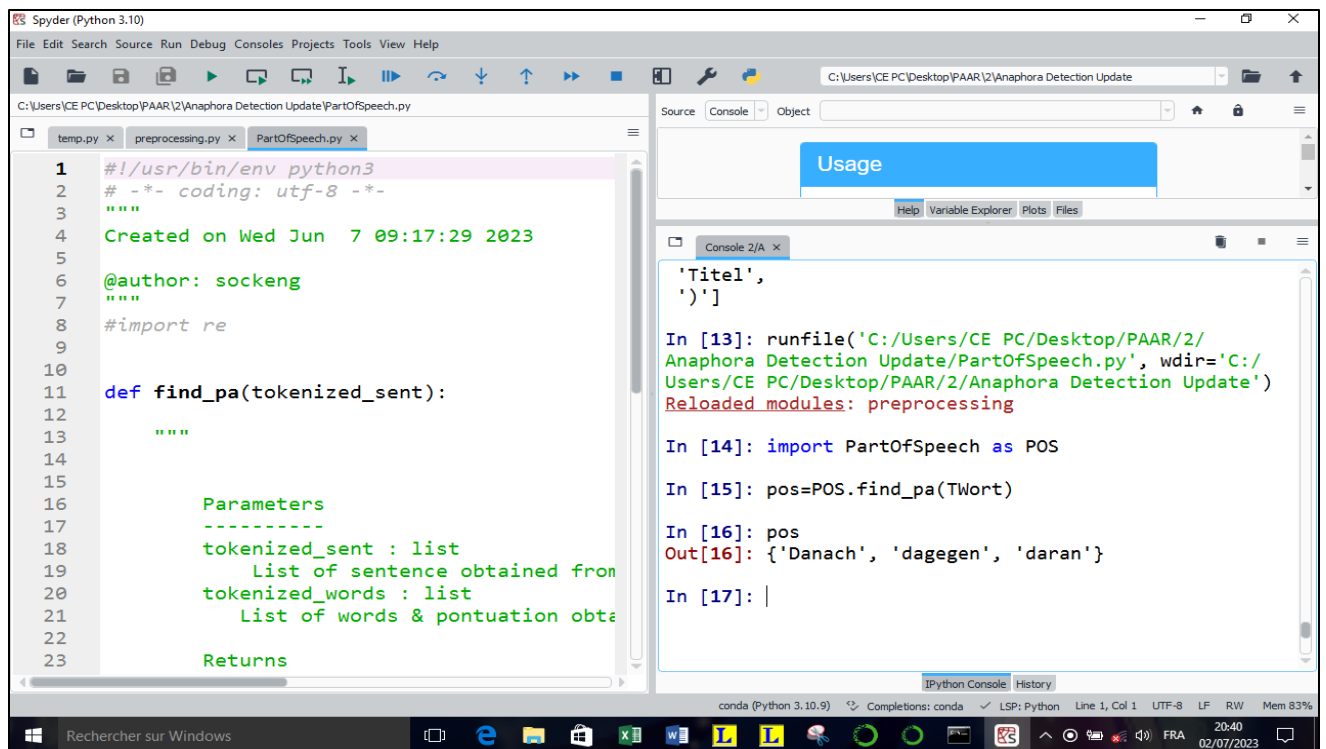


Abbildung 35: Aufruf des Moduls „Part-of-Speech“(3/3)



The screenshot shows the Spyder Python IDE interface. The left pane displays the code in `PartOfSpeech.py`, which includes a docstring with author information and a function `find_pa` that takes a `tokenized_sent` list as input. The right pane shows the IPython console with the following interactions:

```
In [13]: runfile('C:/Users/CE PC/Desktop/PAAR/2/Anaphora Detection Update/PartOfSpeech.py', wdir='C:/Users/CE PC/Desktop/PAAR/2/Anaphora Detection Update')
Reloaded modules: preprocessing

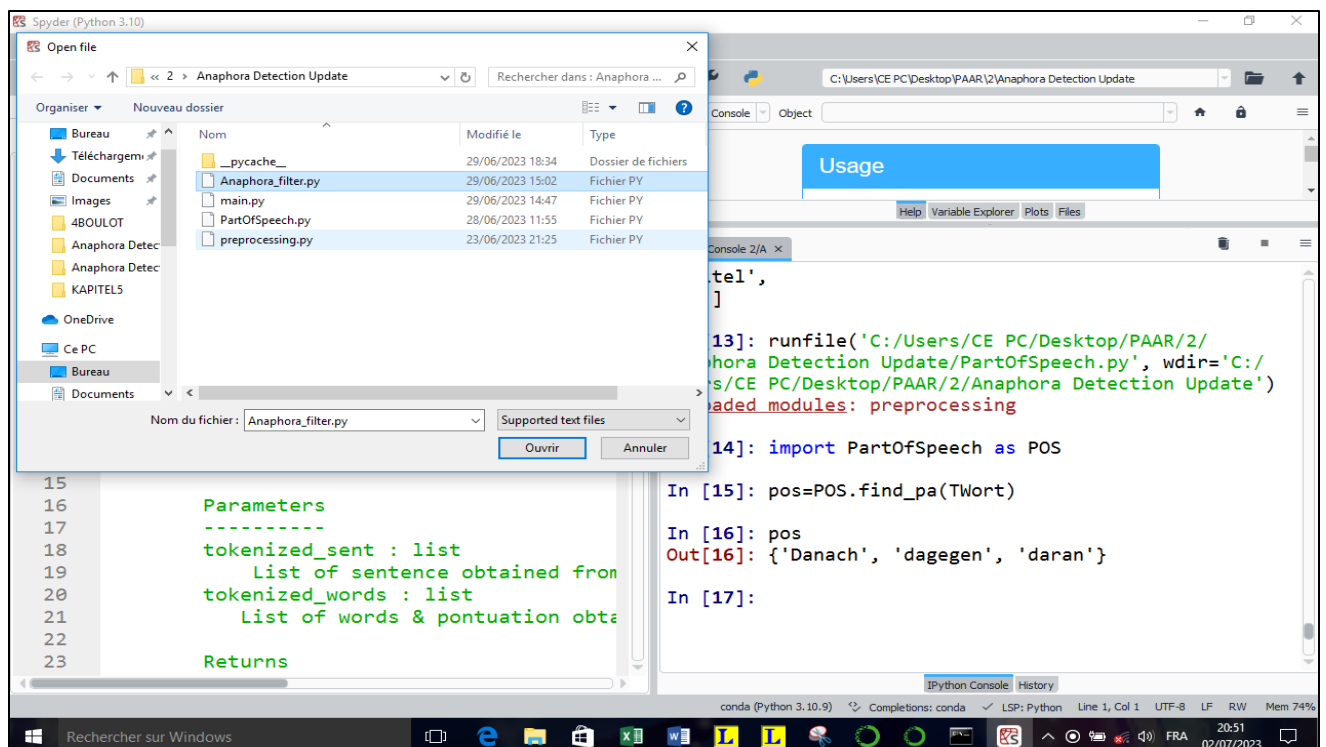
In [14]: import PartOfSpeech as POS

In [15]: pos=POS.find_pa(TWort)

In [16]: pos
Out[16]: {'Danach', 'dagegen', 'daran'}
```

Abbildung 36: Aufruf der Klasse „find pa“

Sobald alle PA des Textes aufgefunden worden sind, importieren wir nun das Modul „**Anaphora_filter**“ und nennen wir es als (AF) (Siehe hierfür **Input [17] & [18]** in Abbildungen 32-39). Diesem Modul wurde die Aufgabe geschrieben, kataphorische und deiktische PA auszufiltern und im Anschluss daran die übrig verbliebenen PA als anaphorisch zu klassifizieren, wobei zugleich auch deren jeweiligen Antezedenzen wiedergegeben werden müssen.



The screenshot shows the Spyder Python IDE interface with an 'Open file' dialog box open, displaying the contents of the `Anaphora Detection Update` folder. The files listed are `_pycache_`, `Anaphora_filter.py`, `main.py`, `PartOfSpeech.py`, and `preprocessing.py`. The `Anaphora_filter.py` file is selected. The background shows the same code and console output as in the previous image, with the console showing the import of `PartOfSpeech` and the execution of `find_pa`.

Abbildung 37: Aufruf des Moduls „Anaphora filter“(1/8)

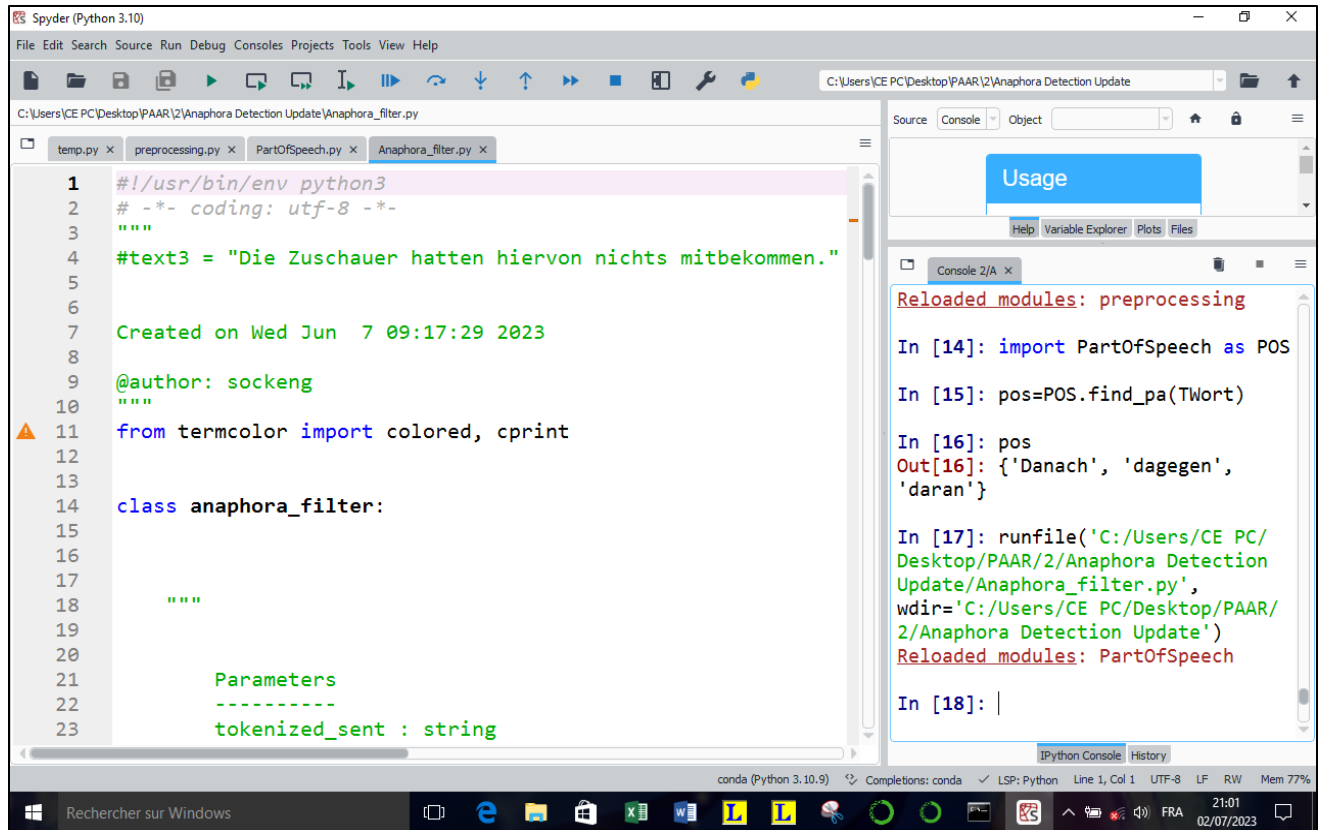


Abbildung 38: Aufruf des Moduls „Anaphora filter“(2/8)

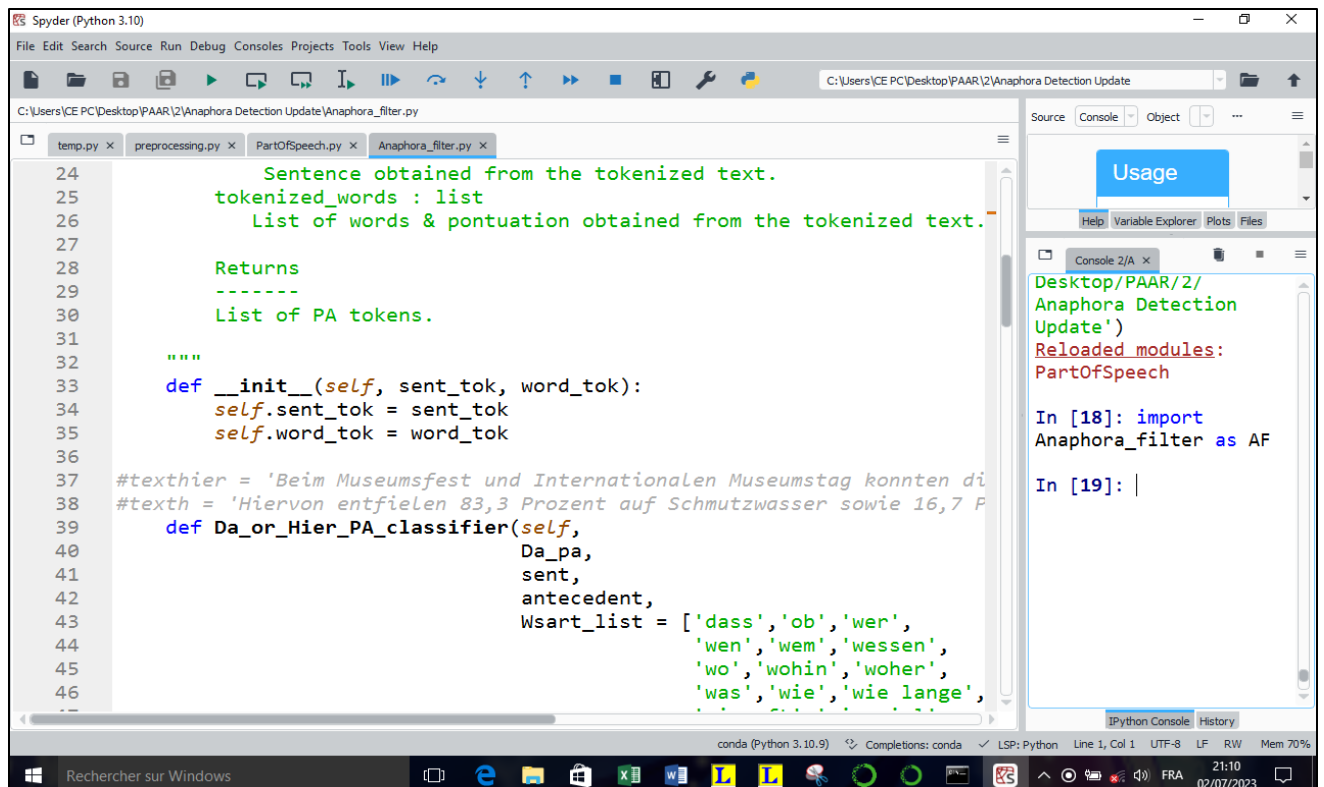


Abbildung 39: Aufruf des Moduls „Anaphora filter“(3/8)

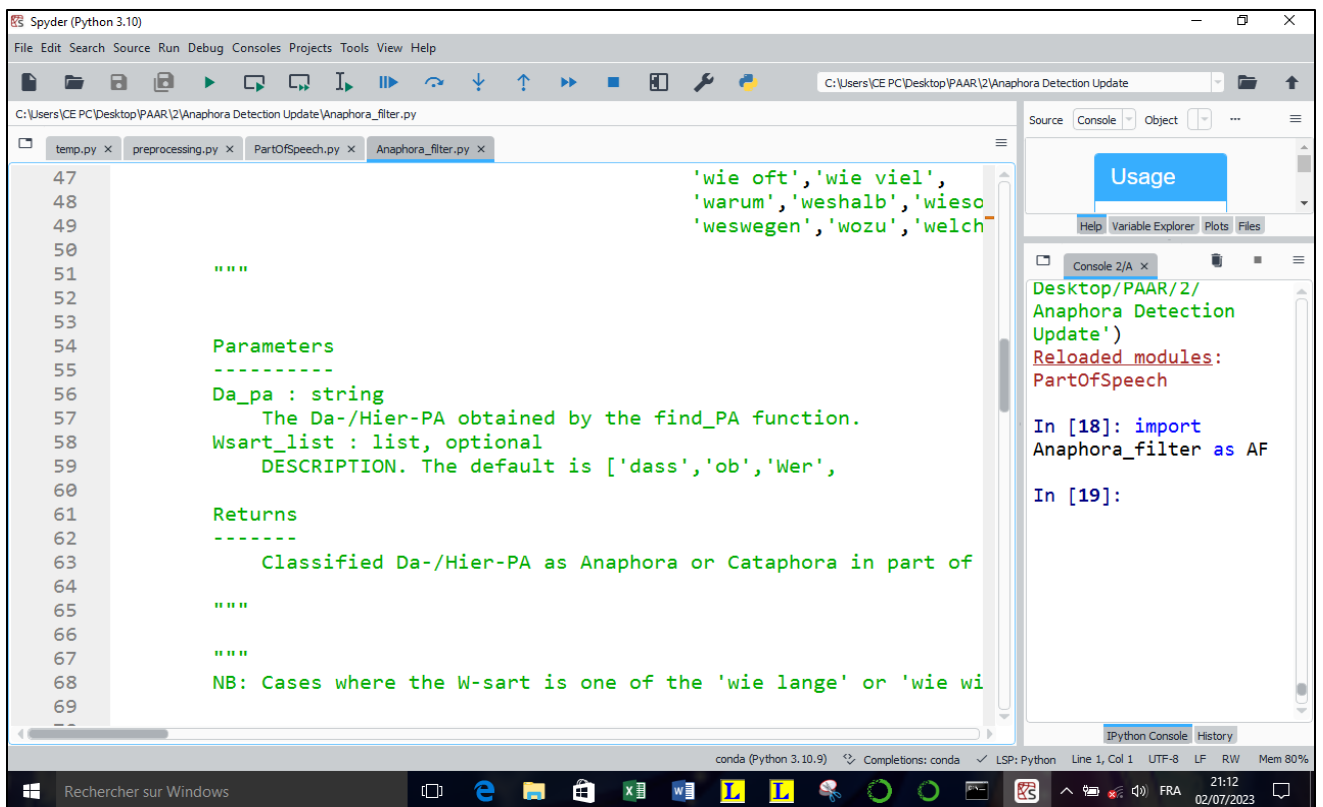


Abbildung 40: Aufruf des Moduls „Anaphora filter“(4/8)

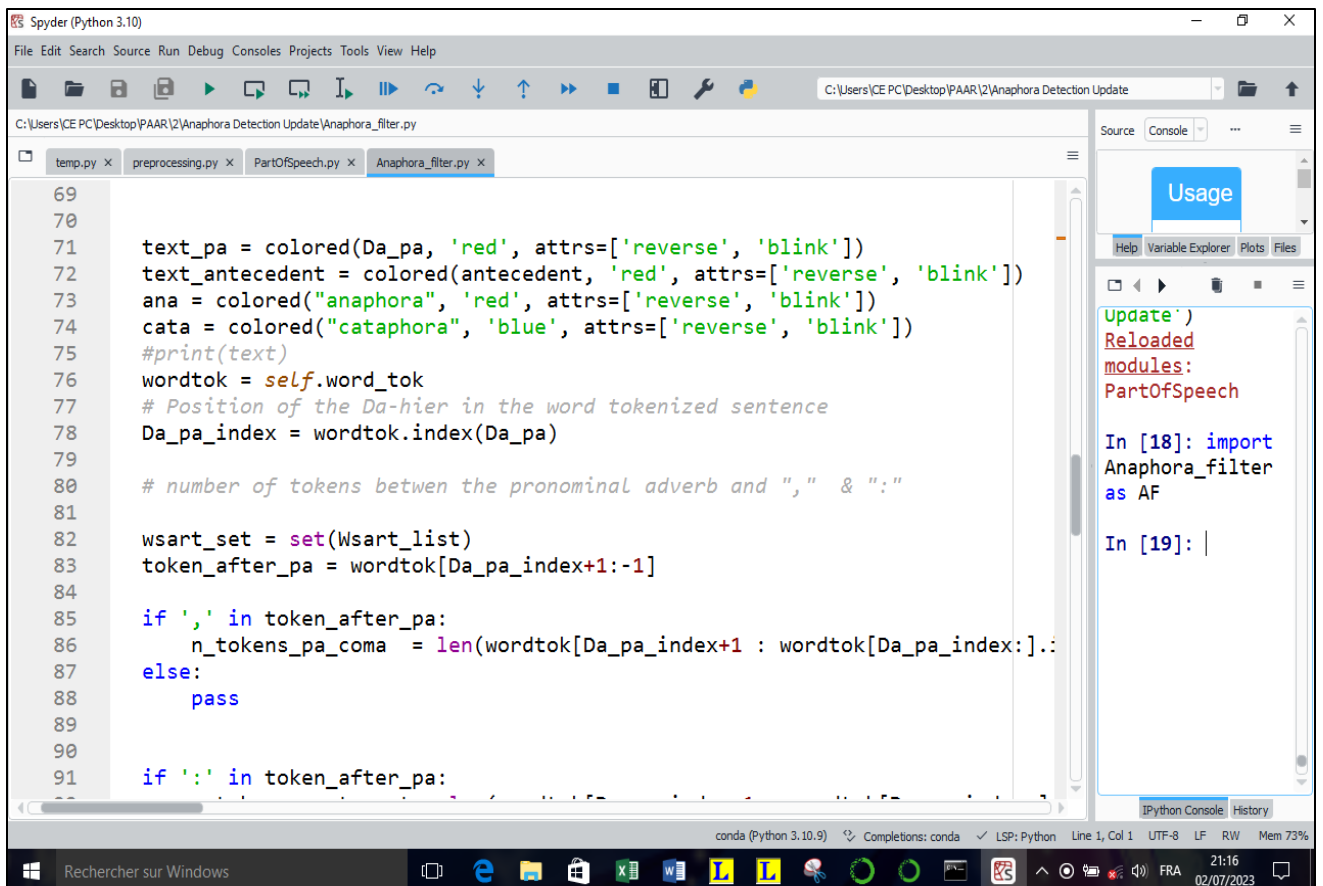


Abbildung 41: Aufruf des Moduls „Anaphora filter“(5/8)

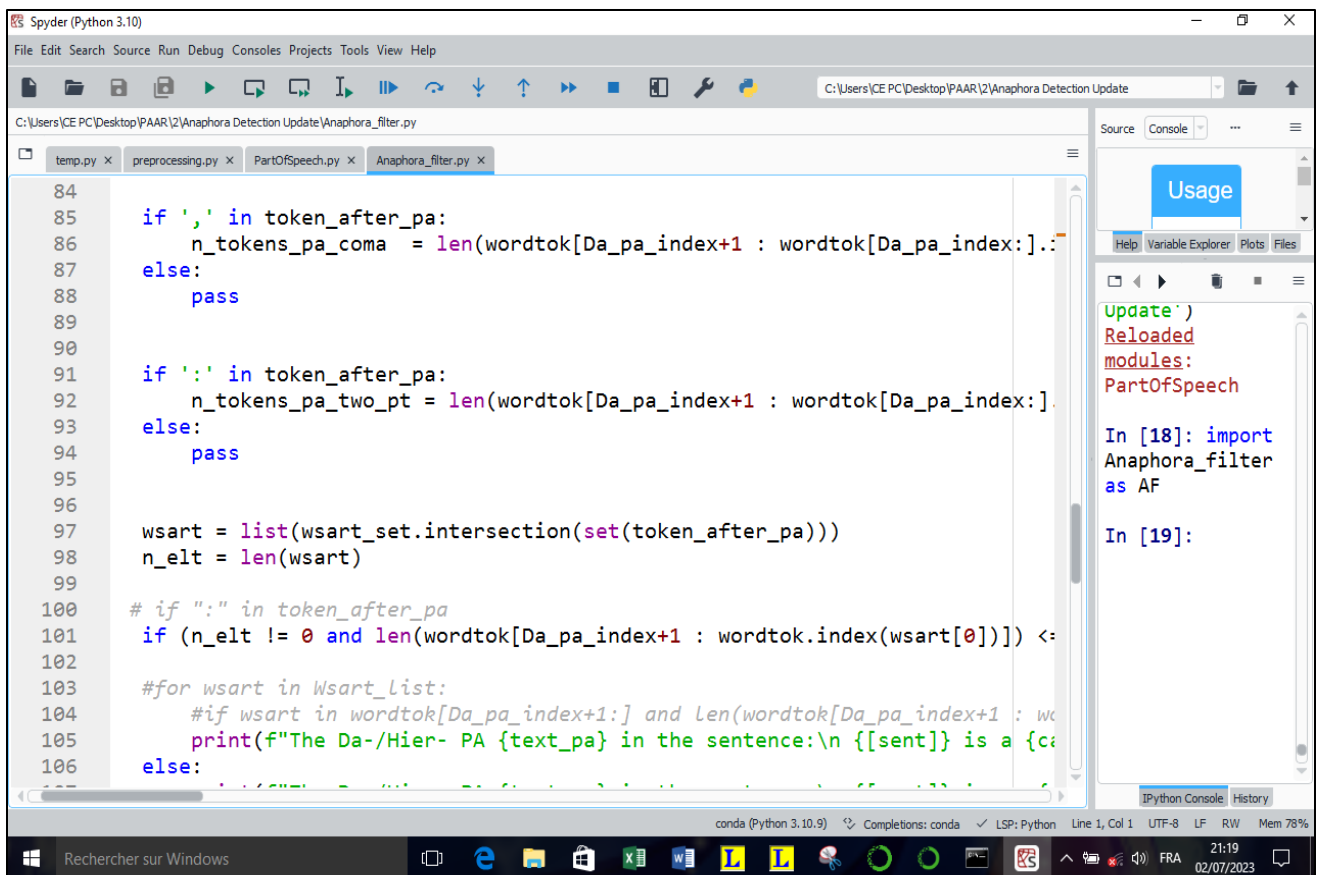


Abbildung 42: Aufruf des Moduls „Anaphora filter“(6/8)

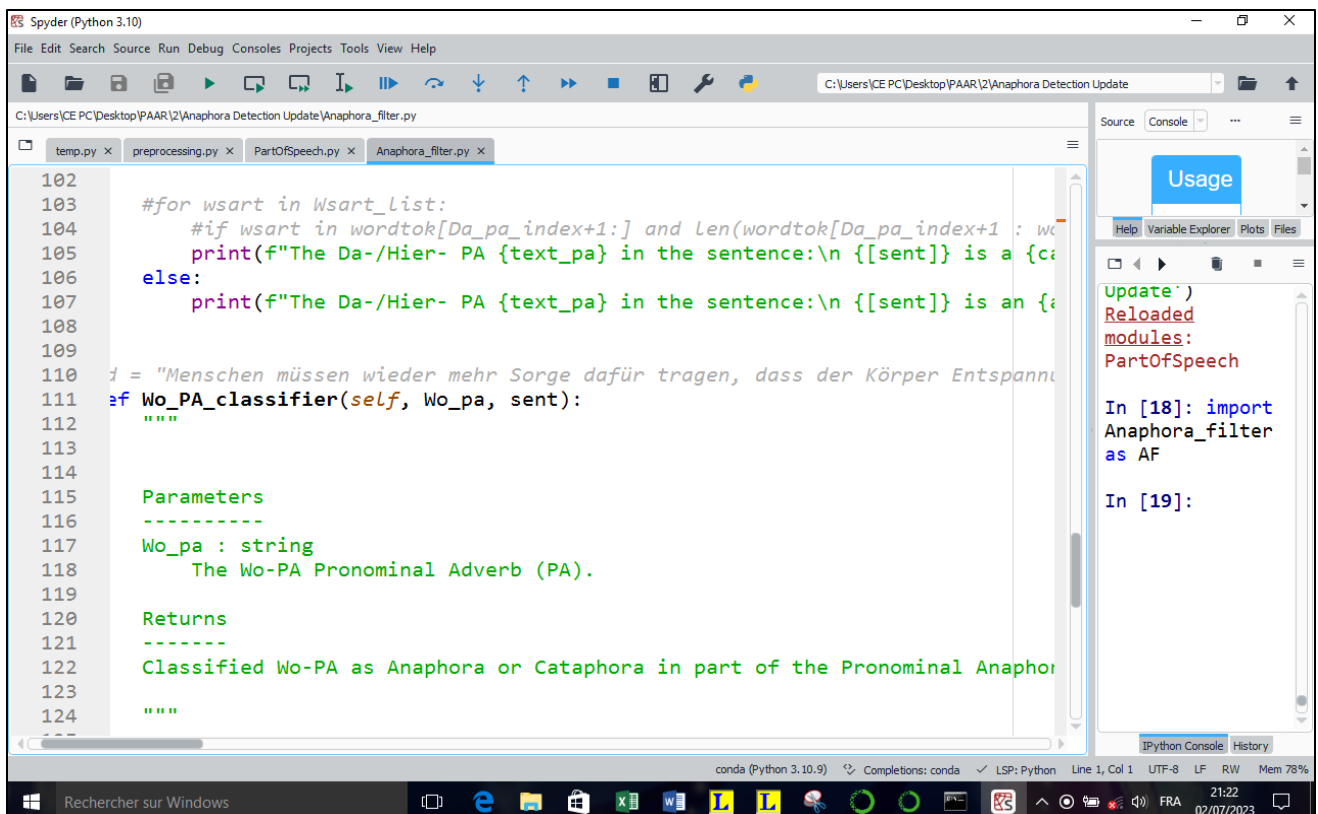


Abbildung 43: Aufruf des Moduls „Anaphora filter“(7/8)

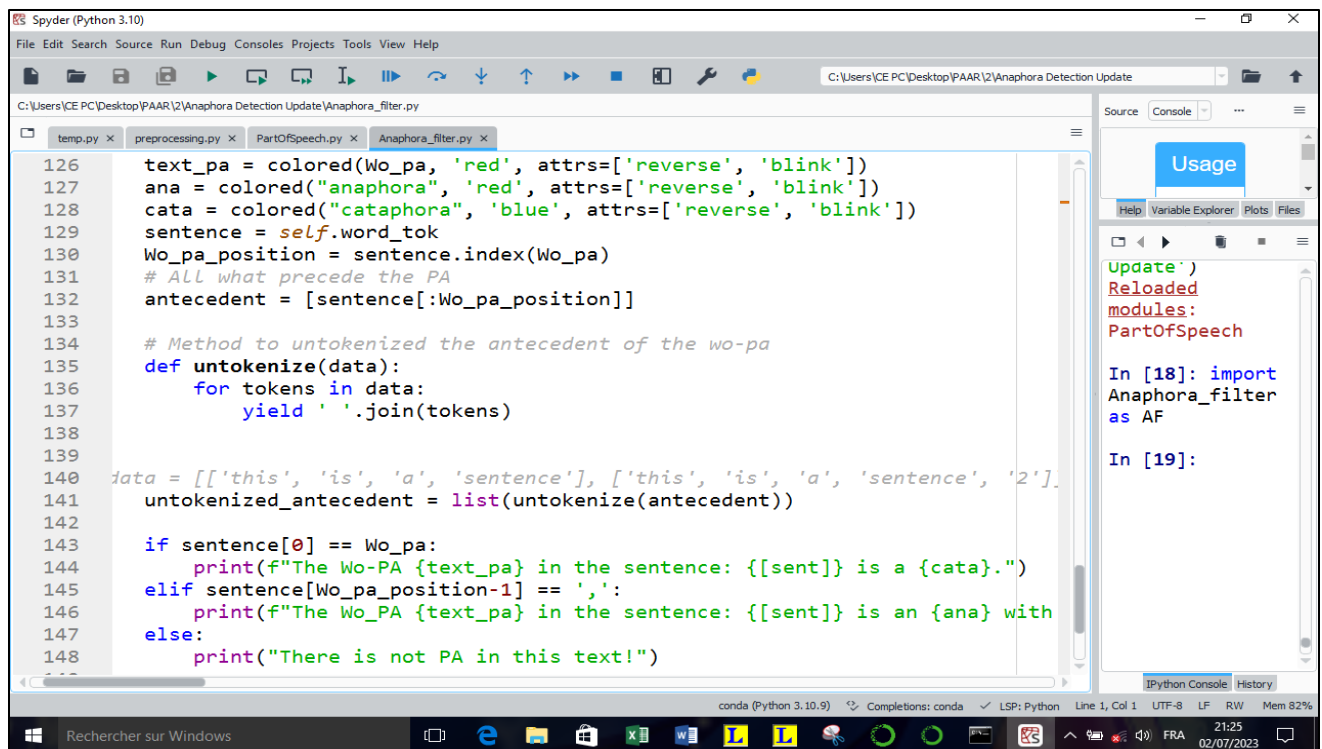


Abbildung 44: Aufruf des Moduls „Anaphora filter“(8/8)

Greifen wir nun auf die Klasse „anaphora_filter“ zu. Wir nennen diesen Zugriff als (af) (Siehe Input [19]. Und da wir im (T) nur PA („Danach“, „dagegen“, „daran“) haben, die zur Kategorie der Da-/Hier-PA gehören, sollen wir nun nur die Methode „Da_or_Hier_PA_classifier“ aufrufen.

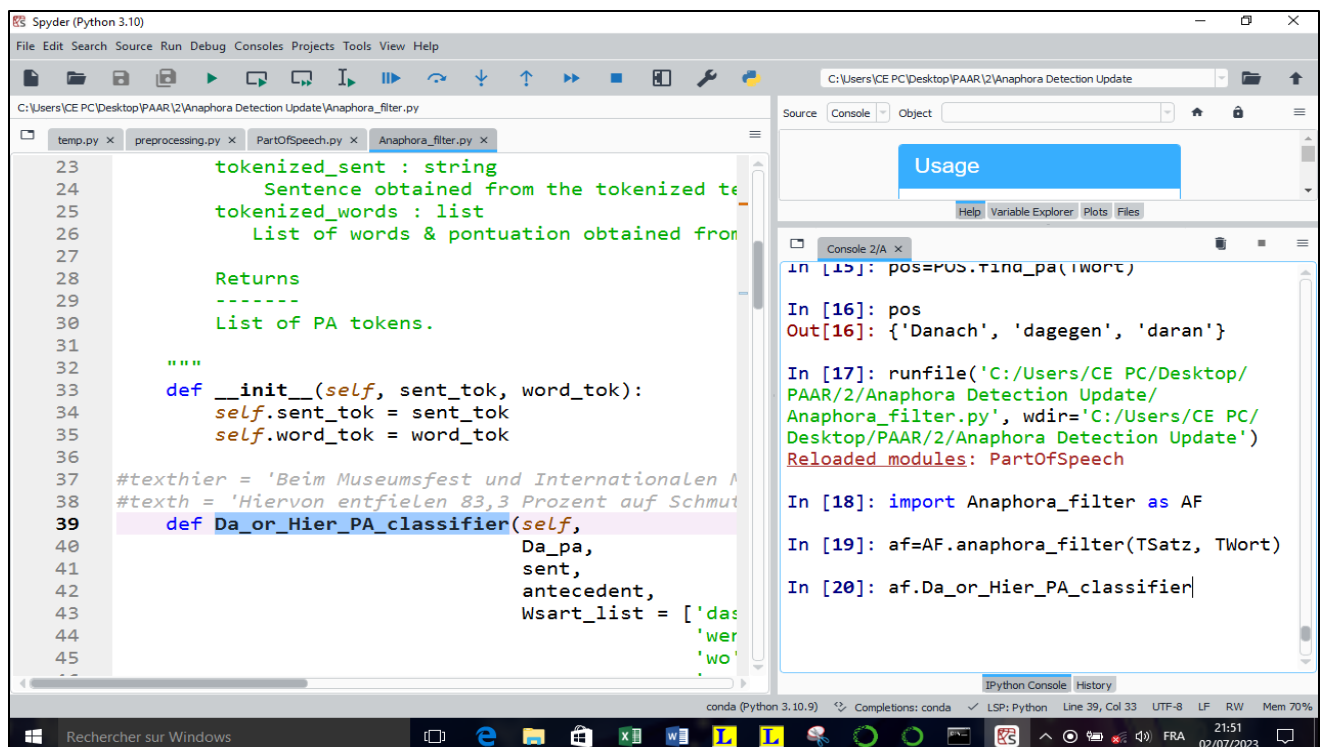


Abbildung 45: Aufruf der Klasse „anaphora filter“ und der Methode „Da or Hier PA classifrier“

Ab diesem Moment kommen wir nun zum Punkt an, an dem die jeweiligen Antezedenzen der im (T) aufgefundenen PA („Danach“, „dagegen“, „daran“) verzeichnet werden müssen (Auflösung von anaphorischen PA). Und da diese Herausforderung nur erheben werden kann, wenn all die 4 zuvor hochgefahrenen Module zusammen exekutiert werden, wollen wir nun das ganze **System PAAR** aufrufen (Abbildungen 41) und es zugleich einsetzen. (Abbildungen 42-44)

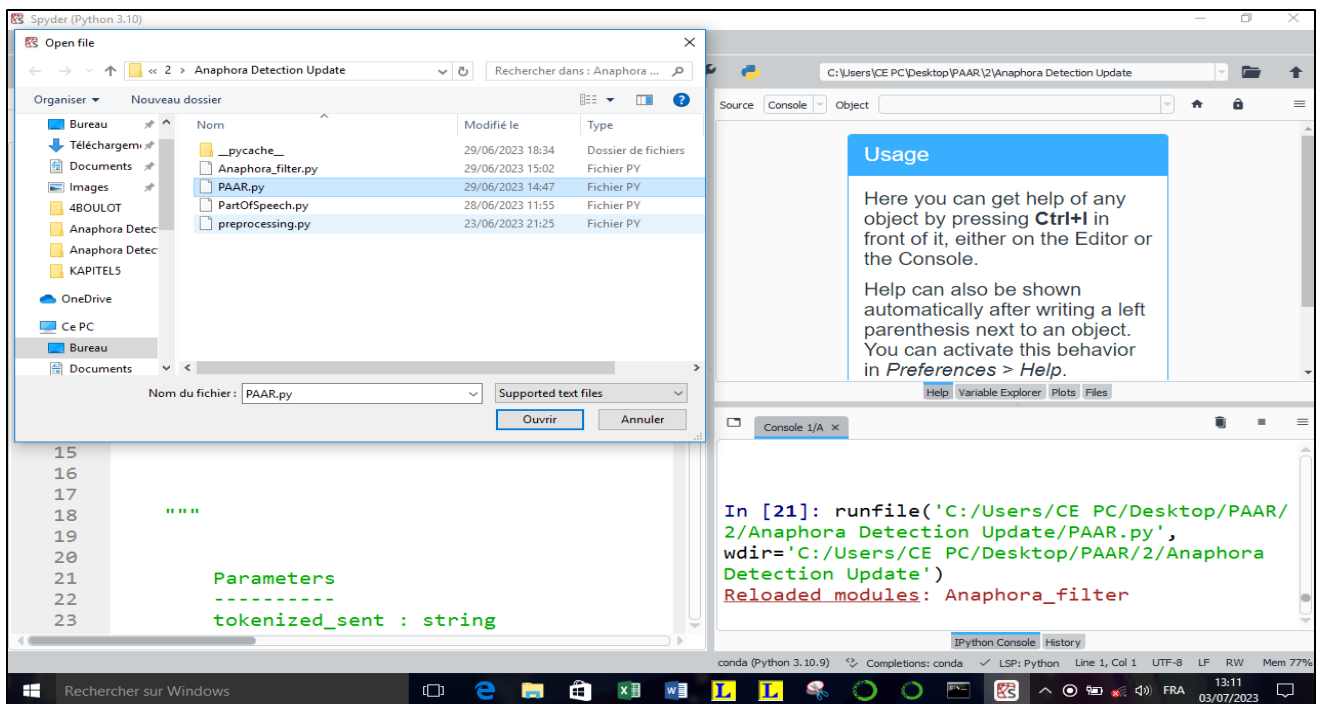


Abbildung 46: Aufruf des Hauptmoduls: Aufruf vom PAAR

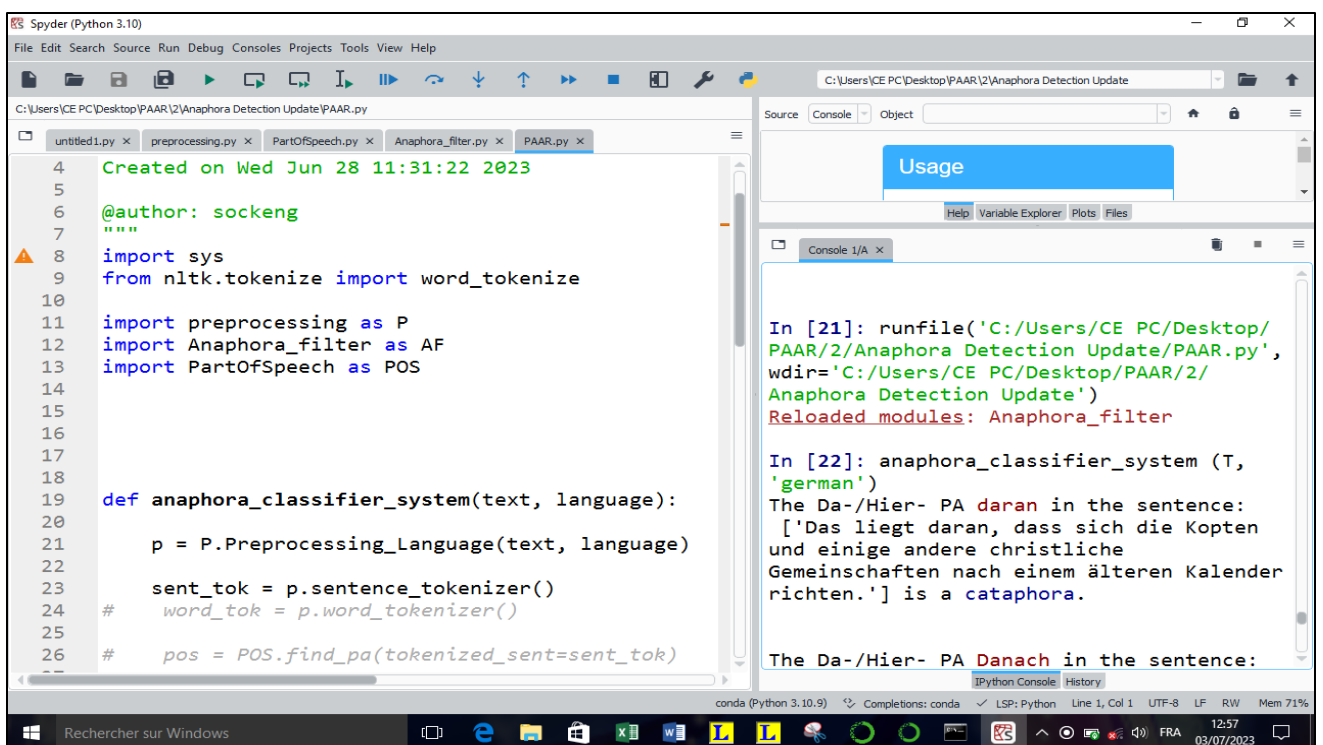


Abbildung 47: Einsatz vom PAAR zur Auflösung der erfundenen PA (1/3)

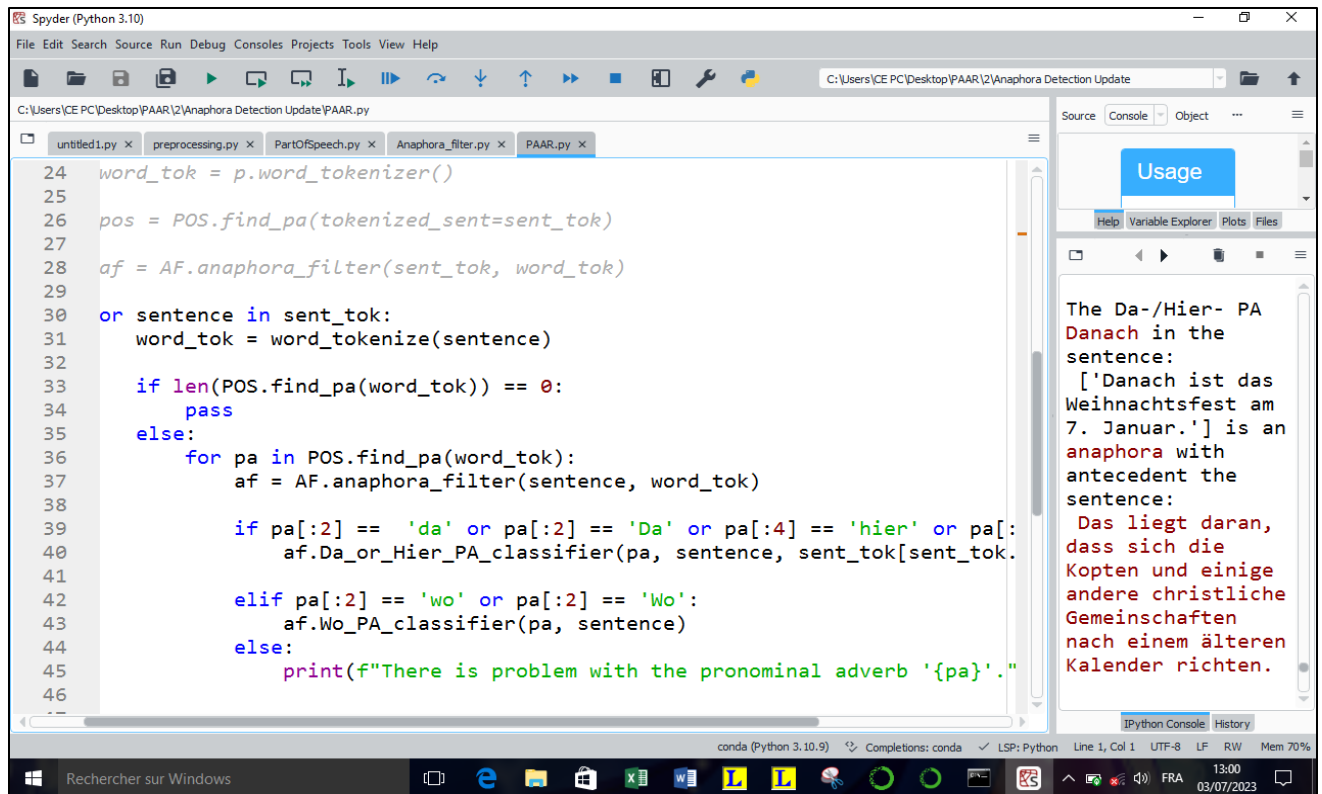


Abbildung 48: Einsatz vom PAAR zur Auflösung der erfundenen PA (2/3)

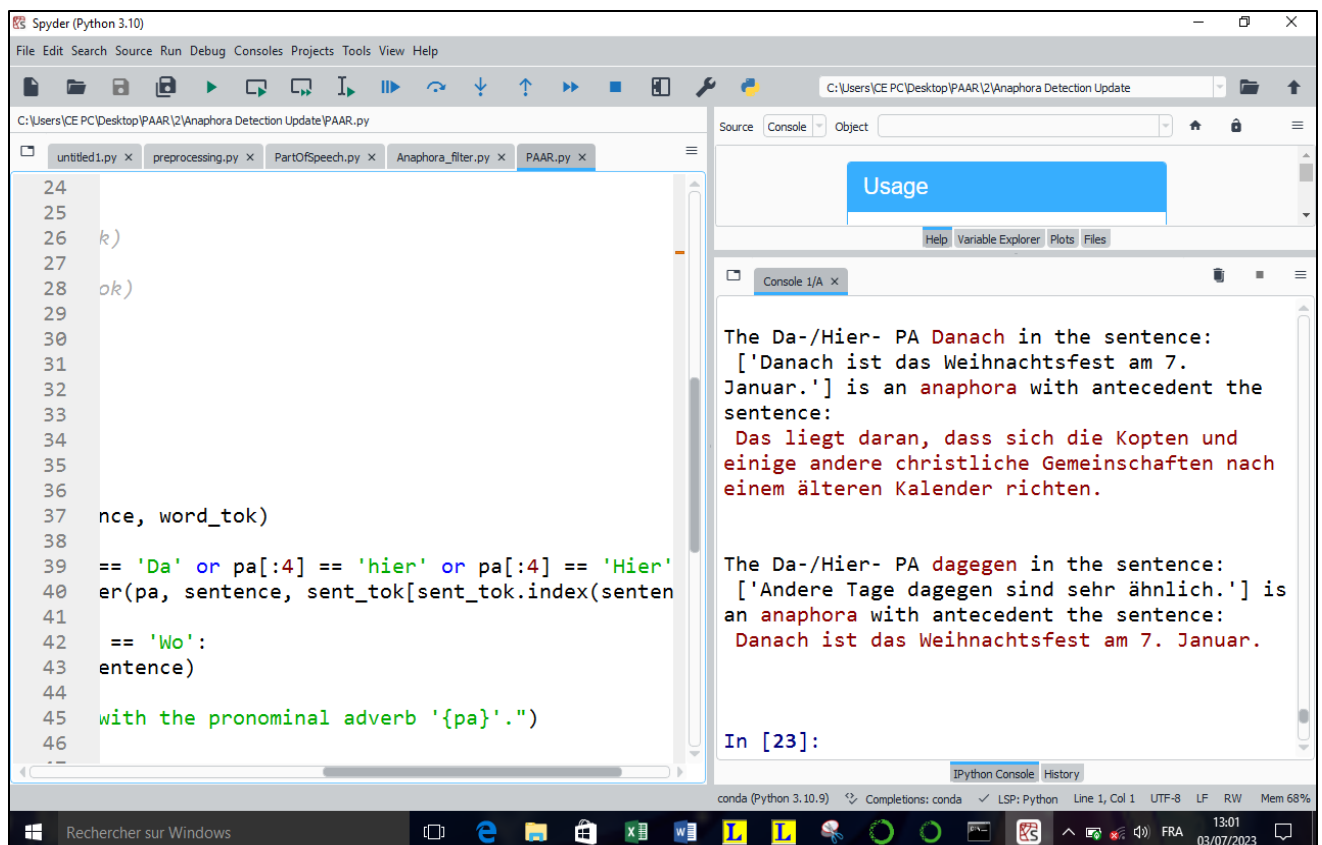


Abbildung 49: Einsatz vom PAAR zur Auflösung der erfundenen PA (3/3)

Aus den Abbildungen (40-42) steht es nämlich fest, dass PAAR sowohl die Natur als auch, bei gegebenen Fällen, die Antezedenzen der jeweiligen PA richtig wiedergegeben hat. Bevor wir nähere Angaben über die Analyse und Interpretationen der Ergebnisse vom PAAR bezüglich des Texts N°1 vom PAK geben, wollen wir zunächst noch einmal diesen Inputtext wiedergeben.

- 1- *"Schade, Weihnachten ist schon vorbei." Das denken bei uns gerade einige Leute. Für andere Menschen steht das Weihnachtsfest allerdings kurz bevor. Etwa für die Kopten, die in Ägypten leben. Sie feiern Weihnachten erst Ende dieser Woche, am 7. Januar. Das liegt **darán**, dass sich die Kopten und einige andere christliche Gemeinschaften nach einem älteren Kalender richten. **Danach** ist das Weihnachtsfest am 7. Januar. Andere Tage **dagegen** sind sehr ähnlich. Sonntags zum Beispiel gehen viele Kopten zum Gottesdienst in eine Kirche. Einige Kinder besuchen an dem Tag auch spezielle Schulen, die Sonntagsschulen. Dort lernen sie etwas über ihre Religion. dpa (SAZ11/JAN.00262 Saale Zeitung, 04.01.2011, S. 30, Originalressort: Kinderseite; ohne Titel)*

Bei Näherbetrachtung dieses Textes stellen wir zunächst nämlich fest, dass er, wie PAAR es herausgefunden hat, nur 3 PA enthält. Weiterhin können wir auch feststellen, dass die Bezugseinheit des PA „darán“ tatsächlich vorne liegt. Das heißt, das PA „darán“ im Text N°1 verweist auf sprachliche Einheiten, die, von seiner Position aus, vorwärts stehen. Daher ist „darán“ hier eine Katapher. Diesem entgegen, liegt die Bezugseinheit des PA „Danach“ Rückwärts und ist nämlich die Präpositionalphrase „nach einem älteren Kalender“. Hier merken wir nämlich, dass das System PAAR zwar die Natur des PA richtig wiedergegeben hat (anaphorisches PA); aber was das Antezedens betrifft, hat PAAR die Antwort nur zum Teil richtig gefunden. Und zwar weil das PA „danach“ seinen Bezug nicht auf den ganzen vorherigen Satz, wie PAAR es meint, nimmt (Siehe Hierfür Abbildung 44) sondern eher, wie zuvor erwähnt nur auf die Präpositionalphrase „nach einem älteren Kalender“. Daher werden wir PAAR bezüglich dieses Mangels nur einen halben Punkt geben. (Mehr dazu in den folgenden Tabellen)

Was das PA „dagegen“ betrifft, stellen wir, nach einer gründlichen Analyse, fest, dass es sich auch um ein anaphorisches PA, mit dem Bezug auf alle rückwärtsliegenden Aussagen, die auf den Tag des Weinachtfest bei Kopten referieren, handelt. Allerdings stellen wir weiterhin auch

fest, dass diese Information mehrmals im vorherigen Textteil vorkommt. PAAR entschied sich aber für den unmittelbar stehenden vorherigen Satz (Siehe Hierfür Abbildung 44). Etwas das grundsätzlich betrachtet, weder falsch noch völlig richtig ist. Denn das Verständnis dieses ausgewählte Antezedens (*Danach ist das Weihnachtsfest am 7. Januar.*) erfordert unbedingt den Aufruf von weiteren hinterlegenden Sätzen, In dieser Hinsicht wollen wir behaupten, dass PAAR das PA „dagegen“ richtig aufgelöst hat. Dennoch behaupten wir noch gleichzeitig, dass der Textteil „Schade, Weinachten..... Weinachfest am 7. Januar“ ein umfassendes Antezedens wäre.

Zusammenfassend, können wir sagen, dass das System PAAR bezüglich der Bearbeitung des Texts N°1 die folgenden Leistungen aufweist.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
3	3	3/3

Tabelle 20: Performanz vom PAAR bei Erkennung von PA im Text N°1

PA	Anzahl	Natur	Ergebnisse
Daran	1	Katapher (Keine Anapher)	1/1
Danach	1	Anapher	1/1
Dagegen	1	Anapher	1/1
TOTAL	3		3/3

Tabelle 21: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°1

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Danach	0,5/1
Dagegen	1/1
TOTAL	1,5/2

Tabelle 22: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N

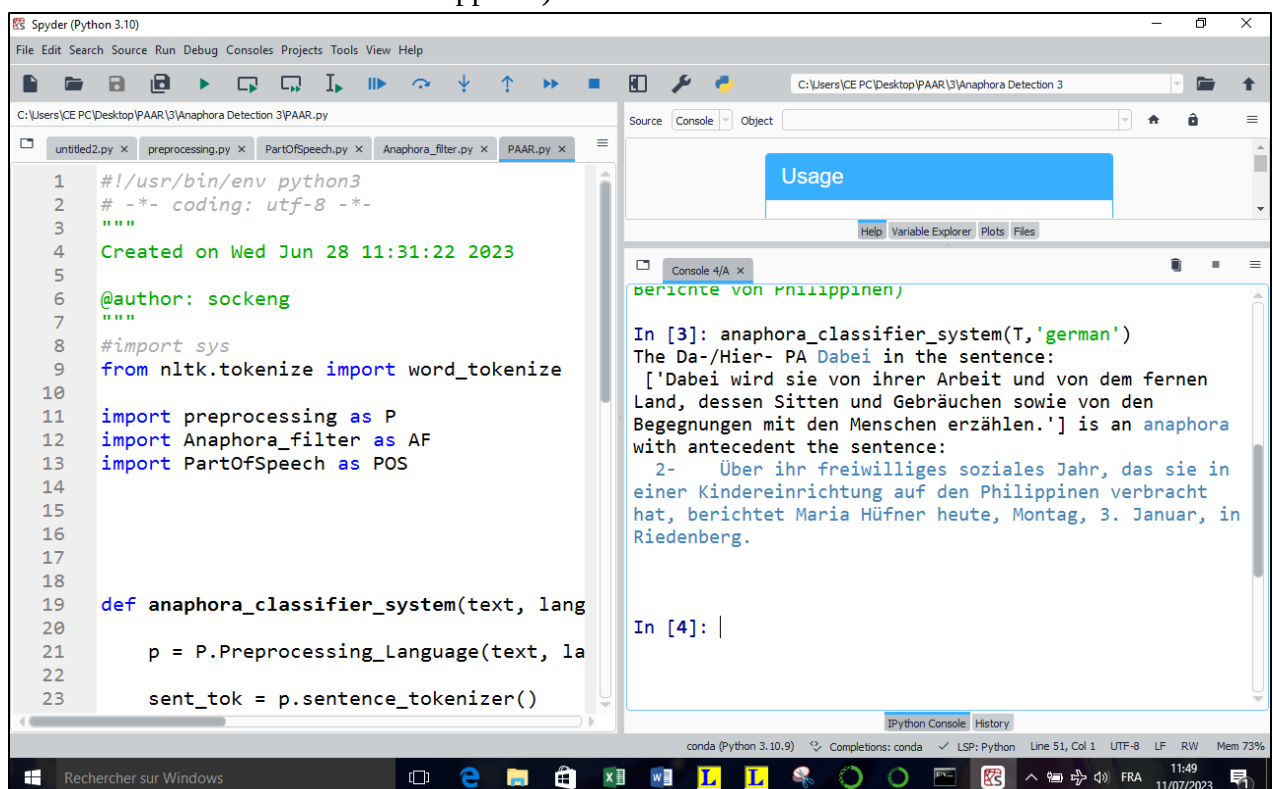
KAPITEL 6: AUSWERTUNG DES PAAR-EINSATZES: ANALYSEN, INTERPRETATIONEN, ERGEBNISSE UND HYPOTHESENPRÜFUNG.

Die maximale Zielsetzung dieses Kapitels ist es, einerseits, das entwickelte System zur Auflösung von Pronominaladverbienanaphern (PAAR) anhand Beispieltex te zu testen, und andererseits, seine Performanz zu etablieren. Dafür werden wir nämlich die Angaben, die wir nach dem PAAR-Einsatz gewonnen haben, gründlich analysieren und interpretieren. Somit werden wir in die Lage versetzt, die voretablierten Hypothesen prüfen zu können. In einem weiteren Schritt werden wir Schlüsse bezüglich der Relevanz der Arbeit sowohl für die Grundlageforschung als auch für das alltägliche Leben ziehen. Und letztendlich werden Ansätze zur Verbesserung des Systems formuliert sowie weitere zukünftige Forschungsperspektive aufgezählt.

6.1. Analyse und Interpretation des Einsatzes vom PAAR auf das Textkorpus

6.1.1. EINSATZ DES PAAR AUF DEN TEXT N°2

- 2- Über ihr freiwilliges soziales Jahr, das sie in einer Kindereinrichtung auf den Philippinen verbracht hat, berichtet Maria Hüfner heute, Montag, 3. Januar, in Riedenberg. *Dabei* wird sie von ihrer Arbeit und von dem fernen Land, dessen Sitten und Gebräuchen sowie von den Begegnungen mit den Menschen erzählen. Beginn ist um 19 Uhr im Kuratiesaal. *red* (SAZ11/JAN.00003 Saale-Zeitung, 03.01.2011, S. 14; Bilder und Berichte von Philippinen)



The screenshot shows the Spyder Python IDE interface. The left pane displays the PAAR.py script, which includes a shebang, encoding declaration, creation date, author information, and imports for sys, nltk.tokenize, preprocessing, Anaphora_filter, and PartOfSpeech. The right pane shows the console output, which includes the usage of the anaphora_classifier_system function and the analysis of the sentence: "Dabei wird sie von ihrer Arbeit und von dem fernen Land, dessen Sitten und Gebräuchen sowie von den Begegnungen mit den Menschen erzählen." The console identifies "Dabei" as an anaphora with the antecedent sentence: "Über ihr freiwilliges soziales Jahr, das sie in einer Kindereinrichtung auf den Philippinen verbracht hat, berichtet Maria Hüfner heute, Montag, 3. Januar, in Riedenberg."

```
1 #!/usr/bin/env python3
2 # -*- coding: utf-8 -*-
3 """
4 Created on Wed Jun 28 11:31:22 2023
5
6 @author: sockeng
7 """
8 #import sys
9 from nltk.tokenize import word_tokenize
10
11 import preprocessing as P
12 import Anaphora_filter as AF
13 import PartOfSpeech as POS
14
15
16
17
18
19 def anaphora_classifier_system(text, lang
20
21     p = P.Preprocessing_Language(text, la
22
23     sent_tok = p.sentence_tokenizer()
```

Usage

berichte von Philippinen)

In [3]: anaphora_classifier_system(T,'german')
The Da-/Hier- PA Dabei in the sentence:
['Dabei wird sie von ihrer Arbeit und von dem fernen Land, dessen Sitten und Gebräuchen sowie von den Begegnungen mit den Menschen erzählen.'] is an anaphora with antecedent the sentence:
2- Über ihr freiwilliges soziales Jahr, das sie in einer Kindereinrichtung auf den Philippinen verbracht hat, berichtet Maria Hüfner heute, Montag, 3. Januar, in Riedenberg.

In [4]: |

Abbildung 50: Einsatz des PAAR auf den Text N°2

Nach dem Einsatz vom PAAR auf den Text N°2 ergibt es sich, dass der Text nur aus einem PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR dieses PA als anaphorisch mit dem Textteil „Über Riedenberg“ als Antezedens. Etwas das auch richtig ist; denn das PA „Dabei“ im Text 2 referiert auf die Handlung, die durch diesen Textteil verschriftlicht wurde. In dieser Hinsicht fassen wir folgende Angaben über den Einsatz vom PAAR auf den Text N°2

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
1	1	1/1

Tabelle 23 Performanz vom PAAR bei der Erkennung von PA im Text N°2

PA	Anzahl	Natur	Ergebnisse
Dabei	1	Anapher	1/1
TOTAL	1		1/1

Tabelle 24: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°2

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dabei	1/1
TOTAL	1/1

Tabelle 25: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°2

6.1.2. EINSATZ DES PAAR AUF DEN TEXT N°3

3- *Mechanisch Ob kleiner Wecker oder große Kirchturmuh — mechanische Uhren funktionieren im Prinzip gleich: in der Uhr arbeitet ein Räderwerk, ein Set aus verschiedenen Zahnrädern. Die werden in Bewegung gesetzt. Gewicht Bei großen Uhren übernimmt in der Regel ein Gewicht diese Aufgabe. Bei einer Armbanduhr geschieht das meist, indem man außen an der Uhr ein Rädchen dreht. Man zieht die Uhr auf. Dadurch wird im Gehäuse eine Feder gespannt, die neue Energie liefert.* (SAZ11/JAN.00119 Saale-Zeitung, 03.01.2011, S. 26; Uhrenkunde)

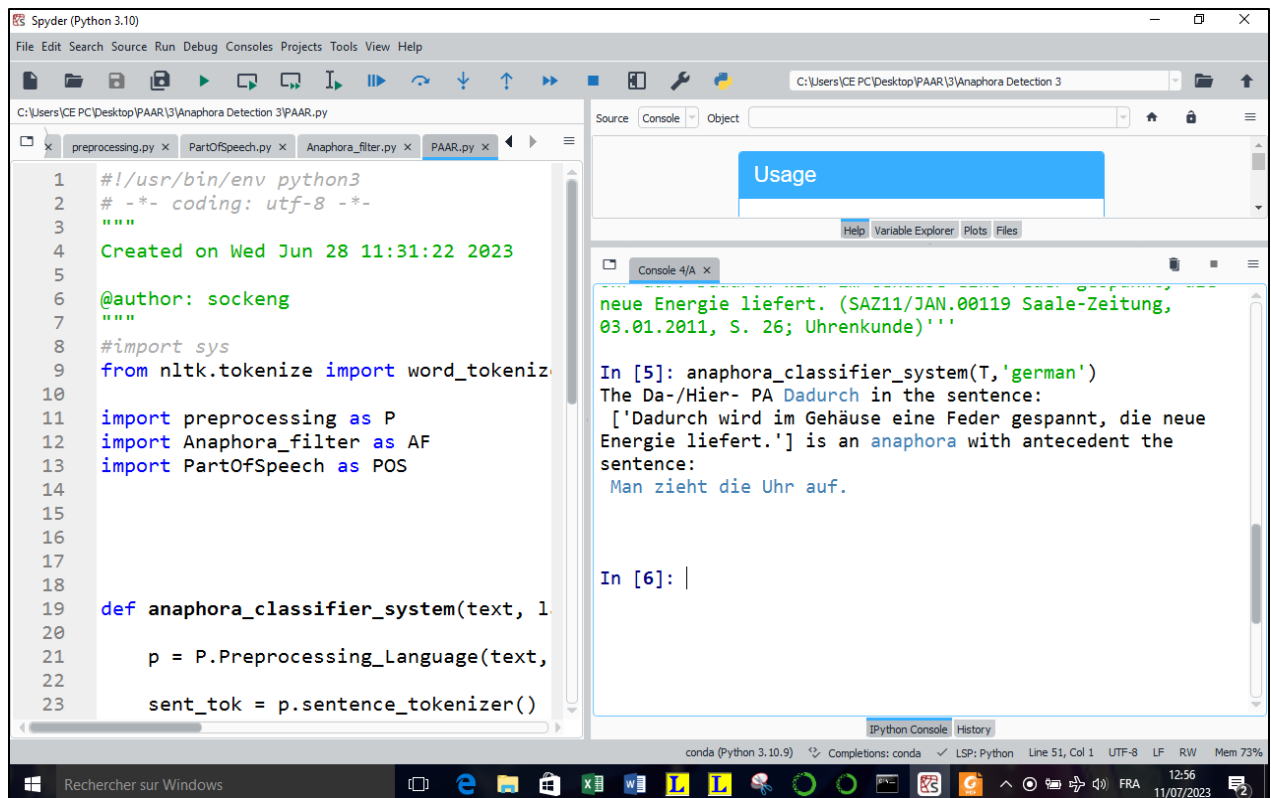


Abbildung 51: Einsatz des PAAR auf den Text N°3

Nach dem Einsatz vom PAAR auf den Text N°3 ergibt es sich, dass der Text nur aus einem PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR dieses PA als anaphorisch mit dem Textteil „Man zieht die Uhr auf“ als Antezedens. Etwas das auch richtig ist, denn das PA „Dadurch“ im Text 3 referiert auf die Handlung, die durch diesen Textteil verschriftlicht wurde. In dieser Hinsicht fassen wir folgende Angaben über den Einsatz vom PAAR auf den Text N°3

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
1	1	1/1

Tabelle 26 Performanz vom PAAR bei der Erkennung von PA im Text N°3

PA	Anzahl	Natur	Ergebnisse
Dadurch	1	Anapher	1/1
TOTAL	1		1/1

Tabelle 27: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°3

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dadurch	1/1
TOTAL	1/1

Tabelle 28: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°3

6.1.3. EINSATZ DES PAAR AUF DEN TEXT N°4

- 4- Die Atomuhr ist die genaueste Uhr, die bisher erfunden wurden. Atome sind die kleinsten Bausteine der Natur. Alles besteht aus Atomen — wir auch. Atomuhren bringen bestimmte Atome *dazu*, dass sie extrem oft ihren Energiezustand ändern. *Dadurch* schwingen die Atome pro Sekunde knapp zehn Milliarden Mal regelmäßig hin und her. Diese Schwingungen werden gezählt. Daher kann die Atomuhr genau angeben, wann eine Sekunde vorbei ist. Zum Vergleich: Eine Pendeluhr schwingt einmal pro Sekunde hin und her. Wenn eine Schwingung unregelmäßig ist, wirkt sich das viel stärker aus. dpa (SAZ11/JAN.00119 Saale-Zeitung, 03.01.2011,S.26;Uhrenkunde)

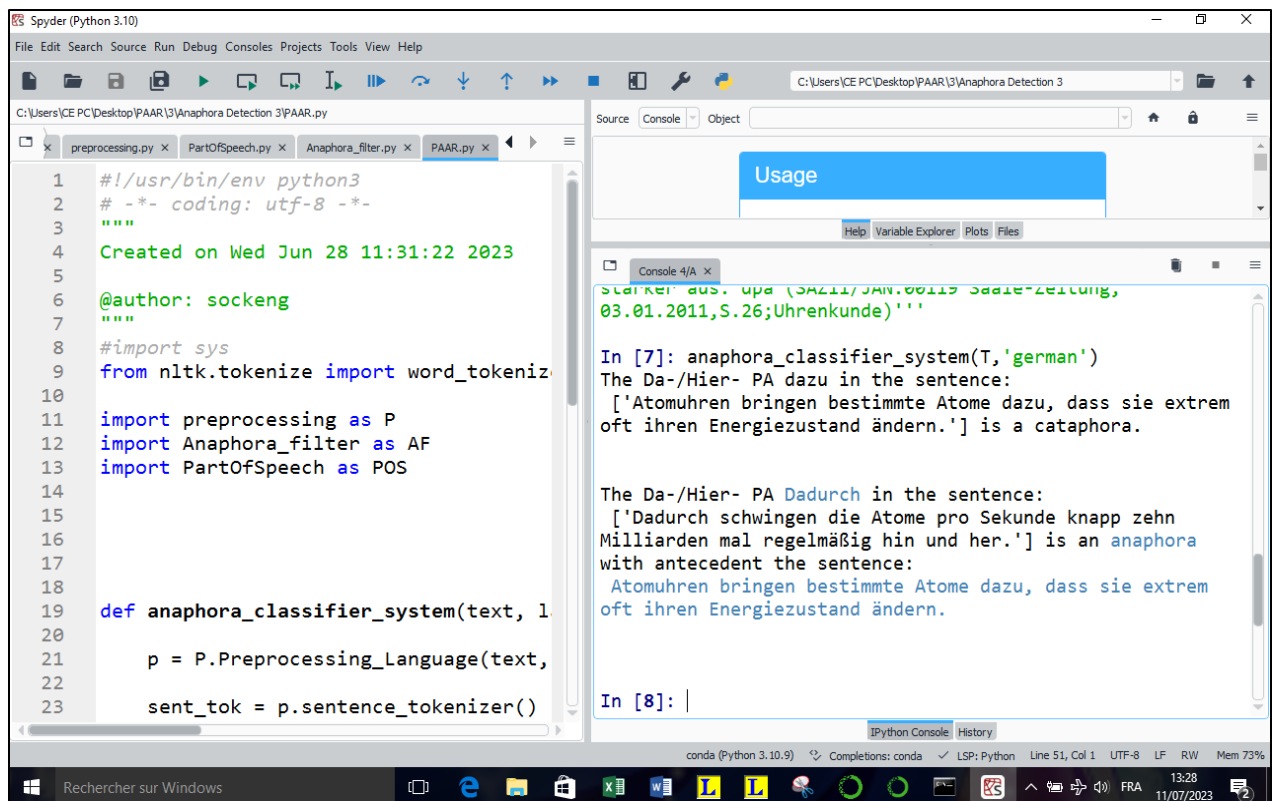


Abbildung 52: Einsatz des PAAR auf den Text N°4

Nach dem Einsatz vom PAAR auf den Text N°4 ergibt es sich, dass der Text aus zwei PAn besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR das erste erkannte PA („Dazu“) als eine Katapher; etwas, das auch richtig ist, denn „Dazu“ verweist hier auf vorwärtsstehende sprachliche Einheiten. Was das zweite PA („Dadurch“) betrifft, klassifiziert PAAR es als eine Anapher mit dem Textteil „Atomuhren bringen bestimmte Atome *dazu*, dass sie extrem oft ihren Energiezustand ändern.“ als Antezedens. Etwas das auch richtig ist, denn das PA „Dadurch“ im Text 4 referiert auf die Handlung, die durch diesen Textteil verschriftlicht wurde. In dieser Hinsicht fassen wir folgende Angaben über den Einsatz vom PAAR auf den Text N°4

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2	2	2/2

Tabelle 29 Performanz vom PAAR bei der Erkennung von PA im Text N°4

PA	Anzahl	Natur	Ergebnisse
Dazu	1	Katapher	1/1

Dadurch	1	Anapher	1/1
TOTAL	2		2/2

Tabelle 30: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°4

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dadurch	1/1
TOTAL	1/1

Tabelle 31: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°4

6.1.4. EINSATZ DES PAAR AUF DEN TEXT N°5

- 5- *Damit* frische Blüten und Blätter sich entfalten können, müssen vergilbte Pflanzenteile entfernt werden. "*Dazu* darf keine Schere verwendet werden", warnt Legrand. "Auch wenn es schwer fällt: Reißen Sie alte Blüten und gelbe Blätter direkt an der Knolle mit einer drehenden Bewegung und einem kräftigen Ruck nach unten raus." *Dadurch* wird vermieden, dass sich Fäulnis bildet. (SAZ11/JAN.00258 SaaleZeitung, 04.01.2011, S. 17; Kühler Platz für bunte Blumen)

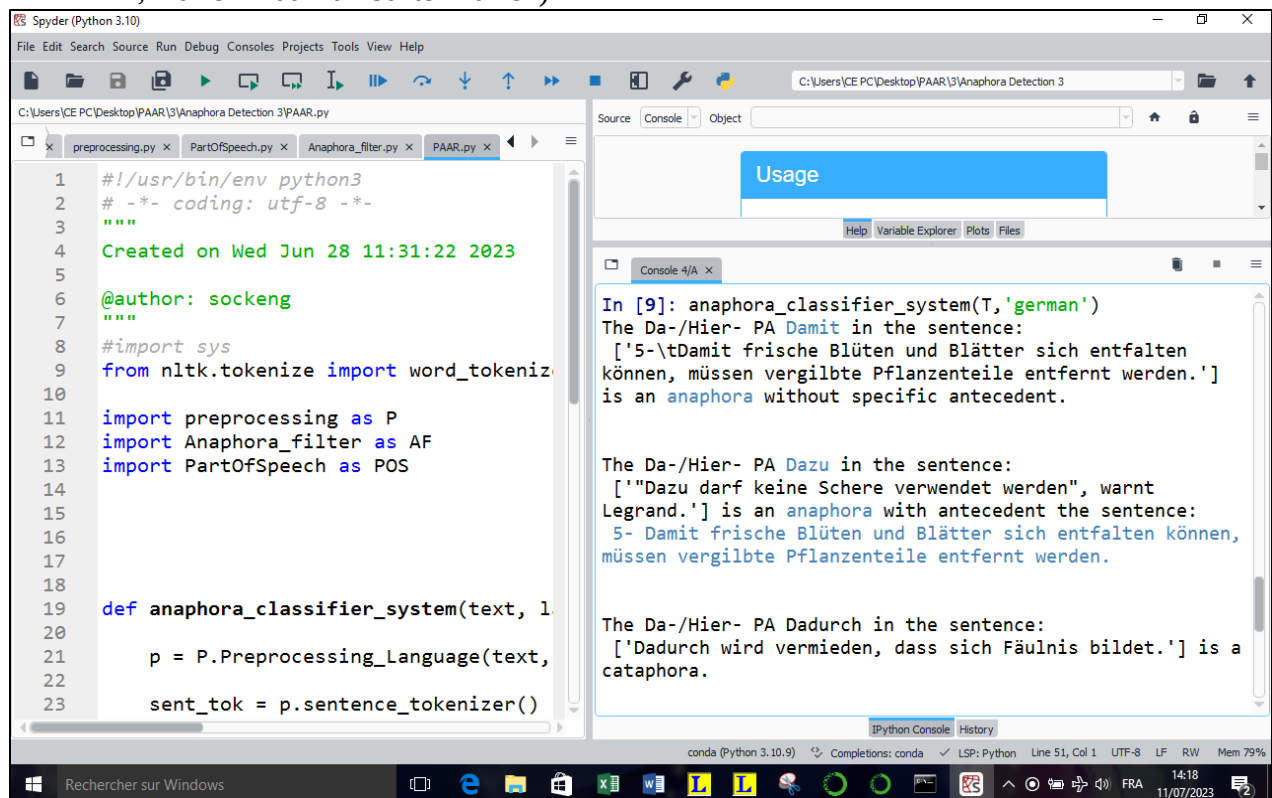


Abbildung 53: Einsatz des PAAR auf den Text N°5

Nach dem Einsatz vom PAAR auf den Text N°5 ergibt es sich, dass der Text aus drei PA besteht. Etwas, das nach Näherbetrachtung des Textes nicht richtig ist. Denn der Text N°5 besteht zwar aus 3 Wörtern, die zur Kategorie der Da-Erscheinungen gehören. Dennoch erweisen nur 2 davon die Natur eines Pronominaladverbs („Dazu“ und „Dadurch“). Die letzte von den drei Da-Erscheinungen („Damit“) ist eher eine Subjunktion und zwar, weil sie erfordert

ein Verb am Positionsende; würde sie aber als PA erscheinen dann würde das Verb eher in zweiter Position stehen müssen. Wir können also sagen, dass PAAR 3 PA im Text N°5 erkannt hat, obwohl dieser nur aus 2 PA besteht (Siehe dazu Tabelle 30).

Was die Natur der bestehenden PA angeht, PAAR verzeichnet das PA „Dazu“ als ein anaphorisches PA mit dem Textteil „*Damit frische Blüten und Blätter sich entfalten können, müssen vergilbte Pflanzenteile entfernt werden.*“ als Antezedens. Etwas, das eigentlich richtig ist, denn das PA „Dazu“ verweist hier auf rückwärtsstehende sprachliche Einheiten. Was das PA „Dadurch“ angeht, wurde dieses vom PAAR leider als kein anaphorisches PA verzeichnet und dies obwohl „Dadurch“ hier auf eine Handlung referiert, die zuvor im vorangehenden Textteil schon versprochen worden ist. Angesichts all dieser Angaben wollen wir durch die folgenden Tabellen die Leistungen des PAAR bezüglich der Verarbeitung des Textes N°5 zusammenfassen.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2 PA	3 (2PA+ 1 SUBJUNKTION)	3/2 ??

Tabelle 32 Performanz vom PAAR bei der Erkennung von PA im Text N°5

PA	Anzahl	Natur	Ergebnisse
Dazu	1	Anapher	1/1
Dadurch	1	Katapher	0/1
TOTAL	2		1/2

Tabelle 33: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°5

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dazu	1/1
TOTAL	1/1

Tabelle 34: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°5

6.1.5. EINSATZ DES PAAR AUF DEN TEXT N°6

- 6- Die niederländischen Wissenschaftler, die dem Phänomen einen Namen gegeben haben, stellten fest, dass "Leisure sickness" vor allem Menschen mit hohem Arbeitspensum und Verantwortungsgefühl betrifft. Psychologen machen auch Bewegungsmangel und einen hohen Medienkonsum für die "Freizeit-Krankheit" verantwortlich. "Menschen müssen wieder mehr Sorge *dafür* tragen, dass der Körper Entspannung findet", meint Psychotherapeut Soljan. Allerdings sollte man analysieren, was wirklich guttut. "Nicht bloß fernsehen", rät der Experte. Stattdessen solle man lieber in Ruhe ein Bad nehmen, einen Tee trinken oder mal wieder nett den Frühstückstisch decken. Soljan hält die Entwicklung des Phänomens "Leisure sickness" für gravierend: "Wenn man den Prozess nicht unterbricht, kann es im schlechtesten Fall zu psychosomatischen Erkrankungen kommen." (SAZ11/JAN.00276 Saale-Zeitung, 04.01.2011, S. 34; Mit der Ruhe kommt die)

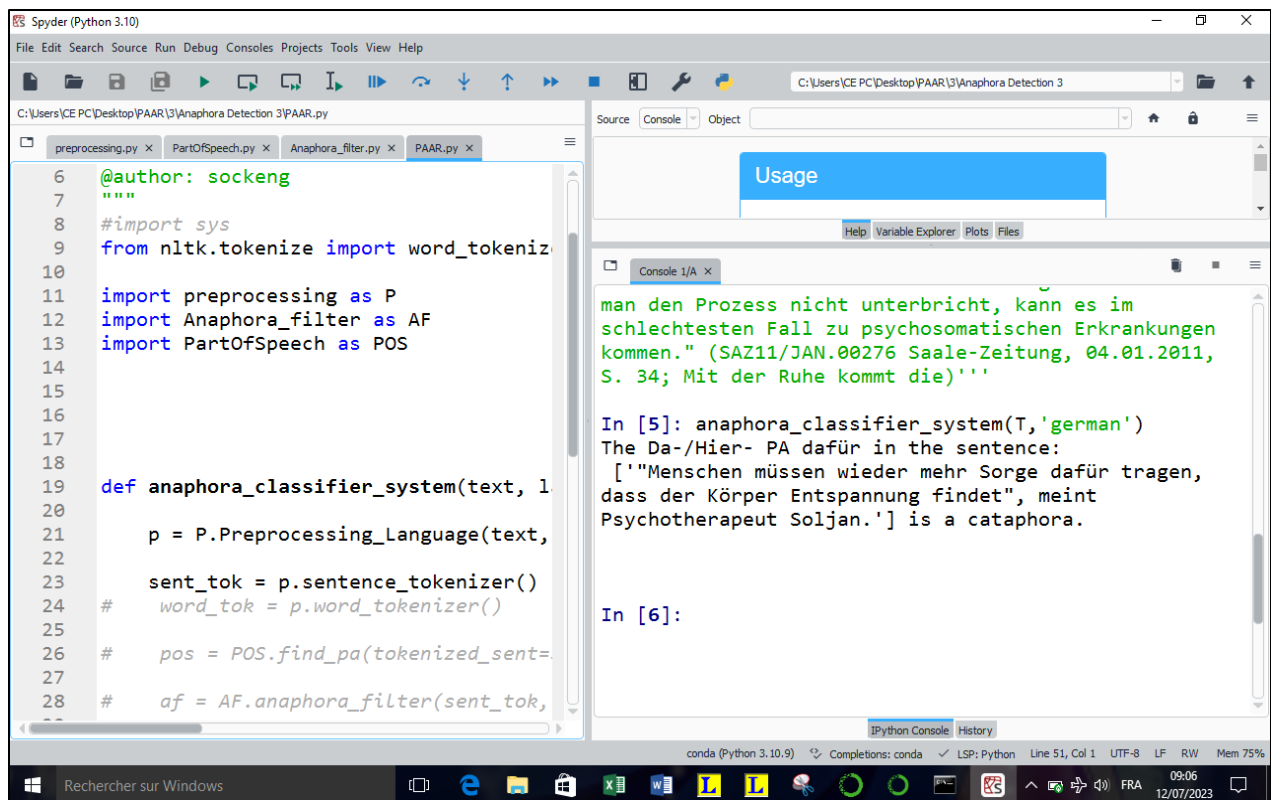


Abbildung 54: Einsatz des PAAR auf den Text N°6

Nach dem Einsatz vom PAAR auf den Text N°6 ergibt es sich, dass der Text aus einem PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR dieses PA („dafür“) als eine Katapher; etwas, das auch richtig ist, denn „dafür“ verweist hier auf vorwärtsstehende sprachliche Einheiten. In dieser Hinsicht fassen wir folgende Angaben über den Einsatz vom PAAR auf den Text N°6

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
1	1	1/1

Tabelle 35 Performanz vom PAAR bei der Erkennung von PA im Text N°6

PA	Anzahl	Natur	Ergebnisse
Dafür	1	Katapher	1/1
TOTAL	1		1/1

Tabelle 36: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°6

6.1.6. EINSATZ DES PAAR AUF DEN TEXT N°7

- 7- Martin Schmitt stürzte von Wolke sieben ins Jammertal, "Überflieger" Simon Ammann träumte nach seinem Höhenflug beim chaotischen Neujahrsspringen *dagegen* schon vom ersten Vierschanzentournee-Triumph. *Dafür* rangen Österreichs vom Winde verwehte Ski-Adler nach der schweren Schlappe in Garmisch-Partenkirchen mühsam um ihre Fassung, die Schmitt trotz seines Ausscheidens in der Qualifikation für das dritte Springen in Innsbruck nicht verlor. (SAZ11/JAN.00011 Saale-Zeitung, 03.01.2011, S. 16; Schmitt nach Hoch im Jammertal)

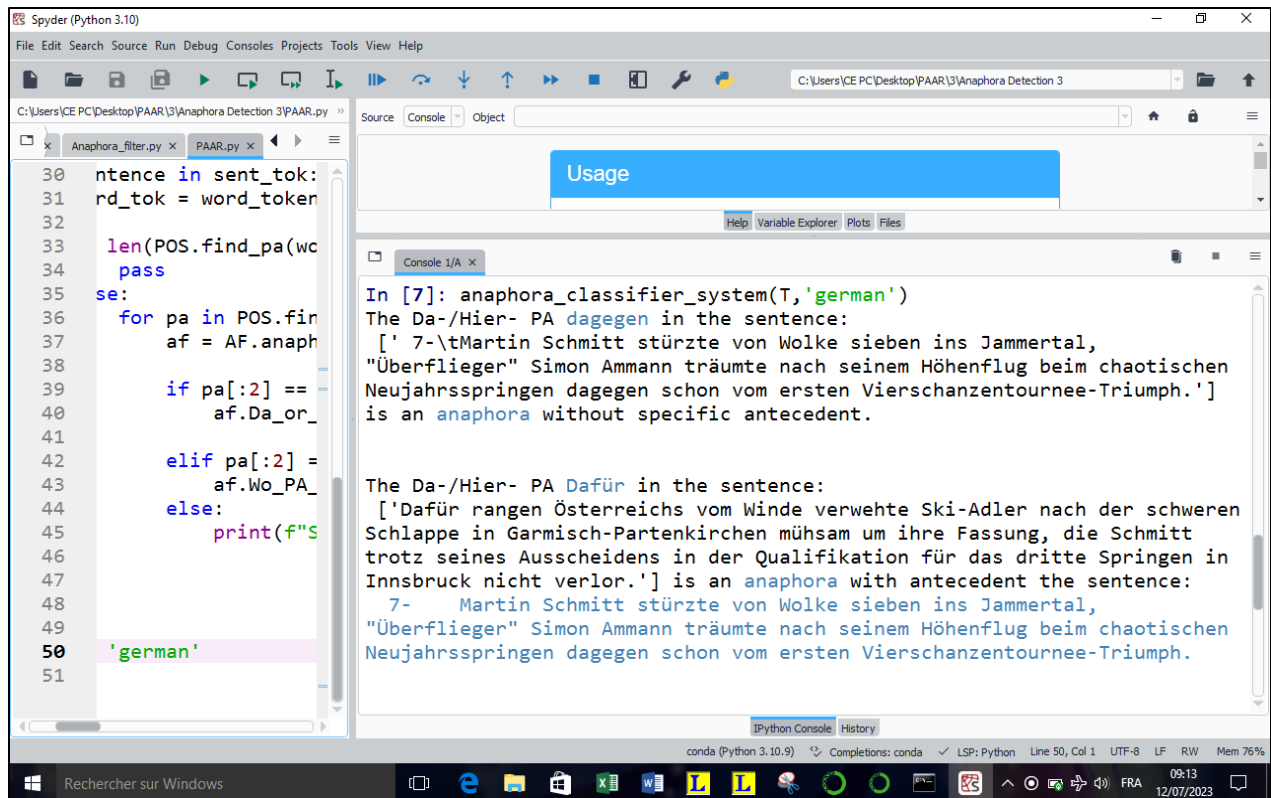


Abbildung 55: Einsatz des PAAR auf den Text N°7

Nach dem Einsatz vom PAAR auf den Text N°7 ergibt es sich, dass der Text aus zwei PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR das erste erkannte PA („dagegen“) als eine Anapher ohne spezifiziertes Antezedens. Etwas, das nur zum Teil richtig ist, denn das PA „dagegen“ ist in der Tat eine Anapher und verweist auf den rückwärtsstehenden Textteil „*Martin Schmitt stürzte von Wolke sieben ins Jammertal*“. In dieser Hinsicht werden wir PAAR keine Punkte bezüglich der Auflösung dieses PA geben (Siehe hierzu Tabelle 37). Was das zweite erkannte PA betrifft, klassifiziert PAAR es als eine Anapher mit dem Bezug auf den Satz „*Martin Schmitt stürzte von Wolke sieben ins Jammertal, "Überflieger" Simon Ammann träumte nach seinem Höhenflug beim chaotischen Neujahrsspringen* **dagegen** schon vom ersten Vierschanzentournee-Triumph.“. Etwas, das auch nur zum Teil richtig ist, denn das anaphorische PA „Dafür“ referiert hier auf die Informationen, die zuvor durch den Textteil „*"Überflieger" Simon Ammann träumte nach seinem Höhenflug beim chaotischen Neujahrsspringen* **dagegen** schon vom ersten Vierschanzentournee-Triumph“ versprochen worden ist. Allerdings steht es nämlich fest, dass man den kompletten Informationsgehalt dieses Textteils nur entschlüsseln kann, wenn man das PA „dagegen“ schon zuvor aufgelöst hat. In dieser Hinsicht können wir sagen, dass PAAR auch hier das Antezedens vom „Dafür“ richtig auserwählt hat. Hier eine Zusammenfassung dieser Einsatzaktion.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2	2	2/2

Tabelle 37: Performanz vom PAAR bei Erkennung von PA im Text N°7

PA	Anzahl	Natur	Ergebnisse
dagegen	1	Anapher	1/1
Dafür	1	Anapher	1/1

TOTAL	2	2/2
-------	---	-----

Tabelle 38: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°7

PA	Noten der Auflösung von anaphorischen PA durch PAAR
dagegen	0/1
Dafür	1/1
TOTAL	1/2

Tabelle 39: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°7

6.1.7. EINSATZ DES PAAR AUF DEN TEXT N°8

- 8- Bei den Weltcup-Rennen in Adelboden (Herren) und Altenmarkt-Zauchensee (Damen) nimmt der Deutsche Skiverband wieder Podestplätze in Angriff. *Daran* änderte auch Rang 16 von Neureuther am Donnerstagabend im Flutlicht beim Sieg des Schweden Andre Myhrer nichts. "Vom Selbstvertrauen war das Rennen trotzdem gut. Ich weiß, ich bin gut *drauf* und lass' mich nicht rausbringen", betonte der zweimalige Weltcup-Sieger Neureuther, der im WMWinter noch ohne "Stockerl" ist. "Ich habe mich heute nicht so gut bewegt, das werde ich am Sonntag mit Sicherheit besser machen." Positiv gestimmt reiste er nach Adelboden — neben Garmisch der einzige Ort, auf dem er schon zweimal auf dem Podest stand. (SAZ11/JAN.00650 Saale-Zeitung, 08.01.2011, S. 22, Sachgebiet: Sport, Originalressort: Sport gesamt; Neureuther hakt Platz 16)

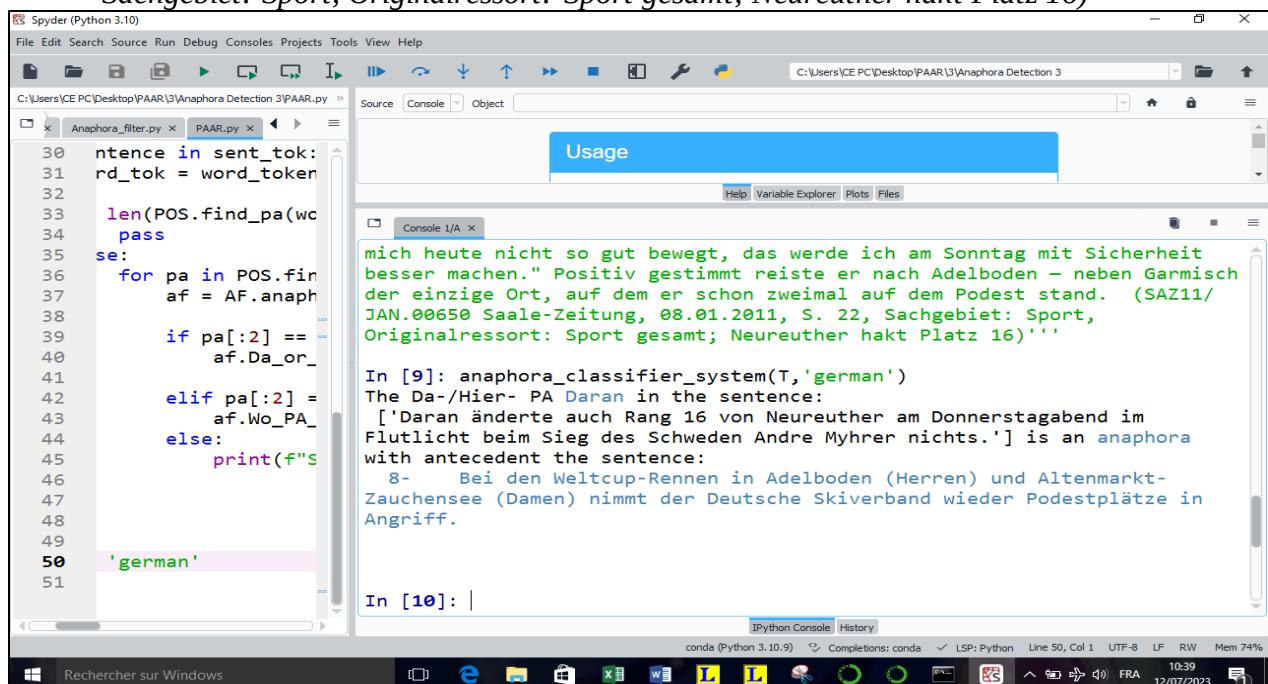


Abbildung 56: Einsatz des PAAR auf den Text N°8

Nach dem Einsatz vom PAAR auf den Text N°8 ergibt es sich, dass der Text aus einem PA besteht, etwas das aber nach Näherbetrachtung des Textes nicht richtig ist. Denn der Text besteht aus zwei PA („Daran“ & „drauf“). Weiterhin klassifiziert PAAR das erste erkannte PA („Daran“) als eine Anapher mit dem Textteil „Bei den Weltcup-Rennen in Adelboden (Herren) und Altenmarkt-Zauchensee (Damen) nimmt der Deutsche Skiverband wieder Podestplätze in Angriff“ als Antezedens. Etwas, das auch nachvollziehbar ist, denn das PA „Daran“ referiert hier auf Informationen, die zuvor durch das Antezedens versprochen worden sind. Die folgenden Tabellen fassen den Einsatz vom PAAR auf den Text N°8 zusammen.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2	1	1/ 2

Tabelle 40: Performanz vom PAAR bei Erkennung von PA im Text N°8

PA	Anzahl	Natur	Ergebnisse
Daran	1	Anapher	1/1
TOTAL	1		1/ 1

Tabelle 41: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°8

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Daran	1/1
TOTAL	1/ 1

Tabelle 42: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°8

6.1.8. EINSATZ DES PAAR AUF DEN TEXT N°9

- 9- Die Bundesbürger müssen 2011 laut DIHK-Schätzung für Reisen wieder tiefer in die Tasche greifen. Gründe für den Preisanstieg seien vor allem die neue Flugverkehrsabgabe und die in vielen Städten beschlossene "Bettensteuer". Besonders Flüge und Pauschalreisen dürften sich verteuern. *Dagegen* blieben die Preise in der Gastronomie weitgehend stabil, sagte der Präsident des Deutschen Industrie- und Handelskammertages (DIHK), Hans Heinrich Driftmann. Insgesamt seien jedoch die Geschäftserwartungen der Branche deutlich besser als noch vor einem Jahr. Nach dem jüngsten DIHK-Tourismusreport wollen rund 40 Prozent der Reisebüros und 29 Prozent der Reiseveranstalter die Kostensteigerungen an ihre Kunden weitergeben. dpa (SAZ11/JAN.00017 Saale-Zeitung, 03.01.2011, S. 18; DIHK:Reisen werden

```

30 sentence in sent_tok:
31 rd_tok = word_token
32
33 len(POS.find_pa(wc
34 pass
35 se:
36 for pa in POS.fin
37 af = AF.anaph
38
39 if pa[:2] ==
40 af.Da_or_
41
42 elif pa[:2] =
43 af.Wo_PA_
44 else:
45 print(f"S
46
47
48
49
50 'german'
51

```

Usage

Console 1/A x

```

stabil, sagte der Präsident des Deutschen Industrie- und
Handelskammertages (DIHK), Hans Heinrich Driftmann. Insgesamt seien
jedoch die Geschäftserwartungen der Branche deutlich besser als noch vor
einem Jahr. Nach dem jüngsten DIHK-Tourismusreport wollen rund 40 Prozent
der Reisebüros und 29 Prozent der Reiseveranstalter die Kostensteigerungen
an ihre Kunden weitergeben. dpa (SAZ11/JAN.00017 Saale-Zeitung,
03.01.2011, S. 18; DIHK:Reisen werden teurer)'''

In [11]: anaphora_classifizier_system(T,'german')
The Da-/Hier- PA Dagegen in the sentence:
['Dagegen blieben die Preise in der Gastronomie weitgehend stabil, sagte
der Präsident des Deutschen Industrie- und Handelskammertages (DIHK),
Hans Heinrich Driftmann.'] is an anaphora with antecedent the sentence:
Besonders Flüge und Pauschalreisen dürften sich verteuern.

In [12]:

```

Abbildung 37: Einsatz des PAAR auf den Text N°9

Nach dem Einsatz vom PAAR auf den Text N°8 ergibt es sich, dass der Text aus einem PA besteht, etwas das auch nach Näherbetrachtung des Textes wahr ist. Weiterhin klassifiziert PAAR das erkannte PA („Dagegen“) als eine Anapher mit dem Textteil „Gründe für den Preisanstieg seien vor allem die neue Flugverkehrsabgabe und die in vielen Städten

beschlossene "Bettensteuer". Besonders Flüge und Pauschalreisen dürften sich verteuern.“ als Antezedens. Etwas, das auch nachvollziehbar ist, denn das PA „Dagegen“ referiert hier auf Informationen, die zuvor durch das Antezedens versprochen worden sind. Die folgenden Tabellen fassen den Einsatz vom PAAR auf den Text N°9 zusammen.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
1	1	1/1

Tabelle 43: Performanz vom PAAR bei Erkennung von PA im Text N°9

PA	Anzahl	Natur	Ergebnisse
Dagegen	1	Anapher	1/1
TOTAL	1		1/1

Tabelle 44: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°9

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dagegen	1/1
TOTAL	1/1

Tabelle 45: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°9

6.1.9. EINSATZ DES PAAR AUF DEN TEXT N°10

10- Der Dirigent, der gegen den Zusatz "Star" nichts einzuwenden hat, behauptet *dagegen*, man habe einvernehmlich Zärtlichkeiten ausgetauscht und *dabei* sei die junge schüchterne Geigerin ganz schön aus sich "raus gegangen". Wang, von 2006 bis 2008 Generalmusikdirektor der Stadt Würzburg, hat gestern die vermutlich letzte Chance "verspielt", ohne Urteil aus den Schlagzeilen herauszukommen. (SAZ11/JAN.00221 Saale-Zeitung, 04.01.2011, S. 8; Entschuldigung abgelehnt)

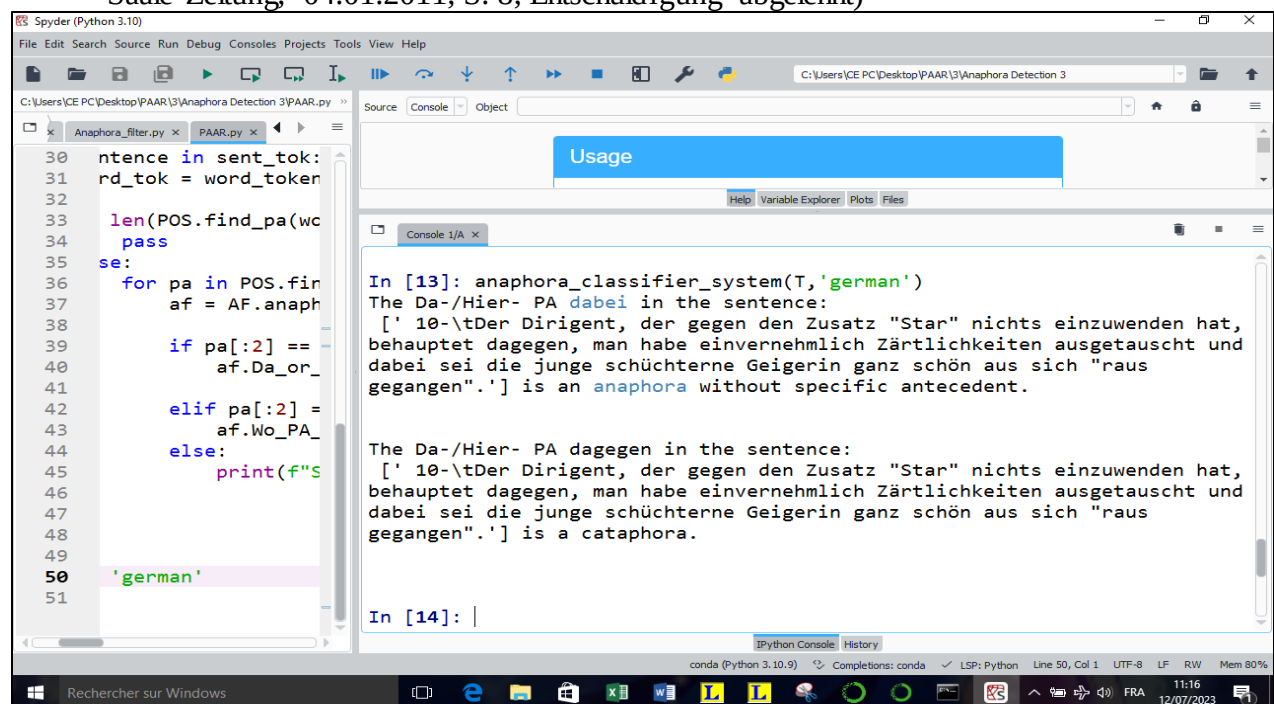


Abbildung 58: Einsatz des PAAR auf den Text N°10

Nach dem Einsatz vom PAAR auf den Text N°10 ergibt es sich, dass der Text aus zwei PA besteht, etwas das auch nach Näherbetrachtung des Textes wahr ist. Weiterhin klassifiziert

PAAR das erkannte PA („Dagegen“) als eine Katapher. Etwas das leider nicht richtig ist, denn „dagegen“ verweist hier auf eine Äußerung, die zuvor im Text gemacht wurde. Zwar hat man durch diesen Aufschnitt des Textes keinen direkten Zugang zu der entsprechenden Äußerung, aber dies soll längst nicht heißen, dass wir mit einer Katapher zu tun haben. In dieser Hinsicht werden wir dem PAAR keine Punkte bezüglich der Erkennung der Natur geschweigen denn der Auflösung des PA „dagegen“ geben.

Was das PA „dabei“ betrifft, PAAR erkennt es als eine Anapher mit dem Textteil „Der Dirigent, der gegen den Zusatz "Star" nichts einzuwenden hat, behauptet **dagegen**, man habe einvernehmlich Zärtlichkeiten ausgetauscht und **dabei** sei die junge schüchterne Geigerin ganz schön aus sich "raus gegangen" als Antezedens. Etwas, das nur zum Teil richtig ist. Es handelt sich zwar um ein anaphorisches PA aber mit Bezug auf den Textteil „man habe einvernehmlich Zärtlichkeiten ausgetauscht“. So betrachtet, lässt sich merken, dass PAAR hier komplett falsche Einheiten ausgewählt hat. Wie lässt sich nämlich klären, dass PAAR in den sprachlichen Einheiten, die er als Antezedens für eine Anapher auswählt auch die aufzulösende Anapher sich befindet. Linguistisch betrachtet, ergibt ein solcher OUTPUT keinen Sinn. Daher vermuten wir, dass das Problem an den durch die Programmiersprache Python entwickelten Zwischenmodulen liegen würde. In dieser Hinsicht prüften wir zunächst das Modul „Preprocessing“. Geprüft wurde nämlich die Zerlegung des Textes in Sätzen und nachher in Wörtern und es ergibt sich, dass man einen ganzen anderen OUTPUT bekommt, wenn man das orthographische Zeichen (Einführungszeichen) innerhalb des Textes löscht. Nach dieser Manöver PAAR erkennt immer die beiden PA und klassifiziert sie immer als Anapher (dabei) und Katapher (dagegen); aber was das Antezedens von „dabei“ angeht, wird dieses Mal im Gegensatz zum letzten Mal keine Angaben gegeben. Etwas das eigentlich besser ist, denn, obwohl es falsch ist, es ist trotzdem nicht Absurd, wie es beim ersten OUTPUT der Fall war.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2	2	2/2

Tabelle 46: Performanz vom PAAR bei Erkennung von PA im Text N°10

PA	Anzahl	Natur	Ergebnisse
Dabei	1	Anapher	1/1
Dagegen	1	Katapher	0/1
TOTAL	2		1/2

Tabelle 47: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°10

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dabei	0,25/1
TOTAL	0,25/1

Tabelle 48: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°10

6.1.10. EINSATZ DES PAAR AUF DEN TEXT N°11

*11- Allein, dass die Verunreinigung von belasteten Würstchen oder Burgern herrührt, hält Krahl für "eher unwahrscheinlich". Dies setze voraus, "dass dort tonnenweise dioxinverseuchtes Fleisch gebraten worden ist". Ungleich plausibler sei **dagegen**, dass das verarbeitete Speiseöl bereits tierische Fette enthalten habe: "Und die **dafür***

verarbeiteten Tiere könnten wiederum mit dioxinverseuchtem Futter gefüttert worden sein." Was Krahel hier nur vage andeutet, wäre nichts anderes als ein aberwitziger Kreislauf der Kontaminierung. Mit dioxinverseuchtem Futter genährte Tiere kontaminieren Speiseöl und geben ihr Gift später, dann in einem weiteren Produktionsschritt zu Futtermittel verarbeitet, an andere Tiere weiter. (SAZ11/JAN.00372 Saale-Zeitung, 05.01.2011, S. 24; ohne Titel)

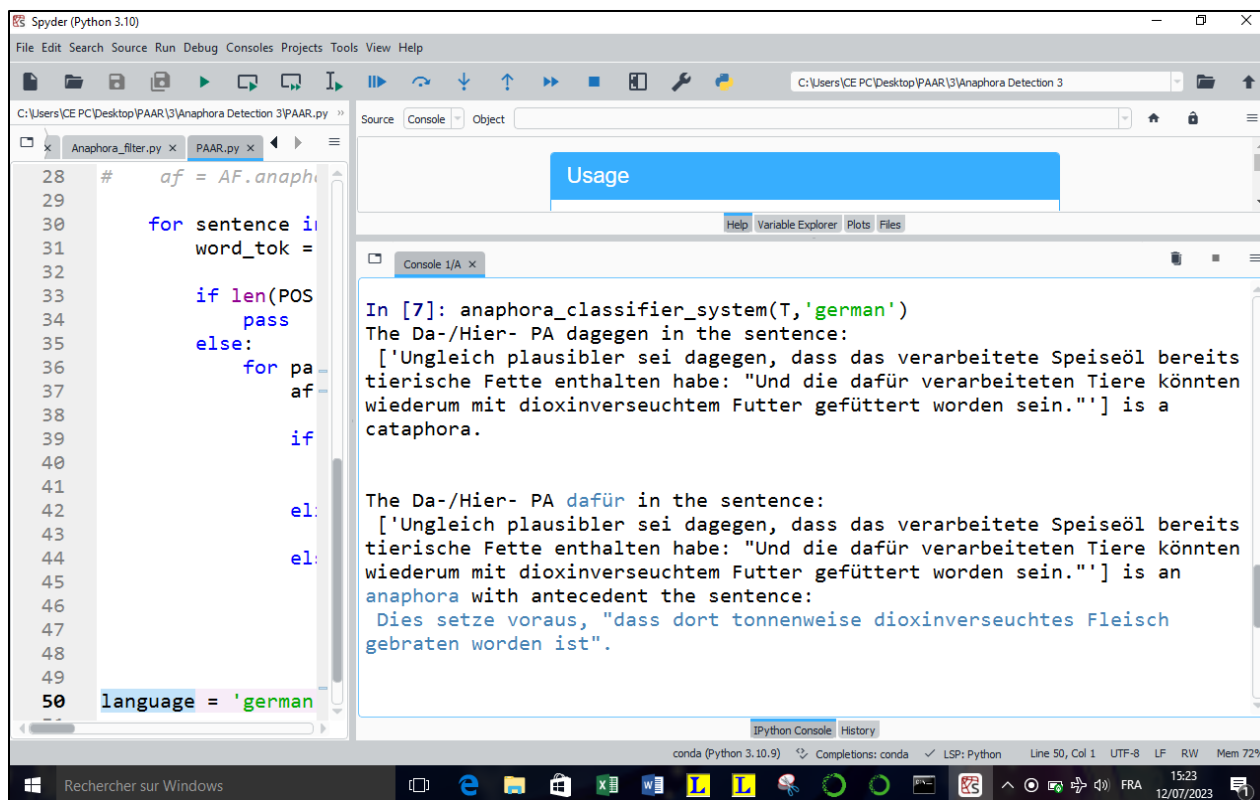


Abbildung 59: Einsatz des PAAR auf den Text N°11

Nach dem Einsatz vom PAAR auf den Text N°11 ergibt es sich, dass der Text aus zwei PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR das erkannte PA („Dagegen“) als eine Katapher. Etwas das auch richtig ist, denn „dagegen“ verweist hier auf vorwärtsstehende sprachliche Einheiten. Was das PA „dafür“ betrifft, PAAR erkennt es als eine Anapher mit dem Textteil „Dies setze voraus, "dass dort tonnenweise dioxinverseuchtes Fleisch gebraten worden ist". " als Antezedens. Etwas, das auch richtig ist.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
2	2	2/2

Tabelle 49: Performanz vom PAAR bei Erkennung von PA im Text N°11

PA	Anzahl	Natur	Ergebnisse
Dafür	1	Anapher	1/1
Dagegen	1	Katapher	1/1
TOTAL	2		2/2

Tabelle 50: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°11

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dafür	1/1

TOTAL	1/1
--------------	------------

Tabelle 51: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°11

6.1.11. EINSATZ DES PAAR AUF DEN TEXT N°12

12- Drei Fahrzeuge waren an einem Unfall auf der Staatsstraße 2290 zwischen den Abzweigungen Gefäll und Premich beteiligt. Der 53-jährige Fahrer des vordersten Fahrzeugs verminderte seine Geschwindigkeit stark, als er durch die tiefstehende Sonne geblendet wurde. Der *dahinter* fahrende 51-Jährige bremste ebenfalls ab. Eine 25-jährige Frau bemerkte das zu spät und fuhr trotz Vollbremsung auf den mittleren Wagen auf, der dann noch leicht auf den ersten Pkw geschoben wurde. Fünf von acht Fahrzeuginsassen wurden leicht verletzt in Krankenhäuser gebracht. Die drei Fahrzeugführer blieben unverletzt. Es entstand ein Gesamtschaden von 26 500 Euro. Die Unfallverursacherin erwartet ein Verfahren wegen fahrlässiger Körperverletzung. pol (SAZ11/JAN.00213 Saale-Zeitung, 04.01.2011, S. 6; Drei Autos)

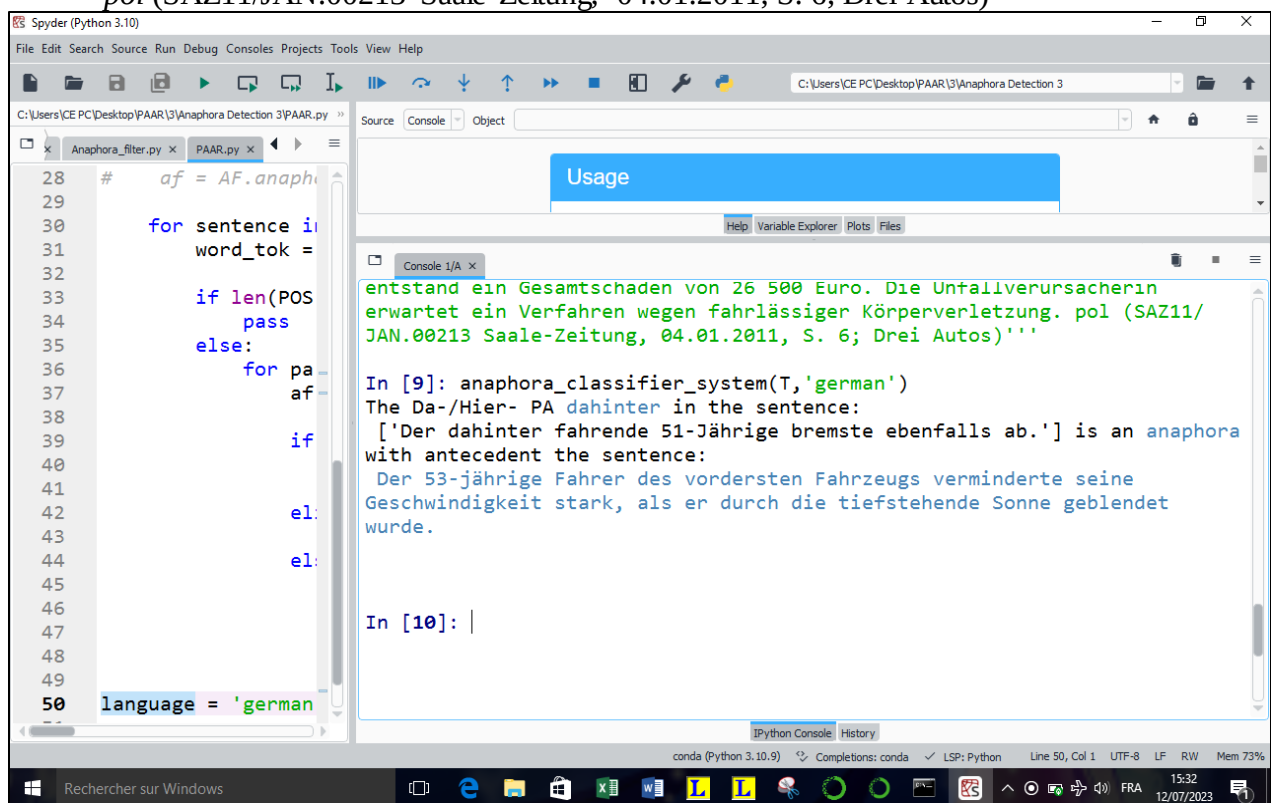


Abbildung 60: Einsatz des PAAR auf den Text N°12

Nach dem Einsatz vom PAAR auf den Text N°12 ergibt es sich, dass dieser Letzte aus einem PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR das erkannte PA („dahinter“) als eine Anapher mit dem Textteil „Der 53-jährige Fahrer des vordersten Fahrzeugs verminderte seine Geschwindigkeit stark, als er durch die tiefstehende Sonne geblendet wurde.“ Als Antezedens. Etwas das auch richtig ist.

Im Text enthaltene PA	Durch PAAR gefundene PA	Ergebnisse
1	1	1/1

Tabelle 52: Performanz vom PAAR bei Erkennung von PA im Text N°12

PA	Anzahl	Natur	Ergebnisse
Dahinter	1	Anapher	1/1
TOTAL	1		1/1

Tabelle 53: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°12

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dahinter	1/1
TOTAL	1/1

Tabelle 54: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°12

6.1.12. EINSATZ DES PAAR AUF DEN TEXT N°13

13- Auch für mobile Geräte gibt es Lernspiele. Cornelsen bietet etwa für iPhone, iPod touch und iPad "Lernsnacks". *Dahinter* verbergen sich Matheübungen zu Themen wie Zins- oder Prozentrechnung. Die Basis-App für 79 Cent enthält 200 Aufgaben zur Prozentrechnung. Jedes weitere Häppchen wird für knapp 3 Euro in Apples Online-Laden, dem iTunes App Store, angeboten. Die Erweiterungen sollen monatlich erscheinen. Die App "MyBricks Zahlen" von Terzio richtet sich an die Jüngsten. Kinder werden *dazu* animiert, mit großen und bunten Bausteinen Zahlen zusammenzubauen. Das soll sie spielerisch an deren Laut und Form heranführen. (SAZ11/JAN.01205 Saale-Zeitung, 13.01.2011, S. 29; ohne Titel)

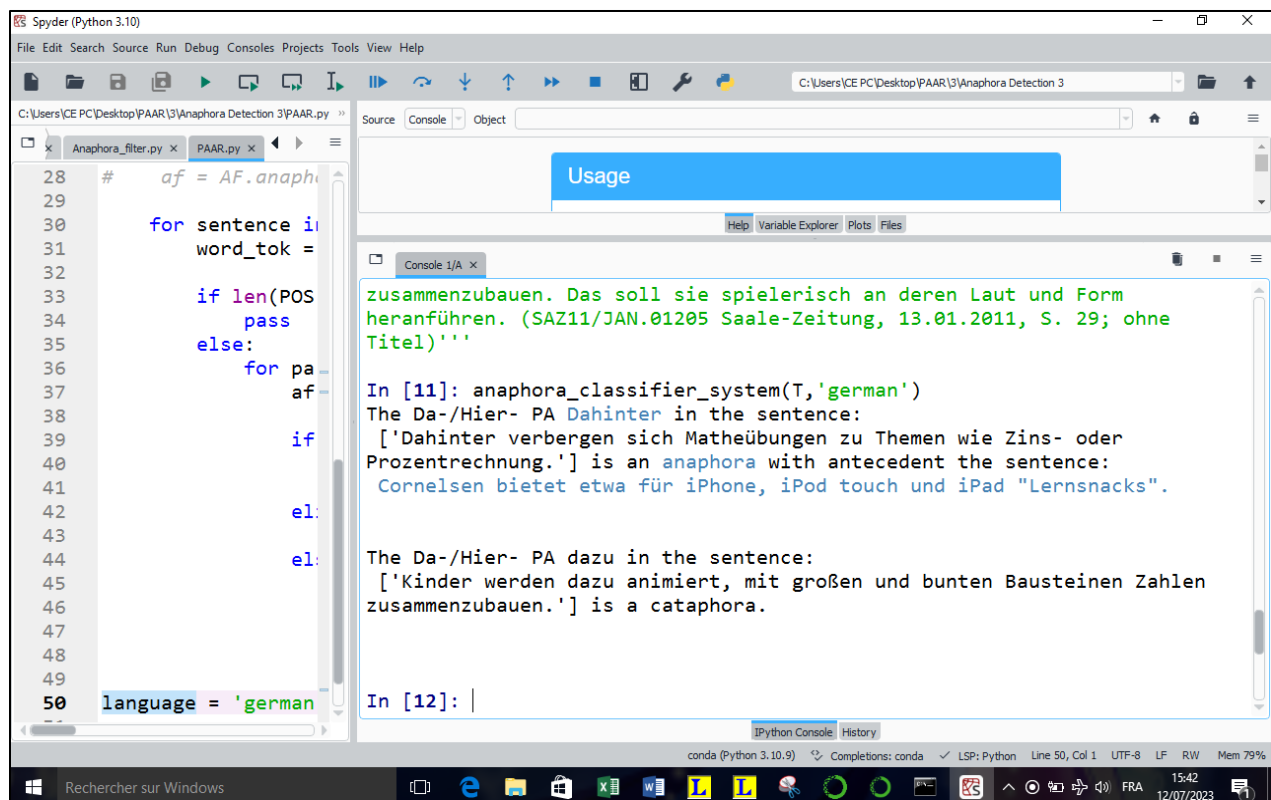


Abbildung 61: Einsatz des PAAR auf den Text N°13

Nach dem Einsatz vom PAAR auf den Text N°13 ergibt es sich, dass dieser Letzte aus zwei PA besteht, etwas das auch nach Näherbetrachtung des Textes richtig ist. Weiterhin klassifiziert PAAR das erkannte PA („dahinter“) als eine Anapher mit dem Textteil „Auch für mobile Geräte gibt es Lernspiele. Cornelsen bietet etwa für iPhone, iPod touch und iPad "Lernsnacks"“ als

Antezedens. Etwas das auch richtig ist. Was das PA „dazu“ angeht klassifiziert PAAR es als eine Katapher. Etwas das auch richtig ist, denn „dazu“ verweist hier auf vorwärtsstehende sprachliche Einheiten.

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
2	2	2/2

Tabelle 55: Performanz vom PAAR bei Erkennung von PA im Text N°13

PA	Anzahl	Natur	Ergebnisse
Dahinter	1	Anapher	1/1
Dazu	1	Katapher	1/1
TOTAL	1		2/2

Tabelle 56: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°13

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Dahinter	1/1
TOTAL	1/1

Tabelle 57: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°13

6.1.13. EINSATZ DES PAAR AUF DEN TEXTABSCHNITT (N°14-20)

- 14- Mancher Autofahrer dürfte gestern die kurzen Wartezeiten auf den Einfallstraßen nach Bad Brückenau genossen haben, verhiessen doch die Behinderungen am Morgen eine Schonung der Stoßdämpfer am Abend: "Wir bleiben *dahinter*, dass die gravierendsten Schäden schnell beseitigt sind", berichtet Stefan Zehe, Leiter der Straßenmeisterei Hammelburg, dass gestern eine große Aktion auf den Bundesstraßen rund um Bad Brückenau startete. (SAZ11/FEB.00653 Saale-Zeitung, 05.02.2011, S. 19; Nicht die tiefsten, aber die größten Löcher)
- 15- Waldfenster Die Jahreshauptversammlung des Musikvereins Waldfenster findet am Freitag, 14. Januar, um 19.30 Uhr im Gasthaus "Krone" statt. Auf der Tagesordnung stehen neben den Berichten des Vorsitzenden und der Abteilungen und den Regularien Neuwahlen und eine Satzungsänderung. Anträge zur Generalversammlung müssen schriftlich bis spätestens 8. Januar beim 1. Vorsitzenden eingereicht werden, *damit* sie in der Versammlung behandelt werden können. red (SAZ11/JAN.00025 Saale-Zeitung, 03.01.2011, S. 6; Neuwahlen und eine Änderung)
- 16- Nicht nur das Jahr 2010 wurde von der "Egerländer Blasmusik" bei einem Konzert im Bad Bockleter Kursaal musikalisch verabschiedet, sondern auch der langjährige Dirigent Rudi Bock. *Damit* ging nach 16 Jahren eine Ära zu Ende, die die Kapelle und auch deren Musikstil entscheidend mit geprägt hat. (SAZ11/JAN.00073 Saale-Zeitung, 03.01.2011, S. 5; Eine Ära ging zu Ende)
- 17- Die Feuerwehr Untererthal führt am Donnerstag, 6. Januar, ihre Jahreshauptversammlung durch. Beginn ist um 19 Uhr im Schulungsraum des Feuerwehrgerätehauses. Auf der Tagesordnung stehen unter anderem ein Bericht des Kommandanten und ein Jahresrückblick. Im Anschluss werden vereinsinterne Ehrungen und Neuaufnahmen von Mitgliedern vorgenommen. *Danach* wählen die Floriansjünger ihren Kommandanten sowie einen 1. und 2. Vorsitzenden. mr (SAZ11/JAN.00052 Saale-Zeitung, 03.01.2011, S. 9; Wahlen bei der Wehr)
- 18- Der FCN-Fanclub trifft sich am Dreikönigstag, Donnerstag, 6. Januar, im Gasthaus Brust in Obererthal. Um 15 Uhr ist zunächst Kegeln angesagt, *danach* wird Bingo gespielt und es gibt ein gemütliches Beisammensein. (SAZ11/JAN.00082 Saale-Zeitung, 03.01.2011, S. 7; ohne Titel)
- 19- Das Feuerwehrhaus wurde in dieser Zeit umgebaut und ein Schulungsraum eingerichtet, der Atemgeräteschutz eingeführt und die Vereinsfahne angeschafft.

Daneben wurde unter seiner Federführung der Fuhrpark aufgerüstet. (SAZ11/JAN.00604 Saale-Zeitung, 08.01.2011, S. 8; Aufgabe in jüngere Hände gelegt)
 20- Am 1. Januar kurz nach 13 Uhr bemerkte der Eigentümer einer Scheune in Rothhausen starke Rauchentwicklung, die aus der Scheune kam. Kurz *darauf* stand das Gebäude schon in Flammen. Die Feuerwehren aus Rothhausen, Thundorf, Münnerstadt, Theinfeld, Maßbach, Stadtlauringen, Weichtungen sowie die Drehleiter aus Bad Kissingen waren mit einem Großaufgebot zur Brandbekämpfung angerückt. Die Maßbacher Feuerwehr kam allerdings erst auf Umwegen zum Brandort, da man sie nach Thundorf geschickt hatte, so Einsatzleiter Holger Ulrich. (SAZ11/JAN.00034 Saale-Zeitung, 03.01.2011, S. 12; Die Flammen schlugen hoch)

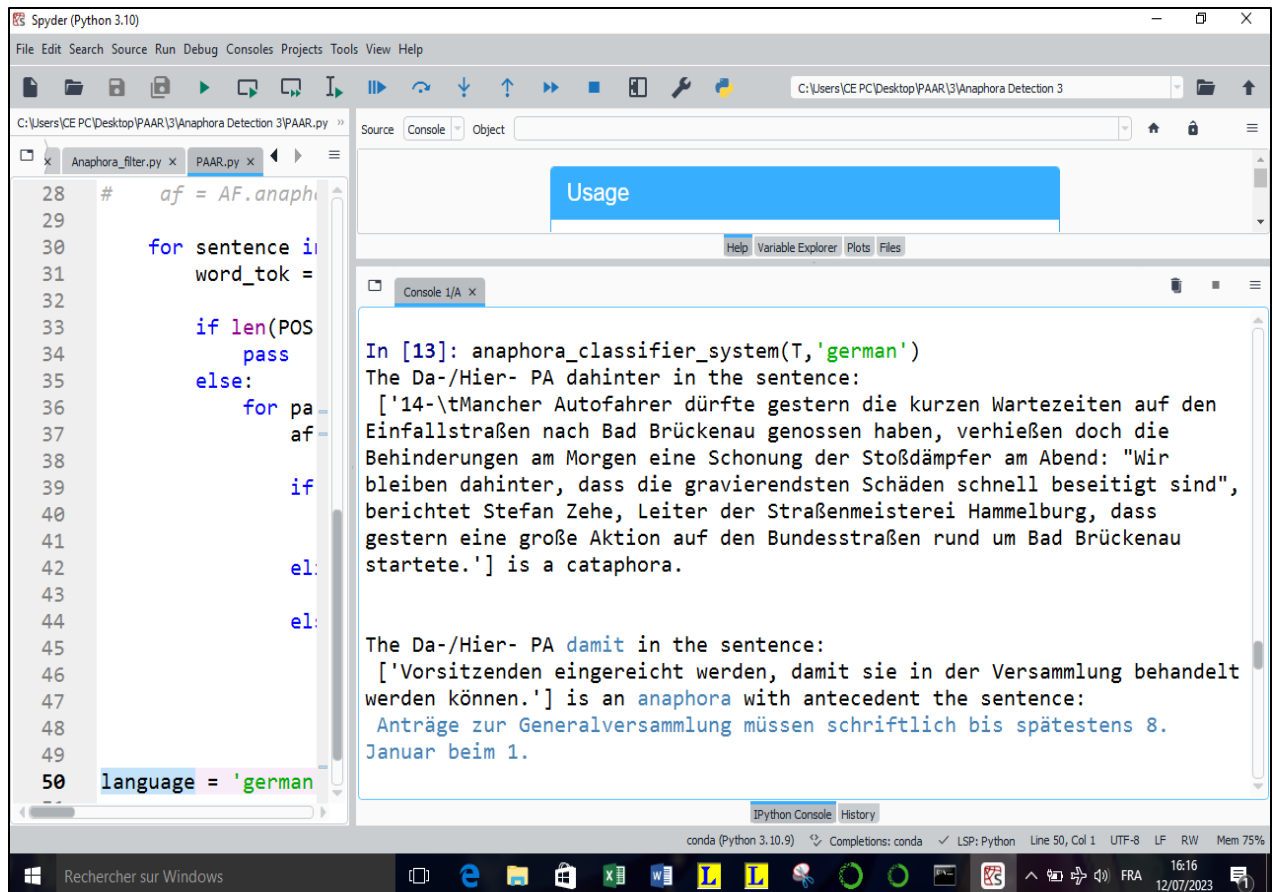


Abbildung 62: Einsatz des PAAR auf den Textabschnitt (N°14-20) (1)

Die Abbildung 62 zeigt nämlich, dass PAAR die beiden ersten Da-Formen des Textausschnitts, der vom Text N°14 bis Text N°20 geht, als Pronominalanaphorien kategorisiert hat. Etwas das nicht hundertprozentig richtig ist. Denn das Wort „dahinter“ ist zwar ein kataphorisches PA, aber was das Wort „damit“ angeht, handelt es sich hier eher um eine Subjunktion und um kein PA.

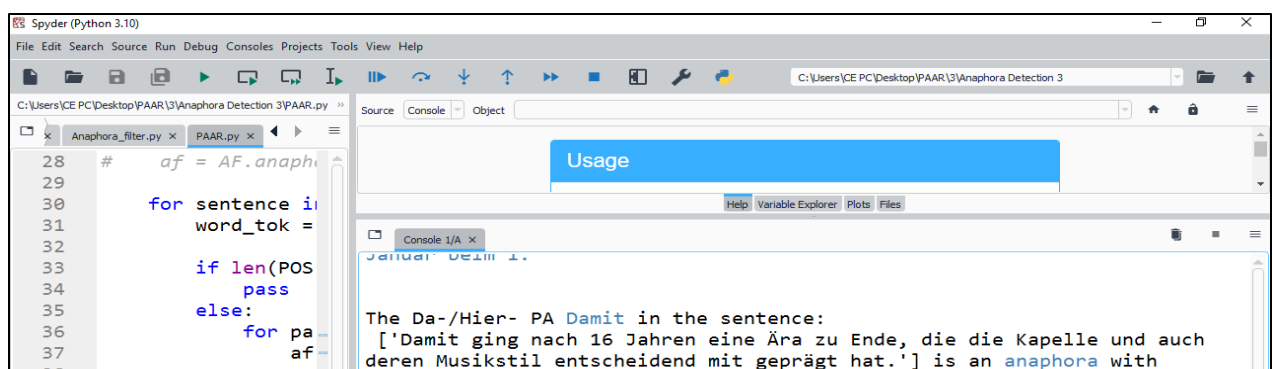


Abbildung 63: Einsatz des PAAR auf den Textabschnitt (N°14-20) (2)

In der Abbildung 63 stellt man nämlich fest, dass PAAR weitere Da-Erscheinungen als PA klassifiziert hat. Etwas das eben auch richtig ist, denn die Wörter „damit“ und „danach“ sind hier anaphorische PA. Was die jeweiligen Antezedenzen angeht, hat PAAR es auch richtig herausgefunden.

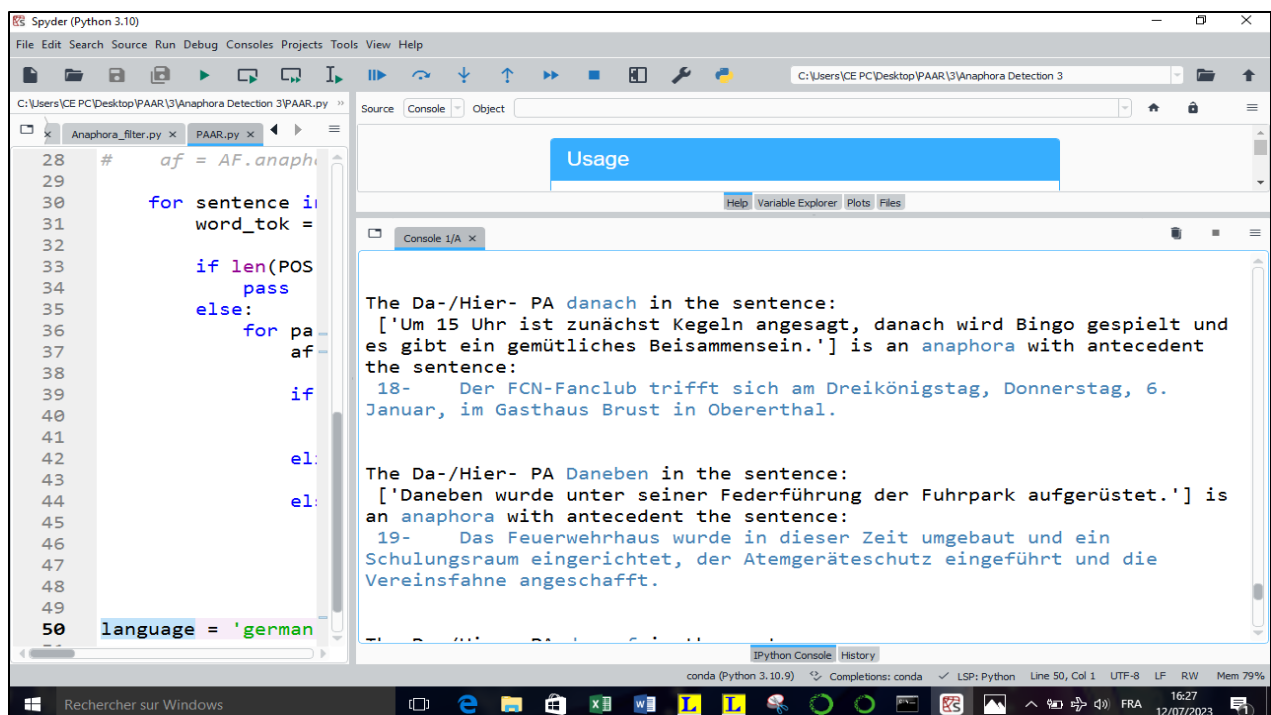


Abbildung 64: Einsatz des PAAR auf den Textabschnitt (N°14-20) (3)

Aus der Abbildung 64 ergibt es sich, dass PAAR die beiden Da-Formen als anaphorische PA klassifiziert hat. Allerdings wurde ein falsches Antezedens dem PA „danach“ zugewiesen. Das

richtige Antezedens ist „Um 15 Uhr ist zunächst Kegeln angesagt“. Im Gegensatz zur Auflösung des PA „danach“ wurde das anaphorische PA „Daneben“ richtig aufgelöst.

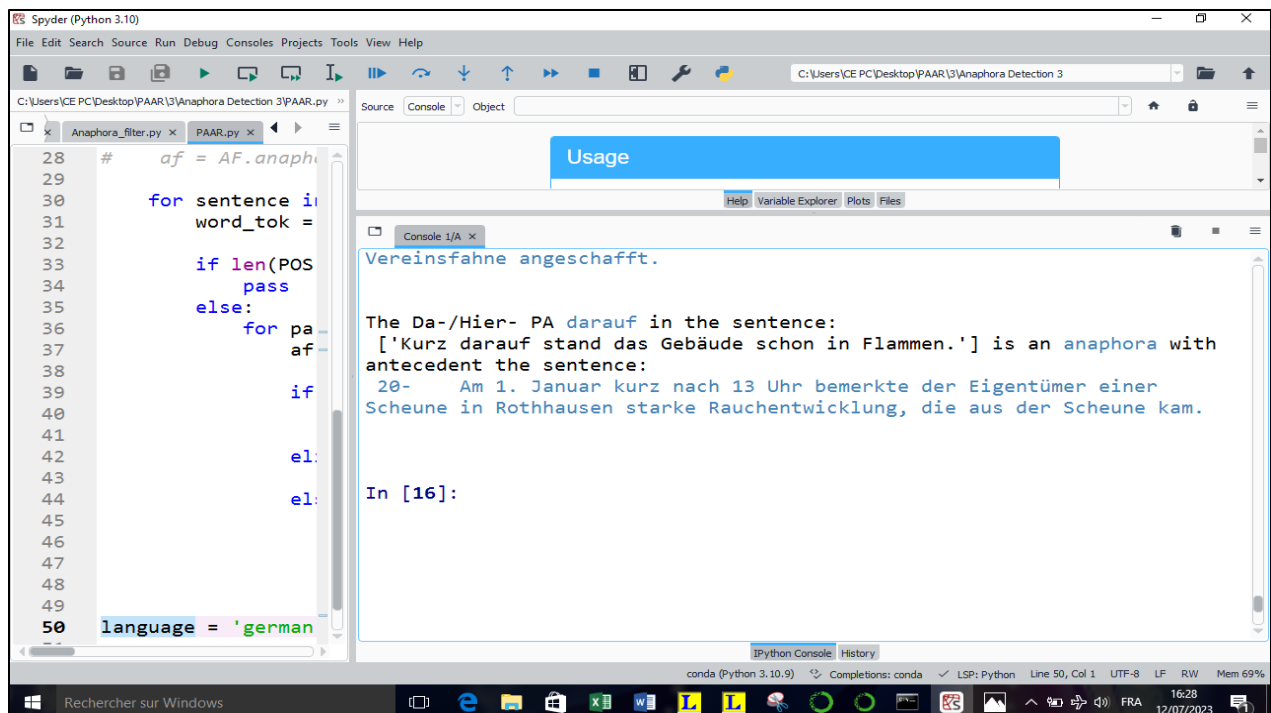


Abbildung 65: Einsatz des PAAR auf den Textabschnitt (N°14-20) (4)

Aus der Abbildung 65 lässt sich nämlich ablesen, dass PAAR das Wort „darauf“ als ein anaphorisches PA verzeichnet hat. Wobei das entsprechende Antezedens auch richtig ausgewählt wurde. Hier eine Zusammenfassung dieses Einsatzes.

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
6	7 (6+ Eine Subjunktion)	7/6 ??

Tabelle 58: Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°14-20)

PA	Anzahl	Natur	Ergebnisse
Dahinter	1	Katapher	1/1
Damit	1	Anapher	1/1
Danach	2	Anaphern	2/2
Daneben	1	Anapher	1/1
Darauf	1	Anapher	1/1
TOTAL	6		6/6

Tabelle 59: Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°14-20)

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Damit	1/1
Danach	1/2
Daneben	1/1
Darauf	1/1
TOTAL	4/5

Tabelle 60: Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°14-20)

6.1.14. EINSATZ DES PAAR AUF DEN TEXTABSCHNITT (N°21-30)

- 21- Angeblich mault der Franke gern, er macht gerne nieder. Am liebsten sich selbst und das, was er hat. Der Kissinger mault über die Kur. "Die Kur ist tot", lautet ein oft gehörter Satz, **wobei** offenbleibt, was da mit Kur gemeint ist. Die private Vorsorge, die Stationäre Rehabilitation oder die Anschlussheilbehandlung? Schaut man auf die Statistik, hat die Übernachtungszahl bei den Versicherungsgästen sogar leicht zugenommen. Es kommt **darauf** an, was der Standort Bad Kissingen aus seiner Kompetenz macht. Zum Beispiel die Klinik am Kurpark mit ihrer Kira-Studie. Die "Kur" ist nicht tot. (SAZ11/JAN.00062 Saale-Zeitung, 03.01.2011, S. 1; Die Kur ist)
- 22- Die Denkpause für Westerwelle ist nun beendet. Wird er in Stuttgart nun seinen Rückzug vom Parteivorsitz ankündigen? Wird er einen Neuanfang der FDP mit einer jüngeren Mannschaft um Generalsekretär Christian Lindner in Aussicht stellen? Nichts deutet **darauf** hin, dass sich der vor wenigen Tagen 49 gewordene Westerwelle kampflos vom politischen Acker machen wird. Im Gegenteil: Die meisten FDP-Auguren sind sich **darin** einig, dass Westerwelles einzige Chance die politische Offensive ist. Befragungen der Parteibasis bestätigen dieses Bild: 77 Prozent der FDP-Anhänger wollen den Super-Wahlkämpfer aus Bonn, der seiner Partei bis 2009 Wahlsiege in Serie verschaffte, auch weiter als Chef behalten. (SAZ11/JAN.00096 Saale-Zeitung, 03.01.2011, S. 22; Westerwelle kämpft um sein Partei-Amt)
- 23- Mit viel Herzblut und Engagement hatte Fridolin Fritz ein Jahrzehnt lang den Ortsverband Schondra als Vorsitzender geleitet. **Dafür** war er kürzlich zum Ehrenvorsitzenden ernannt worden (die Heimatzeitung berichtete). **Darauf** ist er sichtlich stolz, die Urkunde hängt in seinem Arbeitszimmer an der Wand. Aus gesundheitlichen Gründen und auch vom Alter her, wie er sagt, wollte er das Zepter aus der Hand geben und er freut sich, dass er jüngere Nachfolger gefunden hat. (SAZ11/JAN.00248 Saale-Zeitung, 04.01.2011, S. 15; Als der Kassierer noch durchs Dorf lief...)
- 24- Das Auto wurde dann mit dem Rettungssatz vorschriftsmäßig zerlegt. Im Rahmen eines Fachhochschul-Projektes wurde die Übung aufgezeichnet. **Daraus** soll ein Werbefilm für Feuerwehren entstehen. (SAZ11/JAN.00979 Saale-Zeitung, 12.01.2011, S. 6; Weiter unter bewährter Führung)
- 25- Hoffnungen setzt der SVN-Vorstand nicht nur in die Zukunft der aktiven Mannschaft in der aktuell laufenden Oberliga-Runde, sondern auch **darein**, dass das Projekt Umbau Westpfalzstadion (aus dem Hartplatz soll ein Kunstrasenplatz werden) noch in diesem Jahr realisiert werden kann. mw (SBZ10/AUG.09167 Saarbrücker Zeitung, 23.08.2010; "Einmal SVN, immer SVN")
- 26- Jedes Team berät sich ein letztes Mal - "Auf die Plätze, fertig, los!" - und jedes Team beginnt seine Arbeit: Lisdorf klebt den Nagel an die Dachlatte, spannt das Seil über das Feld hin und zurück, wie bei einer Seilbahn. Sie befüllen die Flaschen mit Wasser, blasen die Luftballons auf und füllen das Wasser **darein**. Die Luftballons werden an das Seil gehängt und über die Gießkanne gezogen. Nun wird mit dem Nagel der Ballon zerstoßen und das Wasser spritzt in die Gießkanne. Die anderen Stadtteile haben ihre eigene Technik: Picard baut eine Art Rohr aus der Dachlatte und dem Klebeband. Die effektivste Technik, die die Lisdorfer die Punkte ihres eigenen Spieles kostet, denn Picard hatte auch eine gute Kräuterfee. (SBZ11/AUG.09368 Saarbrücker Zeitung, 23.08.2011; Die Kräuterfee brachte den Pokal im StadtteileWettkampf)
- 27- Sie können mal ausprobieren, Kartoffeln in Tonnen anzubauen. **Dazu** nehme man eine alte Mülltonne oder ein ähnliches Gefäß mit bis zu einem Sechstel gefüllt mit Erde. Dann pflanze man **darein** eine oder mehrere Kartoffeln. Beim Wachsen des Krautes füllt man immer wieder vor der Blüte Erde nach, bis die Tonne voll ist. Zur Erntezeit einfach nur Ausleeren und sich über die reiche Kartoffelzahl freuen. Für einen Jahresvorrat

wird es zwar nicht reichen, aber Idealisten genießen die eigenen Knollen dann trotzdem sehr. (SKU01/APR.15749 Südkurier, 25.04.2001; Kartoffeln in der Mülltonne)

- 28- Das Problem lag **darin**, dass sich die Organisten beim Spiel an den rund 25 Meter auseinander stehenden Instrumenten nicht sehen konnten und dennoch präzise spielen mussten. Das gelang ausgezeichnet, und die Zuhörer konnten auf einer Leinwand beiden auf Finger und Füße schauen. (SAZ11/JAN.00070 Saale-Zeitung, 03.01.2011, S. 4; Musik für)
- 29- Auf Leverkusen und den FSV Mainz 05 haben die Dortmunder nach der Hinrunde zehn Punkte Vorsprung. "Wir müssen aufpassen, dass der Mannschaft nicht zu hohe Erwartungen aufgebürdet werden. Denn **darin** sehe ich die größte Gefahr für die Rückrunde", sagte BVB- Geschäftsführer Hans-Joachim Watzke den "Ruhr Nachrichten". (SAZ11/JAN.00156 Saale-Zeitung, 04.01.2011, S. 21; Zum Training ins warme Spanien)
- 30- CSU-Chef Horst Seehofer sagte: "Beim Bildungspaket können wir **darüber** reden, ob der Kreis der Empfänger ausgeweitet wird und wie man das möglichst unbürokratisch gestaltet. Sinnvoll sind treffsichere Lösungen bei Mindestlöhnen in bestimmten Branchen, vor allem bei der Zeitarbeit." (SAZ11/JAN.00029 Saale-Zeitung, 03.01.2011, S. 19; DGB rät zum Widerspruch gegen Hartz-IV-Bescheide)

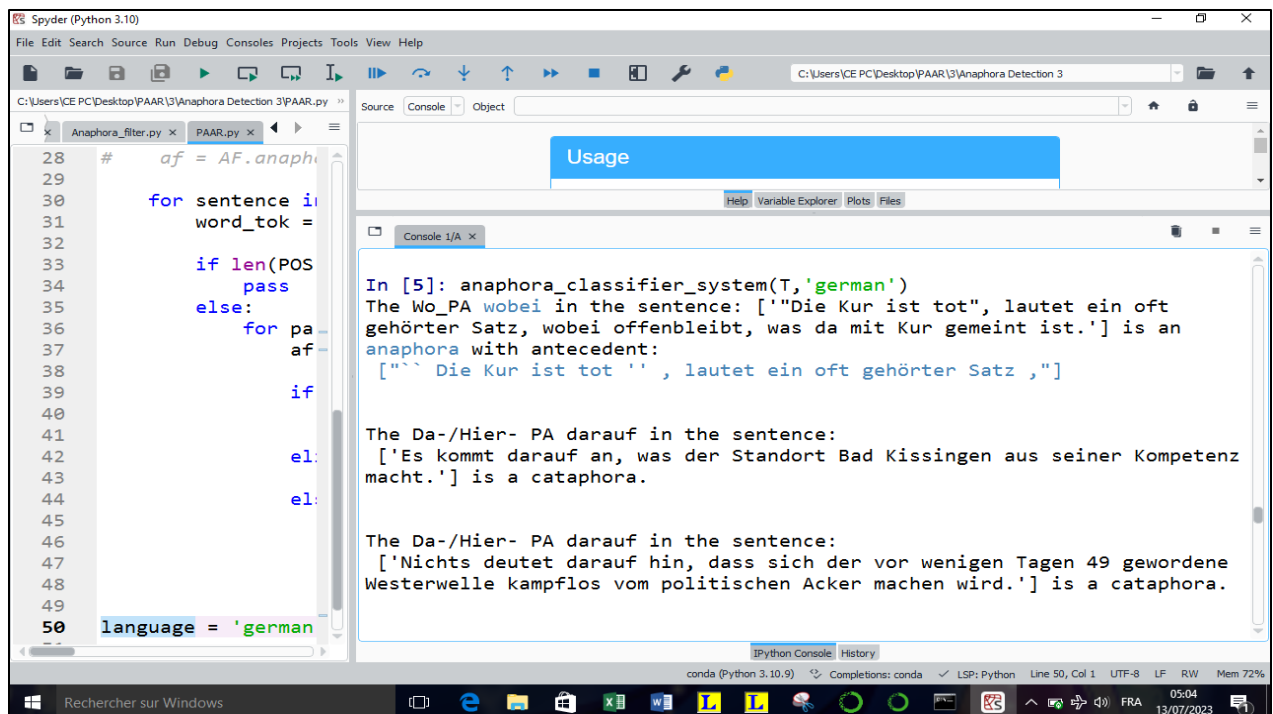


Abbildung 66: Einsatz des PAAR auf den Textabschnitt (N°21-30) (1)

Aus der Abbildung 66 lässt sich nämlich ablesen, dass PAAR das Wort „wobei“ als ein anaphorisches PA verzeichnet hat. Zu merken ist auch nämlich, dass die Auflösung von „wobei“ erfolgreich realisiert wurde. Weiterhin hat PAAR auch zwei Erscheinungen vom Wort „darauf“ verzeichnet hat, und er hat sie alle beiden als kataphorische PA markiert. Etwas, dass auch nachvollziehbar ist.

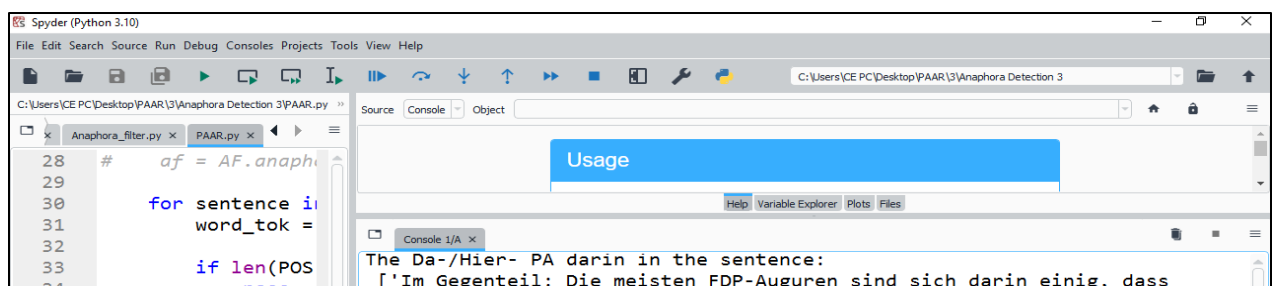


Abbildung 67: Einsatz des PAAR auf den Textabschnitt (N°21-30) (2)

In der Abbildung 67 stellt man nämlich fest, dass PAAR 3 PA aufgefunden hat. Er klassifiziert das erste „darin“ als eine Katapher, während er die beiden anderen „Dafür“ und „Darauf“ als anaphorische Erscheinungen verzeichnet. Was die Auflösung von anaphorischen PA angeht lässt sich nämlich merken, dass PAAR für alle beiden PA die jeweiligen richtigen Antezedenzen aufgefunden hat.

Im Gegensatz zur Ausbildung 67 widerspiegelt die Abbildung 68 die Auffindung durch PAAR von zwei PA. Das erste wurde als eine anaphorische Erscheinung von „Daraus“ verzeichnet, wobei das richtige Antezedens auch nur zum Teil aufgefunden wurde. In der Tat verweist das PA „Daraus“ hier nur auf „Im Rahmen eines Fachhochschul-Projektes“. Was das zweite PA betrifft handelt es sich, PAAR nach, um eine kataphorische Erscheinung „darein“.

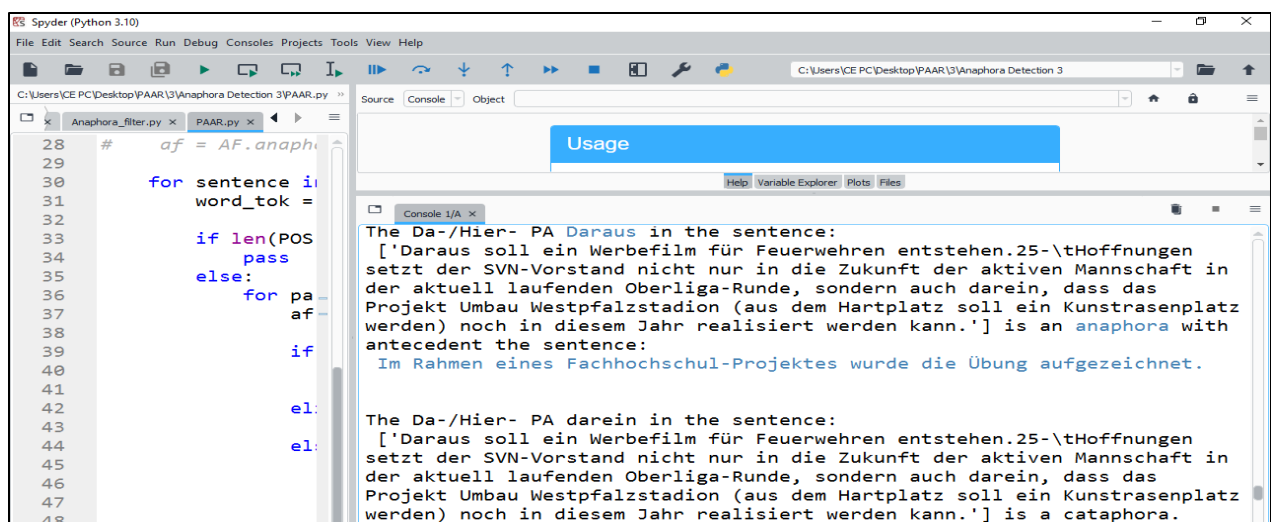


Abbildung 68: Einsatz des PAAR auf den Textabschnitt (N°21-30) (3)

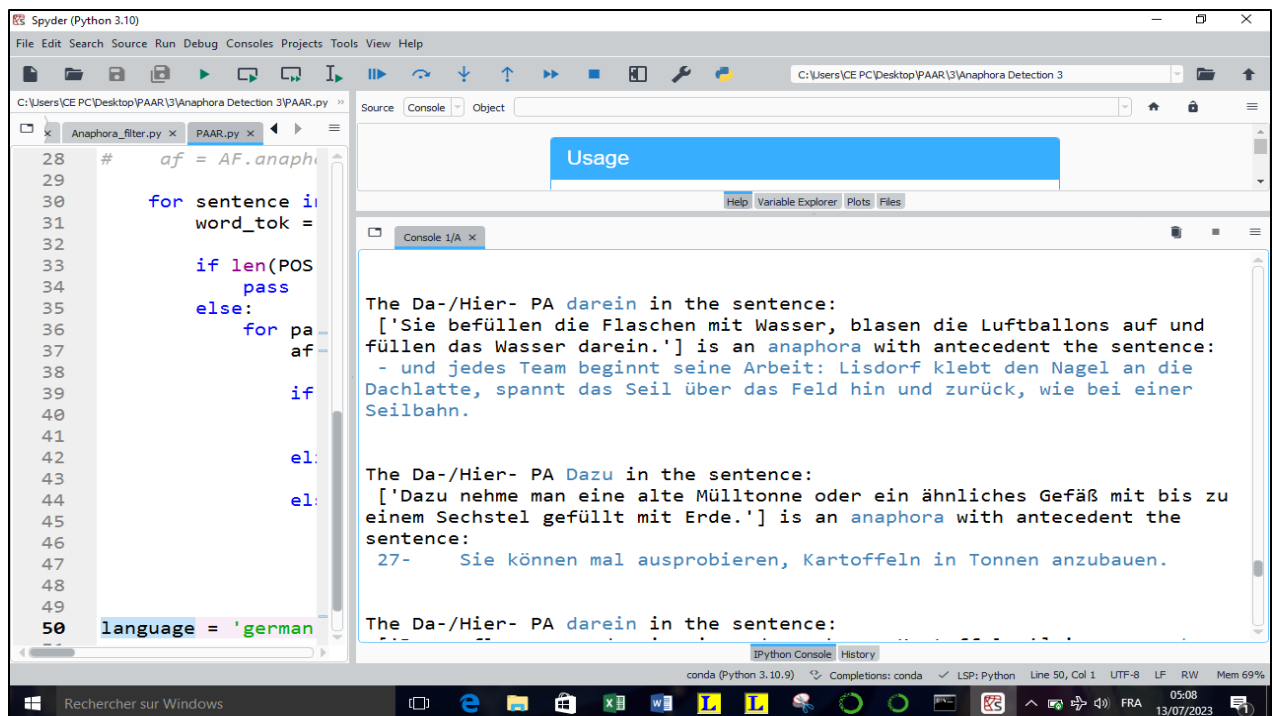


Abbildung 69: Einsatz des PAAR auf den Textabschnitt (N°21-30) (4)

In der Abbildung 69 lässt sich merken, dass PAAR zwei PA aufgefunden hat, das erste („darein“) klassifiziert er als eine Anapher. Etwas das richtig ist; aber er löst sie falsch aus. Denn „darein“ hier verweist nicht auf den vom PAAR aufgezeichneten Textteil, sondern eher nur auf die NP „Die Luftballons“. Weiterhin klassifiziert PAAR das PA „Dazu“ eben als eine Anapher und löst es aber zugleich auch richtig auf.

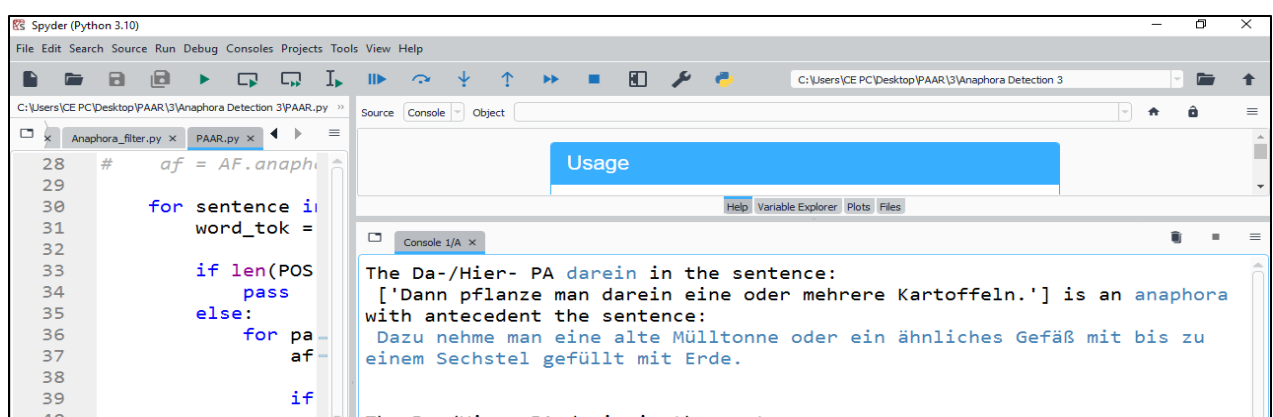


Abbildung 70: Einsatz des PAAR auf den Textabschnitt (N°21-30) (5)

Aus der Abbildung 70 lässt sich ablesen, dass PAAR Zwei PA erfunden hat. Das erste „darein“ klassifiziert er als eine anaphorische Erscheinung. Etwas, das nach Näherbetrachtung des Textauschnitts auch richtig ist. Was die Auflösung vom „darein“ betrifft, wollen wir behaupten, dass die Auflösung nur Teilweise richtig absolviert ist. Denn „darein“ hier referiert nicht auf die ganze Handlung, die zuvor beschrieben worden ist, sondern eher nur auf eine Komponente dieser Handlung. Und diese Komponente ist „eine alte Mülltonne oder ein ähnliches Gefäß“. Das zweite erfundene PA ist „darin“. PAAR hat es als eine Katapher klassifiziert, etwas, das auch nachvollziehbar ist, denn „darin“ hier erweist auf vorwärtsstehende sprachliche Einheiten.

Was die Abbildung 71 betrifft, lässt sich ablesen, dass PAAR zwei PA aufgefunden hat. Das erste („darin“) wurde als eine Anapher klassifiziert, wobei auch das richtige Antezedens nur zum Teil richtig wiedergegeben wurde. In der Tat ist das richtige Antezedens nur den Nebensatz, das mit dem Relativpronomen „dass“ anfängt und nicht den ganzen Satz (Haupt & Nebensatz) wie es PAAR aufgezeichnet hat. Was das zweite betrifft, wurde es als eine Katapher verzeichnet. Etwas, das auch eben richtig ist, denn „darüber“ hier erweist auf vorwärtsstehende sprachliche Einheiten.

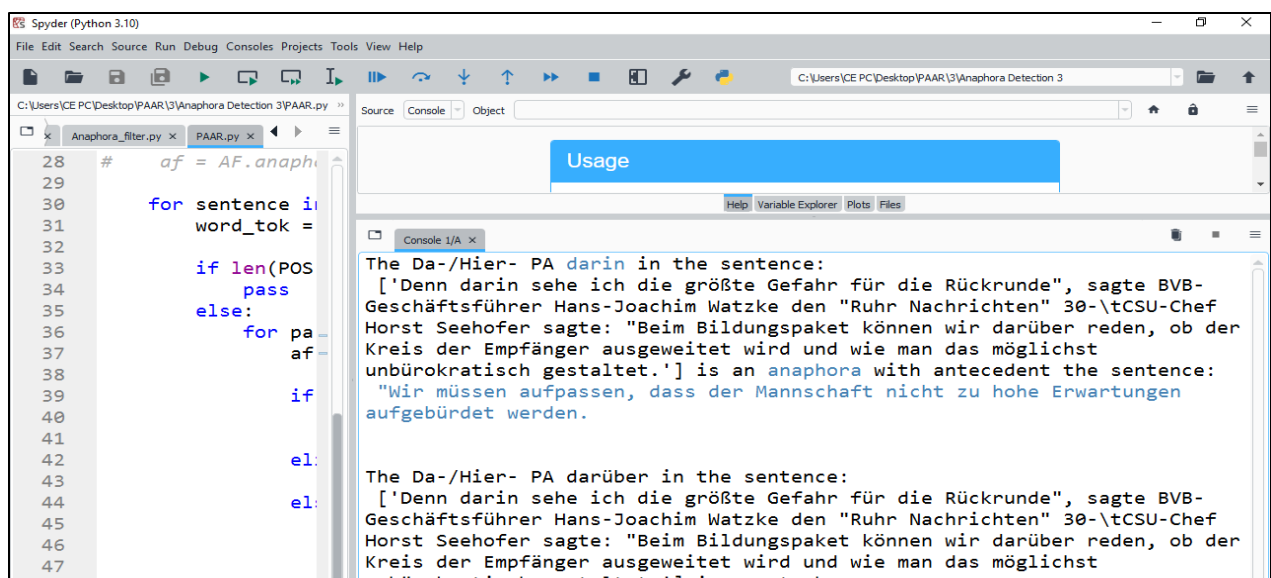


Abbildung 71: Einsatz des PAAR auf Texte N°21-30 (6)

Hier eine Zusammenfassung der Analyse von den Angaben, die nach dem Einsatz vom PAAR auf den Textausschnitt (**N°21-30**) gewonnen wurden.

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
14	14	14/14

Tabelle 61: Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°21-30)

PA	Anzahl	Natur	Ergebnisse
Wobei	1	Anapher	1/1
Darauf	2	Katapher	2/2
Darin	2	Katapher	2/2
Dafür	1	Anapher	1/1
Darauf	1	Anapher	1/1
Daraus	1	Anapher	1/1
Darein	1	Katapher	1/1
Darein	2	Anapher	2/2
Dazu	1	Anapher	1/1
Darin	1	Anapher	1/1
Darüber	1	Katapher	1/1
TOTAL	14		14

Tabelle 62: Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°21-30)

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Wobei	1/1
Dafür	1/1
Darauf	1/1
Daraus	0,5/ 1
Darein	0,5/2
Dazu	1/1
Darin	0,5/1
TOTAL	5,5/8

Tabelle 63: Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°21-30)

6.1.15. EINSATZ DES PAAR AUF DEN TEXTABSCHNITT (N°31-50)

- 31- Edgar Bartl Bad Kissingen Geht es nach dem bayerischen Ministerrat und der Unterfränkischen Regierung, werden die Kommunen künftig viel enger zusammenarbeiten. Auch und gerade bei den Standesämtern. **Darüber** informierten bei einer Dienstbesprechung Peter Müller (Regierung) und Robert Bromma (Amt für Landwirtschaft) die Bürgermeister aus dem Landkreis. (SAZ11/JAN.00192 Saale-Zeitung, 04.01.2011, S. 4; Kein "Kahlschlag" bei den Standesämtern)
- 32- Aufbruch der Jugend Das zweite große Thema betraf die Jugend. Der Aufbruch nach dem Zweiten Weltkrieg war plötzlich in eine Krise geraten. "Und auch wir als Jugendseelsorger von der ersten Garde her waren nicht mehr fähig, dies aufzufangen. **Darum** musste man fragen: Wie gehen wir mit der heutigen Jugend um, nachdem die Jugend plötzlich ganz andere Ziele hatte, andere Aufgaben sah, beispielsweise die Öko-Frage oder den Frieden." (SAZ11/JAN.00547 Saale-Zeitung, 06.01.2011, S. 33; Die Synode hat den Anstoß zur Ökumene gegeben)
- 33- Nach 1945 konnten viele Forscher der Gesellschaft ihre Karrieren fortsetzen. "Da wurde unterschieden, ob jemand ein guter oder ein schlechter Forscher war", sagt Rürup. "Es ging nicht **darum**, ob jemand Nazi gewesen war oder nicht." Die Kaiser-Wilhelm-Gesellschaft änderte ihren Namen. Nobelpreisträger Max Planck, der wie Otto Hahn als integer und unbelastet galt, übernahm im hohen Alter noch einmal die Präsidentschaft. Die Gesellschaft erhielt seinen Namen. (SAZ11/JAN.00869 Saale-Zeitung, 11.01.2011, S. 34; ohne Titel)
- 34- 17 593 Entleihungen konnten die ehrenamtlichen Mitarbeiterinnen der katholischen öffentlichen Bücherei in Garitz im Jahr 2010 registrieren. Das teilte Büchereileiterin Silvia Ammersbach mit. Demnach sind in der Bücherei insgesamt 6985 Medien (Bücher, Spiele, CDs, Tonkassetten, CD Roms, Filme, Zeitschriften) vorhanden. 449 Medien kamen im vergangenen Jahr neu hinzu, 583 mussten wegen Überalterung oder Verschleiß aussortiert werden. 494 Leser nutzten das kostenlose Angebot der Einrichtung. **Darunter** finden sich nicht nur Garitzer Bürger, so Silvia Ammersbach, sondern auch eifrige Leser aus anderen Stadtteilen und den umliegenden Orten. Viele Leserwünsche konnten durch die Nutzung der Katholischen Büchereiarbeit (KBA) erfüllt werden. Zur Ergänzung des eigenen Bestandes werden jedes Jahr rund 1000 Bücher von der KBA ausgeliehen. Diese "Austauschbücher" bleiben sechs Monate in Garitz. (SAZ11/JAN.00288 Saale-Zeitung, 05.01.2011, S. 2; 494 nutzten 6985 Medien 17 593 Mal)
- 35- Im April 1953 heiratete er die belgische Prinzessin Joséphine- Charlotte, die Schwester der belgischen Könige Baudouin I. und Albert II. Aus der Ehe gingen fünf Kinder hervor, **darunter** der derzeit amtierende, 54 Jahre alte Großherzog Henri. 1964 bestieg Jean den Thron, nachdem seine Mutter Charlotte abgedankt hatte. (SAZ11/JAN.00407 Saale-Zeitung, 05.01.2011, S. 31; Luxemburgs Garant für Freiheit und Unabhängigkeit)
- 36- "Die Rahmenbedingungen sind im kommenden Jahr grundsätzlich so, dass Banken gute Gewinne machen können", sagt Bankenprofessor Hans-Peter Burghof. Die Banken profitieren **davon**, dass Notenbanken weiter billiges Geld zur Verfügung stellen. Die Europäische Zentralbank (EZB) dürfte beim Leitzins im Euro-Raum frühestens Ende 2011 das Rekordtief von 1,0 Prozent verlassen. Selbstbewusst plant die Deutsche Bank einen Vorsteuergewinn von 10 Milliarden Euro. **Damit** würde operativ der Rekord des Jahres 2007 (8,7 Mrd Euro) eingestellt, obwohl der Branchenprimus noch die Postbank-Eingliederung stemmen muss. (SAZ11/JAN.00018 Saale-Zeitung, 03.01.2011, S. 18; Banken erster und zweiter Klasse)
- 37- Leutheusser-Schnarrenberger räumte ein, der FDP werde nicht zu Unrecht vorgeworfen, dass sich von ihren Wahlkampfversprechen zu wenig wiederfinde im Regierungshandeln. Als Außenminister habe Westerwelle klar Erfolge zu verzeichnen. Seine schlechten Umfragewerte hätten mit der Leistung der Partei insgesamt zu tun. Beim Bundesparteitag im Mai werde die Führung neu gewählt. "Vieles hängt **davon** ab, was Guido Westerwelle selbst möchte. Wir sollten ihn ganz persönlich entscheiden

- lassen, ob er noch einmal als Vorsitzender antritt", sagte die Ministerin. (SAZ11/JAN.00031 Saale-Zeitung, 03.01.2011, S. 19; Druck auf Westerwelle wächst)
- 38- Gesundheitsminister Rösler zeigte sich **davon** überzeugt, dass der Parteichef weiterkämpfen wird: "Guido Westerwelle ist der beste Wahlkämpfer, den wir haben. Er wird uns erfolgreich in die Wahlkämpfe des Superwahljahres 2011 führen", sagte der FDP-Politiker. Niebe warf den Kritikern Westerwelles vor, sie wollten "ihr persönliches Süppchen kochen". (SAZ11/JAN.00031 Saale-Zeitung, 03.01.2011, S. 19; Druck auf Westerwelle wächst)
- 39- 18 Taufen 18 Kinder wurden getauft, neun Jugendliche feierten Konfirmation. Trauungen gab es fünf, Bestattungen neun. Vier Gemeindemitglieder erklärten ihren Austritt aus der Kirche, zwei Personen traten in die Gemeinde ein. 1306 Mitglieder werden in der Gemeinde betreut, so Pfarrer Pennig. **Davon** wohnen 656 in der Kernstadt und 650 in den Außenorten. Knapp 15 Prozent der Gemeindemitglieder sind Jugendliche, rund 60 Prozent sind erwerbsfähige Erwachsene und 15 Prozent Rentner. (SAZ11/JAN.00043 Saale-Zeitung, 03.01.2011, S. 11; Besinnlicher Jahreswechsel)
- 40- Auch im jetzt begonnenen Jahr stehen wieder zahlreiche Wanderungen auf dem Programm. **Dazu** gehören wie stets die beliebten Wanderungen mit Kurgästen. Die erste wird es am 5. April gegeben, Ziel ist das Schloss Saaleck. (SAZ11/JAN.00005 Saale-Zeitung, 03.01.2011, S. 14; Mit Fackeln auf den Volkersberg)
- 41- In der Realität sehe es heute so aus, dass viele junge Leute abends erst nach Essen, Sport und Duschen ausgehen. "Dann ist es oft 23 Uhr, und der Besuch einer Veranstaltung, bei der um 2 Uhr die Musik abgedreht wird, würde sich dann nicht mehr lohnen, wenn auch noch für den Eintritt sechs oder acht Euro zu zahlen sind", machte Fenn deutlich. Die Folge: Die jungen Menschen würden dann lieber gleich in eine größere Stadt fahren, wo es keine Sperrstunde gebe. Daher könnte die Regelung **dazu** führen, dass die professionelle Gastronomie gegenüber den Veranstaltern auf dem Land, darunter viele örtliche Vereine, bevorteilt werde. (SAZ11/JAN.00079 Saale-Zeitung, 03.01.2011, S. 7; "Sperrstunde bedroht Jugendkulturszene")
- 42- Dem soll die Kira-Studie vorbeugen. Nach vier und nach acht Monaten werden die Patienten für jeweils eine Woche in die Klinik am Kurpark zurückkehren. **Dazwischen** halten sie ständig telefonischen Kontakt mit ihrem Kissinger Physiotherapeuten. (SAZ11/JAN.00061 Saale-Zeitung, 03.01.2011, S. 1; Kira als Weg zur Genesung)
- 43- **Wozu** die DSV-Läufer dennoch imstande sind, zeigte der Frankenhainer Jens Filbrich. Auf dem klassischen Teilstück diktierte er das Rennen mit. "Ich habe mich richtig gut gefühlt. Doch dann kam ein Sturz **dazwischen**, und schon waren 20, 25 Plätze verloren. Zudem habe ich mir das Schienbein richtig wehgetan und konnte im Skating nicht so, wie ich wollte", sagte der Thüringer. (SAZ11/JAN.00148 Saale-Zeitung, 04.01.2011, S. 20; Teichmann steigt nach Rang 56 aus)
- 44- Zum Glück hat Henry eine schlaue Schulfreundin. "Sie schreibt für mich in der Schule mit und erklärt es mir später." Trotzdem haben sich seine Noten im vergangenen Jahr verschlechtert. Henry weiß warum: "Wegen der Weltmeisterschaft war so viel **dazwischen**, dass ich ab und zu eine Vier kassieren musste. Bisher hatte ich meistens Einsen und Zweien." Momentan geht Henry in die achte Klasse. Den Schulabschluss will er auf jeden Fall schaffen — und danach will er Motocross-Berufsfahrer werden. (SAZ11/FEB.01913 Saale-Zeitung, 15.02.2011, S. 30; ohne Titel)
- 45- Die Fundamente, die aus mit Beton ausgegossenen Stahllarmierungen bestehen, sind rund 350 Tonnen schwer und geben den Kolossen den nötigen Halt. Eine Anlage vom Typ V90 wiegt nach Vestas-Angaben 331 Tonnen. Bereits komplett unter der Erde verlegt sind die Kabel, die den Strom von Gauaschach nach Hundsbach fließen lassen, wo sich ein Umspannwerk der Firma Wotan befindet. **Hieran** sind noch zwei weitere Anlagen des Hamburger Unternehmens angeschlossen. Dort wird die elektrische Energie dann ins öffentliche Netz eingespeist. (SAZ11/JAN.00444 Saale-Zeitung, 06.01.2011, S. 9; Strom aus Wind für rund 3300 Haushalte)

- 46- Beim Kauf von Pflanzen und Saaten achtet der Hobbygärtner noch viel zu wenig auf Qualität. Vor allem bei Billigangeboten ist kritisch vorzugehen. Ein wenig Fachwissen und Ruhe beim Einkauf gehört schon **dazu**, die Spreu vom Weizen zu trennen. So hat der Bund Deutscher Baumschulen ein Markensiegel herausgegeben, welches auf den Etiketten zu sehen ist. **Hierauf** müssen auch die botanischen Namen der Gehölze, die Sorten und Größen sowie Hinweise zur Pflanzung angegeben sein. Ähnlich kennzeichnen auch die Staudengärtner ihre Produkte. (SAZ11/APR.01451 Saale-Zeitung, 12.04.2011, S. 33; Beim Einkaufen auf die Qualität achten)
- 47- Um rund 120 000 Euro blieb die Gewerbesteuer hinter den Erwartungen zurück. Die Verschuldung stieg im vergangenen Jahr um etwa 85 000 Euro auf 1,2 Millionen Euro an. Aus den laufenden Einnahmen und Ausgaben konnten dem Vermögenshaushalt 292000 Euro zugeführt werden. **Hieraus** lässt sich ablesen, dass die Kommune trotz angespannter finanzieller Lage noch Mittel für kommende Investitionen "erwirtschaften" konnte. Die Feststellung der Jahresrechnung und die Entlastung der Gemeindeverwaltung erfolgten einstimmig. bit (SAZ11/JUN.01069 Saale-Zeitung, 10.06.2011, S. 17; ohne Titel)
- 48- Stadtrat Christian Fenn begründete den Antrag mit den verhältnismäßig großen Waldflächen der Stadt Hammelburg. **Hieraus** erwachse der Kommune eine große Verantwortung gegenüber dem Forst. (SAZ11/JUN.01185 Saale-Zeitung, 11.06.2011, S. 9; "Großer Wald bedeutet")
- 49- Als dieser am Lkw vorüberfuhr, begann dieser aufgrund seines Gewichtes zu rutschen und schob sich seitlich auf den Pkw. **Hierbei** entstand leichter Sachschaden. red (SAZ11/JAN.00037 Saale-Zeitung, 03.01.2011, S. 12; Lkw rutscht auf Pkw)
- 50- Chorleiter im Amt bestätigt Die Bestätigung bei den Wahlen für drei Jahre erhielt ebenso Chorleiter Peter Hartmann. Er bedankte sich bei den Sängerinnen und Sängern für ihr Engagement und Treue zum Gesang; Vorsitzende Göbel wiederum für seine Gabe der Motivation, man werde sich jedoch um Verstärkung bei den Männerstimmen bemühen. Sie lobte den Probenbesuch, möchte sich aber um vermehrt öffentliche Auftritte kümmern, **wozu** sich das Jubiläum bereits als Einstieg anbiete. **Hierbei** ist an ein Singen für Lebende und Verstorbene mit anschließendem Fastenessen im April und Gestaltung von Gottesdiensten gedacht. Ein Ehrenabend im Juli, Singen mit Gastchören und Sommerfest, sowie ein Konzert mit dem Musikverein im November sind die Planungen im Jubiläumsjahr 2011. (SAZ11/JAN.00924 Saale-Zeitung, 11.01.2011, S. 13; Bewegender Abschied aus dem Vereinsvorstand)

The screenshot shows the Spyder Python IDE interface. On the left, a file explorer shows the project structure with files like 'Anaphora_filter.py' and 'PAAR.py'. The main editor displays a Python script with the following code:

```

30 for sentence in
31     word_tok =
32
33     if len(POS
34         pass
35     else:
36         for pa
37             af
38
39             if
40
41
42     el:
43
44     el:
45
46
47
48
49
50 language = 'german'
51

```

On the right, the 'Console' tab shows the output of the script. It displays the results of the 'anaphora_classifier_system' function for the German language. The output includes the following text:

```

In [4]: anaphora_classifier_system(T,'german')
The Da-/Hier- PA Darüber in the sentence:
['Darüber informierten bei einer Dienstbesprechung Peter Müller
(Regierung) und Robert Bromma (Amt für Landwirtschaft) die Bürgermeister
aus dem Landkreis.'] is an anaphora with antecedent the sentence:
Auch und gerade bei den Standesämtern.

The Da-/Hier- PA Darum in the sentence:
['Darum musste man fragen: Wie gehen wir mit der heutigen Jugend um,
nachdem die Jugend plötzlich ganz andere Ziele hatte, andere Aufgaben sah,
beispielsweise die Öko-Frage oder den Frieden.']] is an anaphora with
antecedent the sentence:
"Und auch wir als Jugendseelsorger von der ersten Garde her waren nicht
mehr fähig, dies aufzufangen.

The Da-/Hier- PA darum in the sentence:

```

Abbildung 72: Einsatz des PAAR auf den Textabschnitt (N°31-50) (1)

In der Abbildung 72 stellen wir fest, dass PAAR 2 anaphorische PA erkannt und aufgelöst hat. Im Gegensatz zur Auflösung vom „Darum“ die fehlerfrei realisiert wurde, ist die Auflösung vom „Darüber“ komplett falsch realisiert worden. Und zwar weil „darüber“ hier Bezug auf den Satz „Edgar Bartl Bad Kissingen Geht es nach dem bayerischen Ministerrat und der Unterfränkischen Regierung, werden die Kommunen künftig viel enger zusammenarbeiten“ nimmt.

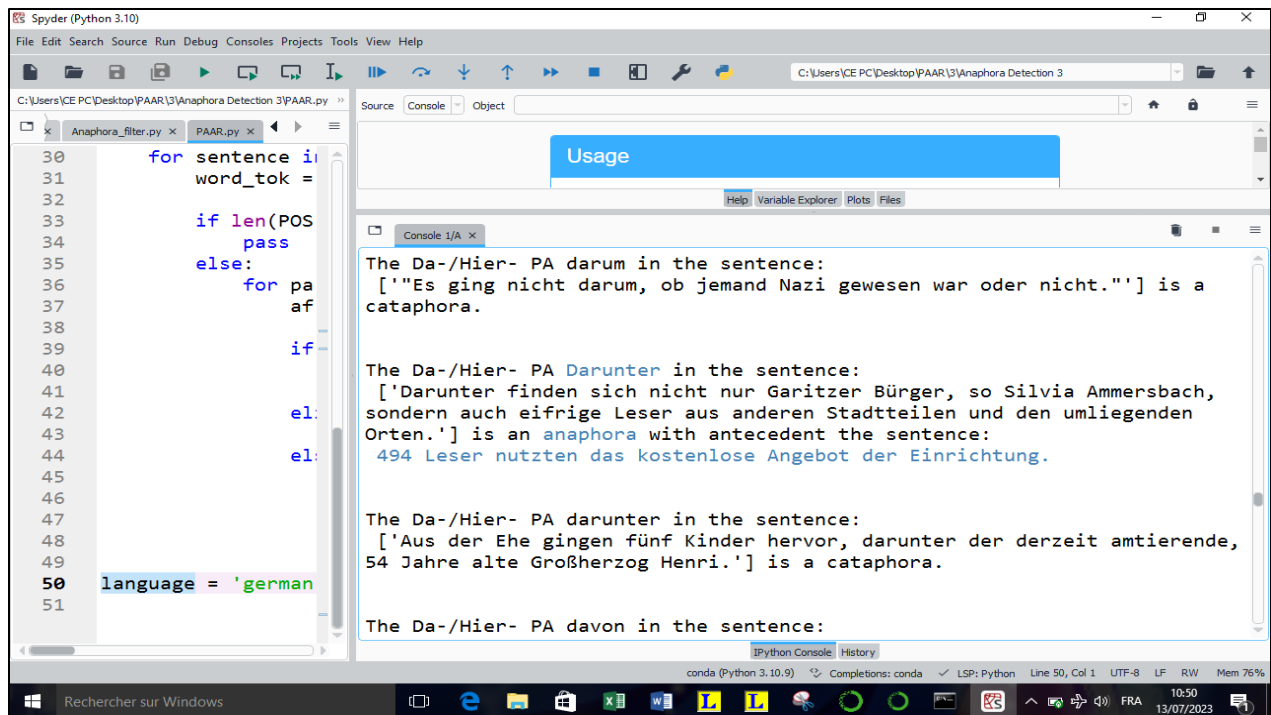


Abbildung 73: Einsatz des PAAR auf den Textabschnitt (N°31-50) (2)

In der Abbildung 73 lässt sich nämlich ablesen, dass PAAR 3 PA aufgefunden hat. Er verzeichnet das PA „Darum“ als eine anaphorische Erscheinung. Dabei wurde die entsprechende Bezugseinheit auch richtig ausgewählt. Was die beiden Erscheinungen vom PA „Darunter“ angeht, wurde die erste als ein anaphorisches PA anerkannt. Etwas das eben richtig ist. Allerdings stellt man nämlich fest, dass das entsprechende Antezedens nur teilweise richtig aufgefunden wurde. Denn das PA „Darunter“ dort verweist nur auf die NP „494 Leser“ nicht auf den ganzen Satz, wie PAAR es aufgezeichnet hat. Weiterhin klassifiziert PAAR die zweite Erscheinung vom PA „Darunter“ als Katapher. Etwas das eben auch nicht richtig ist. Denn dieses „Darunter“ nimmt Bezug auf sprachliche Einheiten („fünf Kinder“), die rückwärts stehen.

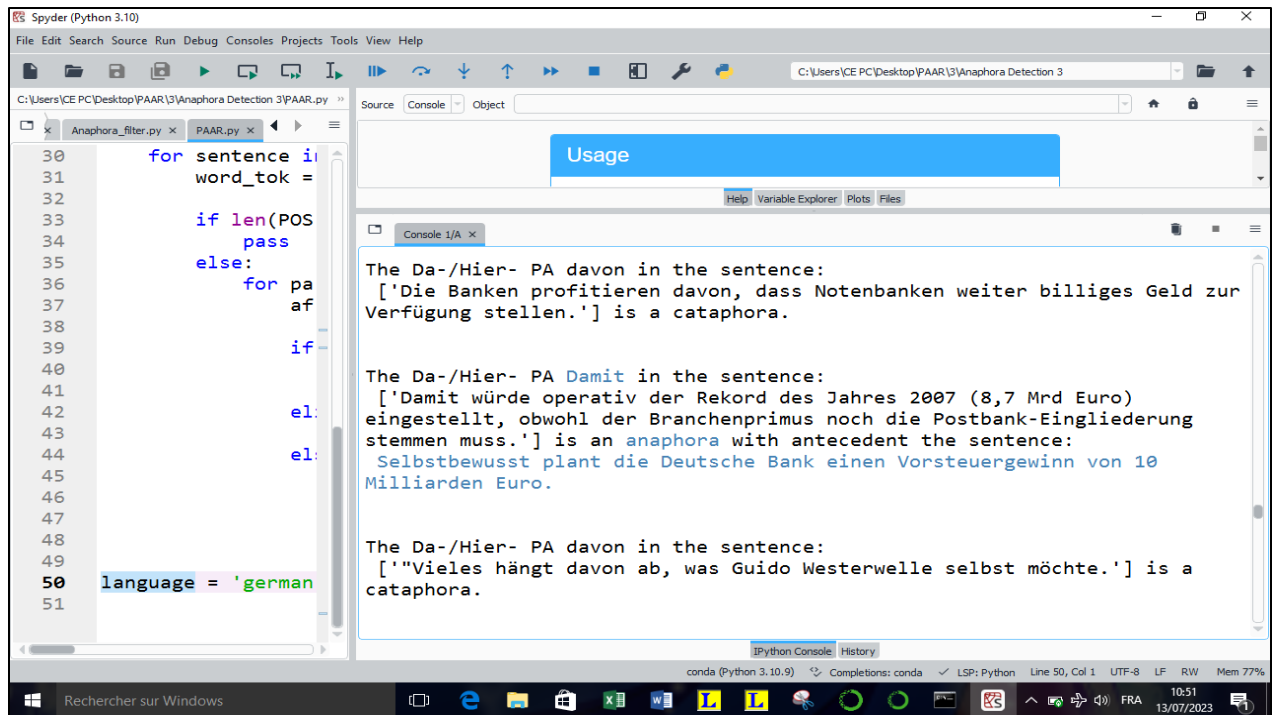


Abbildung 74: Einsatz des PAAR auf den Textabschnitt (N°31-50) (3)

Von der Abbildung 74 lässt sich ablesen, dass PAAR 3 PA erkannt hat: 2 Erscheinungen von „Davon“ und eine Erscheinung von „Damit“. Während er die beiden „davon“ als kataphorische Erscheinungen klassifiziert, etwas, das eben auch richtig ist, klassifiziert er das PA „Damit“ als Anapher und gibt dabei das passende Antezedens.

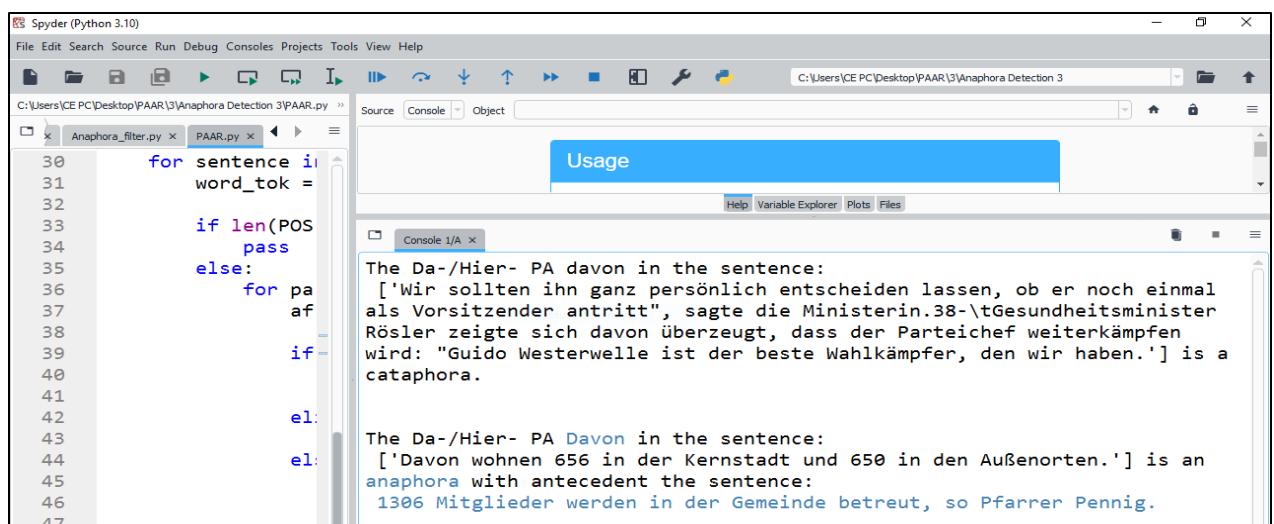


Abbildung 75: Einsatz des PAAR auf den Textabschnitt (N°31-50) (4)

Laut Informationen aus der Abbildung 75 stellen wir nämlich fest, dass PAAR zwei weitere PA („davon“) aufgefunden hat. Während das erste wurde als kataphorisch aufgezeichnet wurde, wurde die zweite Erscheinung eher als Anapher klassifiziert. Etwas, das auch richtig ist. Allerdings müssen wir darauf hinweisen, dass das vom PAAR ausgewählte Antezedens nur zum Teil richtig ist. Und zwar, weil „Davon“ hier verweist nur auf die NP „1306 Mitglieder“. Weiterhin erkennt PAAR das PA „dazu“ und klassifiziert es als eine Katapher, etwas das auch nicht richtig ist, denn diese Erscheinung vom PA „Dazu“ ist eine Anapher mit dem Antezedens „Auch im jetzt begonnenen Jahr stehen wieder zahlreiche Wanderungen auf dem Programm.“

Aus der Abbildung 76 merken wir, dass PAAR 3 PA erkannt hat. Das erste ist „Dazu“ und ist als Katapher kategorisiert worden. Etwas, das auch richtig ist. Was das zweite erfundene PA angeht, handelt es sich um das PA „Darunter“. PAAR hat es als Katapher verzeichnet. Etwas, dass nicht richtig ist, denn „darunter“ verweist hier auf den Textteil „auf dem Land,“. Letztendlich erkennt PAAR ein weiteres PA („Dazwischen“) und verzeichnet es als Anapher. Etwas das auch richtig ist. Zugleich wurde auch das passende Antezedens gefunden.

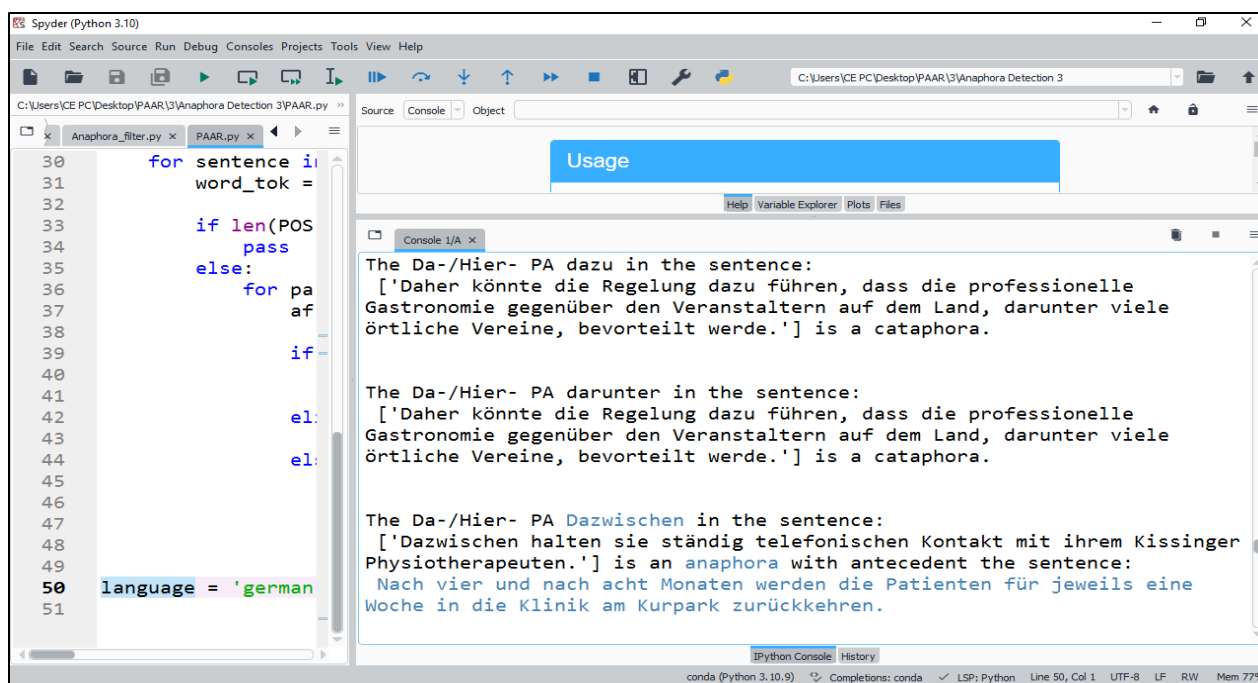


Abbildung 76: Einsatz des PAAR auf den Textabschnitt (N°31-50) (5)

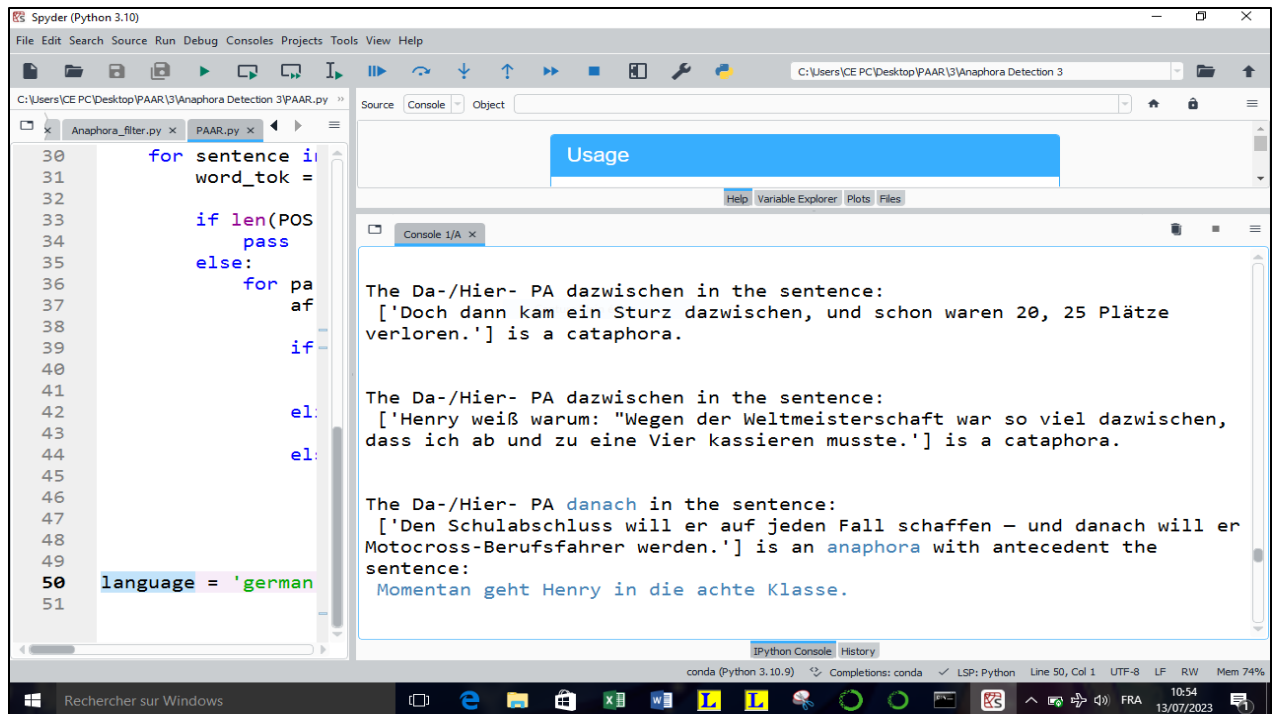


Abbildung 77: Einsatz des PAAR auf den Textabschnitt (N°31-50) (6)

Die Abbildung 77 gibt bekannt, dass PAAR 3 PA erfunden hat; zwei davon klassifiziert er als Katapher und die letzte als Anapher. Dabei hat er nur die Natur von einem PA richtig erkannt. Denn die erste Erscheinung vom „Dazwischen ist keine Katapher, sondern eher eine Anapher mit Bezugspunkt auf „Ich habe mich richtig gut gefühlt“. Was die zweite Erscheinung von „dazwischen“ angeht, handelt es sich um keine Anapher, sondern eher um ein Deixis. Denn „Dazwischen“ hier referiert auf Realitäten, die weder im Vor- noch im Nachtext versprachlicht worden sind. Allerdings haben wir bei der Erstellung des Systems nur zwei große Kategorien bezüglich der Natur des PA erstellt: ANAPHER und NICHT ANAPHER (Katapher & Deixis). Daher werden wir die Erkennung der Natur vom zweiten „Dazwischen“ nicht als ein Fehler betrachten. Was das dritte erkannte PA „danach“ angeht, können wir sagen, dass PAAR es korrekt aufgelöst hat, und dies obwohl die Top Lösung der Satzteil „Den Schulabschluss will er auf jeden Fall schaffen“ wäre. Allerdings stellt man nach Näherbetrachtung des Abschnittes, dass PAAR ein PA („Wozu“) versehen hat.

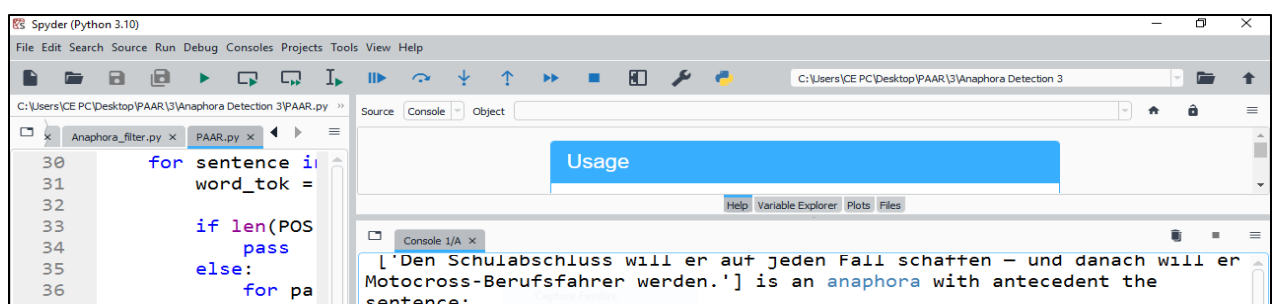


Abbildung 78: Einsatz des PAAR auf den Textabschnitt (N°31-50) (7)

Bei Näherbetrachtung der Angaben, die in der Abbildung 78 abzulesen sind, stellen wir nämlich fest, dass PAAR zwei PA („Hieran“ & „dazu“) erkannt hat, wobei die jeweiligen richtigen Naturen und die entsprechenden Antezedenzien bei gegebenen Fällen zugeschrieben wurden.

In der Abbildung 79 dagegen merkt man, dass PAAR die Naturen von allen bestehenden PA nicht richtig annotiert hat. Dies ist Beispielweise der Fall fürs PA „Hierauf“. Dieses ist keine Katapher, sondern eher eine Anapher mit dem Satz „Aus den laufenden Einnahmen und Ausgaben konnten dem Vermögenshaushalt 292000 Euro zugeführt werden“ als Bezugspunkts, also als Antezedens. Was das erste PA der Reihe angeht, handelt es sich tatsächlich um ein anaphorisches „Hierauf“. Allerdings steht es nämlich fest, dass PAAR das falsche Antezedens ausgewählt hat, denn „Hierauf“ hier bezieht sich nur auf die NP „den Etiketten“ und nicht auf

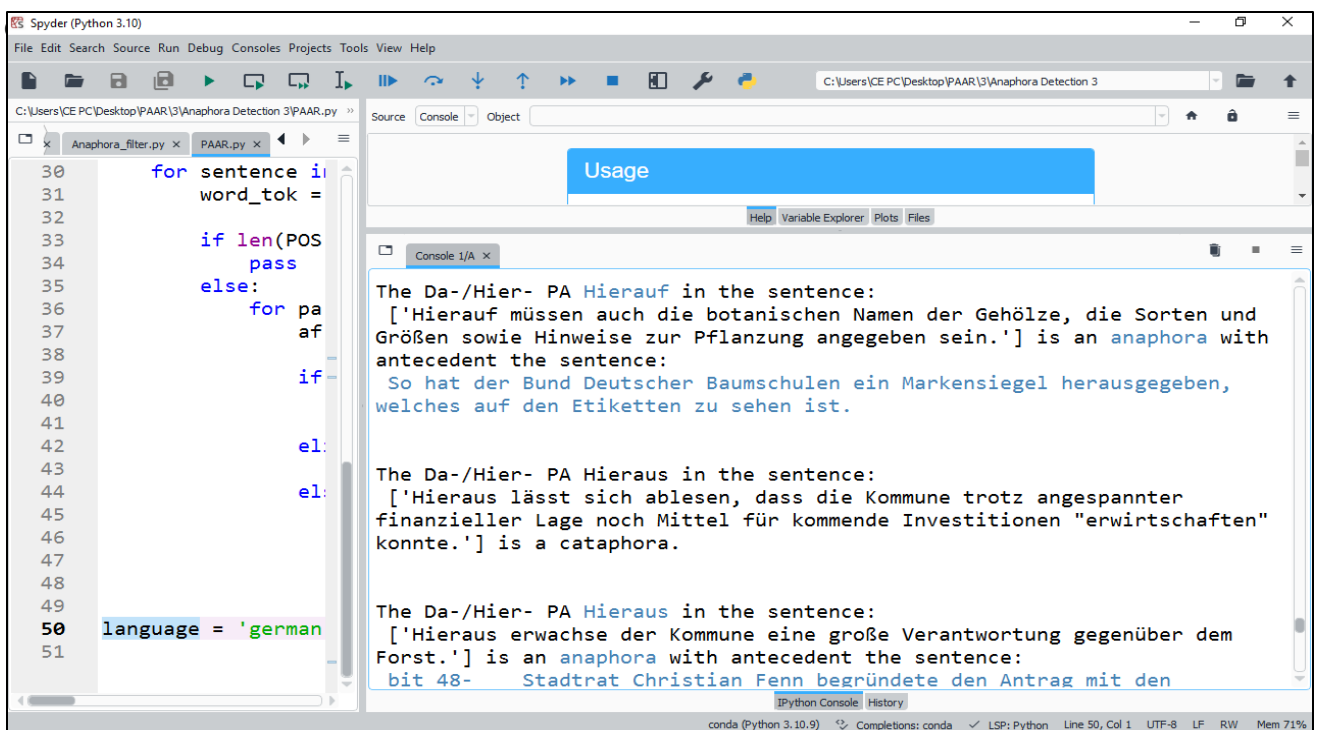


Abbildung 79: Einsatz des PAAR auf den Textabschnitt (N°31-50) (8)

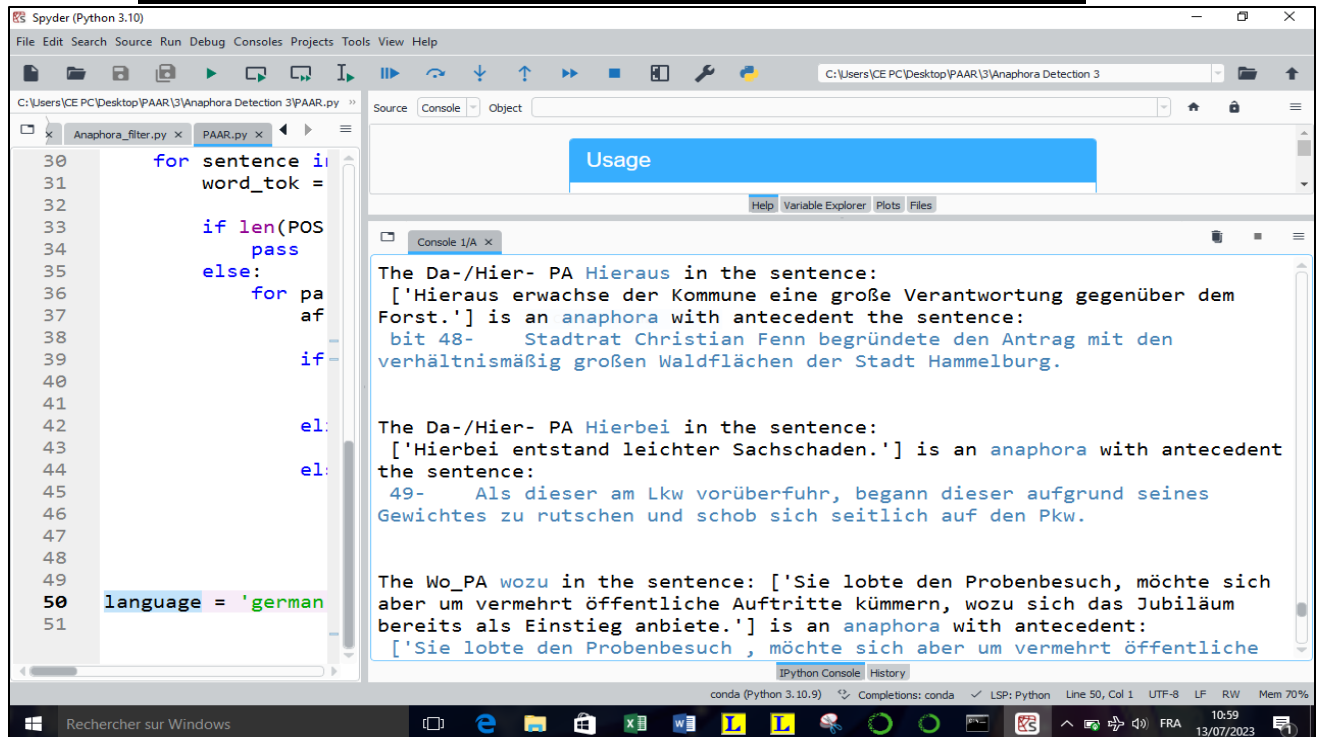


Abbildung 80: Einsatz des PAAR auf den Textabschnitt (N°31-50) (9)

Bei Näherbetrachtung der Angaben, die in der Abbildung 80 abzulesen sind, stellen wir nämlich fest, dass PAAR zwei PA („Hieraus“ & „Hierbei“) erkannt hat, wobei die jeweiligen richtigen Naturen und entsprechenden Antezedenzen richtig wiedergegeben wurden.

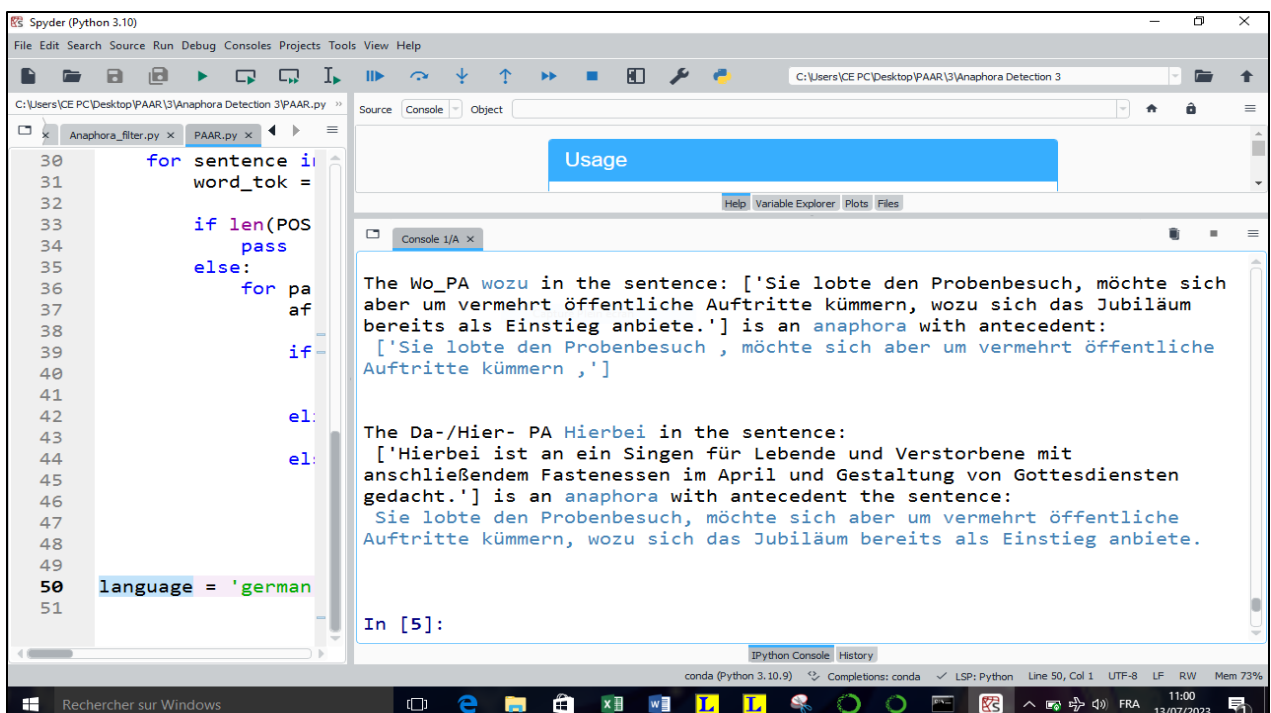


Abbildung 81: Einsatz des PAAR auf den Textabschnitt (N°31-50) (10)

Bei Näherbetrachtung der Angaben, die in der Abbildung 81 abzulesen sind, stellen wir nämlich fest, dass PAAR zwei PA („Wozu“ & „Hierbei“) erkannt hat, wobei die jeweiligen richtigen Naturen und entsprechenden Antezedenzen richtig wiedergegeben wurden. Darunter folgt eine Zusammenfassung der Analysen und Interpretationen der Einsatzaktion von PAAR sowohl auf den Textabschnitt (31-50) als auch auf das ganze PAK. Für die Zusammenfassung der Analysen und Interpretationen bezüglich der Einsatzaktion auf die Textabschnitte (51-75); (76-100) und (101-154) siehe bitte den Anhang.

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
26	25	25/26

Tabelle 64: Performanz vom PAAR bei Erkennung von PA im Text N°31-50

PA	Anzahl	Natur	Ergebnisse
Darüber	1	Anapher	1/1
Darum	2	Anapher	2/2
Darunter	1	Anapher	1/1
Darunter	2	Katapher	0/2
Davon	3	Katapher	3/3
Damit	1	Anapher	1/1
Davon	1	Anapher	1/1
Dazu	1	Anapher	1/1
Dazu	2	Katapher	2/2
Dazwischen	1	Anapher	1/1
Dazwischen	2	Katapher (Nicht Anapher)	1/2
Danach	1	Anapher	1/1
Hieran	1	Anapher	1/1
Hierauf	1	Anapher	1/1
Hieraus	1	Katapher	0/1
Hieraus	1	Anapher	1/1
Hierbei	2	Anapher	2/2
Wozu	1	Anapher	1/1
TOTAL			21/25

Tabelle 65: Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°31-50)

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Darüber	0/1
Darum	2/2
Darunter	0,5/2
Damit	1/1
Davon	0,5/1
Dazu	0/1

Dazwischen	1/1
Danach	0,5/1
Hieran	1/1
Hierauf	0,25/1
Hieraus	1/1
Hierbei	2/2
Wozu	1/1
TOTAL	10,25/16

Tabelle 66: Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°31-50)

6.2. Ergebnisse der Einsatzaktion: Berechnung der Präzision von PAAR

Darunter wird die Performanz vom PAAR bezüglich 3 Kriterien, wie etwa, die Erkennung von PA, deren jeweiligen Natur sowie auch deren Auflösung, wenn es sich um anaphorische Erscheinungen angeht, zusammengefasst. Weiterhin wird die Performanz vom PAAR bezüglich jeder PA-Hauptgruppe (Da-; Hier- und Wo-PA) berechnet.

Im PAK enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
261	245(239 PA+6SUBJUNKTIONEN)	245/261

Tabelle 76: Performanz vom PAAR bei Erkennung von PA im PAK

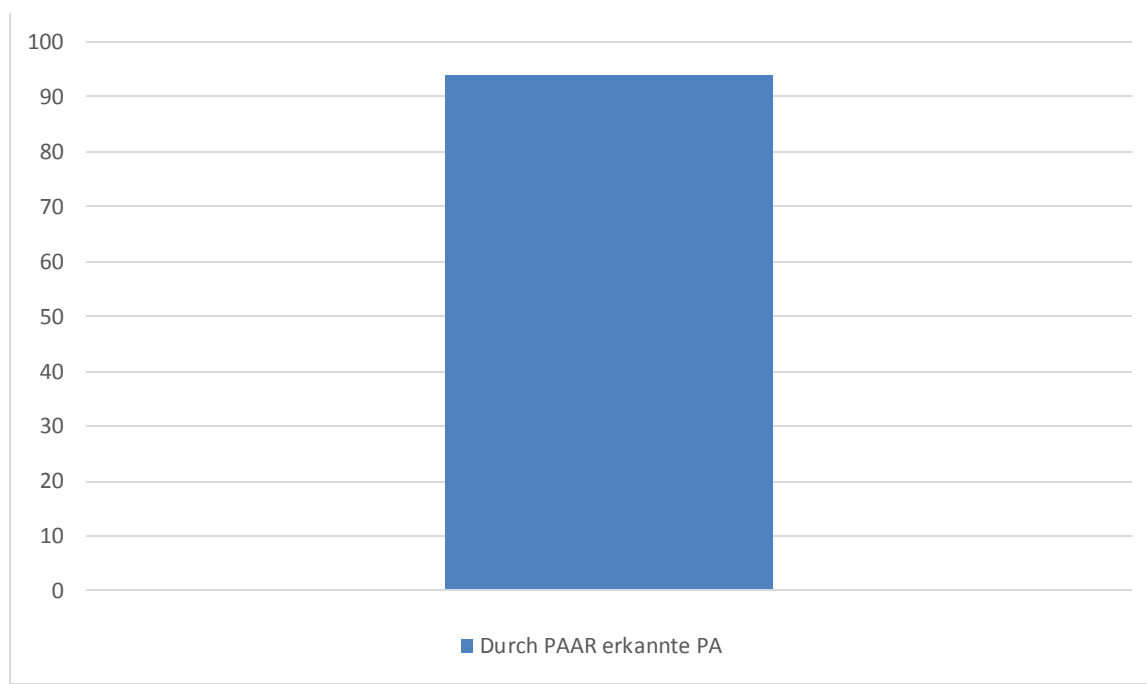


Abbildung 155: Performanz vom PAAR bei der Erkennung von PA im PAK

Diese Graphik zeigt uns nämlich, dass PAAR einen Durchschnitt von 93,8% bei der Erkennung von im Text (PAK) erhaltene PA erzielt hat. Somit wurde alle Wörter, die die Morphologie

einer der 60 PA aus der PA-Liste (Siehe Hierfür Kapitel 5) aufweist, erkannt. Allerdings stellten wir auch während der Analyse fest, dass Wörter wie DAMIT zwar diese Morphologie aufweisen, tauchen manchmal aber als Subjunktionen und nicht als Pronominaladverbien (PA). Und dies könnte als einer der Gründe sein, warum PAAR bei der Erkennung vom PA nicht 100% geleistet hat. Ein weiterer Grund, weshalb PAAR nicht alle im Text erhaltene PA erkannt hat, geht auf die Performanz des ausgewählten Parsers bzw. des ausgewählten Wort-Parsers zurück. Man merkte nämlich während der Analyse, dass manche PA einfach bei der Erkennung ignoriert wurden, weil der Text zuvor falsch zerlegt wurde. Es kam manchmal vor, dass wenn PA direkt nach einer Nummer stehen, wurden sie nicht als solches erkannt und war weil der Parser die beiden Einheiten als eine aufgezeichnet hat (Siehe Hierfür die Verarbeitung des Textes 108 in Abbildungen 121&122). Darüber hinaus merkten wir, dass manche Erscheinungen von PA in abgekürzten Formen (drauf, drüber, drunter....) nicht erkannt wurde. Genauso auf diesem Aspekt muss PAAR noch verbessert werden. Schauen wir uns nun die Graphik, die Angaben über die Performanz vom PAAR bezüglich der Erkennung der Natur (Anapher/ Keine Anapher (Deixis/Katapher)) von PA gibt.

PA	Anzahl	Natur	Ergebnisse
Daran	3	Katapher ²²⁹	3/3
Daran	1	Anapher	1/1
Danach	1	Katapher	1/1
Danach	6	Anapher	6/6
Dagegen	2	Katapher	1/2
Dagegen	4	Anapher	4/4
Darauf	5	Katapher	5/5
Darauf	4	Anapher	4/4
Davon	6	Anapher	6/6
Davon	10	Katapher	10
Damit	8	Anapher	8/8
Damit	5	Katapher	2/5
Daraus	1	Anapher	1/1
Darüber	1	Katapher	1/1
Darüber	2	Anapher	2/2
Dabei	7	Anapher	7/7
Dabei	5	Katapher	1/5
Daneben	1	Anapher	1/1
Darein	1	Katapher	1/1
Darein	2	Anapher	2/2
Darin	3	Katapher	3/3

²²⁹Beim Gebrauch des Ausdrucks KATAPHERS wird sowohl auf kataphorisch als auch deiktisch funktionierende sprachliche Einheiten referiert.

Darin	1	Anapher	1/1
Dahinter	1	Katapher	1/1
Dahinter	3	Anapher	3/3
Dazu	8	Anapher	8/8
Dazu	10	Katapher	5/10
Dafür	3	Anapher	3/3
Dafür	3	Katapher	2/3
Darum	2	Anapher	2/2
Darunter	1	Anapher	1/1
Darunter	4	Katapher	0/4
Dadurch	4	Anapher	4/4
Dadurch	2	Katapher	1/2
Dazwischen	3	Anapher	2/3
Dazwischen	3	Katapher	1/3
Hierbei	3	Anapher	3/3
Hierein	3	Anapher	3/3
Hieraus	1	Katapher	0/1
Hieraus	1	Anapher	1/1
Hierdurch	4	Anapher	4/4
Hierfür	3	Anapher	3/3
Hiergegen	1	Anapher	1/1
Hierauf	2	Anapher	2/2
Hieran	1	Anapher	1/1
Hierhinter	1	Anapher	1/1
Hierin	2	Anapher	2/2
Hiermit	1	Anapher	1/1
Hiernach	2	Anapher	2/2
Hierneben	1	Anapher	0/1
Hierneben	1	Katapher	0/1
Hierüber	1	Anapher	1/1
Hierum	1	Anapher	1/1
Hierunter	2	Anapher	2/2
Hierunter	1	Katapher	0/1
Hiervon	1	Anapher	1/1
Hierzu	8	Anapher	8/8
Hierzwischen	1	Anapher	1/1
Wohinter	1	Katapher	0/1
Wohinter	2	Anapher	2/2
Wogegen	2	Anapher	2/2
Wofür	1	Katapher	1/1
Wofür	1	Anapher	1/1
Womit	2	Katapher	0/2
Womit	3	Anapher	3/3
Wonach	3	Anapher	3/3
woran	1	Deixis	1/1
Woran	1	Anapher	0/1
Worauf	3	Anapher	2/3
Worauf	2	Katapher	2/2

Woraus	1	Anapher	0/1
Woraus	2	Katapher	1/2
Worein	3	Anapher	3/3
Worin	3	Anapher	1/3
Worüber	3	Anapher	3/3
Worum	3	Anapher	0/3
Worum	2	Katapher	2/2
Worunter	4	Anapher	4/4
Worunter	1	Katapher	0/1
Wovon	2	Anapher	1/2
Wozu	4	Anapher	1/4
Wozu	1	Katapher	1/1
Wozwischen	4	Anapher	4/4
Wobei	1	Katapher	0/1
Wobei	10	Anapher	10/10
Wodurch	4	Anapher	4/4
TOTAL	239		195/239

Tabelle 77: Performanz vom PAAR bei Erkennung der Natur von PA im PAK

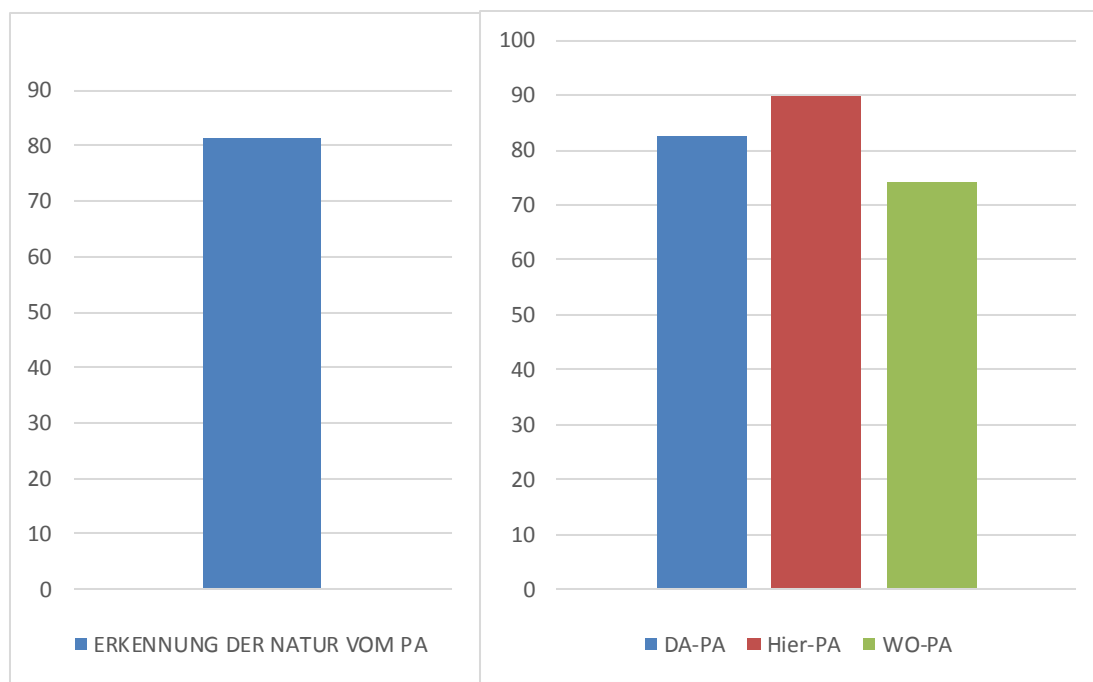


Abbildung 156: Performanz vom PAAR bei der Erkennung der Natur von im PAK erhaltenen PA

Die erste Graphik zeigt uns die allgemeine Performanz von PAAR bei der Zuweisung der richtigen Natur zu den zuvor erkannten PA. Dabei hat PAAR einen Durchschnitt von 81,58% bekommen. Dies bedeutet, dass die von uns entwickelten Katapher-Filtern relativ gut funktioniert haben. Dadurch konnte PAAR nämlich eine Performanz von 90% bei der

Sortierung von kataphorischen und anaphorischen Erscheinungen von Hier-PA (2te Graphik) erzielen. Und was dieselbe Aufgabe bezüglich der Da-PA angeht, bekommt PAAR einen Durchschnitt von 82,3%, während es bei Wo-PA eher eine Leistung von etwa 74% erzielt. All dies bedeutet vor allem, dass PAAR in der Lage war, PA zu erkennen und ihnen ihre jeweilige Natur zuzuweisen. Etwa, dass für den Weiterverlauf der Auflösungsprozedur von zentraler Relevanz ist. Als nächster Schritt der Prozedur kommt die konkrete Auflösung von PA, die als Anapher klassifiziert worden sind. Die folgenden Tabellen und Graphiken fassen die Performanz vom PAAR dazu zusammen.

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Danach	3/5
Daran	1/1
Dahinter	2,25/5
Daneben	1/1
Darauf	3/4
Darüber	0/1
Darum	2/2
Darunter	0,5/2
Damit	5,75/8
Dazu	5,5/8
Dazwischen	2/2
Dadurch	3/4
Davon	2/5
Dabei	4,25/5
Dagegen	3/4
Darin	0,5/1
Dafür	3/3
Darein	0,5/2
Daraus	0,5/1
Darüber	1/1
Hiermit	1/1
Hierein	2/3
Hieran	1/1
Hierauf	1,25/2
Hieraus	1/1
Hierbei	3/3
Hierdurch	3/4
Hierfür	1,5/3
Hiergegen	1/1
Hierhinter	1/1
Hierin	0,5/2
Hiernach	1,5/2
Hierüber	1/1
Hierum	1/1
Hierunter	1/2
Hiervon	0,5/1

Hierzu	2/8
Hierzwischen	1/1
Wohinter	1,5/2
Wofür	1/1
Wogegen	1,5/2
Wobei	8,75/10
Womit	2,25/3
Wonach	1,5/3
Worauf	1,75/2
Wodurch	3,5/4
Worein	1,5/3
Worin	0,5/1
Wozu	1/1
Worüber	2,5/3
Worunter	2/4
Wovon	1/1
Wozwischen	3,5/4
TOTAL	101,75/141

Tabelle 78: Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK

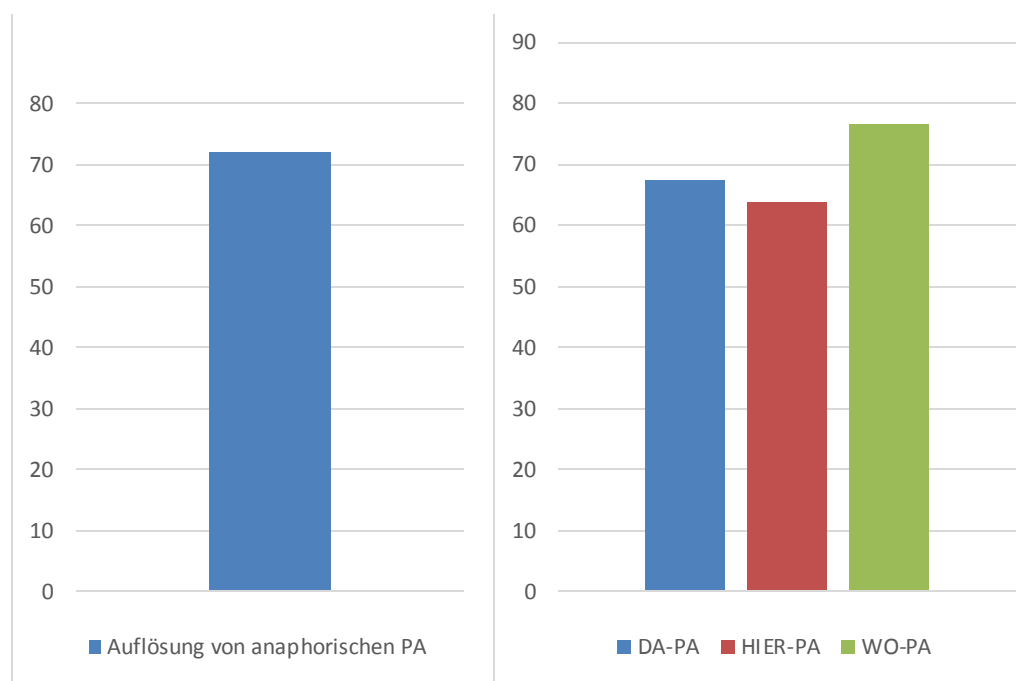


Abbildung 157: Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK

Aus den beiden Graphiken lässt sich nämlich ablesen, dass PAAR eine allgemeine Performanz von 72,16% bei der Auflösung von anaphorischen Pronominaladverbien erzielt hat. Betrachtet man nun seine Performanz nur bezüglich jeder PA-Varianten, dann stellen wir fest, dass PAAR fast 80% bei der Auflösung von Wo-PA erreicht hat, wogegen nur 62,3% und 67,8% jeweils

für Hier- & Da-PA erreicht worden sind. Weiterhin wollen wir noch zur besseren Illustrierung der Performanz vom PAAR bei der Auflösung von anaphorischen PA, eine Leistungswelle bezüglich jedes einzelnen PA erstellen.

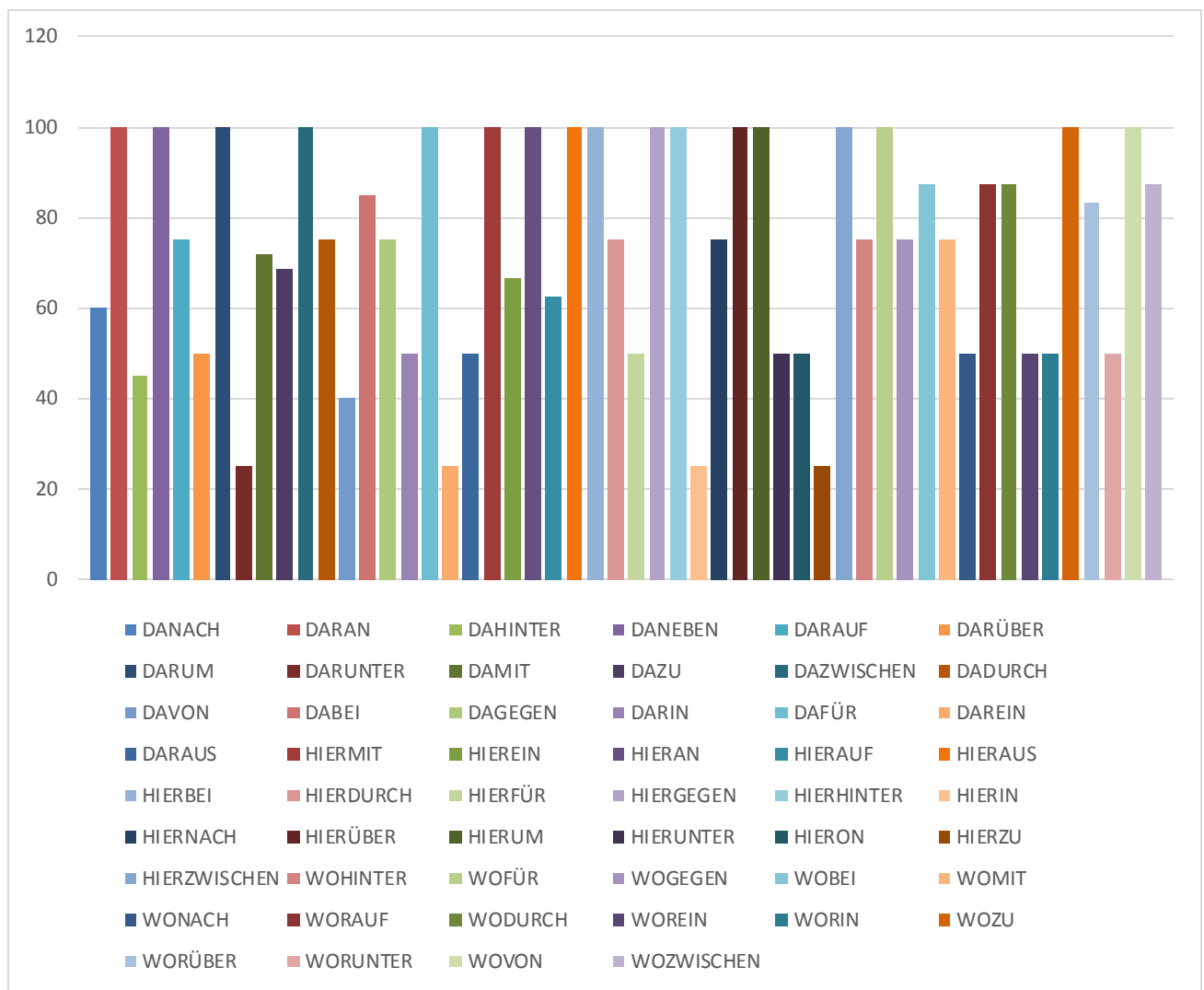


Abbildung 158: Performanz von PAAR bei der Auflösung von anaphorischen PA im PAK bezüglich aller aufgelösten PA

6.3. Hypothesenprüfung; Schlussbetrachtung und Ausblicke

Die sich abschließende Forschungsarbeit hatte als maximales Ziel, ein System zur Resolution von Pronominaladverbienanaphern zu entwickeln. Zur Erreichung dieses Ziels mussten jedoch Zwischenaufgaben erledigt werden. Doch bevor wir in Einzelheiten über die Natur und die Nützlichkeit jedes dieser Zwischenstopps berichten, wollen wir zunächst einmal hier noch einige Auskünfte über den Anreiz (Motivation) dieser Arbeit geben.

Der Forschungsbereich, in dem und für den dieses Forschungsprojekt ausgedacht und realisiert wurde, ist der Computerlinguistik bzw. der NLP (Natural Language Processing) / TALN (Traitement Automatique du Langage Naturel). Wir gingen nämlich von der Feststellung aus, dass bestehende Systeme zur Auflösung von anaphorischen Einheiten (AR-Systeme) nicht in der Lage waren, sprachliche Einheiten, wie Adverbien, zu verarbeiten bzw. aufzulösen. Wir interessierten uns daher auf die Gründe, die diesen Mangel erklären könnten. Und uns wurde nämlich klar, dass die Auflösung solcher sprachlichen Einheiten mit klassischen Ansätzen (heuristischen Ansätzen) überhaupt nicht möglich wäre. Und zwar, weil der Aufbau semantischer Module dafür benötigt würden. Jedoch sind solche Module nur dank höchstkomplexer mathematischer Prozesse (Maschinelearning) erstellbar. Trotzdem gingen wir mit der Forschung weiter und entdeckten, dass es eine Art von Adverbien gibt, die sich möglicherweise auch durch heuristische Ansätze verarbeiten lassen könnte und wofür es noch kein AR-System besteht. Wir entschieden uns daher diesem Forschungsweg nachzugehen, um letztendlich in der Lage zu sein, unsere Vermutungen zu prüfen.

Auf diesem Forschungsweg mussten wir, wie gesagt, mehrere Zwischenstationen machen. Die erste davon ist der Überblick, den wir uns im Kapitel 1 bezüglich des Gebrauchs von solchen Systemen gemacht haben. Es gibt sich nämlich heraus, dass solche Systeme wichtige Komponente für umfangreiche informatische Systeme sind, worin mit natürlichen Sprachen operiert wird. Dies ist beispielweise der Fall bei Systemen zur automatischen Textgenerierung; zur automatischen Textzusammenfassung oder noch zur automatisierten Informationssuche. Sobald wir uns einen groben Überblick über das Wesen (Gestalt bzw. Gegenstände; Aufgabenbereichen; akademische, berufliche und gesellschaftliche Implikationen) des Forschungsgebiets COMPUTERLINGUISTIK gemacht haben; machten wir, in einem weiteren Zwischenstopp (Kapitel 1), das Gleiche für das Forschungsgebiet TEXTLINGUISTIK. Das Hauptziel hierfür war den Begriff „Anapher“ so gründlich und umfassend, wie möglich zu bestimmen. Und sobald wir einige Kenntnisse über die Natur und das Funktionieren von anaphorischen sprachlichen Einheiten gewonnen hatten, interessierten wir uns, in einem weiteren Schritt (Kapitel 3), für die sprachliche Einheit, die als Gegenstand unserer Forschungsarbeit gilt: Das Pronominaladverb. Nachdem wir diese Wortkategorie als eine Variante von Adverbien eingeordnet haben, interessierten wir uns für seine Anaphorizität, wobei wir anhand einiger sich wiederholenden Merkmale Regeln entworfen konnten, welche die Auffindung des Antezedens eines anaphorischen Pronominaladverbs ermöglichen werden. Im Anschluss daran gelang es uns, einen Algorithmus auf Beine zu stellen, der diese Aktion

automatisieren sollte. Und angesichts der Art und Weise, mit der dieser Algorithmus entworfen worden ist, konnten wir unsere zweite Hypothese bezüglich der Nützlichkeit der zu berücksichtigen linguistischen Parameter zwischen Algorithmen zur Resolution von anaphorischen NP/Pronomen mit denen der anaphorischen Adverbien, vor allem, wenn es sich um Pronominaladverbien geht, verifizieren.

Sobald wir wussten, wie wir theoretisch vorgehen werden, um anaphorische Pronominaladverbien aufzulösen, blieb es uns nur noch Wege zu erforschen, um herauszufinden, wie wir dieses Wissen (Auflösung von Pronominaladverbienanaphern) dem Computer beibringen könnten. Dafür haben wir einen weiteren Zwischenstopp (Kapitel 4) gemacht, der uns geholfen hat, die Architektur von bestehenden AR-Systemen gründlich zu untersuchen. Daraus konnten wir nämlich feststellen, dass manche Module, wie Beispielsweise der PARSE und Morphosyntaktisch-Filter ständig vorhanden sind. Die Besonderheiten dieser Module liegen in ihrer Fähigkeit, dem Computer zu ermöglichen, anaphorische Einheiten zu identifizieren und diese Letzten den weiteren Modulen des Systems für die Auflösung bereitzustellen. Solche Module sind also zum Aufbau unseres AR-Systems unabdingbar. In dieser Hinsicht lässt sich auch unsere erste Hypothese verifizieren. Diese besagt nämlich, dass manche Module von AR-Systemen für Pronomen/ NP auch in AR-Systeme für Pronominaladverbien vorhanden würden. Weiterhin gingen wir der Erforschung der anerkannten Architekturen von AR-Systemen tiefer nach, wobei es uns am Ende gelang, eine passende Architektur für das angestrebte System zu erstellen. Und sobald wir diese Architektur eines Systems zur Auflösung von anaphorischen Pronominaladverbien hergestellt haben, haben wir uns, in einem weiteren Schritt (Kapitel 5), die Hilfe eines Informatikers (Programmierer) geholt. Dieser war nämlich dafür zuständig, diese Architektur (den Algorithmus) in Computersprache umzuwandeln und somit den Algorithmus vom Computer gebräuchlich zu machen. Dabei stellten wir nämlich fest, dass die Programmiersprache par Exzellenz für Angelegenheiten im Bereich der NLP (Natural Language Processing) PYTHON ist. Zudem kamen wir zur Erkenntnis, dass beim Entwurf solcher computerlinguistischen Systeme alle Komponente in der selben und gleichen Programmiersprache realisiert werden müssen. Somit wurde unsere letzte Hypothese sowohl verifiziert als auch zugleich falsifiziert. Denn der Gebrauch von Modulen eines AR-Systems in einem anderen ist durchaus möglich aber setzt voraus, dass die beiden Systeme mit der gleichen Programmiersprache codiert werden.

Nachdem das angestrebte AR-System, namens PAAR (Pronominaladverbienanapher Resolver), entwickelt wurde, testierten wir es anhand eines für das Manöver konstituierten Textkorpus. Mit einer Präzision von etwa 73% bei der Auflösung von anaphorischen Pronominaladverbien lässt sich PAAR jedoch in vielerlei Hinsichten verbessern. Aus den Analysen der verschiedenen OUTPUTS vom PAAR, ergibt es sich nämlich, dass manche Auflösungsverfahren fehlgeschlagen sind, nicht weil das Hauptmodul an seine Grenze gestoßen ist, sondern eher, weil der Parser falsch funktioniert hat. In der Tat gab es manche PA, die entweder gar nicht erkannt werden konnten und zwar, weil die Segmentierung fehlerhaft abgelaufen war. Sogar, wenn sie erkannt wurden, war das ausgewählte Antezedens eben auf Grund fehlerhafter Segmentierung nicht volltreffend. Weiterhin stellte es sich heraus, dass PAAR hauptsächlich Schwierigkeiten erfunden hat, wenn das Antezedens des anaphorischen PA kein Satz war, sondern eher eine NP im vorangehenden Textteil. Zur Erhebung dieses Mangels und somit zur Verbesserung der Performanz unseres AR-Systems beabsichtigen wir in den zukünftigen Forschungen, Ansätze zur Eruierung dieses Mangels zu entwickeln. Allerdings können wir schon, von vornherein, sagen, dass wir entweder Fall für Fall vorgehen werden, um zu sehen, welche PA sich am meisten so verhalten hat und im Anschluss daran analysieren, welche Auffälligkeiten aufkommen und von daher unser Programm /CODE entsprechend aufbauen. Wir könnten aber auch in einer weiteren Forschungsperspektive, die Methode des Maschinellernens gebrauchen und somit einen Grundalgorithmus (Lernalgorithmus)entwerfen bzw. codieren und ihm dann eine gute Menge an ausgewählten und zielgerecht vorverarbeiteten Daten (Anapher-Antezedens annotierte Textdaten) auszusetzen. Ziel dieses Schulungsmanövers wäre, dass der Algorithmus sich selbst Regularitäten merkt, um beim Einsatz auf beliebige Texte in der Lage zu sein, fehlerfrei passende Antezedenzen auszuwählen und somit eine deutlich höhere Performanz, im Gegensatz zu der ersten „Version“ von PAAR, zu erzielen.

Literaturverzeichnis

Literaturverzeichnis

1. Bücher

- ADVAITH SIDDHARTHAN. 2003. Resolving Pronouns Robustly: Plumbing the Depths of Shallowness. In Proceedings of the 2003 EACL Workshop on The Computational Treatment of Anaphora.
- ALPAYDIN, E., 2004: Introduction to machine learning, *Alpaydin, Ethem: Introduction to machine learning*, MIT Press, Cambridge, Mass.
- ALPAYDIN, E., 2019, Maschinelles Lernen. Berlin, Boston: De Gruyter Oldenbourg. <https://doi.org/10.1515/9783110617894>
- ALTMANN, Hans, 1981, Formen der »Herausstellung« im Deutschen. Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen. Tübingen.
- ALTMANN, Hans/KEMMERLING, Silke, 2000, Wortbildung fürs Examen. Wiesbaden.
- ARTHUR L. SAMUEL, 1959: Some studies in machine learning using the game of checkers. IBM J Res Dev 3:210–229. doi:10.1147/rd.33.0210
- ASHER, Nicholas, 1993, Reference to abstract objects in discourse. Dordrecht, Kluwer.
- BALZERT, H 1998: Lehrbuch der Softwaretechnik 2. Auflage
- BEAUGRANDE, R. A. DE/ DRESSLER, W. U., 1981. Einführung in die Textlinguistik. Tübingen, Niemeyer.
- BECK, KATHRIN. 2013. TüBa-D/Z Release 8. www.sfs.unituebingen.de/ascl/ressourcen/corpora/tueba-dz.html (am 22.11.2012).
- BIBER, DOUGLAS. 1995. Dimensions of register variation: A cross-linguistic comparison. Cambridge: Univ. Press.
- BOBROW, DANIEL GUREASKO. 1964. „Natural Language Input for a Computer Problem Solving System“. AI Technical Reports 66.
- BREINDL, Eva, 1989, Präpositionalobjekte und Präpositionalobjektsätze im Deutschen. Berlin: De Gruyter (Linguistische Arbeiten, 220).
- BRINKER, Klaus, 1992. Textlinguistik. Heidelberg, Groos.
- BRINKER, K. et al. (Hg.), 2000. Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung. 1. Halbband: Textlinguistik. Berlin, New York, de Gruyter.
- BRINKER, K., 2001, Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden, Berlin.
- BRINKER, Klaus. 2005, Linguistische Textanalyse: eine Einführung in Grundbegriffe und Methoden. Berlin, Erich Schmidt.
- BRINKER, Klaus, 2010a. Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden, Berlin, Schmidt.
- BRINKER Klaus & SAGER Sven F. 2006, Linguistische Gesprächsanalyse: Eine Einführung (Grundlagen der Germanistik (GrG), Band 30) ISBN: 3503079815, Erich Schmidt Verlag GmbH & Co.
- BRINKER, Klaus/CÖLFEN, Hermann/PAPPERT, Stefen, 2014, Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden, Berlin.

BYSTRIVA Bohušová, Zuzana, 2005 : Kontrastive Phonetik Slowakisch-Deutsch. Segmente. Assimilation. Banská Bystrica: UMB

CAROLA STROBL, 2019, Darum sind Pronominaladverbien eine Herausforderung für Deutschlerner: Eine korpusbasierte kontrastive Interimssprachenanalyse. Hierzu Germanistische Mitteilungen, Jahrgang 45 (2019), Ausgabe 1&2 © 2019 Universitätsverlag WINTER GmbH Heidelberg Germanistische Mitteilungen 45.1 | 2019 89

CHOMSKY, Noam, 1981, Lectures on Government and Binding, Dordrecht.

CONSTEN, M., 2004. Anaphorisch oder deiktisch? Zu einem integrativen Modell domänengebundener Referenz. Tübingen: Niemeyer.

CORNISH, F., 1999. Anaphora, Discourse, and Understanding: Evidence from English and French. Oxford: Clarendon Press.

DANIEL JURAFSKY, JAMES H. MARTIN, 2008, Speech and Language Processing-An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition 2. Auflage. Prentice Hall, Upper Saddle River, New Jersey ISBN 978 -0-13-187321-6 <https://de.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03)

DENBER MICHEL 1998, Automatic Resolution of Anaphora in English, Technical report, Eastman Kodak Co; http://www.wlv.ac.uk/~le1825/anaphora_resolution_papers/denber.ps

DETLEF H. ROST, 2009, Intelligenz: Fakten und Mythen. Beltz Psychologie Verlags Union, , ISBN 978-3-621-27646-7.

DETLEF H. ROST, 2013: Handbuch Intelligenz. Beltz, Weinheim, ISBN 978-3-621-28044-0.

DÖBEL, INGA ; LEIS, DR. M. ; VOGELSANG, MANUEL M. ; NEUSTROEV, DMITRY ; PETZKA, DR. H. ; RÜPING, DR. S. ; VOSS, DR. A. ; WEGELE, MARTIN ; WELZ, DR. J. 2018: *Maschinelles Lernen - Kompetenzen, Anwendungen und Forschungsbedarf*. München : Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. Elektronische Publikation, 2018. URL [https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer Studie ML 201809.pdf](https://www.bigdata.fraunhofer.de/content/dam/bigdata/de/documents/Publikationen/Fraunhofer_Studie_ML_201809.pdf). (09.01.2019)

DRESSLER, Wolfgang, 1972, Einführung in die Textlinguistik, Tübingen.

DRESSLER, Wolfgang (Hrsg.) 1978. Textlinguistik, Darmstadt, Wissenschaftliche Buchgesellschaft.

DREYER, Hilke/SCHMITT, Richard, 2012, Lehr- und Übungsbuch der deutschen Grammatik aktuell. 2. Aufl. Ismaning: Hueber Verlag (Die Gelbe aktuell).

EICHINGER, Ludwig M., 1989, Raum und Zeit im Verbwortschatz des Deutschen. Eine valenzgrammatische Studie. Tübingen.

EICHINGER, Ludwig M., 2000, Deutsche Wortbildung. Eine Einführung. Tübingen.

EISENBERG, Peter, 1994, Grundriss der deutschen Grammatik. 3., überarb. Aufl. Stuttgart Metzler.

EISENBERG, Peter, 2004, Grundriss der deutschen Grammatik Der Satz. 2., überarbeitete und aktualisierte Auflage. 2 Bände. Stuttgart, Weimar: Metzler (2).

EISENBERG, Peter, 2006a, Grundriss der deutschen Grammatik. Das Wort. 3., durchges. Aufl. 2 Bände. Stuttgart, Weimar: J.B. Metzler (1).

EISENBERG, Peter, 2006b, Grundriss der deutschen Grammatik • Der Satz. 3., durchges. Aufl. 2 Bände. Stuttgart, Weimar: J.B. Metzler (2).

ENGEL, Ulrich, 1991, Deutsche Grammatik. 2., verb. Aufl. Heidelberg: Groos.

ENGEL, Ulrich, 2004, Deutsche Grammatik. Neubearb. München: Ludicum.

EROMS, Hans-Werner, 1986, Funktionale Satzperspektive, Tübingen, Niemeyer, Germanistische Arbeitshefte 34.

EROMS, Hans-Werner, 2000, Syntax der deutschen Sprache. Berlin, New York, Walter de Gruyter.

- FELIPE MARTINS, 2014: IPython Notebook Essentials. Packt Publishing, Birmingham [u. a.] 2014, ISBN 978-1-78398-834-1, S. 190.
- FILLMORE, Charles J, 1973: Deixis I [and] Deixis II. Unpublished lectures, mimeo.
- FLEISCHER, Jürg, 2002, Die Syntax von Pronominaladverbien in den Dialekten des Deutschen. Eine Untersuchung zu Preposition Stranding und verwandten Phänomenen. Unbearb. zugl.: Zürich, Univ., Diss., 2001-2002. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik Beihefte 123).
- FLEISCHER, Wolfgang, 1969, Wortbildung der deutschen Gegenwartssprache. Leipzig.
- FLEISCHER, WOLFGANG/BARZ, IRMHILD, 2007, Wortbildung der deutschen Gegenwartssprache. Unter Mitarbeit von Marianne Schröder. Tübingen
- GANSEL, Christina/JÜRGENS, Frank, 2009, Textlinguistik und Textgrammatik. Eine Einführung Neuausg. der 2. überarb. u. erg. Aufl. Göttingen: Vandenhoeck & Ruprecht (UTB Sprachwissenschaft, 3265).
- GETHMANN, C. F., BUXMANN, P., DISTELRATH, J., HUMM, B. G., LINGNER, S., NITSCH, V., SCHMIDT, J. C. & SPIECKER, I., 2022. Künstliche Intelligenz in der Forschung (Bd. 48) [E-Book]. Springer. <https://doi.org/10.1007/978-3-662-63449-3>
- GISA Rauh, 1978, Linguistische Beschreibung deiktischer Komplexität in narrativen Texten. Zugl.: Göttingen, Diss., 1978. Tübingen: Narr (Tübinger Beiträge zur Linguistik, 106)
- GIVÓN, T., 1983, *Topic continuity in discourse : a quantitative cross-language study*. Amsterdam : John Benjamins Publishing
- GORELICK MICHA, OZSVALD IAN, 2014: High performance Python. O'Reilly, Sebastopol, CA, ISBN 978-1-4493-6159-4, S. 370.
- GÖTZE, LUTZ/HESS-LÜTTICH, ERNEST W. B., 1999, Grammatik der deutschen Sprache. Sprachsystem und Sprachgebrauch. Aktualisierte Neuausg. Gütersloh: Bertelsmann-Lexikon-Verl. (Nachschlagewerke zur deutschen Sprache im Bertelsmann-Lexikon-Verlag, 4).
- GRAMMIS-PROJEKTGRUPPE, 2012, Pronominaladverbien in Konnektorfunktion. Institut für Deutsche Sprache. Mannheim [Online verfügbar unter http://hypermedia.ids-mannheim.de/call/public/sysgram.ansic?ht?vtyp=d&v_id=2602], zuletzt aktualisiert am 03.01.2012, zuletzt geprüft am 22.02.2014. http://hypermedia.ids-mannheim.de/call/public/termwb.ansicht?v_a_pp=g&v_id=110], zuletzt aktualisiert am 08.09.2009, zuletzt geprüft am 22.02.2014.
- GRAMMIS-PROJEKTGRUPPE, 2012, Pronominaladverbien in Konnektorfunktion. Institut für Deutsche Sprache. Mannheim [Online verfügbar unter http://hypermedia.ids-mannheim.de/call/public/sysgram.ansic?ht?vtyp=d&v_id=26021], zuletzt aktualisiert am 03.01.2012, zuletzt geprüft am 22.02.2014.
- GROSZ, BARBARA J. AND CANDACE L. SIDNER. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):1
- GROSZ, BARBARA J., ARAVIND K. JOSHI, AND SCOTT WEINSTEIN. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):20
- HAGEN HIRSCHMANN, 2019, Korpuslinguistik, J.B. Metzler, Stuttgart, ISBN 978-476-02643-9
- HAIDER Hubert, 2010, The Syntax of German (= Cambridge Syntax Guides). Cambridge University Press, Cambridge (UK).
- HALLIDAY, Michael Alexander Kirkwood & HASAN. Ruqaiya, 1976, Cohesion in English. London, Longman.
- HARWEG, Roland, 1968, Pronomina und Textkonstitution. München.
- HARWEG, Roland, 1979, Pronomina und Textkonstitution. 2., verbesserte und erg. Auflage. München.
- HARWEG, Roland, 1990, Perspektiven der Textlinguistik, Vortrag in Shanghai.

- HEINEMANN, Wolfgang & DIETER Viehweger, 1991, Textlinguistik. Eine Einführung, Tübingen, Niemeyer.
- HELBIG, Gerhard/BUSCHA, Joachim, 1991, Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. Leipzig: Verlag- Enzyklopädie.
- HELBIG, Gerhard/BUSCHA, Joachim, 2001, Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht. Neubearbeitete., [Nachdr.]. Berlin: Langenscheidt.
- HELBIG, Gerhard, 2002, Linguistische Theorien der Moderne, Berlin, Weidler Bucherverlag.
- HENTSCHEL, Elke/WEYDT, Harald, 1990, Handbuch der deutschen Grammatik. Berlin: De Gruyter,
- HENTSCHEL, Elke (Hrsg.), 2010, Deutsche Grammatik. Berlin: De Gruyter (De Gruyter Lexikon).
- HENTSCHEL, Elke/WEYDT, Harald, 2013, Handbuch der deutschen Grammatik. 3., völlig neu bearb. Aufl. Berlin: De Gruyter (de Gruyter Studienbuch).
- HENTSCHEL, Elke/WEYDT, Harald, 2013, Handbuch der deutschen Grammatik. 4., vollst. überarb. Aufl. Berlin: De Gruyter (De Gruyter Studium).
- HEUER, Walter/FLÜCKIGER, Max/GALLMANN, Peter, 2008, Richtiges Deutsch. Vollständige Grammatik und Rechtschreiblehre unter Berücksichtigung der aktuellen Rechtschreibreform. 28., überarb. Aufl. Zürich: Verl. Neue Zürcher Zeitung.
- HIRST, GRAEME JOHN, 1976A, Artificial intelligence and computational linguistics. In Methodology and problems. Unpublished report, Department of Computer Science, University of British Columbia.
- HIRST, GRAEME JOHN 1977A. Focus in reference resolution in natural language understanding. Paper presented at the Language and Speech Conference, Melbourne,
- HIRST, GRAEME JOHN 1977B: Cohesive discourse transitions and reference resolution: The cinema metaphor and beyond into the transfinite. Unpublished manuscript
- HIRST, GRAEME JOHN 1978B: A set of primitives for discourse transitions. Unpublishable manuscript.
- HIRST, GRAEME JOHN, 1982, Humorous information processing: Why AI must consider burnout. In preparation.
- HOBBS, JERRY R. 1974. "A Model for Natural Language Semantics, Part I: The Model." Research Report #36, Department of Computer Science, Yale University.
- HOBBS, JERRY R., 1975. "A General System for Semantic Analysis of English and its Use in Drawing Maps from Directions". American Journal of Computational Linguistics, Microfiche 32.
- HOBBS, JERRY R., 1976a "A Computational Approach to Discourse Analysis". Research Report 76-2, Department of Computer Sciences, City College, City University of New York.
- HOBBS, JERRY R., 1976b "Pronoun Resolution". Research Report 76-1, Department of Computer Sciences, City College, City University of New York. August 1976. [pdf, 2.2MB]
- HOBBS, JERRY R. 1977c "From 'Well-Written' Algorithm Descriptions into Code", Research Report 77-1, Department of Computer Sciences, City College, City University of New York.
- HOLMLANDER, Inger, 1979, Zur Distribution und Leistung des Pronominaladverbs. Das Pronominaladverb als Bezugselement eines das Verbergänzenden Nebensatzes/Infinitivs. Zugl.: Uppsala, Univ., Diss., 1979. Stockholm: Almqvist & Wiksell (Acta Universitatis Upsaliensis Studia Germanistica Upsaliensia, 21).
- JANN RAVELING, 2022 -Was ist künstliche Intelligenz? DIGITALISIERUNG / INDUSTRIE 4.0, <https://www.wfb-bremen.de/de/page/stories/digitalisierung-industrie40/was-ist-kuenstliche-intelligenz-definition-ki> (09.08.2023)
- JUNG, Thomas, 2005, Darstellung und Anwendung der Centering-Theorie nach Walker, München, GRIN Verlag, <https://www.grin.com/document/61851> [21/012021].

JUNG, WALTER, 1988, Grammatik der deutschen Sprache. Bearb. von Günter Starke. 9. KAI-UWE CARSTENSEN, CHRISTIAN EBERT, CORNELIA EBERT, SUSANNE JEKAT, RALF KLABUNDE UND HAGEN LANGER (HRSG.), 2010, Computerlinguistik und Sprachtechnologie: Eine Einführung, 3. Auflage, Heidelberg, Spektrum Akademischer Verlag, ISBN 978-3-8274-2023-7

KAI-UWE CARSTENSEN, 2017, Sprachtechnologie: Ein Überblick. Homepage: <http://www.kai-uwe-carstensen.de>. URL dieses PDF: <http://www.kai-uwe-carstensen.de/Publikationen/Sprachtechnologie.pdf> Email: contact at kai-uwe-carstensen.de Version 2.2

unveränderte Auflage, Leipzig: Bibliographisches Institut.

KALLMEYER, KLEIN, MEYER-HEMAN, NETZER, SIEBERT, 1974, Lekturkolleg zur Textlinguistik-Bd.1: Einführung: 45, Athenäum Fischer Taschenbuch Verlag, 380722050X

KARIN PITTNER, 1999, Adverbiale im Deutschen (= Studien zur deutschen Grammatik. Band 60). Stauffenburg Verlag, Tübingen.

KARIN PITTNER, JUDITH BERMAN, 2010, Deutsche Syntax. Ein Arbeitsbuch. 4. Auflage. Narr Verlag, Tübingen

KARS, Jürgen/HÄUSSERMANN, Ulrich, 1997, Grundgrammatik Deutsch. Unter Mitarbeit von Gerhard Koller/Sieglinde Gruber. 5. Aufl. Frankfurt am Main: Dienste-Neg.

KATIA, Iacono, Bakk, 2010, Kohäsion beim Simultandolmetschen aus dem Deutschen ins italienische und Spanische: Eine deskriptive korpusbasierte Analyse. Masterarbeit der Universität Wien

KLEIBER, G. 2001, L' anaphore associative, Paris: PUF.

KLIUKAITĖ Augustina, 2006, zu den Begriffen Kohärenz und Kohäsion, Universität Vilnius, <https://www.google.com/amp/s/augustinakliukaite.wordpress.com/2010/10/02/zuden-begriffen-koharenz-und-kohasion/amp/> [12.11.2020].

KOCH Ann-Kristin, 2001, Grammatische Bedingungen von Textkohärenz, München, GRIN Verlag, <https://www.grin.com/document/107853> [10.11.2020].

KONRAD Ehlich, 2007, Sprache und sprachliches Handeln: Band 1: Pragmatik und Sprachtheorie. Band 2: Prozeduren des sprachlichen Handelns. Band 3: Diskurs - Narration - Text - Schrift Berlin, New York: De Gruyter. <https://doi.org/10.1515/9783110922721.3.fm>

KREUTZER, R. T. & SIRRENBORG, M. 2019 : Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey (German Edition) (1. Aufl. 2019 Aufl.). Springer Gabler. <https://doi.org/10.1007/978-3-658-25561-9>

LEE, J. W. 2002, Textkohärenztypologie. Ein Beitrag zur Textlinguistik. Seul.

LURJIDZE. N, 2007, Grundlagen der Textlinguistik der geschriebenen Sprache, München, GRIN Verlag, [<https://www.grin.com/document/159663/>].

MCCARTHY, J.; MINSKY, M. L.; ROCHESTER, N.; SHANNON, C. E. 1955: A proposal for the Dartmouth summer research project on artificial intelligence.

MEYER JOSEF AND DALE ROBERT 2002: Mining a Corpus to Support Associative Anaphora Resolution,; Centre for Language Technology Macquarie University, Sydney, Australia

MITKOV, R., 1994, An integrated model for anaphora resolution. *COLING 1994 Volume 2: The 15th International Conference on Computational Linguistics*.

MITKOV, R., 2002, *Anaphora Resolution*. Longman

MITKOV Ruslan, 2002a *Anaphora Resolution*, Longman, London, UK.

MOSKALSKAJA, Olga Ivanovna 1984, Textgrammatik. Leipzig, Bibliographisches Institut. (russisch 1981. Grammatika teksta. Moskau: Wysschaja Schkola).

NEGELE, Michaela (2012b): Varianten der Pronominaladverbien im Neuhochdeutschen. Grammatische und soziolinguistische Untersuchungen. Berlin: De Gruyter (Studia linguistica Germanica, 108).

- NITIN INDURKHYA, FRED J. DAMERAU, 2010, *Handbook of Natural Language Processing*. 2. Auflage. Hauptman and Hall/CRC, 2010, ISBN 978-1-4200-8592-
<https://de.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03)
- NORVIG PETER, STUART RUSSELL, 2012; *Künstliche Intelligenz: ein moderner Ansatz*. Pearson, Higher Education, (Always learning). – ISBN 978-3-868-94098-557
- PÄL UZONYI, 1994 *Einführung in die Computerlinguistik*, Germanistisches Institut der Eötvös-Loránd-Universität Budapest, Budapest ISSN 0138-9055
- QUINLAN, J. ROSS., 1993, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- REY, GÜNTER D. 2017: *Neuronale Netze Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung*. Hogrefe, – URL http://neuronalesnetz.de/downloads/neuronalesnetz_de.pdf. – Letzter Zürich: 09.01.2019 – ISBN 978-3-456-84513-5
- RICHERT, WILLI 2013: *Building Machine Learning Systems with Python*. Packt Publishing (Community experience distilled). – 290 S. – ISBN 978-1-782-16141-7
- RÜTTENAUER Martin, 1978, Vorkommen und Verwendung der adverbialen Proformen im Deutschen. Hamburg: Buske (Papiere zur Textlinguistik, 16).
- SANDBERG, Bengt, 2004, Pronominaladverbien und finale damit-Sätze. Kritische, Korpusbasierte Anmerkungen. Frankfurt am Main: Lang (Germanistische Schlaglichter, N.F., Bd. 1).
- SCHECKER, M., WUNDERLI, P. (Hrsg.), 1975, Beiträge zum Problem der Textualität, Tübingen, Niemeyer.
- SCHECKER MICHAEL, 1996, Zur historischen und systematischen Entfaltung von 'Textualität', in: Michaelis S., Tophinke D. (Hgs.), *Texte - Konstitution, Verarbeitung, Typik*, München: Lincom Verlag.
- SCHERER CARMEN, 2006: *Korpuslinguistik*, Winter, Heidelberg, ISBN 3-8253-5164-5
- SCHMIDT, Siegfried J. 1976. *Texttheorie. Probleme einer Linguistik der sprachlichen Kommunikation*, 2. Auflage UTB 202. München, Fink.
- SCHMITZ Ulrich, 1992, *Computerlinguistik: Eine Einführung*, Opladen, Westdeutscher Verlag GmbH, ISBN 978-3-531-12350-9 ISBN 978-3-322-92483-4 (eBook) DOI 10.1007/978-3-322-92483-4
- SCHULZ, Dora/GRIESBACH, Heinz, 1972, *Grammatik der deutschen Sprache*. 9. neubearb. Aufl. München i.e. Ismaning: Hueber (Sprachen der Welt).
- SCHWARZ Monika, 1992: *Einführung in die kognitive Linguistik*, Francke, Tübingen
- SCHWARZ, Monika, 2000a, Indirekte Anaphern in Texten. Studien zur domänen gebundenen Referenz und Kohärenz im Deutschen. Tübingen. Im Netzwerk verfügbar unter <http://www.degruyter.com/viewbooktoc/Product/159259> [13.07.2020].
- ŠIMEĚKOVÁ, Alena, 1994, *Untersuchungen zum »trennbaren« Verb im Deutschen I*. Univerzita Karlova. Praha.
- STANFORD UNIVERSITY, 2016, *Artificial Intelligence and life in 2030. One hundred year study on artificial intelligence*. Abgerufen von https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai_100_report_0831fnl.pdf (abgerufen am 19.10.2019).
- STRUBE MICHAEL AND UDO HAHN 1995, *Parse Talk about Sentence- and Text-Level Anaphora*, Computational Linguistics Research Group, Freiburg University, Germany; <http://acl.ldc.upenn.edu/E/E95/E95-1033.pdf>
- TAYLOR, BROCK HAROLD, 1975, *A case-driven parser*. Unpublished MSc thesis, Department of Computer Science, University of British Columbia,
- THOMAS MITCHELL 1997: *Machine Learning*. Mcgraw-Hill, London, ISBN 978-0-07-115467-3.
- VATER, H, 1986, *Einführung in die Referenzsemantik*, Köln: Universität Köln (KLAGE 11)

- VATER, H, 1994, Einführung in die Textlinguistik, München, Fink (UTB 1660).
- VATER HEINZ, 2001, Einführung in die Textlinguistik. Struktur und Verstehen von Texten. 3.überarb. Aufl. München: Fink (UTB für Wissenschaft Uni-Taschenbücher Linguistik, Literaturwissenschaft, 1660).
- VATER HEINZ, 2005, Referenz-Linguistik. München: Wilhelm Fink-verlag (UTB Sprachwissenschaft, 2685).
- WALKER, Donald E, 1976. Speech understanding research. SRI Project 4762, Final technical report, SRI International, October.
- WALKER, Donald E, 1978. Understanding spoken language. (The computer science library, Artificial intelligence series 5), New York: North Holland,
- WALKER M.A. , JOSHI A.K. , E.F. PRINCE 1998 , Centering Theory in Discourse, Oxford Uni-versity Press,
- WEINRICH, Harald, 1971, Tempus. Besprochene und erzählte Welt. 2. Auflage. Stuttgart.
- WEINRICH, Harald, 1993, Textgrammatik der deutschen Sprache, Mannheim/Leipzig/Wien/Zürich, Dudenverlag.
- WEINRICH, Harald, 2003, Textgrammatik der deutschen Sprache. 2. Aufl. Wiesbaden.
- WEINRICH Harald, 2007, Textgrammatik der deutschen Sprache. Unter Mitarbeit von Maria Thurmair/Eva Breind Eva-Maria Willkop. 4.,rev- Aufl. Hildesheim: Olms.
- WERLICH, Egon, 1975, Typologie der Texte. Entwurf eines textlinguistischen Modells zur Grundlegung einer Textgrammatik. Heidelberg.
- WOODS, William A; KAPLAN, Ronald M and NASH-WEBBER, Bonnie Lynn, 1972: The Lunar Science Natural Language Information System: Final report. Report 2378, Bolt Beranek and Newman Inc, Cambridge, Massachusetts,
- ZIFONUN, Gisela/STRECKER, Bruno/HOFFMANN, Ludger, 1997, Grammatik der deutschen Sprache. Im Original erschienen 1997. Berlin, New York: De Gruyter (Schriften des Instituts für Deutsche Sprache, 7).

2. ARTIKEL

- ARTEMIS ALEXIADOU, 2013, Adverbial and adjectival modification. In Marcel den Dikken (Hrsg.): The Cambridge Handbook of Generative Syntax. Cambridge University Press, Cambridge (UK), pp. 458–484.
- ARIEL, M., 1990, *Accessing noun-phrase antecedents*. (Croom Helm Linguistics Series.) London & New York : Routledge. [Siddharthan 2003] Siddharthan Advait, “Resolving Pronouns Robustly: Plumbing the Depths of Shallowness,” in: *Proceedings of the Workshop on Computational Treatments of Anaphora, 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2003)*, pp. 7-14, Budapest, Hungary, 2003. [Stand: 22.8.2005] http://www1.cs.columbia.edu/~as372/anaphora_EACL03.pdf
- ARIEL Mira, 1996, Referring expressions and the +/- coreference distinction. In: Thorstein Fretheim and Jeanette K. Gundel (eds.): *Reference and Referent Accessibility*, Benjamins, Amsterdam, Netherlands, pp. 13-35
- ARIEL Mira, 1998, Referring and accessibility. In: *Journal of Linguistics*, Vol. 24, No. 1, Cambridge University Press, pp. 65-87,
- ARIEL, M., 2001, Accessibility theory: An overview. Sanders T, Schilperoord J., Spooren W., (dir.), *Text Representation*, Amsterdam, Benjamins, 29-87.
- ASHER, N., 2000. Events, Facts, Propositions and Evolutive Anaphora. In: HIGGINBOTHAM, J./ PIANESI, F./ VARZI, A.C. (Hg.). *Speaking of Events*. Oxford: Oxford University Press, 123-150.
- AUFMKOIK, Johanna Andrea, 2014, Sententiale Pro-Formen als Satz determinierer. Elektronische Publikationen Universitätsbibliothek Wuppertal [Online verm gbar unter

[http://elpub.bib.uwuppertal.de/servlets/DocumentServlet?id=3771 & lang=en](http://elpub.bib.uwuppertal.de/servlets/DocumentServlet?id=3771&lang=en) , zuletzt aktualisiert am 12.08.2014 , zuletzt geprüft am 12.04.2017.

A. V. Abyzov, A. G. Rys, 2017, WAS IST COMPUTERLINGUISTIK? In UDK 81-1195 Bildungseinrichtung „Militärakademie der Republik Belarus“, Minsk, S. 148-149

BALDWIN, B., 1997, CogNIAC: high precision coreference with limited knowledge and linguistic resources. *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, Association for Computational Linguistics 38-45,

BALZERT, HELMUT 2011: Nichtfunktionale Anforderungen, in: Balzert, Helmut (Hrsg.): *Lehrbuch der Softwaretechnik: Entwurf, Implementierung, Installation und Betrieb*, Spektrum, Heidelberg [u.a.], S. 45-65

BEATA KOUCHNIR 2004 A Machine Learning Approach to German Pronoun Resolution in: *Proceedings of the ACL 2004 workshop on Student research* S. 31-37

BRAZIL, David, 1985, Phonology: Intonation in Discourse. In: *Handbook of Discourse Analysis* 1-4. DIJK, Teun A. (Hrsg.), S. 57-75.

BREINDL, Eva, 2004, Konnektoren in Übungsgrammatiken. In: Peter Kühn (Hrsg.): *Übungsgrammatiken Deutsch als Fremdsprache. Linguistische Analysen und didaktische Konzepte* 1. Aufl. Regensburg: Fachverband Deutsch als Fremdsprache (Materialien Deutsch als Fremdsprache, 66), S. 426-458.

BREINDL, Eva ,2012, Pronominaladverbien in Konnektorfunktion. Unter Mitarbeit von Elke Donalies. Institut für Deutsche Sprache. Mannheim [Online verfügbar unter [http://hypermedia.ids-mannheim.de/ca/ll/public/sysgram.ansicht?v_typ=d &v_id=2602](http://hypermedia.ids-mannheim.de/ca/ll/public/sysgram.ansicht?v_typ=d&v_id=2602)], zuletzt aktualisiert am 03.01.,2012 11:08., zuletzt geprüft am 22.02.2014.

BREINDL, Eva, 2013, Präpositionalobjektsätze. In: Jörg Meibauer/Markus Steinbach/Hans Altmann (Hg.): *Satztypen des Deutschen*, Berlin/Boston: De Gruyter (De Gruyter Lexikon), S.458-481

BREINDL, EVA/WASSNER, ULRICH H., 2006, Syndese vs. Asyndese: Konnektoren und andere Wegweiser zur Interpretation semantischer Relationen in Texten. In: Hardarik Blühndorn/Eva Breindl/Ulrich H. Waßner (Hg.): *Text Verstehen: Grammatik und darüber hinaus*. Berlin: W. De Gruyter, S. 46-70.

BRENNAN S.E. , FRIEDMAN M.W. , POLLARD C.J. 1987 , A centering approach to pronouns, in: *Proceedings of the 25th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, , pp. 155–162

BRINKER, Klaus, 1979, Zur Gegenstandsbestimmung und Aufgabenstellung der Textlinguistik. In *Text vs Sentence. Basic Questions of Text Linguistics*. Petöfi, János (Hrsg.), S. 3-12.

BRINKER, Klaus, 2010b, Textlinguistik. In: Klaus Weimar/Georg Braungart/Klaus Grubmüller et al. (Hrsg.): *Reallexikon der deutschen Literaturwissenschaft Neubearbeitung des Reallexikons der deutschen Literaturgeschichte*, Bd. 3. Unter Mitarbeit von Armin Schulz, [3rd ed.]. 3 Bände. Berlin, New York: De Gruyter (III), S. 607—611.

BRONCKART JEAN-PAUL, ECATERINA BULEA BRANCKART 2011, Clause, Phrase, Proposition ? des Enjeux d'un débat conceptuel in *Dans un système Linguistique aux actions langagieres* S. 209-219

CARBONELL J.G. ,BROWN R.D. ,1988: Anaphora resolution: a multi-strategy approach, in: *Proceedings of the 12th Conference on Computational Linguistics-Volume 1*, Association for Computational Linguistics, , pp. 96–101 .

CARSTENSEN, S.JEKAT UND R.KLABUNDE, 2010, Computerlinguistik – Was ist das? In: Kai-Uwe Carstensen, Christian Ebert, Cornelia Ebert, Susanne Jekat, Ralf Klabunde und Hagen Langer (Hrsg.), *Computerlinguistik und Sprachtechnologie: Eine Einführung*, 3. Auflage, Heidelberg, Spektrum Akademischer Verlag, S. 1-23

- CHOMSKY, NOAM. 1956. Three Models for the Description of Language. *Information Theory, IRE Transactions on Information Theory* IT, 2(3), S.113–124.
- CHOMSKY, NOAM. 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
- CONNOLLY, D., BURGER, J.-D./ DAY, D.-S., 1994, A machine learning approach to anaphoric reference. *Proceedings of International Conference on New Methods in Language Processing*, 255–261
- CONSTEN, M., 2003, Towards a unified model of domain-bound reference. In: F. Lenz (Hg.), *Deictic Conceptualisation of Space, Time and Person*. Amsterdam/ Philadelphia: John Benjamins, 223-248.
- DANEŠ, František, 1978, Zur linguistischen Analyse der Textstruktur. In: *Textlinguistik*, Wolfgang (Hrsg.), 1970, Dressler, Wege der Forschung 427 S. 185-192.
- DENBER Michel, 1998, Automatic Resolution of Anaphora. In: English, Internal Report, Eastman Kodak Co.
- DIJK, TEUN. A. Van, 1990, Introduction: Text in the next decade. In *TEXT an interdisciplinary journal for the study of discourse* 10 (1/2): Discourse analysis in the 1990s: Statements by Board Members of TEXT, S. 1-3.
- DIPPER, Stefanie, Heike ZINSMEISTER, 2011, Annotating abstract anaphora. In: *Language Resources and Evaluation*. 46(1), S. 37-52.
- DONČEVA Kosdatinka, 1980, Zu einigen funktionalen Wesenszügen der rück-weisenden d-Pronominaladverbien im System des heutigen Deutsch In: *Deutsch als Fremdsprache* 17 (4), S. 239
- DONČEVA Kosdatinka, 1982, Zum syntaktischen Status des d-Pronominaladverbs als Korrelat- In: *Deutsch als Fremdsprache* 19 (4), s. 221-224.
- DRESSLER, Wolfgang, 1985, Morphology. In: *Handbook of Discourse Analysis* 1-4. DIJK, Teun A. (Hrsg.), S. 76-86.
- DREYER KRUMM, HANS-JÜRGEN, 200L, Lehrwerkproduktion, Lehrwerkanalyse, Lehrwerkkritik. In: Gerhard Helbig/Lutz Götze/Gert Henrici et al- (Hg.): *Deutsch als Fremdsprache. Ein internationales Handbuch*. Berlin: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft/HSK, 19.2), 10 29-1041.
- EICHINGER, Ludwig M., 2006, Wortbildung – ein Haus mit drei Nachbarn. In: Proost, Kristel/Winkler, Edeltraud (Hrsg.): *Von Intentionalität zur Bedeutung konventionalisierter Zeichen. Festschrift für Gisela Harras zum 65. Geburtstag*. Tübingen, S. 179–196.
- EISENBERG, Peter, 2002a, Morphologie und Distribution – Zur Morphosyntax von Adjektiv und Adverb im Deutschen. In: Schmoe, Friederike (Hrsg.): *Das Adverb. Zentrum und Peripherie einer Wortklasse*. Wien, S. 61–76.
- FILLMORE, Charles J, 1968: The case for case. in: Emmon Werner Bach and Robert Thomas Harrns (editors). *Universals: In linguistic theory*. New York: Holt, Rinehart and Winston, 1968, S. 0-88
- FILLMORE, Charles J, 1977: The case for case reopened, in: Peter Cole and Jerrold M Saddock (editors). *Syntax and semantics 8: Grammatical relations*. New York: Academic Press, S. 59-81.
- FILLMORE C.J. 1986: Pragmatically controlled zero anaphora, in: *Annual Meeting of the Berkeley Linguistics Society*, 12, , pp. 95–107 .
- FINKBEINER, Rita, 2018. Deixis und Anapher. In *Handbuch Pragmatik*, Liedtke F., Tuchen A. (Hrsg.), Stuttgart, J.B. Metzler, S. 186-197
- FISSENI Bernhard 2010, Computer sind Dumm- Und daher äusserst Nützlich: Eine kurze Einführung in die komplexe Welt der Computerlinguistik. In: *Kritische Ausgabe Jg. 13 Nr. 18*, S. 85-89.
- FLEISCHER, Wolfgang, 1995, Konfixe. In: Pohl, Inge/Erhardt, Horst (Hg.): *Wort und Wortschatz*. Tübingen, S. 61–68.

- FLEISCHER, Wolfgang, 1997, Zum Status des Fremdelements-ier- in der Wortbildung der deutschen Gegenwartssprache. In: Kessler, Christine/Sommerfeldt, Karl-Ernst (Hg.): *Sprachsystem – Text – Stil*. Frankfurt/M., S. 75–87.
- FOTIS JANNIDIS 1999, Was ist Computerphilologie? In: Karl Eibl, Volker Deubel, Fotis Jannidis (Hg.): *Jahrbuch für Computerphilologie* 1. Paderborn: mentis, S. 39–60. <https://de.m.wikipedia.org/wiki/Computerphilologie> (21.02.2022 um 18h03)
- FRANCIS, G., 1994. Labelling discourse: An aspect of nominal-group lexical cohesion. In: Coulthardt, M. (ed.), *Advances in written text analysis*. London/ New York, 83-101
- GASPERIN Caroline, SALMON-ALT Susanne, VIEIRA Renata, 2004, “How useful are similarity word lists for indirect anaphora resolution?,” in: *Proceedings of DAARC*, Sao Miguel, Azores, <http://www.cl.cam.ac.uk/users/cvg20/daarc04.pdf> [Stand: 22.8.2005]
- GISA RAUH, 2015, Adverbs as a linguistic category (?). In: Karin Pittner u. a. (Hrsg.): *Adverbs. Functional and diachronic aspects*. John Benjamins, Amsterdam, S. 19–45.
- GIVÓN, T., 1992, The grammar of referential coherence as mental processing instructions. *Linguistics*, 30: 5-55.
- GÖTZE, Lutz, 200la, Grammatiken. In: Gerhard Helbig/Lutz Götze/Gert Henrici et al. (Hg.): *Deutsch als Fremdsprache. Ein internationales Handbuch*, Bd. 19.2. Berlin: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft/HSK, 19.2), S 1070-1078.
- GÖTZE, Lutz, 200lb, Linguistische und didaktische Grammatik. In: Gerhard Helbig/Lutz Götze/Gert Henrici et al. (Hg.): *Deutsch als Fremdsprache. Ein internationales Handbuch* Bd. 19.1. Berlin: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft/HSK, 19.1), s. 187-194.
- GROSZ, BARBARA J. 1977. The representation and use of focus in a system for understanding dialogs. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, volume 1 Cambridge, MA. pages 67—76,
- GROSZ, BARBARA J., ARAVIND K. JOSHI, AND SCOTT WEINSTEIN. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting*, Cambridge, MA, June. Association for Computational Linguistics. pages 44—50,
- GROSZ B.J. , WEINSTEIN S. ,JOSHI A.K. 1995, Centering: a framework for modeling the local coherence of discourse, *Comput. Linguist.* 21 (2) S. 203–225 .
- GUERIN, Olivia, 2014, Construction du référent, textualité et genre discursif : Les anaphores génériques dans les séquences encyclopédiques des récits de voyages. In : *Congrès mondial de linguistique française-CMLF 2014*, S. 3091-3110. Erhältlich Unter <http://www.shs-conferences.org> [10.02.2020].
- GUNDEL, Jeanette K., Nancy A. HEDBERG and Ron ZACHARSKI, 2005, “Pronouns Without NP Antecedents: How do we Know when a Pronoun is Referential?,” in: *Branco, António, Tony McEnery and Ruslan Mitkov (eds.): Anaphora Processing*, pp. 351-364, Amsterdam, Netherlands.
- GÜNTHER, Susanne, 2001, "wobei (.) es hat alles immer zwei Seiten. ". Zur Verwendung von wobei im gesprochenen Deutsch. In: *Deutsche Sprache* 4, S. 313—341
- HAHN Udo, Matthias HEYN, Ludwig HITZENBERGER, Bernhard KELLE, Hans-DIETER LUTZ, KLAUS WOTHKE, 1988, *Zur Konturierung des Faches Computerlinguistik*. In *LDV-Forum* Bd. 5, Jg. 1987/88. Nr. 4. September, S. 86-96
- HALLIDAY, Michael Alexander Kirkwood, 1985, *Dimensions of Discourse Analysis: Grammar*. In: *Handbook of Discourse Analysis* 2. van Dijk, Teun A. (Hrsg.), London/New York, S. 29-56.
- HARWEG, Roland, 1974, *Textlinguistik*. In: *Perspektiven der Linguistik Band 2*, KOCH, WALTER A. (Hrsg.), Stuttgart, S. 88-116.

- HELBIG, Gerhard, 2004a, Bemerkungen zu den Pronominaladverbien und zur Pronominalität. In: Gerhard Helbig/Horst Sitta (Hrsg.): *Kleinere Schriften zur Grammatik*. Unter Mitarbeit von Bernd Skibitzki/Johanne s Wenzel/Barbara Wotjak. München: ludicium-Verlag, S. 286—300.
- HELBIG, Gerhard, 2004b, Bemerkungen zu den Pronominaladverbien und zur Pronominalität. In: Gerhard Helbig/Horst Sitta (Hrsg.): *Kleinere Schriften zur Grammatik*, München: ludicium-Verlag, S. 286-300.
- HENTSCHEL/WEYDT 2003, Wenn Partikeln frech werden". In: Held, G. (ed.) (2003): *Partikeln und Höflichkeit*. Bern etc.: 55-72.
- HINRICHS, E., FILIPPOVA, K., AND WUNSCH, H. 2005: What Treebanks Can Do For You: Rule-based and Machine-learning Approaches to Anaphora Resolution in German. In *Proceedings of the Fourth Workshop on Treebanks and Linguistic Theories*, Barcelona, pages 77–88,
- HINRICHS EW, FILIPPOVA K. UND WUNSCH H. 2005: Ein datengetriebener Ansatz zur pronominalen Anapherauflösung im Deutschen. In *Proceedings of the 5th International Conference on Recent Fortschritte in der Verarbeitung natürlicher Sprache (RANLP 2005)*, Borovets, Bulgarien S. 112-125
- HINRICHS, E. W., FILIPPOVA, K., AND WUNSCH, H. 2007: A Data-driven Approach to Pronominal Anaphora Resolution in German. In *Proceedings of the 5th International Conference on Recent Advances in Natural Language Processing*, Borovets. 29-54
- HINRICHS E. W. , HENRICH, V. AND, ,2011: Determining Immediate Constituents of Compounds in GermaNet. In *Proceedings of Recent Advances in Natural Language Processing*, pages 420–426,
- HIRST, GRAEME JOHN 1978A: Report on the Conference on Theoretical Issues in Natural Language Processing-2. A [SJ~ quarterly, number 31, S. 9-11.
- HIRST, GRAEME 1981 Anaphora: In *Natural Language Understanding: A Survey*. Lecture Notes in Computer Science 119.S-30-49
- HIRST, GRAEME, & ST-ONGE, DAVID. 1997. WordNet: An Electronic Lexical Database. The MIT Press. Kap. Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms, s. 305–332.
- Hobbs, Jerry R. 1977a. ``From English Descriptions of Algorithms into Programs'', Proceedings, Annual Conference, Association for Computing Machinery, 323-329,
- Hobbs, Jerry R. 1977b ``What the Nature of Natural Language Tells us about How to Make Natural-language-like Programming Languages More Natural'', Proceedings, Symposium on Artificial Intelligence and Programming Languages, Rochester, New York. pp. 85-93.
- HOBBES, J. 1978 Resolving pronoun references, *Lingua* 44 (4) S. 311–338.
- HOFMANN, Ludger, 2000, Anapher im Text. In: *Text und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. Klaus Brinker/Gerd Antos/Wolfgang Heinemann et al. (Hrsg.), Berlin/New-York, S. 295–304.
- HOFFMANN, Ludger, 2009, Adverb. In: Ludger Hoffmann (Hrsg.): *Handbuch der deutschen Wortarten*. Berlin: De Gruyter; Walter de Gruyter (De Gruyter Lexikon), S. 223-264.
- ISENBERG H., 1968, Überlegungen zur Texttheorie. In ASG-BerichtNr.2, Berlin, Akademie der Wissenschaften, Arbeitsstelle Strukturelle Grammatik. S. 1–18.
- ISENBERG, H. 1970, Der Begriff „Text“ in der Sprachtheorie. In: Bericht 8 der Arbeitsgruppe Strukturelle Grammatik Berlin. 20-36
- ISENBERG, Horst, 1971, Überlegungen zur Texttheorie. In: *Literaturwissenschaft und Linguistik I* IHWE, J. (Hrsg.), Frankfurt, Athenäum S. 155-172.
- ISENBERG, Horst, 1977, 'Text' versus 'Satz'. In: *Probleme der Textgrammatik II* DANEŠ, František, DIETER Viehweger (Hrsg.), *studia grammatica* XVIII, S. 119-145.

- KALLMEYER, Werner & MEYER-Hermann, REINHARD, 1980, Textlinguistik. In: Lexikon der Germanistischen Linguistik. 2. Auflage. ALTHAUS, HANS Peter, HELMUT Henne, HERBERT Ernst Wiegand (Hrsg.). Tübingen, Niemeyer, S. 242-258
- KARIN PITTNER, JUDITH BERMAN, 2012, Eine Art Wortart: Das Adverb im Deutschen. In: Zeitschrift für Sprachwissenschaft 38-2 (2019), 191–241. DOI: <https://doi.org/10.1515/zfs-2019-2004>.
- KENNEDY, C./BOGURAEV, B., 1996, Anaphora for everyone : Pronominal anaphora resolution without a parser. In *Proceedings of the 16th International Conference on Computational Linguistics-Volume (1)*, Association for Computational Linguistics, 113-118.
- KNEES, Mariele, 2006, The German Temporal Anaphor danach — Ambiguity in Interpretation and Annotation. In: Ron Artstein/Massimo Poesio (Hrsg.): Ambiguity in Anaphora Workshop Proceedings. ESSLI 2006. Málaga, Spain, 7—11 August, S. 57—64.
- KÖNIG, Katherina, 2012, Formen und Funktionen von syntaktisch desintegriertem deswegen im gesprochenen Deutsch. In: Gesprächsforschung 13, S. 45-71.
- KUBCZAK, H. 2016, Anaphorische Konstruktion und antezedensloser Rückverweis. In: Jahrbuch der ungarischen Germanistik 2011, ILSE, V. & Szendi, Z. (Hrsg.), Budapest/Bonn, Gondolat Kiadói Kör S. 221-247.
- LAPPIN Shalom and MCCORD Michael, 1990 “Anaphora Resolution in Slot Grammar,” in: *Computational Linguistics*, Vol. 16, No. 4, MIT Press, pp. 197-212.
- LAPPIN Shalom and J. LEASS, Herbert, 1994 “An Algorithm for Pronominal Anaphora Resolution,” *Computational Linguistics*, Vol. 20 No. 4, pp. 535-561,
- LAPPIN Shalom and SHIH, Hsue-Hu, 1996 “A generalized reconstruction algorithm for ellipsis resolution,” in: *Proceedings of COLING-96*, pp. 351-364, University of Copenhagen, Copenhagen, Denmark.
- LUDGER PASCHEN (RUHR-UNIVERSITÄT BOCHUM) 2013 Implementation eines regelbasierten Pronomen-resolutionsalgorithmus für das Deutsche In: Tagungsband der 53. Studentischen Tagung Sprachwissenschaft (StuTS), Cornelia van Scherpenberg, Fleur Szabó, Stefanie Eckmann (Hgg.) S 60-100
- MCCARTHY AND WENDY G. LEHNERT. 1995. Using decision trees for coreference resolution. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pages 1050–1055.
- MCCORD, M. C. 1989 Design of LMT: a Prolog-based machine translation system. *Computational Linguistics* 15, 33-52.
- MCCULLOCH WS, PITTS W., 1943: A logical calculus of the ideas immanent in nervous activity. *Bull Math. Biophys.*, 5, 1943, S. 115-133.
- METZING, Dieter, 2022 Computerlinguistische Modellierung. Sprachspezifische Eigenschaften und sprachtypologische Generalisierung Aus: Willée, Gerd; Schröder, Bernhard; Schmitz, Hans-Christian (Hrsg.): Computerlinguistik. Was geht, was kommt? Sankt Augustin S. 202-207
- MITKOV Ruslan, 1997, "Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches," in: *Proceedings of the ACL'97/EACL'97 Workshop on Operational Factors in Practical, Robust Anaphora Resolution*, pp. 14-21, Madrid, Spain.
- MITKOV, R., 1998, Robust pronoun resolution with limited knowledge. *Proceedings of the 17th international conference on Computational linguistics-Volume (2)*, Association for Computational Linguistics, S. 869-875.
- MITKOV, R., 1999, Multilingual anaphora resolution. *Machine Translation*, 14(3-4), 281-299.

- MITKOV Ruslan, 2001a, "Towards a more consistent and comprehensive evaluation of anaphora resolution algorithms and systems," *Applied Artificial Intelligence: An International Journal*, Vol. 15, No. 3, pp. 253-276.
- MITKOV, R., BOGURAEV, B./ LAPPIN, 2001b, Introduction to the special issue on computational anaphora resolution. *Computational Linguistics*, 27(4), 473– 477.
- MITKOV Ruslan, EVANS Richard and ORASAN Constantin, 2002b "A New, Fully Automatic Version of Mitkov's Knowledge-Poor Pronoun Resolution Method," in: *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)*, Mexico City, Mexico.
- MITKOV Ruslan and BARBU Catalina, 2002c, "Using bilingual corpora to improve pronoun resolution," *Languages. In Contrast*, Vol 4, No. 2, Benjamins, pp. 201-211.
- MITKOV, R., EVANS, R., ORASAN, C./ PEKAR, V., 2007, Anaphora resolution : To what extent does it help NLP applications?. *Discourse Anaphora and Anaphor Resolution Colloquium*, Springer, Berlin, Heidelberg, 179-190
- NEGELE Michaela, 2012a, Regionale Standardvariation im DaF-Unterricht. Pronominaladverbien in Lehrwerken und Standardgrammatiken. In: *Lebende Sprachen, Zeitschrift für interlinguale und interkulturelle Kommunikation* 57 (1), S. 94-121
- OPPENRIEDER Wilhelm, 1991, Preposition Stranding im Deutschen: Da will ich nichts von hören! In: Gisbert Fanselow/W.Sascha Felix (Hrsg.): *Strukturen und Merkmale syntaktischer Kategorien*. Tübingen: Narr (Studien zur deutschen Grammatik, 39), S. 159-173.
- PAUSE, P.F, 1991, Anaphern im Text. In: *Semantik. Ein internationales Handbuch der zeitgenössischen Forschung*, STECHOW, A. von /D.Wunderlich (Hrsg.), Berlin/Newyork, de Gruyter, S. 548-559
- POESIO Massimo, ISHIKAWA Tomonori, SCHULTE Sabine im Walde, and VIEIRA Renata, 2002, "Acquiring lexical knowledge for anaphora resolution," in: *Proceedings of LREC 2002, – Third International Conference on Language Resources and Evaluation*, pp. 1220-1224, Las Palmas de Gran Canaria, Spain. <http://cswww.essex.ac.uk/poesio/publications/LREC02.pdf> [Stand: 22.8.2005]
- POESIO M., STEVENSON R., DI EUGENIO B., HITZEMAN J. M., 2004, Centering: A Parametric Theory and its Instantiations. *Computational Linguistics* 30 3, S. 309–363.
- REHBEIN, Jochen, 1995, Über zusammengesetzte Verweiswörter und ihre Rolle in argumentierender Rede. In: Harald Wohlrapp (Hrsg.): *Wege der Argumentationsforschung. Tagung über Argumentationstheorie*. Stuttgart- Bad Cannstatt: Frommann-Holzbog (Problemata, 135), S. 166-197.
- REHBEIN, Jochen, 1999, Konnektivität im Kontrast: Zuz Struktur und Funktion türkischer Konverbien und deutscher Konjunktionen, mit Blick auf ihre Verwendung durch monolinguale und bilinguale Kinder. In: Lars Johanson/Jochen Rehbein (Hrsg.): *Türkisch und Deutsch im Vergleich*. Wiesbaden: Harrassowitz (Turcologica, 39), S. 189-243.
- RICH, ELAINE AND SUSANN LUPERFOY. 1988. An architecture for anaphora resolution. In *Proceedings of the 2nd Conference on Applied Natural Language Processing*, Austin, TX, February. S. 18-24,
- SALMON-ALT, S. ,2002, Le projet ANANAS : Annotation Anaphorique pour l'Analyse Sémantique de Corpus. *Actes de TALN*, Nancy, 24-27 juin, 163-172.
- SALMON-ALT, S., 2004a, Résolution automatique d'anaphores infidèles en français : Quelles ressources pour quels apports ?. *Actes de TALN*, session poster, Fès, 19-21 avril.
- SALMON-ALT, 2004B, particulière de la relation Assertion-Indice. *Discours*. In : *Revue de linguistique, psycholinguistique et informatique*. URL : <http://discours.revues.org/7739>; DOI : 10.4000/discours.7739.

- SALMON-ALT, S., Bick, E., Romary, L./Pierrel, J.-M. ,2004,. La FREEBANK : vers une base libre de corpus annotés. *Actes de TALN*, Fès, 19-21 avril.
- SCHIEHLEN MICHAEL 2004: Optimizing algorithms for pronoun resolution. In *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva. S. 1134-1164
- SCHRODT, Richard, 1988, Wortstellung und Nebensatz. Die Entstehung von Subjunktoren aus Pronominaladverbien und das Problem der Verbstellung im deutschen Nebensatz. Universität Duisburg-Essen. Essen (LAU D-Linguistic Agency) [Onlineverfügbar unter <http://www.linse.uni-due.de/laud-downloadliste/articles/wortstellung-undnebensatz-die-Entstehung-von-Subjunktoren-aus-Pronominaladverbien-und-das-problem-der-verbstellung-im-deutschen-html>], zuletzt geprüft am 05.04.2016
- SCHWARZ, Monika, 1994 Anaphors without explicit antecedents. Vortrag auf der „International conference on anaphoric relations and (in)coherence“ in Antwerpen, S. 81-102.
- SCHWARZ, Monika, 1995 Referenz as Prozess. In: *Reference in a Multidisciplinary Perspective*, Geiger, R. (Hrsg.), Hildesheim, Olms S. 461-470.
- SCHWARZ, Monika, 1996, Lexikalische und konzeptuelle Restriktionen beim Verstehen direkter und indirekter Anaphern. In: *Lexical Structures and Language Use*, Hundsnurscher, F./E. Weigand (Hrsg.), Tübingen, Niemeyer, S. 399-407.
- SCHWARZ, Monika, 1997, Anaphern und ihre diversen Antezedenten: Koreferenz und Konsorten In: *Linguistik im Fokus. Festschrift für H. Vater, Dürscheid, Ch./K.H. Ramers/M. Schwarz* (Hrsg.), Tübingen, Niemeyer S. 445-456.
- SCHWARZ, Monika, 2000b, Textuelle Progression durch Anaphern- Aspekte einer prozeduralen Thema-Rhema-Analyse. In: *Linguistische Arbeitsberichte*, 74, S. 111-126.
- SCHWARZ, M., 2001, Establishing Coherence in Text. *Conceptual Continuity and Text-world Models. Logos and Language. Vol. II, No. 1* (2001). 15-24.
- SCHWARZ-FRIESEL, Monika/CONSTEN, Manfred/MARX, Konstanze, 2004, Semantische und konzeptuelle Prozesse bei der Verarbeitung von Komplex-Anaphern. In: *Stabilität und Flexibilität in der Semantik*, POHL Inge / KONERDING Klaus-Peter (Hrsg.), Frankfurt a. M., S. 67-88.
- SCHWARZ-FRIESEL, Monika/CONSTEN, MANFRED, 2007, Anapher. In: *Handbuch der deutschen Wortarten*, HOFMAN Ludger (Hrsg.), Berlin, S. 265-292.
- SOON WM, HT NG UND DCY LIM 2001: Ein maschineller Lernansatz zur Koreferenzauflösung von Nominalphrasen. *Computerlinguistik*, 27 (4): 521-544,
- SPOOREN, Wilbert, 1999, Texte strukturieren: Textlinguistik. In: *Sprache und Sprachwissenschaft*, R. Pörings/U. Schmitz (Hrsg.), Tübingen, 187-210.
- STARKE, Günter, 1982, Weiterführende Nebensätze, eingeleitet mit Pronominaladverbien. In: *Deutsch als Fremdsprache* 19 (4), S. 215-220.
- STEINBERGER JOSEF, POESIO MASSIMO, MIJAIL A. KABADJOV, AND KAREL JEZEK. 2007. Two uses of anaphora~ resolution in summarization. *Information Processing & Management*, 43(6) S:1663-1680.
- STEINITZ, Renate, 1968/1974, Nominale Proformen. In: *Lektürekolleg zur Textlinguistik. Band 2*, KALLMEYER, Werner et al. (Hrsg.), Frankfurt, Reader, S. 246-266.
- STRUBE MICHAEL. 1998. Never look back: An alternative to Centering. In *COLING '98 Montreal, Canada*. pages 1251-1257
- STRUBE MICHAEL AND UDO HAHN. 1999. Functional Centering – Grounding Referential Coherence in Information Structure. *Computational Linguistics*, 25(3): S-309-344.
- STRUBE MICHAEL AND MÜLLER CHRISTOPH. 2003. A Machine Learning Approach to Pronoun Resolution: In Spoken Dialogue. In *ACL'03, Sapporo, Japan*. pages 168-175
- STRUBE M., C. MÜLLER, S. RAPP UND, 2002. Anwenden von Co-Training auf Referenzauflösung. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, Philadelphia, PA, Seiten 352-359,

- TELLJOHANN H. , E. HINRICHS UND S. KÜBLER, 2004. DIE Tu-Ba-D/Z-Baumbank: Annotating German with a Context-free backbone. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Seiten 2229–2235, Lissabon, Portugal.
- THOMAS S. MORTON, 1999, Using coreference for question answering. In *Coreference and Its Applications*, pages 85–89, College Park, Maryland. Association for Computational Linguistics
- THOMAS S. MORTON. 2000. Coreference for nlp applications. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL'00)*, Hong Kong. 100-115
- TIETZ, Heike, 1997, Die Zukunft der Textlinguistik. Zusammenfassende Bemerkungen zu einer Diskussion. In: *Die Zukunft der Textlinguistik. Traditionen, Transformationen*, ANTON, GERD, TIETZ Heike (Hrsg.), Reihe Germanistische Linguistik 188, S. 223-230.
- ULRICH EBERL, 2018 was ist Künstliche Intelligenz – was kann sie leisten? In *aus Politik und Zeitgeschichte, Künstliche Intelligenz, Zeitschrift der Bundeszentrale für politische Bildung* 68. Jahrgang, 6–8/2018, 5. Februar S. 8-14
- VATER HEINZ, 1975, Pro-Formen des Deutschen. In: Michael Schecker/Peter Wunderlich (Hrsg.): *Textgrammatik. Beiträge zum Problem der Textualität*. Tübingen: Niemeyer (Konzepte der Sprach- und Literaturwissenschaft, I 7), s. 20-42.
- VATER, H, 1984, Referenz und Determination im Text. In: *Sprache und Pragmatik*, Rosengren, I. (Hrsg.), Lunder Symposium, S. 323-344.
- VILAIN M. , J. BURGER, J. ABERDEEN, D. CONNOLLY UND L. HIRSCHMAN, 1995. Ein modelltheoretisches Koreferenz-Scoring-Schema. In *MUC6 '95: Proceedings of the 6th Conference on Message Understanding*, Columbia, MD, S. 45–52
- V. NG UND C. CARDIE, 2002. Kombination von Stichprobenauswahl und fehlergesteuertem Pruning für maschinelles Lernen von Koreferenzregeln. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-02)*, Philadelphia, PA, 89-107
- WALKER, M.A. 1989 , Evaluating discourse processing algorithms, in: *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, , pp. 251–261 .
- WALKER, MARILYN, 1998, Centering, Anaphora Resolution, and Discourse Structure. In: *Centering Theory in Discourse*, M. WALKER & A. Joshi & E. Prince (Hrsg.), Oxford Clarendon Press. S. 401-435.
- WELLMANN, Hans, 1998, Die Wortbildung. In: *Duden. Grammatik der deutschen Gegenwartssprache*. Hrsg. v. d. Dudenredaktion. 6., neu bearb. Aufl. Mannheim, Leipzig, S. 408–557.
- WINOGRAD, TERRY. 1972. „Understanding Natural Language”. *Cognitive Psychology* 3 (1). S. 1–191
- WOODS, William A, 1968. Procedural semantics for a question-answering machine. *AFIPS conference proceedings*, 33, FJCC, S. 457-471.
- WOODS, William A, 1970, : Transition network grammars for natural language analysis. *Communications of the ACM*, 13(10), S. 591-606.
- WOODS, William A, 1977. Lunar rocks in natural English: Explorations in natural language question answering, in: *Zampolli*, S. 521-569.
- WUNSCH, H. 2006: Anaphora-Auflösung – Was auf Deutsch hilft. In *Proceedings of the International Conference on Linguistic Evidence*, Tübingen, Deutschland, Seiten 101–105,
- ZIMMERMANN, Ralf, 2012, Darum soll es gehen: Der Status von Konnektoren im Deutschen. In: Veronika Kotáľková (Hrsg.): *Perspektiven der Textanalyse*. [interdisziplinäre Konferenz, die im Februar 2011 an der Schlesischen Universität Opava (Tschechische Republik) stattgefunden hat]. Unter Mitarbeit von Gabriela Rykalová. Internationales Treffen

"Perspektiven der Textanalyse". Tübingen: Stauffenburg (Stauffenburg Linguistik, 62), S. 53—59.

3. VORLESUNGEN UND SKRIPTEN

ALEXANDRIS Christina, 2018, Computerlinguistik Lehrinheit 5: Multilinguale Mensch-Maschine-Kommunikation: Linguistische Aspekte und Anwendungen, Nationale Universität Athen, Abteilung der deutschen Sprache und Literatur.

<https://docplayer.org/58621093-Computerlinguistik-lehreinhalt-6-7-ressourcen-dr-christina-alexandris-nationale-universitaet-athen-deutsche-sprache-und-literatur.html> (09/02/2022)

DEMBERG Vera 2014 Einführung in die Computerlinguistik, Universität des Saarlandes Wintersemester. <https://www.google.com/url?sa=t&source=web&rct=j&url=http://www.coli.uni-saarland.de/courses/I2CL->

[14/material/fohlen/CL14fohlen1_Intro.pdf&ved=2ahUKEwj3guy31If3AhWDr6QKHfgFChkQFnoECAYQAAQ&usg=AOvVaw1FBrrwJh2q2opLXXL0to4A](https://www.google.com/url?sa=t&source=web&rct=j&url=http://www.coli.uni-saarland.de/courses/I2CL-14/material/fohlen/CL14fohlen1_Intro.pdf&ved=2ahUKEwj3guy31If3AhWDr6QKHfgFChkQFnoECAYQAAQ&usg=AOvVaw1FBrrwJh2q2opLXXL0to4A) (09/02/2022)

DETMAR MEURERS, 2009, Computerlinguistik –Undevenis? Quovadis? Perspektiven auf die Entwicklung der Computerlinguistik bis ins Jahr 2019, Universität Tübingen, Festakt zum 60. Geburtstag von Prof.Dr. Michael Hess Zürich, 2.

<https://www.google.com/url?sa=t&source=web&rct=j&url=http://www.sfs.uni-tuebingen.de/~dm/handouts/zuerich-09-11-02.pdf&ved=2ahUKEwj3guy31If3AhWDr6QKHfgFChkQFnoECAYQAAQ&usg=AOvVaw3kJaQWqflqVAoIDcMHIP9p> (09/02/2022)

HAHN Udo, 2019, Computerlinguistik I, Lehrstuhl für Computerlinguistik, Friedrich-Schiller-Universität Jena, Institut für germanistische Sprachwissenschaft. <http://www.julielab.de>

https://www.google.com/url?sa=t&source=web&rct=j&url=https://docplayer.org/130939778-Prof-dr-udo-hahn-seminar-im-modul-m-gsw-09-wise-2018-19.html&ved=2ahUKEwiM9c-v0If3AhWtQkEAHX8tCJCQFnoECCEQAAQ&usg=AOvVaw0MPh_nHmpOS34GiBWRHeQR (09/02/2022)

HESS Michael, 2005, Einführung in die Computerlinguistik (I), Universität Zürich, Institut für Computerlinguistik, Wintersemester 2004/2005.

https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.cl.uzh.ch/dam/jcr:ffff-ffff-b2e3-81a5-ffff-ffff7c4f4a7e/ecl1.0.1.pdf&ved=2ahUKEwiL8be42of3AhUKTcAKHZnVD5AQFnoECAUQAQ&usg=AOvVaw1wzCM6zhTeaLgb_jfy0sLY (09/02/2022)

KALLMEYER Laura 2013 Einführung in die Computerlinguistik, Heinrich-Heine-Universität Düsseldorf Sommersemester

https://www.google.com/url?sa=t&source=web&rct=j&url=https://docplayer.org/115408749-Einfuehrung-in-die-computerlinguistik.html&ved=2ahUKEwjzmMlxof3AhUM08AKHSMrA34QFnoECB0QAAQ&usg=AOvVaw1G3FU0omwJ_A05VdWTq2kj (09/02/2022)

MÜLLER Stefan, 2013, Einführung in die Computerlinguistik; Deutsche Grammatik, Freie Universität Berlin, Fachbereich Philosophie und Geisteswissenschaften, Institut für Deutsche und Niederländische Philologie.

https://www.google.com/url?sa=t&source=web&rct=j&url=https://hpsg.huberlin.de/~stefan/P/Scslides.pdf&ved=2ahUKEwjV5vXY5of3AhXNQEEAHYQWCBAQFnoECBgQAAQ&usg=AOvVaw2k3e_fY3dBx1Cnffboqfa5 (09/02/2022)

SCHÜTZE Hinrich, 2018 Einführung in die Computerlinguistik, Ludwig Maximilians Universität München, Centrum für Informations und Sprachverarbeitung, <http://www.cis.unimuenchen.de>, <http://www.cis.unimuenchen.de/fachschafft/>

<https://www.google.com/url?sa=t&source=web&rct=j&url=https://www.cis.uni-muenchen.de/~hs/teach/18w/pdf/01intro.pdf&ved=2ahUKewjkncDl2of3AhXDoFwKHUT5DX8QFnoECAQQAQ&usg=AOvVaw3AmD13C3cDzLrCBbTf51N>. (09/02/2022)

SIMON Clematide 2007, Einführung in die Computerlinguistik I, Universität Zürich, Institut für Computerlinguistik Wintersemester, PDF-Version: <http://www.cl.unizh.ch/siclemat/lehre/ws0607/ec11/script/script.pdf>

TELLJOHANN H. , EW HINRICHS, S. KÜBLER UND H. ZINSMEISTER, 2006. Stylebook für die Tübinger Baubank des Schriftdeutschen (Tu-Ba-D/Z). Seminar für Sprachwissenschaft, Universität Tübingen, Deutschland,

4. LEXIKA UND WÖRTERBÜCHER

BUSSMANN, Hadumod, 1990, Lexikon der Sprachwissenschaft. 2. Auflage. Stuttgart, Kröner.
DUDEN, 1998, Die Grammatik, Das unentbehrliche Standardwerk für richtiges Deutsch. 6., neu bearb. Aufl. Mannheim: Dudenverlag. (Der Duden, 4).

DUDEN, 2003, Deutsches Universalwörterbuch, 5.Auflage. Mannheim, Dudenverlag.

DUDEN, 2005, Die Grammatik, Neue bearbeitete und erweiterte Auflage, Mannheim, Wien, Zürich, Dudenverlag

DUDEN, 2007, Richtiges und gutes Deutsch: Wörterbuch der sprachlichen Zweifelsfälle auf der Grundlage der neuen amtlichen Rechtschreibregeln; [Antwort auf grammatische und stilistische Fragen, Formulierungshilfen und Erläuterungen zum Sprachgebrauch] 6., völlig überarbeitete. Aufl., [Nachd.] Mannheim: Dudenverlag. (Der Duden, 9).

DUDEN, 2009, Die Grammatik, Unentbehrlich für richtiges Deutsch. 8., überarb. Aufl. Mannheim: Dudenverl. (Der Duden, 4).

METZLER LEXIKON SPRACHE, 2005, Stuttgart, J.B. METZLER Verlag.

WIKTIONARY, 2022, Die freie Enzyklopädie. <http://de.wikipedia.org/wiki/> unter der Lizenz CC BY-SA 3.0 License). [02.03.2022]

5. BACHELOR-UND MASTERARBEITEN

BEZNOSKOV Thomas 2019, Klassifikation von Produkttypen mit Hilfe von Machine Learning Methoden zur Automatisierung der Navigationspfade in einem Online-Shop Bachelorarbeit eingereicht im Rahmen der Bachelorprüfung im Studiengang Bachelor of Science Angewandte Informatik am Department Informatik der Fakultät Technik und Informatik der Hochschule für Angewandte Wissenschaften Hamburg Betreuender Prüfer: Prof. Dr. Ulrike Ste!ens Zweitgutachter: Prof. Dr. Chris Biemann

BITTAR, A., 2006. *Un algorithme pour la résolution d'anaphores événementielles*. Mémoire de master, Université Paris, 7, Denis Diderot UFR de Linguistique.

SELMA AKOL GÖTKAS, 2012, Textlinguistische Analyse von Wiederaufnahmestrukturen in deutschen Nachrichtentexten. Übertragung der Ergebnisse auf den DaF-Unterricht. Magisterarbeit, 2012, München, GRIN Verlag, <https://www.grin.com/document/419749>

MAG. DUNJA DARER, 2020, Einsatz Künstlicher Intelligenz zur Entscheidungsfindung in der Personalbeschaffung – eine rechtliche Analyse Masterarbeit zur Erlangung des akademischen Grades Master of Arts in Business an der Fachhochschule Burgenland

SEVCIKOVA Renata, 2005, Schwierigkeiten in der automatischen Verarbeitung von Anaphora, Lizenziatsarbeit des Instituts für Computerlinguistik der Universität Zürich

SOFAC FOGO Nelson 2021, Zum Aufbau von anaphorischen Konstruktionen in der Textsorte Prüfungsarbeiten: Ein Beitrag zur Anaphernresolution, Abschlussarbeit, Universität von Jaunde 1, Deutsche Abteilung,

6. DISSERTATIONEN

DON TUGGENER 2016: Incremental Coreference Resolution for German Thesis presented to the Faculty of Arts and Social Sciences of the University of Zurich for the degree of Doctor of

Philosophy by Accepted in the spring semester 2016 on the recommendation of the doctoral committee: Prof. Dr. Martin Volk (main advisor) PD Dr. Gerold Schneider

HIRST, GRAEME JOHN 1976b: Anaphora and reference in natural language understanding. Unpublished thesis draft, Department of Computer Science, University of British Columbia.

KNEES, Mariele, 2009, Zur Semantik und Referenz des temporalanaphorischen Pronominaladverbs danach. Dissertation. Friedrich-Schiller-Universität, Jena.

LIM, Seong, Woo, 2004, Kohäsion und Kohärenz: Eine Untersuchung zur Textsyntax am Beispiel schriftlicher und mündlicher Texte, Ph.D. Dissertation, der Julius-Maximilians-Universität Würzburg, philosophischen Fakultät II

NAGEL EILERS Kathrin, 2011, Aspekte vager Referenz beim Pronomen it, Ph.D. Dissertation Von der Carl von Ossietzky-Universität Oldenburg, Fakultät III Sprach- und Kulturwissenschaften

SALMON-ALT, S. ,2001, *Référence et Dialogue finalisé : de la linguistique à un modèle opérationnel*. Thèse de Doctorat, Université H. Poincaré, Nancy.

SELMA AKOL GÖTKAS, 2020, Textkohärenz durch Pronominaladverbien: Eine textlinguistische Untersuchung zur rezeptiven Textkompetenz türkischer Daf-Studierender am Beispiel von Pronominaladverbien. Dissertation. Friedrich-Schiller-Universität, Jena.

SELMİ, Afef, 2019, L'ambiguïté anaphorique et la résolution automatique de l'anaphore pronominale Propositions méthodologiques à partir d'un corpus fermé de résumés d'œuvres littéraires en français(RESUMAN) ; Thèse de doctorat de l'université bourgogne Franche-comte de Dijon.

WUNSCH, H. 2009: Regelbasierte und gedächtnisbasierte Pronomenauflösung für das Deutsche: Vergleich und Bewertung von Datenquellen. Dissertation, Universität Tübingen,

WUNSCH, H. 2010. Rule-based and Memory-based Pronoun Resolution for German: A Comparison and Assessment of Data Sources. PhD thesis, University of Tübingen.

ZSÓFIA Haase, 2015, Indirekte Anaphern als Kohärenzerscheinungen mit besonderer Rücksicht auf die indirekten pronominalen Anaphern, Ph.D. Dissertation der Universität Debrecen, Graduiertenkolleg Sprachwissenschaft

7. ERZÄHLUNGEN, ROMANE UND ZEITUNGEN

BAMBECK Joern J, 1979, die Anderen. In Die Anderen, LE BLANC Thomas (Hrsg.), München, Wilhelm Heyne Verlag S. 9-22

HESSE Herman, 1977, Kurgast, Frankfurt am Main, Scuhrkamp, Taschenbuch Verlag

MANN, Thomas, 2007, das Wunderkind und andere Erzählungen, Frankfurt am Main, Fischer Taschenbuch Verlag

SAZ11/JAN.00262 SAALE ZEITUNG, 04.01.2011, S. 30, Originalressort: Kinderseite; ohne Titel)

8. INTERNETQUELLEN

<https://www.deepl.com/translator> & <https://translate.google.com/> (12.03.2022) für – Übersetzungssysteme <http://start.csail.mit.edu/> (12.03.2022) Für Question AnsweringStart-System

<https://de.m.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03)

https://uni-tuebingen.de/fakultaeten/philosophische_fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/ (20.02.2022)

https://www.kleinefaecher.de/kartierung/kleine-faecher-von-a-z.html?tx_dmdb_monitoring%5BdisciplineTaxonomy%5D=26&cHash=016cd76da667e9ee63e6eefeb4acfb4c#overview (20.02.2022)

<https://www.studieren-studium.com/master/studieren/Computerlinguistik-Deutschland#result>
<https://www.studieren-studium.com/studium/studieren/Computerlinguistik-Deutschland#result> (04.03.2022 um 14h30)
<http://iks.cs.ovgu.de/iws/> (20.02.2022 um 13h)
<https://www.uni-koblenz-landau.de/de/koblenz/fb4/icv> (20.02.2022)
<https://www.isi.hhu.de/> (20.02.2022 um 13h)
<https://www.cis.lmu.de/> (20.02.2022 um 13h)
<https://www.philol.uni-leipzig.de/institut-fuer-linguistik> (20.02.2022 um 13h)
<https://www.ims.uni-stuttgart.de/> (20.02.2022 um 13h)
<https://www.uni-kassel.de/fb02/> (20.02.2022 um 13h)
<https://www.ling.uni-konstanz.de/> (20.02.2022 um 13h)
<https://dh.phil-fak.uni-koeln.de/> (20.02.2022 um 13h)
<https://www.gw.uni-jena.de/germanistische-sprachwissenschaft> (20.02.2022 um 13h)
<https://www.linguistik.hu-berlin.de/de> (20.02.2022 um 13h)
<https://astt.fb06.uni-mainz.de/> (20.02.2022 um 13h)
<https://www.uni-saarland.de/fachrichtung/lst.html> (20.02.2022 um 13h)
<https://www.linglit.tu-darmstadt.de/institutlinglit/index.de.jsp> (20.02.2022 um 13h)
<https://www.uni-potsdam.de/en/ling/index> (20.02.2022 um 13h)
<https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft?context=ascl> (20.02.2022 um 13h)
<https://www.uni-trier.de/index.php?id=1175> (20.02.2022 um 13h)
<https://www.uni-giessen.de/fbz/fb05/germanistik/iprof/asclhome> (20.02.2022 um 13h)
<https://www.cl.uni-heidelberg.de/> (20.02.2022 um 13h)
<https://www.uni-bielefeld.de/fakultaeten/linguistik-literaturwissenschaft/studium-lehre/faecher/texttechnologie/> (20.02.2022 um 13h)
<https://www.linguistics.ruhr-uni-bochum.de/> (20.02.2022 um 13h)
<https://www.linguistik.phil.fau.de/> (20.02.2022 um 13h)
<https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/seminar-fuer-sprachwissenschaft/studium-lehre/studiengaenge/faq/was-ist-computerlinguistik/> (29.01.2022 um 10h)
<https://de.m.wikipedia.org/wiki/Computerlinguistik> (21.02.2022 um 18h03)
<https://www.cl.uzh.ch/de/texttechnologies/what-is-cl.html> (12.02.2022 um 17h)
<http://latlapps.unige.ch/Parser> ; <http://alpage.inria.fr/iwpt09/atala/fips.pdf> (21.02.2022 um 18h)
 Association pour le Traitement Automatique des Langues : <http://www.atala.org/> (21.02.2022 um 14h)
<https://www.contify.de/glossar/richtig-schreiben/adverbiale-bestimmung/> (Abgerufen am 01.05.2022)
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr)
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr)
<https://www.spiegel.de/ausland/ukraine-news-am-montag-russland-will-ukrainer-vor-eigenes-tribunal-stellen-a-426c518e-a466-49f4-aab1-259cc1f060f1> (25/07/2022 um 17Uhr)
<https://deutshegrammatik20.de/adektiv/adverb/pronominaladverbien/>
https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829 (25/07/2022 um 17Uhr)
<http://www.mein-deutschbuch.de> (02.08.2022 um 12h).

https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829 (25/07/2022 um 17Uhr)

<http://wolfgang-schindler.userweb.mwn.de/skripte/Textling10-06.pdf> (Texte) Abruf am (24.10.2020)

https://efahrer.chip.de/news/oekostrom-pflicht-ab-2025-warum-das-fuer-mieter-verhaengnisvoll-werden-kann_108829 (25/07/2022 um 17Uhr)

<https://www.spiegel.de/politik/deutschland/news-ukraine-krieg-gas-krise-ampel-olaf-scholz-cum-ex-donald-trump-joerg-meuthen-a-186f7cb5-9200-46c9-a5ac-bcfd3b281d96> (19.08.2022)

<http://glottopedia.org/index.php/Koh%C3%A4renz> Abruf am (24.11.2020)

http://glottopedia.org/index.php/Thematische_Progression [19.11.2020]

<http://glottopedia.org/index.php/Koh%C3%A4renz> [24.11.2020]

<http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html> (04.04.08).

http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html (04.04.07).

www.onlinedoctranslator.com (28.01.2023 à 16h)

<https://blog.neuronation.com/de/definition-der-intelligenz-was-ist-das-eigentlich/> (10.04.2023 um 07h)

<https://www.wfb-bremen.de/de/page/stories/digitalisierung-industrie40/was-ist-kuenstliche-intelligenz-definition-ki> (05.04.2023)

<https://www.computerweekly.com/de/definition/Parser#:~:text=Ein%20Parser%20ist%20ein%20Programm,undere%20Software%20ihn%20verstehen%20kann> (27.05.2023 um 21h)

<https://www.sprache-spiel-natur.de/2020/06/14/textsammlungen-und-datenbanken-zur-verwendung-des-deutschen/> (Abgerufen am 30.05.2023 um 13h)

<https://www.primetraining.global/termes-de-reference-lecons-sur-leur-structure-et-linterpretation-des-resultats/> (15.06.2023 um 9h16)

[https://de.m.wikipedia.org/wiki/Anaconda_\(Python-Distribution\)](https://de.m.wikipedia.org/wiki/Anaconda_(Python-Distribution)) (13.06.2023 um 01h30)

www.spyder-ide.org [https://de.m.wikipedia.org/wiki/Spyder_\(Software\)](https://de.m.wikipedia.org/wiki/Spyder_(Software)) (13.06.2023 um 01h30)

<https://dgd.ids-mannheim.de> (13.06.2023 um 01h30)

<http://wortschatz.uni-leipzig.de/> (13.06.2023 um 01h30)

<http://www.dwds.de/resource/kerncorpus/> (13.06.2023 um 01h30)

<http://www.ids-mannheim.de/kt/corpora.html> (23.05.2023).

<http://hdl.handle.net/11022/0000-0000-7FC7-2> (23.05.2023).

<https://www.sketchengine.eu/dewac-german-corpus/> (23.05.2023).

<http://www.korpora.org/Limas/> (23.05.2023).

<https://www2.ids-mannheim.de/cosmas2/projekt/einsteiger/thema-druck.html?template=/cosmas2/template/print.tpl#was> (09.06.2023 um 06h)

www.kali.uni-hannover.de (25.05.2023).

<http://www.uni-giessen.de/gloning/etexte.htm> (25.05.2023).

http://www.ruhr-uni-bochum.de/wege/archiv_1.htm (25.05.2023).

<https://blogs.uni-paderborn.de/lingutools/oeffentliche-korpora/> (23.05.2023).

<http://mhdbdb.sbg.ac.at> (25.05.2023).

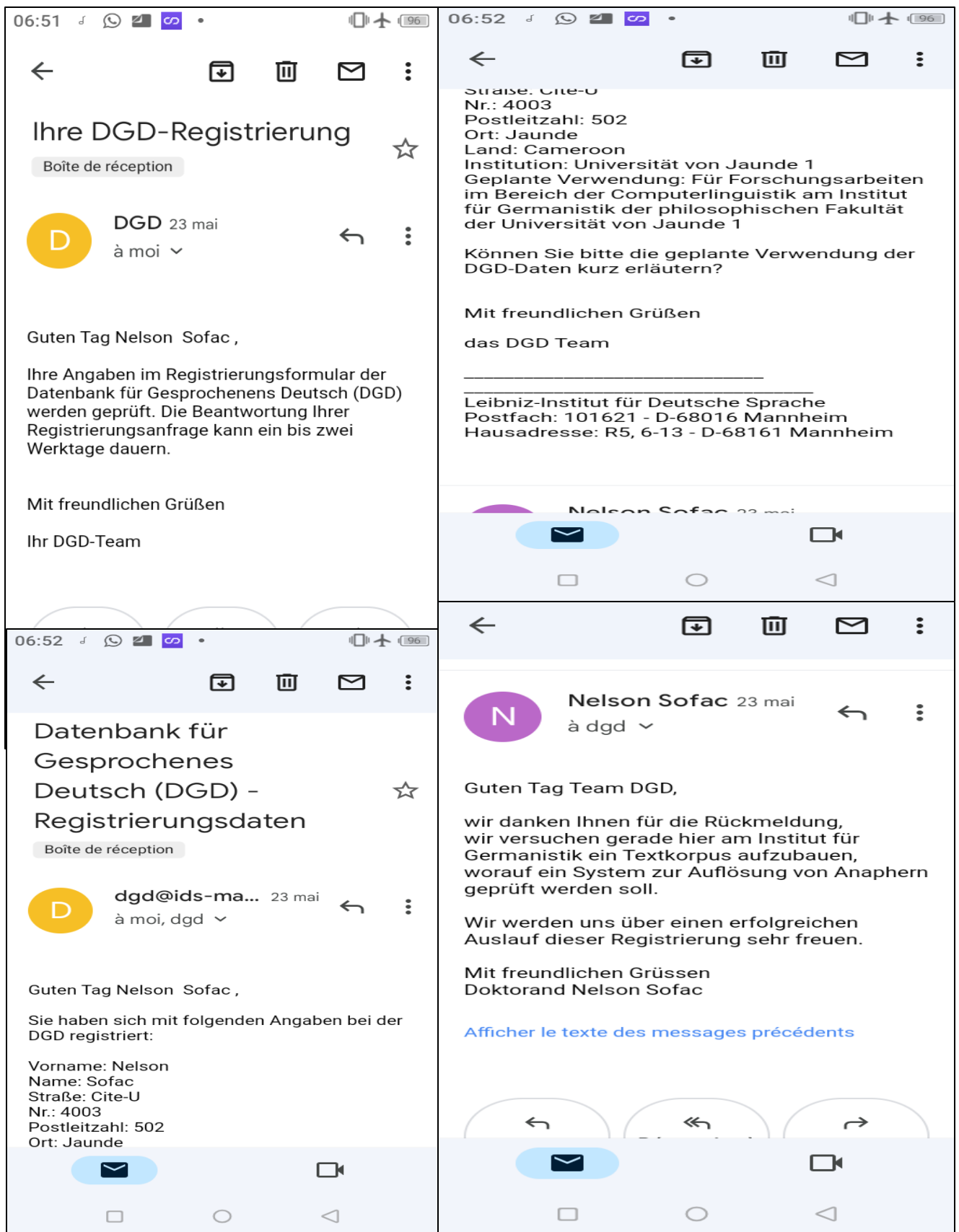
<https://www.sprache-spiel-natur.de/2020/06/14/textsammlungen-und-datenbanken-zur-verwendung-des-deutschen/> (Abgerufen am 30.05.2023 um 13h)

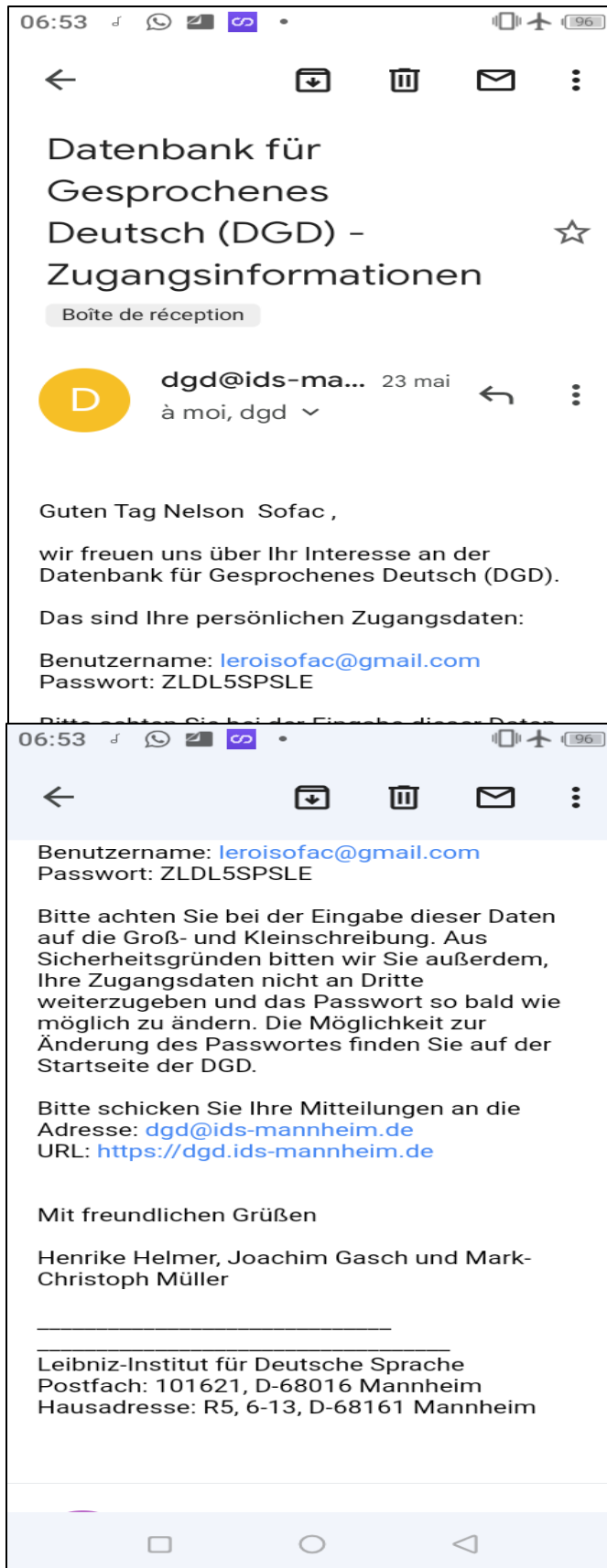
<https://www.sprache-spiel-natur.de/2020/06/14/textsammlungen-und-datenbanken-zur-verwendung-des-deutschen/> (Abgerufen am 30.05.2023 um 13h)

<https://www.wortbedeutung.info/Textkorpus/> (Abgerufen am 23.5.2023)

ANHANG

1. Das mit der DGD-Team (IDS) geführte Gespräch





Abbildungen 1A- 4A: Gespräch mit dem DGD-Team Bild N°1-7

2. Weitere Angaben zum Aufbau des PAK

© Leibniz-Institut für Deutsche Sprache, Mannheim	
COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023 glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica	
Datum	: Donnerstag, den 1. Juni 2023, 20:05:53
Archiv	: W4 - Archiv der geschriebenen Sprache
Korpus	: D-Korpora - Korpora aus Deutschland
Archiv-Release:	Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage	: <i>Dabei</i>
Suchoptionen	: Ei+Ri+Di, Flex
Ergebnis	: 9.192.454 Treffer

Abbildung 5A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DABEI enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim	
COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023 glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica	
Datum	: Freitag, den 2. Juni 2023, 10:43:40
Archiv	: W4 - Archiv der geschriebenen Sprache
Korpus	: D-Korpora - Korpora aus Deutschland
Archiv-Release:	Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage	: <i>Dadurch</i>
Suchoptionen	: Ei+Ri+Di, Flex
Ergebnis	: 1.033.686 Treffer

Abbildung 6A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DADURCH enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim	
COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023 glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica	
Datum	: Freitag, den 2. Juni 2023, 10:57:04
Archiv	: W4 - Archiv der geschriebenen Sprache
Korpus	: D-Korpora - Korpora aus Deutschland
Archiv-Release:	Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage	: <i>Dafür</i>
Suchoptionen	: Ei+Ri+Di, Flex
Ergebnis	: 5.292.172 Treffer

Abbildung 7A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAFÜR enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:05:43
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Dagegen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.761.926 Treffer

Abbildung 8A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAGEGEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:33:20
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Dahinter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 292.565 Treffer

Abbildung 9A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAHINTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:51:07
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Damit*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 8.714.607 Treffer

Abbildung 10A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAMIT enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:57:52
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Danach*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 2.111.335 Treffer

Abbildung 11A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DANACH enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:03:59
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Daneben*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 340.299 Treffer

Abbildung 12A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DANEKEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Mittwoch, den 31. Mai 2023, 9:21:41
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Daran*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.763.793 Treffer

Abbildung 13A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARAN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 19:13:48
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darauf*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 3.004.753 Treffer

Abbildung 14A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARAUF enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 19:38:49
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Daraus*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 606.069 Treffer

Abbildung 15A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARAUS enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:45:00
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darein*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 604 Treffer

Abbildung 16A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAREIN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:37:55
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darin*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 981.172 Treffer

Abbildung 17A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARIN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:09:23
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darüber*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 2.365.720 Treffer

Abbildung 18A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARÜBER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:15:50
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darum*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.083.897 Treffer

Abbildung 19A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARUM enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:31:01
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Darunter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.206.434 Treffer

Abbildung 20A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DARUNTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:31:31
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Davon*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 3.498.234 Treffer

Abbildung 21A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAVON enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:40:30
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Dazu*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 5.709.995 Treffer

Abbildung 22A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAZU enthalten, ausgewählt wurden

Datum : Sonntag, den 4. Juni 2023, 14:50:32
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Dazwischen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 162.536 Treffer

Abbildung 23A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA DAZWISCHEN enthalten, ausgewählt wurden

Datum : Mittwoch, den 31. Mai 2023, 14:17:33
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hieran*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 5.419 Treffer

Abbildung 24A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERAN enthalten, ausgewählt wurden

Datum : Donnerstag, den 1. Juni 2023, 19:22:35
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierauf*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 9.113 Treffer

Abbildung 25A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERFÜR enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 19:49:54
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hieraus*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 9.650 Treffer

Abbildung 26A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERAUS enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 20:07:58
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierbei*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 249.608 Treffer

Abbildung 27A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERBEI enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 10:48:39
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierdurch*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 21.298 Treffer

Abbildung 28A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERDURCH enthalten, ausgewählt wurden

Datum : Sonntag, den 4. Juni 2023, 4:47:19
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierein*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 76 Treffer

Abbildung 29A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIEREIN enthalten, ausgewählt wurden

Datum : Freitag, den 2. Juni 2023, 10:58:56
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierfür*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 225.176 Treffer

Abbildung30A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERFÜR enthalten, ausgewählt wurden

Datum : Freitag, den 2. Juni 2023, 11:09:21
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hiergegen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.919 Treffer

Abbildung 31A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERGEGEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:35:38
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierhinter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 14 Treffer

Abbildung32A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERHINTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:41:04
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierin*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 9.539 Treffer

Abbildung33A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERIN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:53:01
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hiermit*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 28.078 Treffer

Abbildung 34A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERMIT enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:59:59
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hiernach*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 2.598 Treffer

Abbildung35A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERNACH enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:05:50
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierneben*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 4 Treffer

Abbildung36A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERNEBEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:11:54
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierüber*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 11.234 Treffer

Abbildung 37A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERÜBER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:17:38
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierum*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 301 Treffer

Abbildung 38A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERUM enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:27:26
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierunter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 1.805 Treffer

Abbildung 39A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERUNTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:33:48
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hiervon*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 18.707 Treffer

Abbildung 40A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERVON enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:45:49
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierzu*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 229.875 Treffer

Abbildung 41A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERZU enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:52:59
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Hierzwischen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 2 Treffer

Abbildung 42A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA HIERZWISCHEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 20:14:46
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Wobei*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 822.050 Treffer

Abbildung 50A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOBEI enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 10:54:56
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wodurch*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 98.490 Treffer

Abbildung 51A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WODURCH enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:01:58
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wofür*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 138.771 Treffer

Abbildung 52A Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOFÜR enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:29:34
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wogegen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 11.173 Treffer

Abbildung 53A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOGEGEN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Freitag, den 2. Juni 2023, 11:38:46
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wohinter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 298 Treffer

Abbildung 54A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOHLINTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:55:06
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *womit*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 142.722 Treffer

Abbildung 55A Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOMIT enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:01:54
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wonach*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 185.323 Treffer

Abbildung 56A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WONACH enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Mittwoch, den 31. Mai 2023, 14:39:31
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Woran*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 99.821 Treffer

Abbildung 57A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WORAN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 19:30:07
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Worauf*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 144.286 Treffer

Abbildung58A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WORAUF enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Donnerstag, den 1. Juni 2023, 19:57:05
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *woraus*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 15.332 Treffer

Abbildung 59A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WORAUS enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:48:55
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Worein*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 21 Treffer

Abbildung 60A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOREIN enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 4:42:38
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Worin*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 28.451 Treffer

Abbildung61A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOMIT enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:13:59
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *Worüber*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 45.322 Treffer

Abbildung 62A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOMIT enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 5:28:33
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *worum*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 50.559 Treffer

Abbildung 63A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WORUM enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:29:16
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *worunter*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 4.069 Treffer

Abbildung 64A Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WORUNTER enthalten, ausgewählt wurden

© Leibniz-Institut für Deutsche Sprache, Mannheim

COSMAS II-Server, C2API-Version 4.23.5 - 17.05.2023
glemm - a German lemmatizer, version R-2.4.4, 04.07.2022 by Cyril Belica

Datum : Sonntag, den 4. Juni 2023, 14:35:52
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wovon*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 95.862 Treffer

Abbildung 65A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOVON enthalten, ausgewählt wurden

Datum : Sonntag, den 4. Juni 2023, 14:48:04
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *WOZU*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 97.287 Treffer

Abbildung 66A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOZU enthalten, ausgewählt wurden

Datum : Sonntag, den 4. Juni 2023, 14:55:35
Archiv : W4 - Archiv der geschriebenen Sprache
Korpus : D-Korpora - Korpora aus Deutschland
Archiv-Release: Deutsches Referenzkorpus DeReKo-2023-I
Suchanfrage : *wozwischen*
Suchoptionen : Ei+Ri+Di, Flex
Ergebnis : 6 Treffer

Abbildung 67A: Kopf des Dokuments (Ergebnisse einer Suchanfrage im DeReKo) woraus Textabschnitte, die das PA WOZWISCHEN enthalten, ausgewählt wurden

3. Einsatz des PAAR auf den Textabschnitt (N°51-75)

- 51- Auch Veranstaltungen wie "Hammelburg moves" wären kaum noch möglich, weil die Vereine um 2 Uhr die Musik abschalten müssten, während bei den Gastronomen noch gefeiert werden könne, betonte Alexander Stolz. "Zu kulturellem Leben gehört auch ein Nachtleben *dazu*, und das verträgt sich einfach nicht mit einer Sperrstunde", so der Stadtrat. Die Jugendkulturszene werde *hierdurch* bedroht und das Land verliere weiter an Attraktivität. (SAZ11/JAN.00079 Saale-Zeitung, 03.01.2011, S. 7; "Sperrstunde bedroht Jugendkulturszene")
- 52- "Versteigerungen drohen" Doch *davon* sei der Arbeitsmarkt meilenweit entfernt. Es gebe immer mehr Menschen, die nicht angemessen bezahlt würden, zum Beispiel Verkäuferinnen oder Friseurinnen, und trotz Arbeit nicht genug Geld zum Leben hätten. Sie zahlten *hierdurch* auch nur geringe Beträge in die Rentenversicherung ein, *wodurch* ihnen auch im Alter Armut drohe. (SAZ11/JAN.00611 Saale-Zeitung, 08.01.2011, S. 9; Zunahme der Armut hat viele Gründe)
- 53- Aufgrund der öffentlichen Ausschreibung des Bauvorhabens kann frühestens im Oktober mit dem Start gerechnet werden. Die Regierung schlug deshalb vor, einen Teil der Bauarbeiten vorzuziehen, wie zum Beispiel die nichtöffentliche Erschließung auf dem Baugrundstück. *Hierdurch* wäre dann

- der Baubeginn nachgewiesen. Die restlichen Ausschreibungen könnten dann auch zu einem späteren Zeitpunkt erfolgen. (SAZ11/JUN.01579 Saale-Zeitung, 15.06.2011, S. 18; Baubeginn)
- 54- Die Insekten kommen nicht wieder hoch, denn feinste, abwärts gerichtete Haare im oberen Teil des Tunnels halten sie unten. **"Hierein"** werden von den Pflanzen Verdauungsenzyme abgegeben, die auch im menschlichen Magen vorkommen", sagt der Züchter und Buchautor Thomas Carow. Bakterien unterstützen den Zersetzungsprozess. (SAZ11/MAI.02058 Saale-Zeitung, 17.05.2011, S. 33; Vorsicht, gefräßige Fallen!)
- 55- **Damit** sich das nicht wiederholt, hatten sie, ihr Mann Marco sowie Vereinsmitglied Christine Funk die Idee mit der Igelklappe, die seit November vor ihrem Haus in der St. Nikolauserstraße 26 in Großrosseln-Naßweiler steht. Als Vorbild diente ihnen die Babyklappe von Krankenhäusern. "Wir haben einen Kaninchenstall so umgebaut, dass er einen isolierten Schlafbereich hat sowie einen Bereich mit Futter und Wasser", beschreibt Maas die "Klappe". Für die kalte Zeit liegt ein Heizkissen bereit, das sich nach dem Entpacken selbstständig auflädt, denn Igel mögen es warm. **Hierein** kann der Finder den hilfsbedürftigen Igel legen und dann den Kasten verschließen, denn in die Klappe darf nur ein Tier: "Sonst gäbe es Mord und Totschlag. (SBZ12/NOV.01390 Saarbrücker Zeitung, 05.11.2012; Babyklappe für kranke Igel)
- 56- Als Kosten für den Tiefbau, die Verlegung des Rohrs sowie zwei Schächte gibt Enrico Bachmann rund 27 000 Euro an. **Hierein** fällt nicht der Deckenschluss der Straße, da er als Instandsetzung abzurechnen sei, erklärt der Leiter der Tiefbauverwaltung weiter. Der Rat beschloss die Baumaßnahme einstimmig. (so) (SZE16/DEZ.03417 Sächsische Zeitung, 08.12.2016, S. 15; Regenentwässerung beschlossen)
- 57- Um 17.15 Uhr schließt sich die Busabfahrt zur 22. FEN-Auftaktsitzung (Föderation Europäischer Narren) in Knetzgau an. **Hierfür** wurden gleich drei Programmpunkte aus Untererthal nominiert: der Gardetanz der Prinzengarde, der Showtanz der Großen Garde und die Bütt von Ruth Neder und Bettina Hügel. Am 14. Januar fährt das Untererthaler Prinzenpaar, Inge I. und Georg I. mit einer Abordnung zum 2. Prinzenpaar-Treffen nach Albertshofen, und tags **darauf** wird der Jubiläumsabend "50 Jahre KOKAGE Wiesentheid" aufgesucht. (SAZ11/JAN.00616 Saale-Zeitung, 08.01.2011, S. 10; ohne Titel)
- 58- Kartenwünsche für die Bunten Abende des FC Rottershausen, die am 12., 18. und 19. Februar stattfinden, können Interessierte noch bis Samstag, 22. Januar, in die Boxen in der Metzgerei Bieber, der Bäckerei Stahl sowie auch im Sportheim einwerfen. Der Kartenvorverkauf für die Faschingsveranstaltungen findet schließlich am Sonntag, 30. Januar statt. Beginn **hierfür** ist um 16 Uhr im Sportheim des FC. Khw (SAZ11/JAN.01852 Saale-Zeitung, 19.01.2011, S. 7; Kartenwünsche in Boxen werfen)
- 59- Ein großes Problem sei der Mainausbau, gegen den sich Fischer seit Jahren wehren. Jüngst zogen Kitzinger Fischer vor das Bundesverwaltungsgericht, da sie mit den Ausgleichsmaßnahmen unzufrieden waren: "Doch ihre Klage wurde abgewiesen." Aktuell überlegen sie laut Silkenat, ob sie auf europäischer Ebene weiterkämpfen sollen. Ginge es nach den rund 500 Flussfischern in Unterfranken, sollten Freizeitsportler per Gesetz **davon** abgehalten werden, den Main und seine Nebenflüsse allzu intensiv zu nutzen: "Denn der Freizeitdruck hat enorm zugenommen." Immer mehr Menschen finden laut dem bezirklichen Fischereireferenten gefallen am Bootfahren sowie an Wasser- und Jetski. Für Fische bedeute dies Stress pur. Aktuell werde an einem Seitenarm des Mains sogar ein Motorboot-Parcours geplant. **Hiergegen** klagten Würzburger Fischer. (SAZ11/AUG.00298 Saale-Zeitung, 03.08.2011, S. 8; Freizeitsportler machen Fischen Stress)
- 60- **Damit** sich alle Fans dieses Sportes auf die kommende Saison an den Stränden vorbereiten können, findet am 24. Mai 2008 von 9.00 Uhr bis 18.00 Uhr der 1. Grabower-Beachday statt. **Hierhinter** verbirgt sich ein Beachvolleyball-Turnier, auf insgesamt 4 Feldern, mit spannenden Spielen den ganzen Tag. Anmelden können sich begeisterte Amateur Beachvolleyballteams als Mixed oder "Normal"-Team bis zum 13.05. 2008 unter folgenden Telefonnummern: GWaldbad, 03 87 56 / 27 7 12 oder (SVZ08/APR.01056 Schweriner Volkszeitung, 02.04.2008, S. 13; Beachvolleyball für Amateurtteams)
- 61- Die Hammelburger Wehr führte den Lehrgang auch durch, weil das neue Löschfahrzeug LF 20/20 mit einem Notfallrucksack ausgerüstet ist. **Hierin** findet sich ein Großteil der medizinischen Geräte, mit denen auch ein Rettungswagen ausgestattet ist. "Wenn wir solche Geräte haben, brauchen wir natürlich auch Leute, die **damit** umgehen können", betonte Rettungsassistent Maximilian Zoll, der den Kurs gemeinsam mit Frank Karrlein leitete. (SAZ11/FEB.00026 Saale-Zeitung, 01.02.2011, S. 9; Wie Feuerwehrleute Leben retten können)
- 62- Grob gesagt geht es bei einem "Spread Ladder Swap" um eine Wette auf die künftige Zinsentwicklung. Das Problem: Die Bank wettet gegen ihren Kunden. "Der Gewinn der einen Seite ist der Verlust der anderen", sagte der Vorsitzende Richter Ulrich Wiechers bei der

- Urteilsbegründung am Dienstag. **Hierin** liege ein "schwerwiegender Interessenkonflikt". **Darüber** müsse die Bank den Kunden aufklären. (SAZ11/MAR.02932 Saale-Zeitung, 23.03.2011, S. 22; ohne Titel)
- 63- **Hammelburg** Das erste von drei Windrädern nördlich von Gauaschach steht, und nächste Woche soll es Strom liefern. Auch die beiden anderen Windräder sollen bereits eine Woche **darauf** errichtet sein und ans Netz gehen. Die Gesamtanlage wird eine Leistung von sechs Megawatt haben und im Jahr voraussichtlich knapp zwölf Millionen Kilowattstunden (kWh) Strom erzeugen. **Hiermit** können nach Angaben des Betreibers, der Hamburger Firma Wotan, rund 3300 Haushalte mit elektrischer Energie versorgt werden. 9 (SAZ11/JAN.00416 Saale-Zeitung, 06.01.2011, S. 1; Strom für rund 3300 Haushalte)
- 64- Die CSU-Fraktion hat einen Dringlichkeitsantrag für die nächste Stadtratssitzung gestellt. **Hiernach** soll die Stadt ein Grundstück für den eventuellen Neubau einer Fachober- und Berufsoberschule (FOS/BOS) in Hammelburg bereitstellen. (SAZ12/MAR.02173 Saale-Zeitung, 19.03.2012, S. 10; CSU für Fachober- und Berufsschule am Sportzentrum)
- 65- Da der junge Fahrer seinen Führerschein zur Zeit wegen eines Fahrverbots bei der Polizei liegen hat, setzte er seine Fahrt aus Furcht vor den Konsequenzen fort, ohne sich um den Schaden zu kümmern. **Dabei** geriet er jedoch so in Panik, dass er bereits an der nächsten Einmündung erneut von der Straße abkam und ein Straßenschild umfuhr. Auch **hiernach** fuhr er einfach weiter. Pech für ihn, dass Anwohner durch die Unfallgeräusche wach wurden und den Vorfall bemerkten. (SAZ12/APR.00009 Saale-Zeitung, 02.04.2012, S. 3; ohne Titel)
- 66- **Hierneben** kann im Rahmen des öffentlichen Straßenverkehrs auch recht schnell der Straftatbestand der Nötigung verwirklicht werden. Dies kann unter Umständen bereits dann anzunehmen sein, wenn ein Fahrzeug auf der Autobahn auf ein anderes Fahrzeug zu dicht auffährt und dies über längere Zeit anhält. (SVZ07/FEB.10453 Schweriner Volkszeitung, 24.02.2007, S. 26; Vollbremsung mit juristischen Folgen)
- 67- Aus dem Großbetrieb ist noch zu erfahren, dass hierneben deutschen Ausgaben auch englische gedruckt wurden, so für den britischen Verlag Random House, der wie GGP Media zum Bertelsmann-Konzern gehört. Stephen Esson, Direktor bei Random House in Großbritannien, erhielt aber nicht nur sauber gedruckte Exemplare des Dan-Brown-Bestsellers aus Pößneck, sondern als Dankeschön für die gute Zusammenarbeit auch ein Kunstwerk, das wiederum unmittelbar auf den The Da Vinci Code-Auftrag zurückgeht. (OTZ06/MAI.10520 Ostthüringer Zeitung, 27.05.2006, S. OAPO127; Verwobene Zeichen rufen nach der Entschlüsselung - Sakrileg-Druckplatten)
- 68- Noch geht es in Oberbach beschaulich zu. Doch schon am Dienstag nach Ostern wird der Ausbau der Ortsdurchfahrt beginnen. **Hierüber** informierte Bürgermeister Alfred Schrenk (SPD) im Rahmen der jüngsten Gemeinderatssitzung die Kommunalpolitiker. (SAZ11/APR.02731 Saale-Zeitung, 21.04.2011, S. 19; Die Ruhe vor dem Sturm)
- 69- Der slowakische Fahrer, der am Parkplatz Klöffelsberg wartete, konnte keinen gültigen Versicherungsnachweis vorlegen. Daher untersagte die Polizei ihm die Weiterfahrt bis zur Vorlage einer entsprechenden Bescheinigung. **Hierum** wollte sich die Firma kümmern. (SAZ12/SEP.00875 Saale-Zeitung, 08.09.2012, S. 9; Lkw mit defektem Reifen und ohne Versicherung unterwegs)
- 70- Markus Reeh Hammelburg Immer mehr Menschen rutschen in Deutschland unter die Armutsgrenze, und **davon** bleibt natürlich auch der Altlandkreis Hammelburg nicht verschont. Die Zahl derer, die auf Hilfe vom Staat und soziale Einrichtungen wie Kleiderkammer und Tafel angewiesen sind, steigt kontinuierlich. Und **hierunter** sind zunehmend auch ältere Menschen, deren Renten nicht ausreichen, um über die Runden zu kommen. (SAZ11/JAN.00611 Saale-Zeitung, 08.01.2011, S. 9; Zunahme der Armut hat viele Gründe)
- 71- Riskanter Alkoholkonsum Suchterkrankungen sind ein enormes gesellschaftliches und auch sozioökonomisches Problem. Die Zahl der an einer Suchterkrankung leidenden Menschen nimmt stetig zu. Aktuell gehen Experten **davon** aus, dass es in Deutschland etwa 9,3 Millionen Menschen mit riskantem Alkoholkonsum gibt, **hierunter** 2,7 Millionen mit einem schädlichen Alkoholgebrauch und 1,7 Millionen Alkoholabhängige. Etwa 1,5 Millionen Menschen leiden an einer Medikamentenabhängigkeit (größte Teilgruppe: Frauen in der zweiten Lebenshälfte), 150 000 sind abhängig von so genannten harten Drogen. (SAZ11/JAN.01234 Saale-Zeitung, 14.01.2011, S. 38; Ein Medikament gegen Alkoholsucht)
- 72- Viele Polit-Promis getroffen War sie zu Beginn bei öffentlichen Auftritten noch etwas unsicher, so kann sie mittlerweile ohne Probleme eine Rede vor vielen hundert Menschen halten. Stolze 160 Termine hat sie seit 2008 absolviert, **hierunter** auch einige im Ausland, und **dabei** etliche hochrangige Persönlichkeiten kennen gelernt. So traf sie in Bad Kissingen mit Bundeskanzlerin Angela Merkel zusammen, in Hammelburg sprach sie mit Bundespräsident Christian Wulff, und den bayerischen Ministerpräsidenten Horst Seehofer hat sie gleich mehrfach getroffen, zuletzt in

- München. (SAZ11/JUN.01543 Saale-Zeitung, 15.06.2011, S. 9; "Ich würde es immer wieder machen")
- 73- Der Sachausschuss Caritas war sehr zufrieden mit der Haussammlung, bei der wieder mehr als 700 Euro gespendet wurden. **Hiervon** bleiben 30 Prozent in der Pfarrei. **Davon** werden etwa die Weihnachtspäckchen für die Senioren finanziert. Pfarrer Lux gab einen Überblick über Vergangenes und Geplantes in der Pfarreiengemeinschaft. So erwähnte er, dass er die Arbeiten nicht alle alleine bewerkstelligen könne und auf die Hilfe seiner Hauptamtlichen, aber insbesondere auch auf die der Ehrenamtlichen angewiesen sei. (SAZ11/APR.02013 Saale-Zeitung, 16.04.2011, S. 6; Die Arbeit wird auf viele Schultern verteilt)
- 74- Die jährliche Mitglieder-versammlung des Feuerwehrvereins der Wehr Aura findet am kommenden Donnerstag, 6. Januar, in der Festhalle statt. Beginn ist um 19.30 Uhr. **Hierzu** ist die gesamte Ortsbevölkerung, insbesondere die aktiven Dienstleistenden der Wehr sowie sämtliche Mitglieder des Feuerwehrvereins, willkommen. Auf der Tagesordnung stehen unter anderem Berichte der Gruppenführer und des Kommandanten. Im Anschluss werden die Neuwahlen der Vorstandschaft durchgeführt. **Danach** sollen die neuen Mitglieder aufgenommen werden. (SAZ11/JAN.00084 Saale-Zeitung, 03.01.2011, S. 10; Neuwahlen beim Feuerwehrverein)
- 75- Jetzt haben die Wissenschaftler das Testverfahren weiterentwickelt, bei dem ein Teelöffel Blut ausreicht, um im Organismus zirkulierende Tumorzellen schon in sehr frühen Stadien nachzuweisen. **Hierzu** haben die Wissenschaftler einen winzigen Silikonchip entwickelt, der mit mehreren Zehntausend Antikörpern beschichtet ist. Wird dieser Mikrochip mit Blut benetzt, trennen sich die bösartigen Zellen von normalen Zellen und haften an den Antikörpern. Die anhaftenden Krebszellen können mit gewissen Färbemethoden nachgewiesen und zum "Leuchten" gebracht werden. Nach Aussagen der Wissenschaftler ist das Verfahren in der Lage, die minimale Menge von zehn Krebszellen pro Milliliter Blut aufzuspüren. Die neue Methode ähnelt dem ebenfalls gängigen Verfahren der Biopsie, also der Entnahme von verdächtigem Gewebe. Allerdings ist **hierzu** kein chirurgischer Eingriff nötig, wenige Blutstropfen genügen. (SAZ11/JAN.00411 Saale-Zeitung, 05.01.2011, S. 34; Ein Teelöffel Blut für die Diagnose)

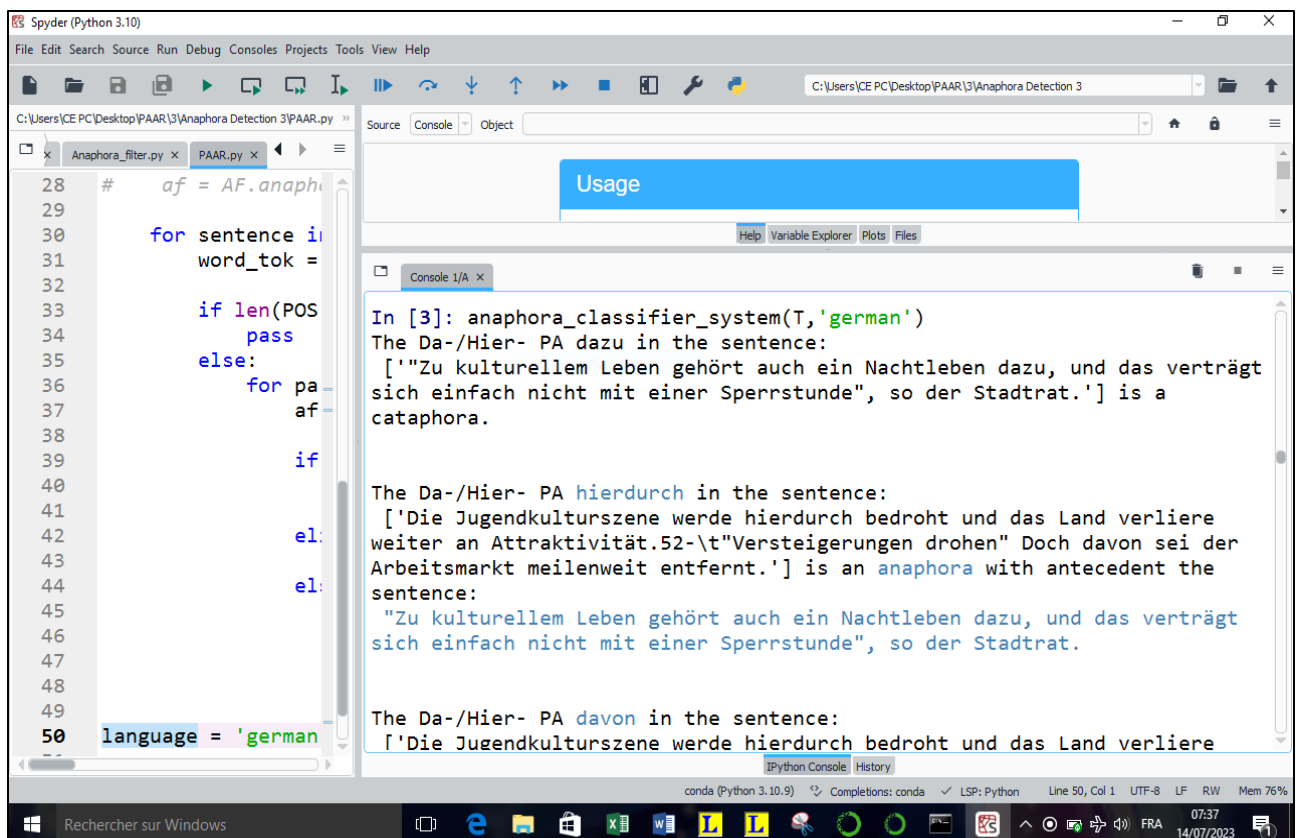


Abbildung 82: Einsatz des PAAR auf Texte N°51-75 (1)

Bei Näherbetrachtung der Angaben, die in der Abbildung 75 abzulesen sind, stellen wir nämlich fest, dass PAAR das PA „dazu“ als kataphorisch verzeichnet hat, etwas, dass nicht richtig ist, denn „dazu“ verweist in diesem Textabschnitt eher auf die präpositionale Phrase („Zu kulturellem Leben“. Was das PA „hierdurch“ angeht, wurde

es vom PAAR als anaphorisch klassifiziert, etwas, das auch richtig ist. Allerdings ist PAAR bei der Auflösung vom „hierdurch“ fehlgeschlagen. Und zwar weil das richtige Antezedens der Satz „Auch Veranstaltungen wie *"Hammelburg moves"* wären kaum noch möglich, weil die Vereine um 2 Uhr die Musik abschalten müssten, während bei den Gastronomen noch gefeiert werden könne, betonte Alexander Stolz.“ ist.

In der Abbildung 76 lässt sich feststellen, dass PAAR zwei PA („davon“ & „hierdurch“) erkannt hat, wobei die jeweiligen richtigen Naturen und die entsprechenden Antezedenzen richtig wiedergegeben wurden. Die gleiche Analyse gilt auch für die PA („Wodurch“ & „Hierdurch“) in der Abbildung 77.

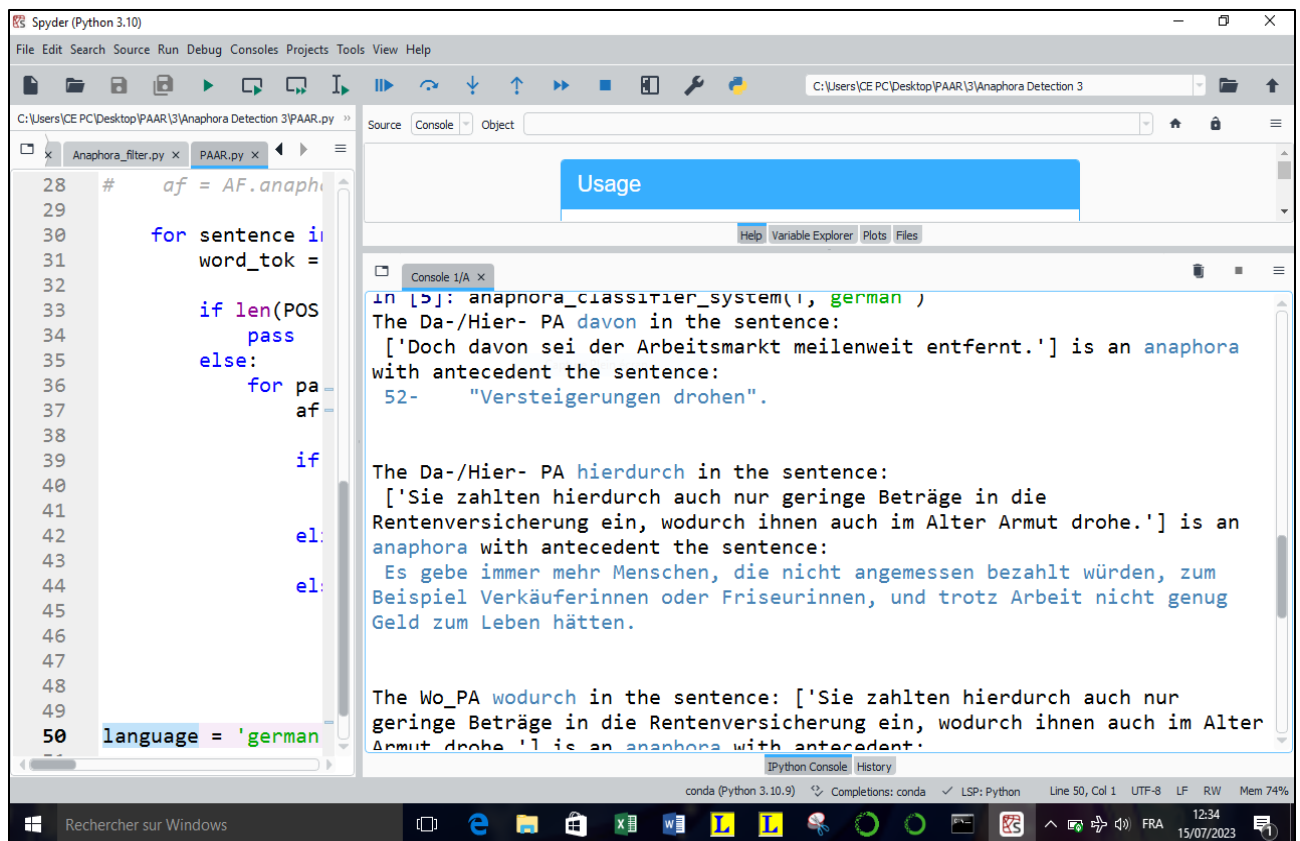


Abbildung 83: Einsatz des PAAR auf Texte N°51-75 (2)

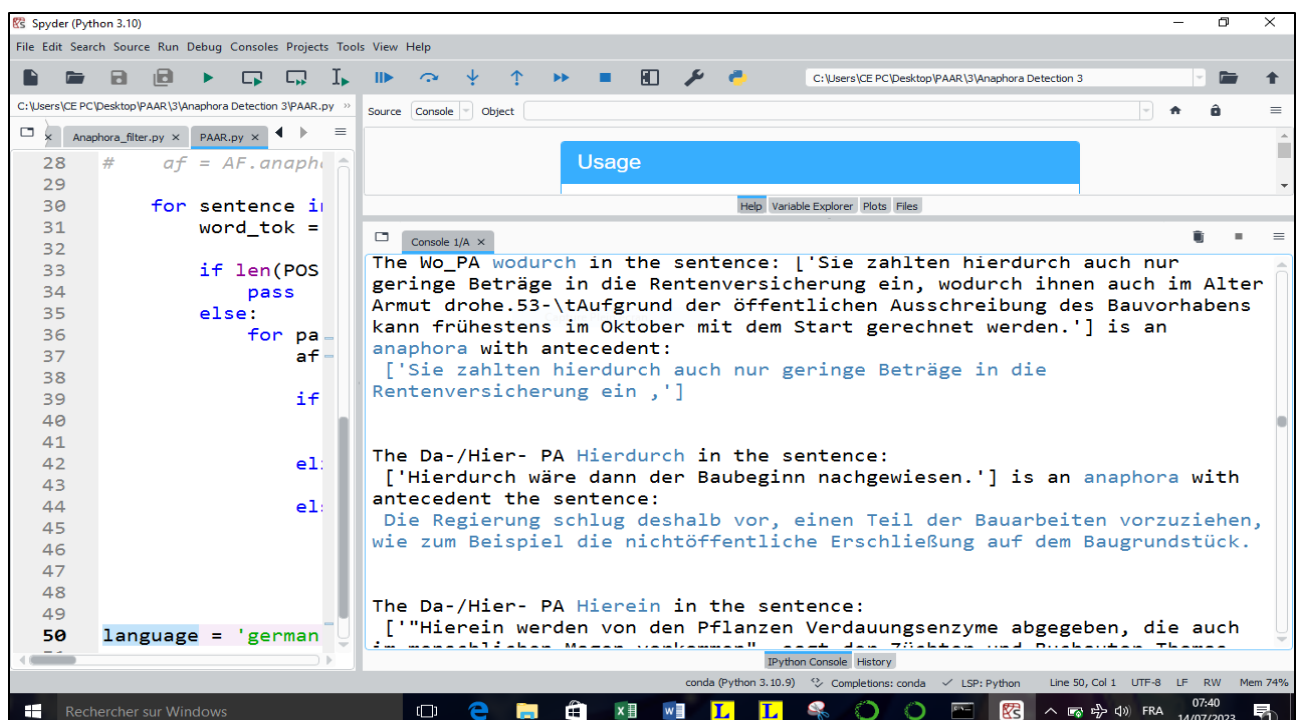


Abbildung 84: Einsatz des PAAR auf Texte N°51-75 (3)

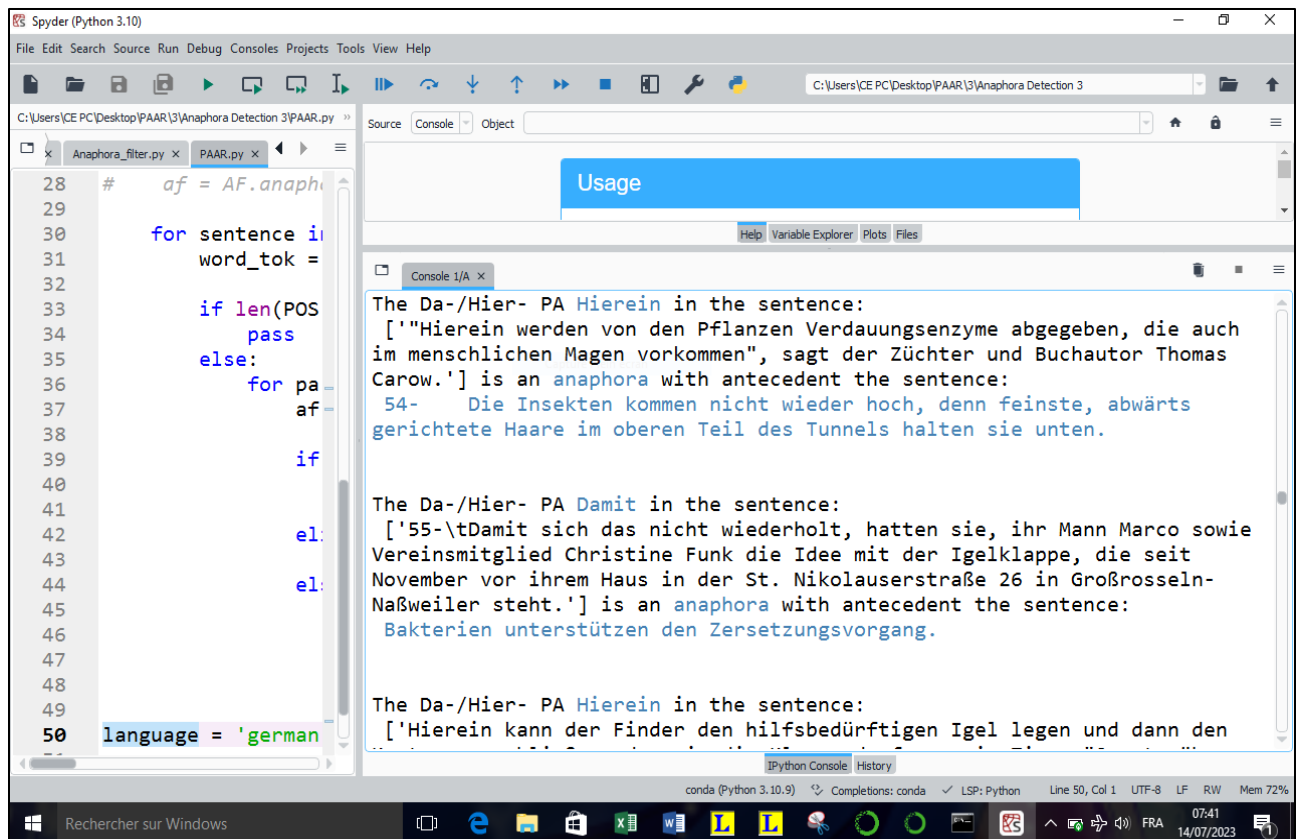


Abbildung 85: Einsatz des PAAR auf Texte N°51-75 (4)

In der Abbildung 78 stellt man fest, dass PAAR 2 PA gefunden hat. Etwas das nach Näherbetrachtung des Textes eigentlich nicht richtig ist. Denn das Wort „Damit“ hier ist kein PA, sondern eher eine SUBJUNKTION. Und was das Wort „Hierein“ angeht, handelt es sich zwar um ein PA aber mit falschen Antezedens gekoppelt. Das richtige wäre eher die PP *„im oberen Teil des Tunnels“*

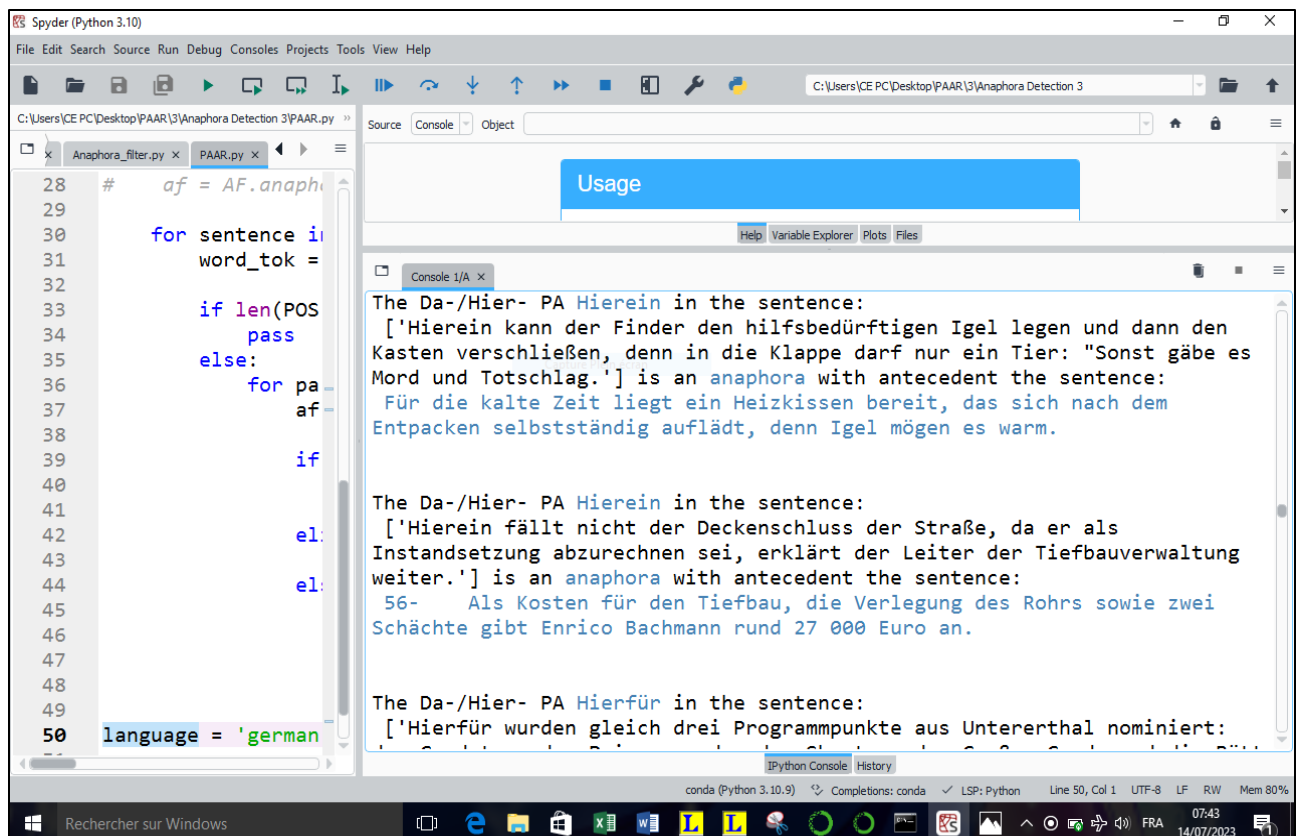


Abbildung 86: Einsatz des PAAR auf Texte N°51-75 (5)

PAAR hat die zwei Erscheinungen vom PA „Hierin“ in der Abbildung 80 als Anapher verzeichnet. Etwa, das eigentlich, nach Näherbetrachtung des Textes richtig ist. Allerdings stellt man nämlich fest, dass das erste PA nicht korrekt aufgelöst worden ist; denn das richtige Antezedens wäre die NP („ein Heizkissen“). Was die zweite Erscheinung angeht, kann gesagt werden, dass dies richtig aufgelöst worden ist. Denn es handelt sich ja im Allgemeinen um die „kosten“.

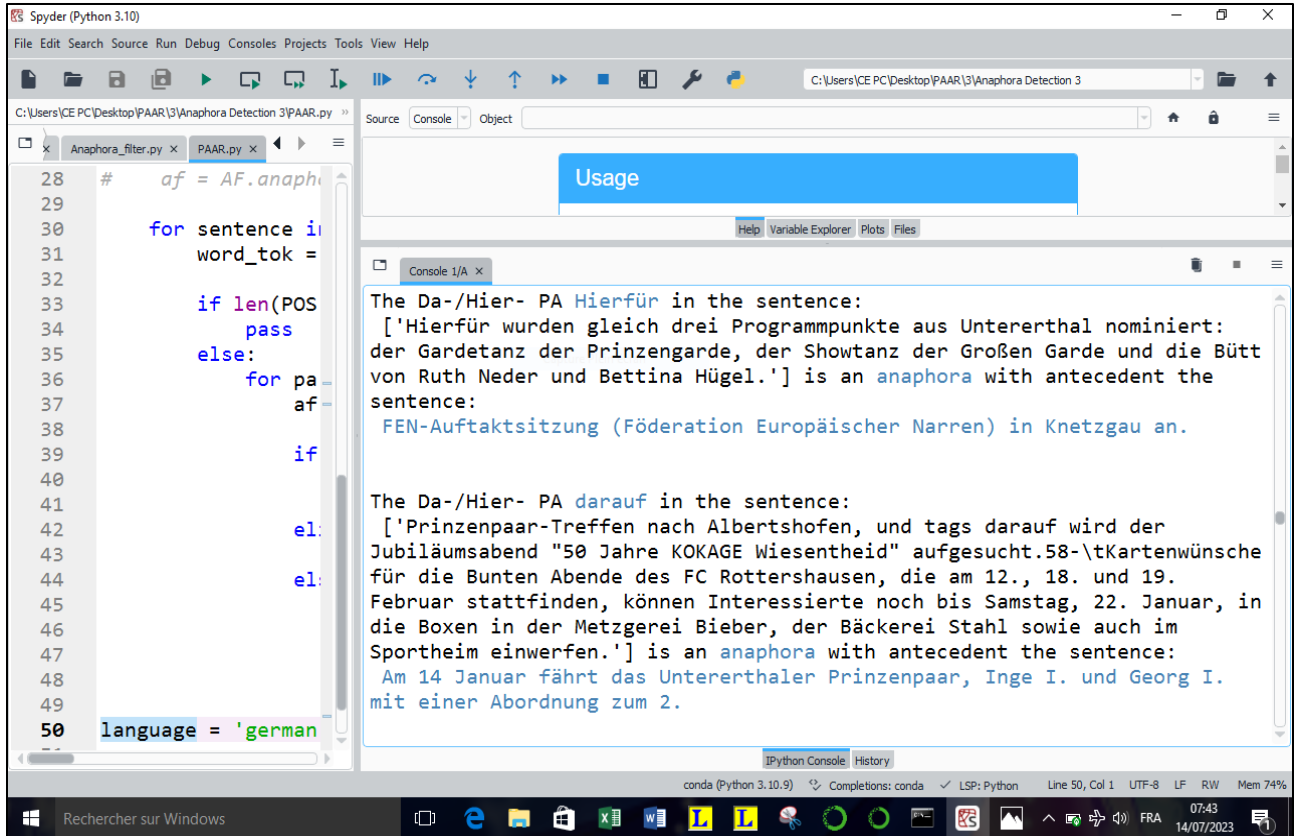


Abbildung 87: Einsatz des PAAR auf Texte N°51-75 (6)

In der Abbildung 80 lässt sich feststellen, dass PAAR zwei PA („Hierdurch“ & „darauf“) erkannt hat, wobei die jeweiligen richtigen Naturen und die entsprechenden Antezedenzen richtig wiedergegeben wurden. Allerdings weist PAAR einige Schwäche bei dem Parser des Textes auf.

Das PA „Hierfür“ in der Abbildung 81 ist zwar eine Anapher, aber das entsprechende Antezedens wurde nur zum Teil richtig gegeben. Denn „Hierfür“ verweist hier nur auf die NP „Der Kartenvorverkauf für die Faschingsveranstaltungen“ und nicht auf den Rest des Satzes. Was die PA „davon“ angeht, handelt es sich, wie PAAR es angegeben hat, um eine Katapher und zwar, weil die Bezugseinheit von „davon“ vorwärts steht.

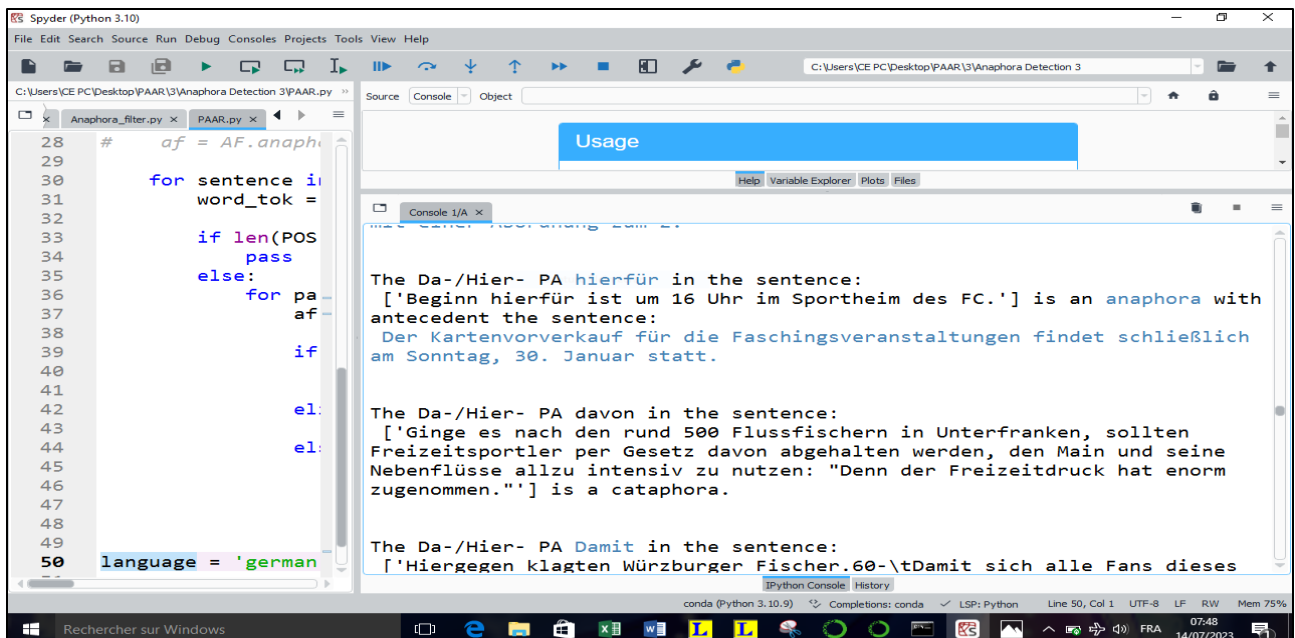


Abbildung 88: Einsatz des PAAR auf Texte N°51-75 (7)

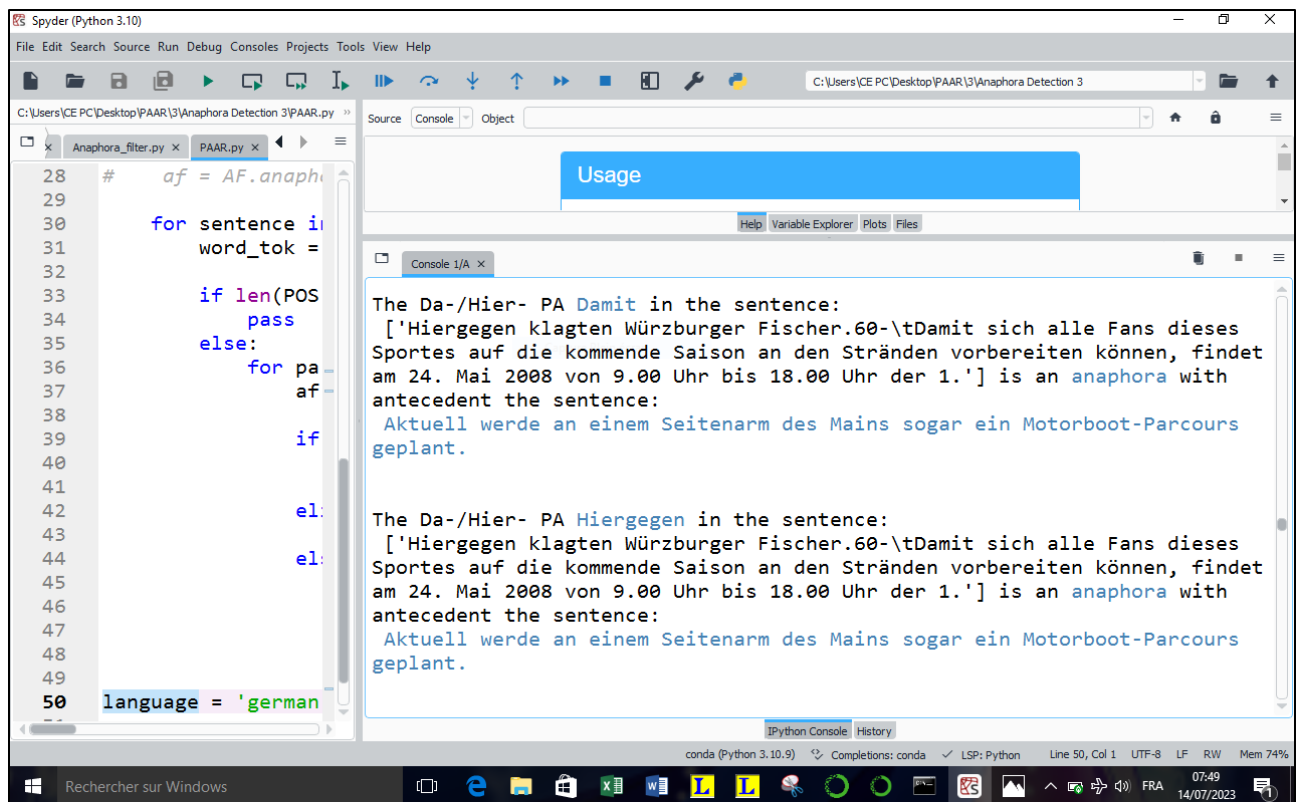


Abbildung 89: Einsatz des PAAR auf Texte N°51-75 (8)

In der Abbildung 82 ist das Wort „damit“ kein anaphorisches PA, sondern eher eine SUBJUNKTION. Und was das PA „hiergegen“ angeht, hat PAAR es richtig aufgelöst.

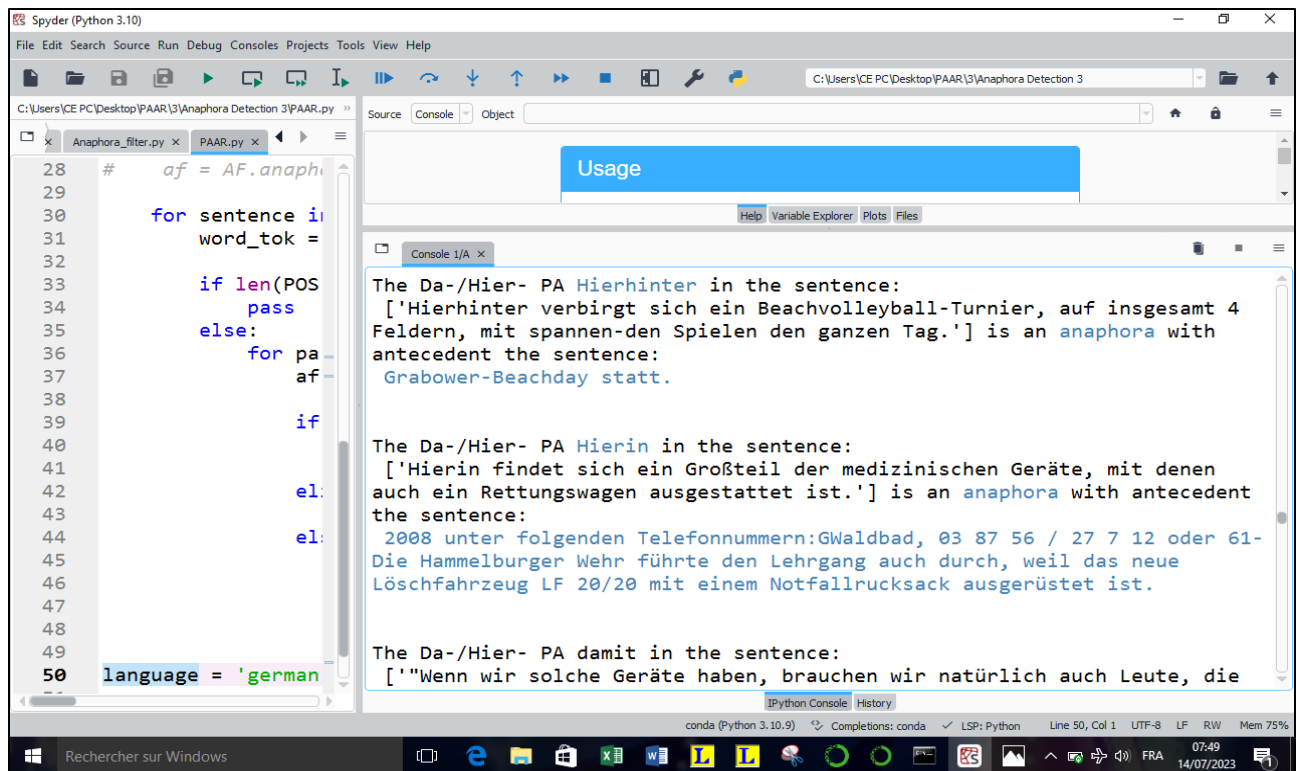


Abbildung 90: Einsatz des PAAR auf Texte N°51-75 (9)

Das PA „hierhinter“ in der Abbildung 83 ist richtig aufgelöst worden. Was das PA „Hierin“ hingegen geht, handelt es sich zwar um ein anaphorisches PA aber das passende Antezedens ist die NP „*einem Notfallrucksack*“

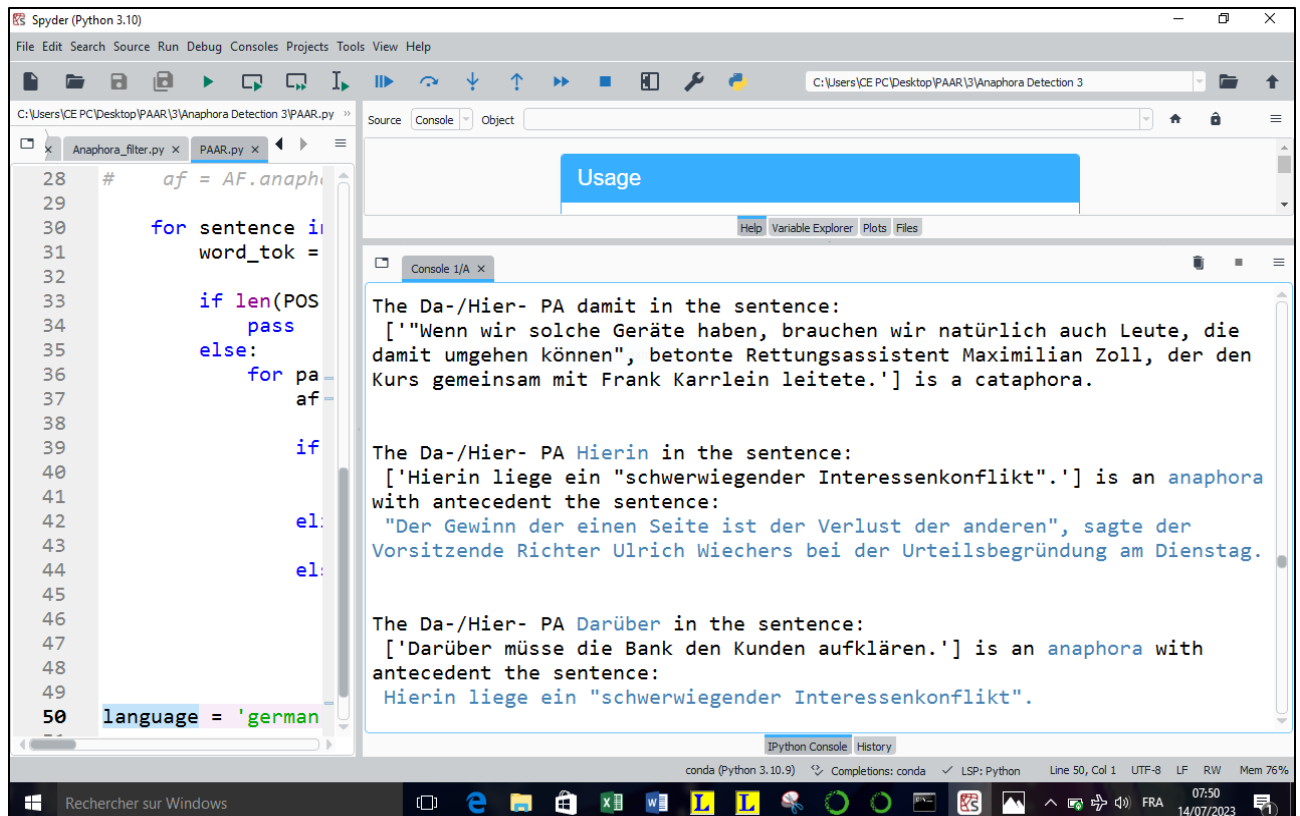


Abbildung 91: Einsatz des PAAR auf Texte N°51-75 (10)

Das PA „Damit“ in der Abbildung 84 ist keine Katapher, wie PAAR es anerkannt hat, sondern eher eine Anapher mit dem Satzteil „*Solche Geräte*“ als Antezedens. Was das PA „Hierin“ angeht handelt es sich zwar um eine Anapher aber das richtige Antezedens ist „*Die Bank wettet gegen ihren Kunden*“. Das anaphorische PA „Darüber“ wurde hingegen gut aufgelöst.

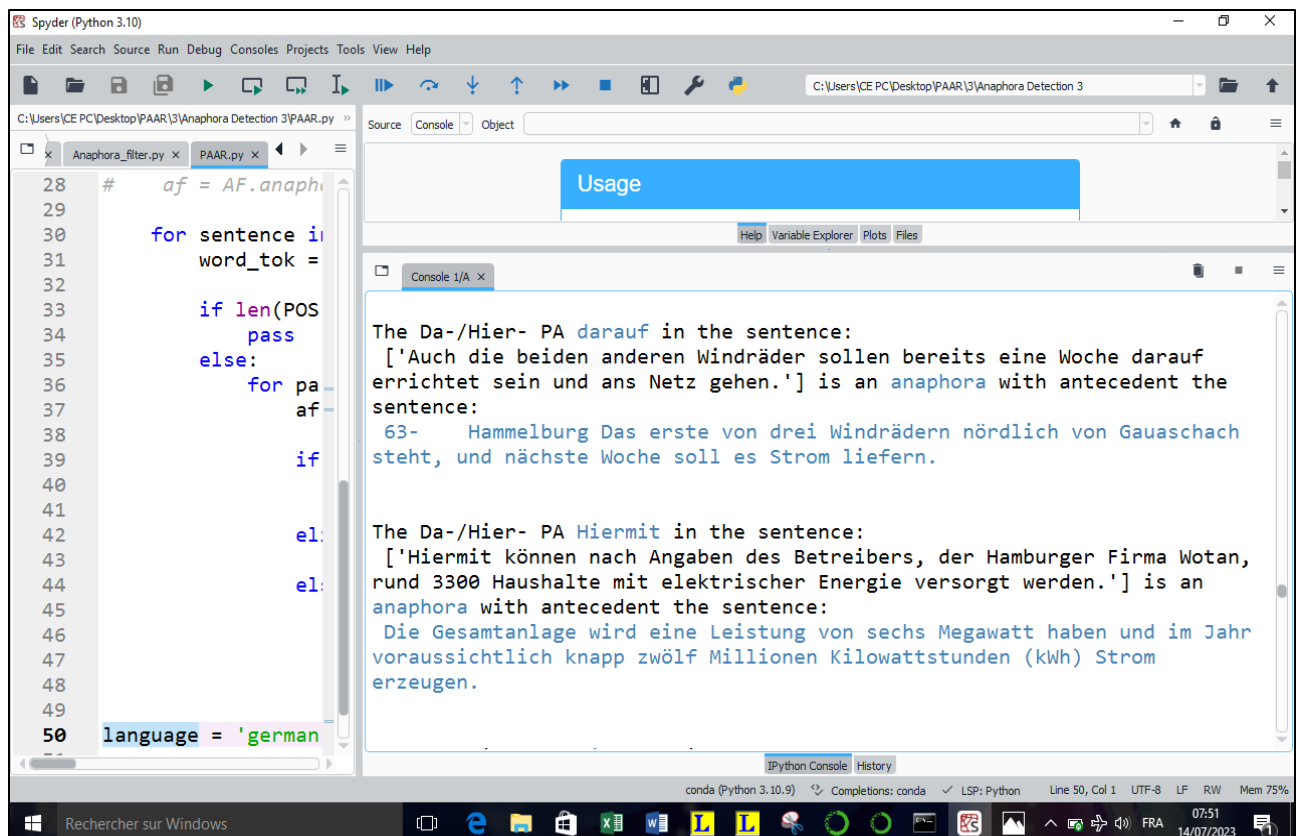


Abbildung 92: Einsatz des PAAR auf Texte N°51-75 (11)

Das PA „darauf“ hier in der Abbildung 85 ist zwar eine Anapher aber mit dem Antezedens „nächste Woche“. Was das zweite PA der Reihe angeht, wurde es vom PAAR richtig aufgelöst.

Das PA „Hiernach“ in der Abbildung 86 ist zwar eine Anapher aber mit dem Antezedens („einen Dringlichkeitsantrag“). Was das zweite PA „Dabei“ angeht, wurde es vom PAAR gut aufgelöst

Das PA „Hiernach“ in der Abbildung 87 ist vom PAAR richtig aufgelöst worden. Was das PA „Hierneben“ angeht, handelt es sich um keine Anapher.

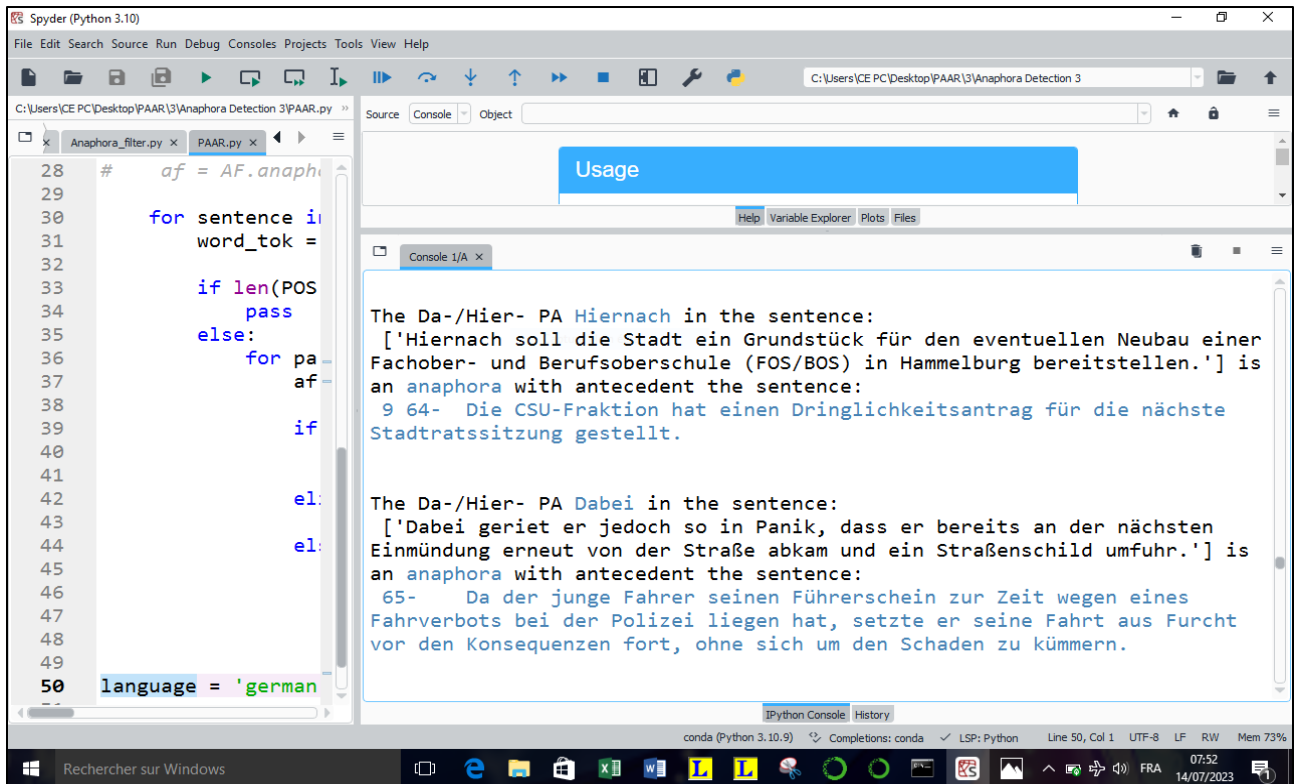


Abbildung 93: Einsatz des PAAR auf Texte N°51-75 (12)

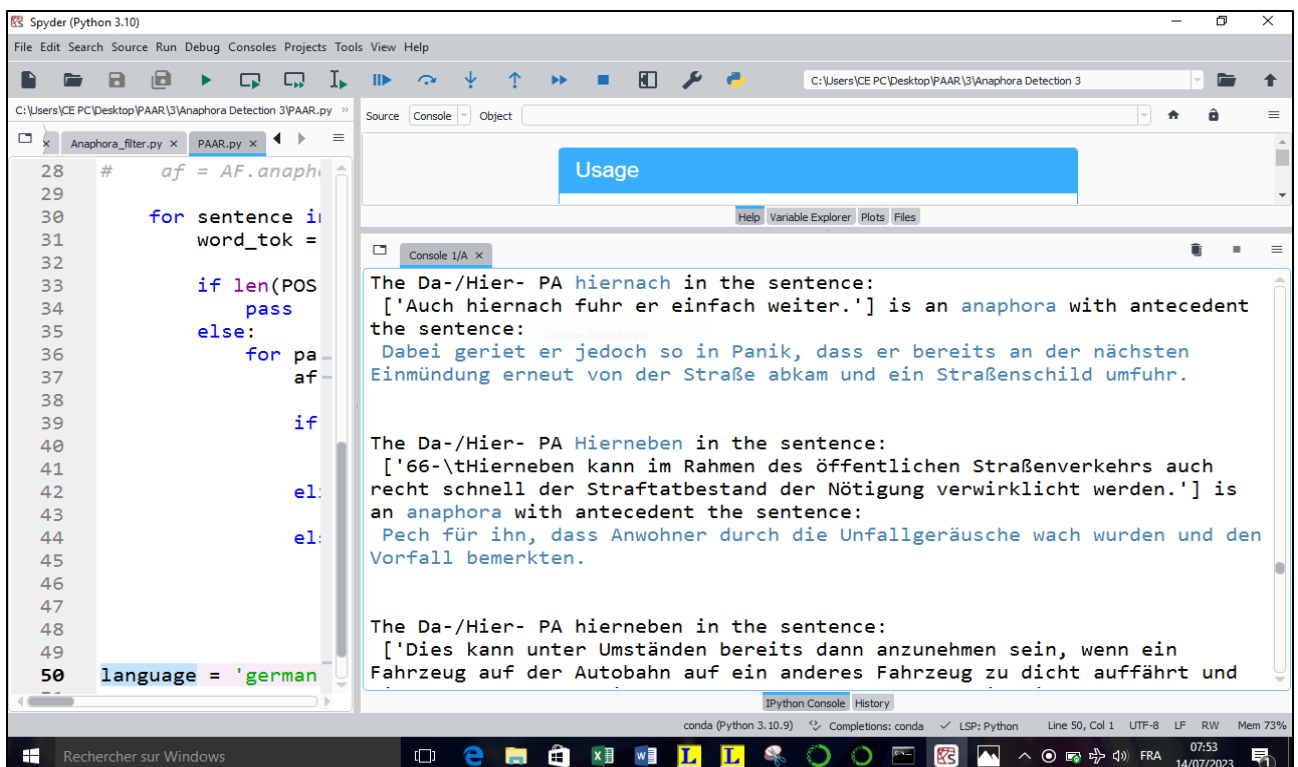


Abbildung 94: Einsatz des PAAR auf Texte N°51-75 (13)

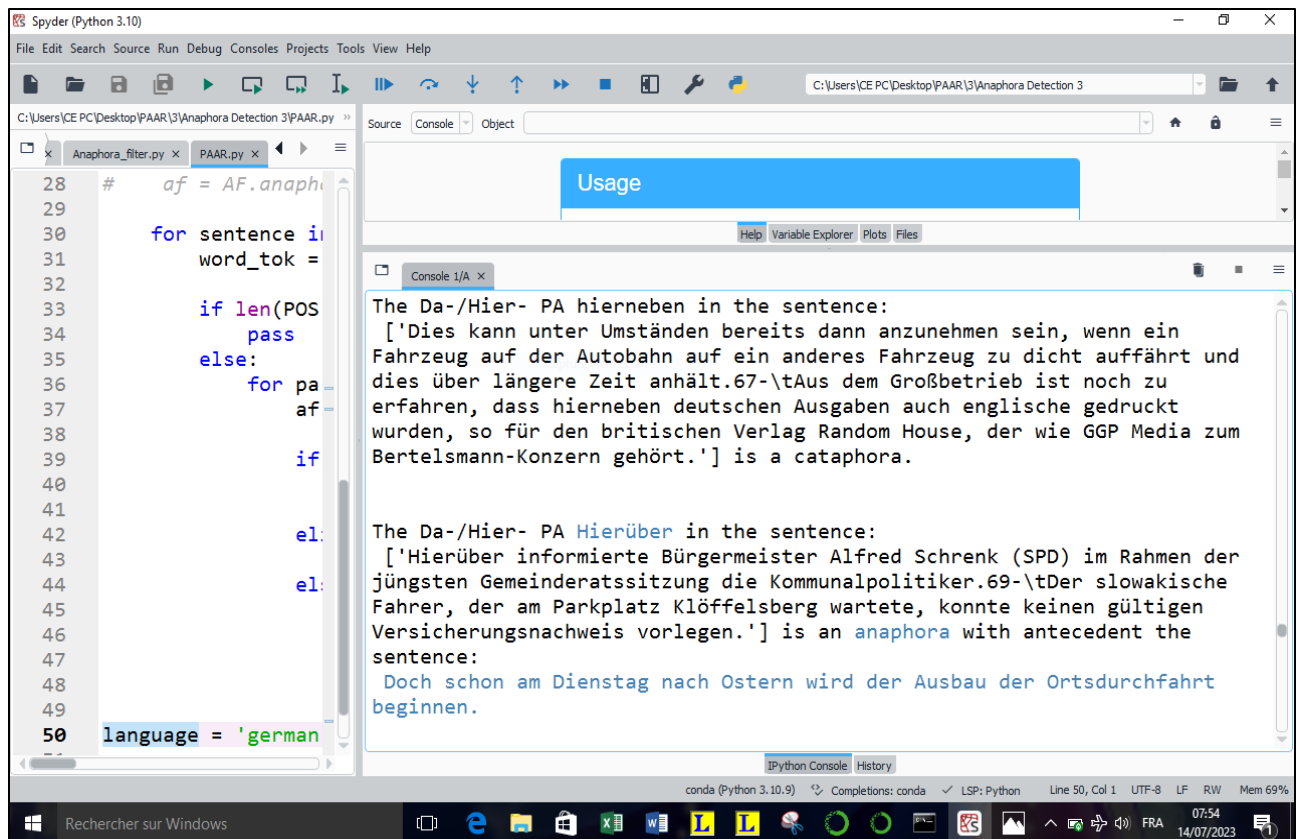


Abbildung 95: Einsatz des PAAR auf Texte N°51-75 (14)

Das PA „Hierneben“ aus der Abbildung 88 ist zwar keine Katapher aber eben auch keine Anapher. Allerdings weisen wir noch darauf hin, dass uns nur die Natur „ANAPHER“ interessiert. In dieser Hinsicht werden wir solche Vorfälle nicht als kritisch betrachten, und zwar, weil es sich um kein anaphorisches PA handelt. Das PA „Hierüber“ ist anaphorisch und wurde gut aufgelöst. Allerdings merkt man einige Abweichungen beim Parsen des Textes.

Das PA „Hierum“ in der Abbildung 89 ist anaphorisch und wurde vom PAAR richtig aufgelöst. Was das zweite PA („davon“) der Abbildung angeht, ist es zwar anaphorisch, aber wurde falsch aufgelöst.

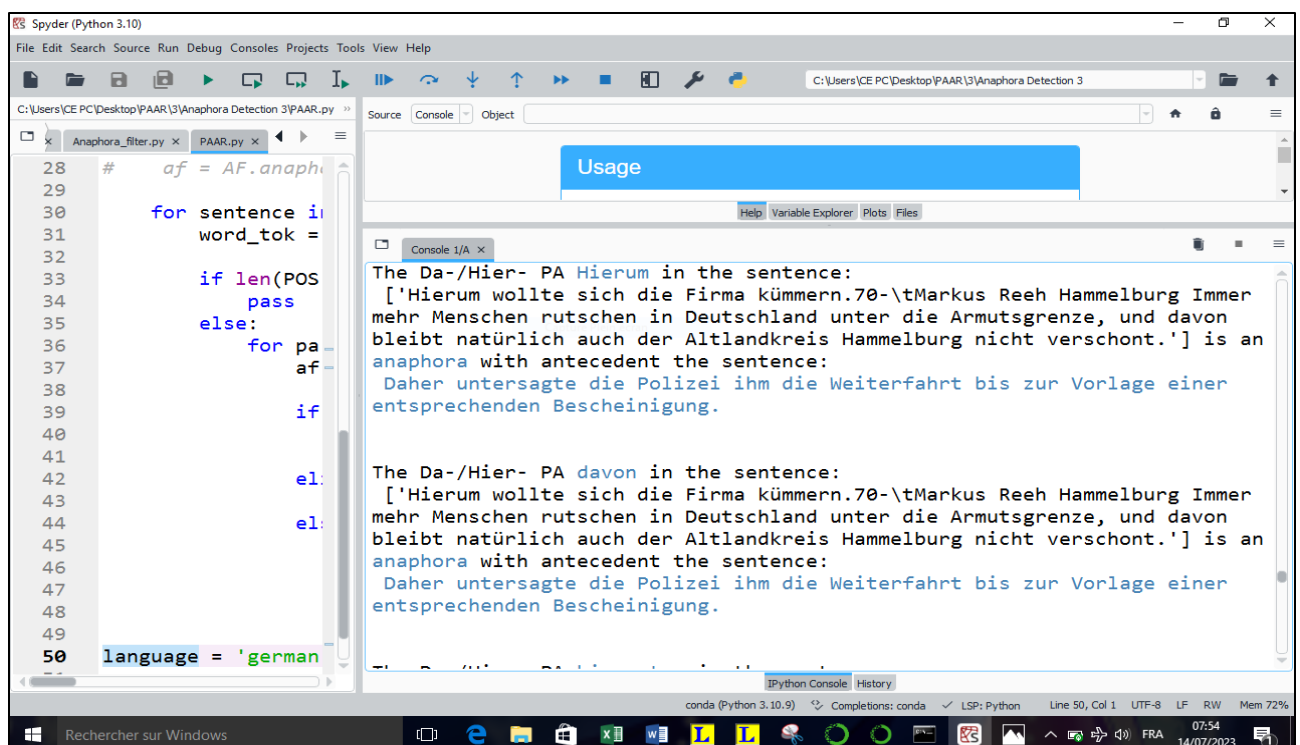


Abbildung 96: Einsatz des PAAR auf Texte N°51-75 (15)

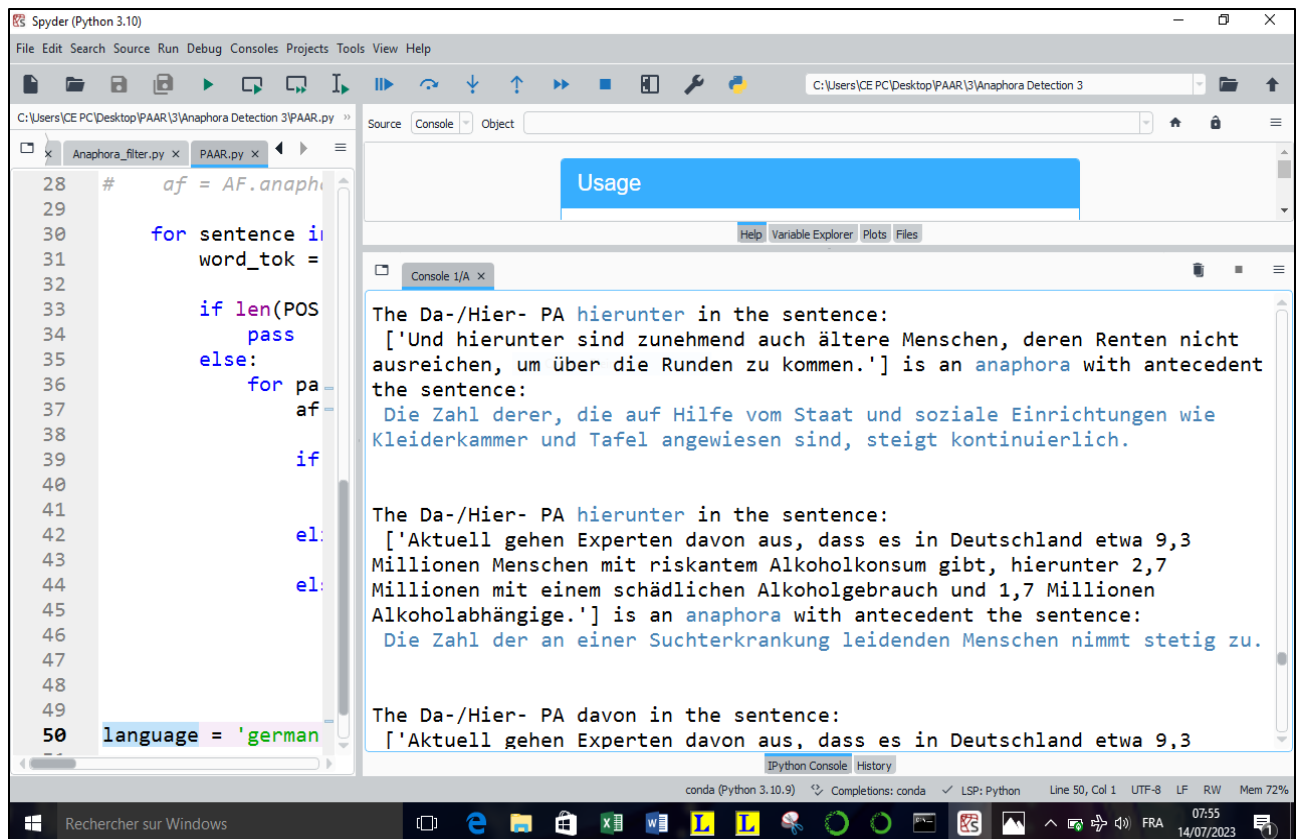


Abbildung 97: Einsatz des PAAR auf Texte N°51-75 (16)

Das PA „Hierunter“ erscheint zwei Mal in der Abbildung 90. PAAR hat die erste Erscheinung richtig aufgelöst. Was die zweite aber angeht, handelt es sich zwar um eine Anapher, aber mit dem Antezedens „9,3 Millionen Menschen mit riskantem Alkoholkonsum“.

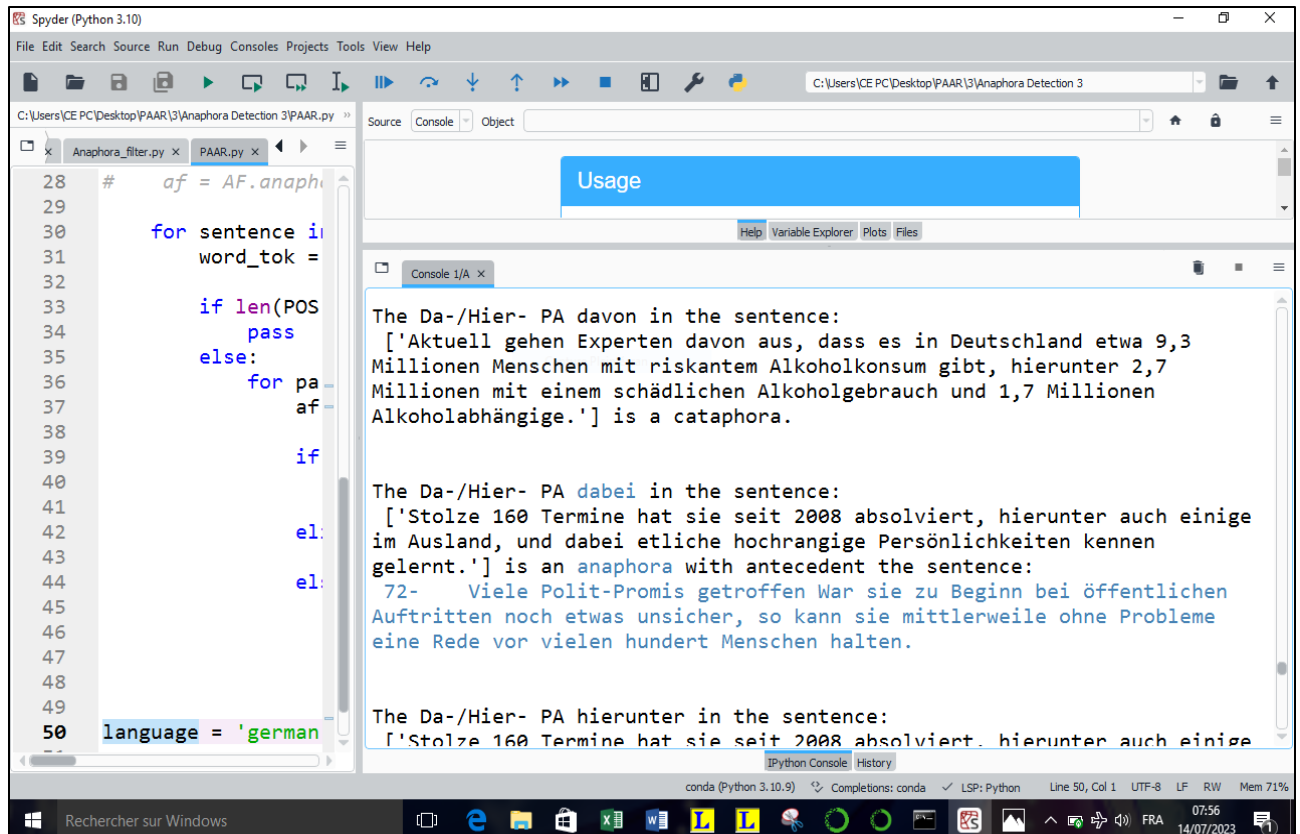


Abbildung 98: Einsatz des PAAR auf Texte N°51-75 (17)

Das PA „davon“ in der Abbildung 91 ist tatsächlich eine Katapher. Was das PA „dabei“ angeht, handelt es sich zwar um eine Anapher aber mit der NP „160 Termine“ als Antezedens.

Das PA „Hierunter“ in der Abbildung 92 ist keine Katapher, sondern eine Anapher, denn es bezieht sich, eben wie „dabei“ auf die NP „160 Termine“. Was das PA „Hiervon“ angeht, wurde es nur zum Teil gut aufgelöst. Denn das passende Antezedens ist der Satzteil „mehr als 700 Euro“. Das PA „davon“ wurde dagegen richtig aufgelöst.

In der Abbildung 93 wurde das PA „Hierzu“ falsch aufgelöst. Das passende Antezedens ist jedoch „Die jährliche Mitglieder-versammlung des Feuerwehrvereins der Wehr Aura findet am kommenden Donnerstag, 6. Januar, in der Festhalle statt“. Was das PA „Danach“ angeht, wurde es vom PAAR richtig aufgelöst. Denn „Danach“ nimmt Bezug auf die Handlung, die vom Verb durchgeführt beschrieben wurde.

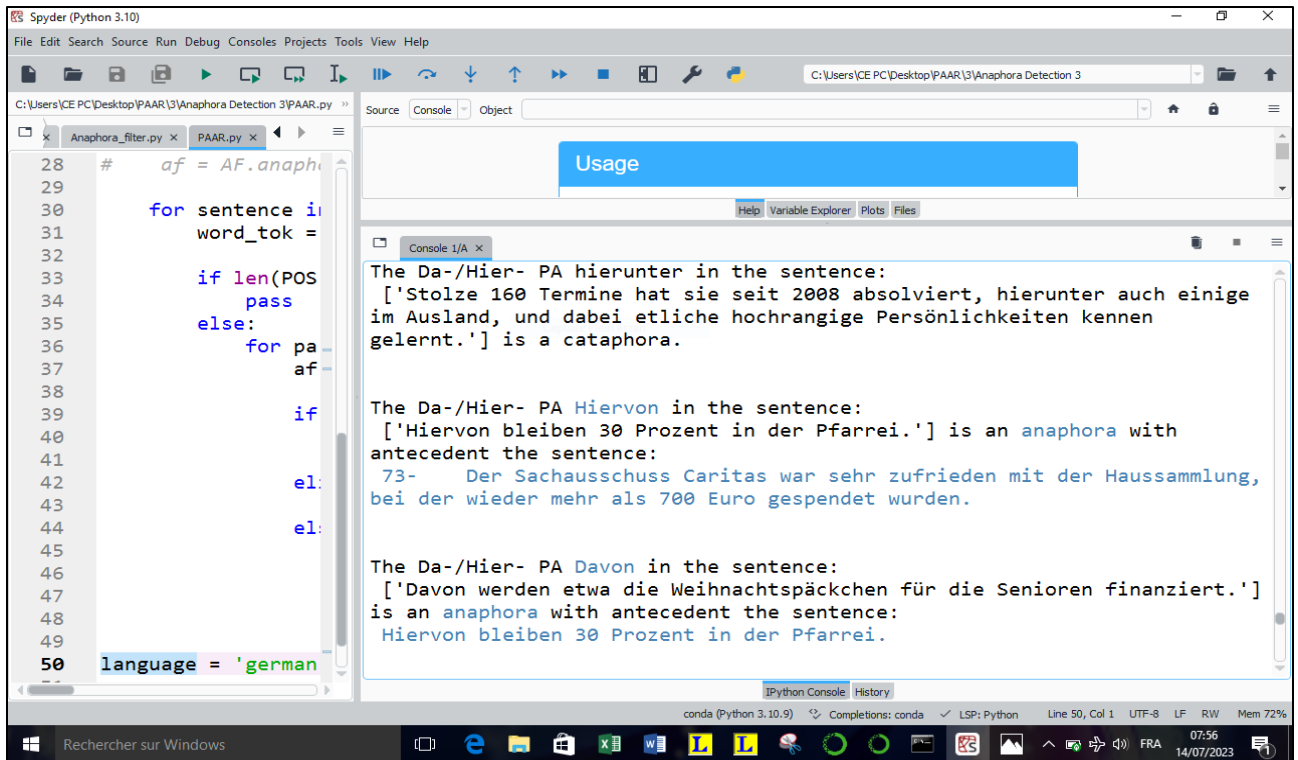


Abbildung 99: Einsatz des PAAR auf Texte N°51-75 (18)

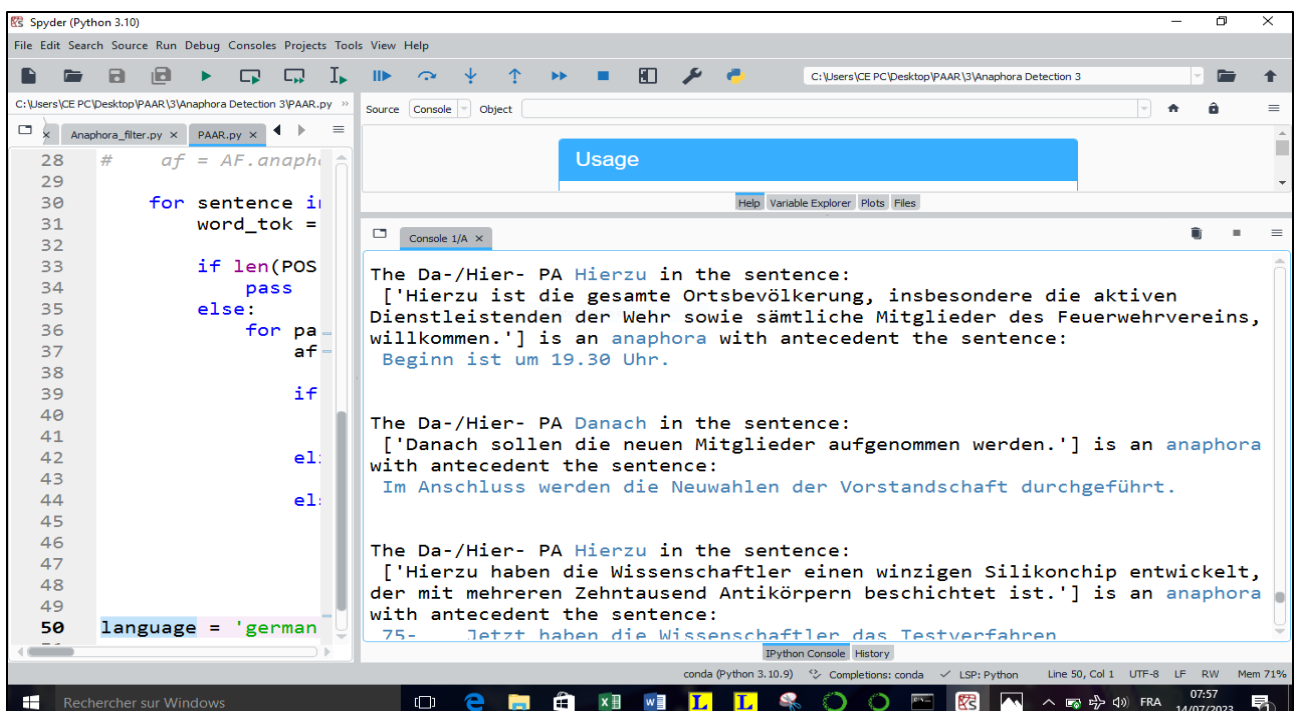


Abbildung 100: Einsatz des PAAR auf Texte N°51-75 (19)

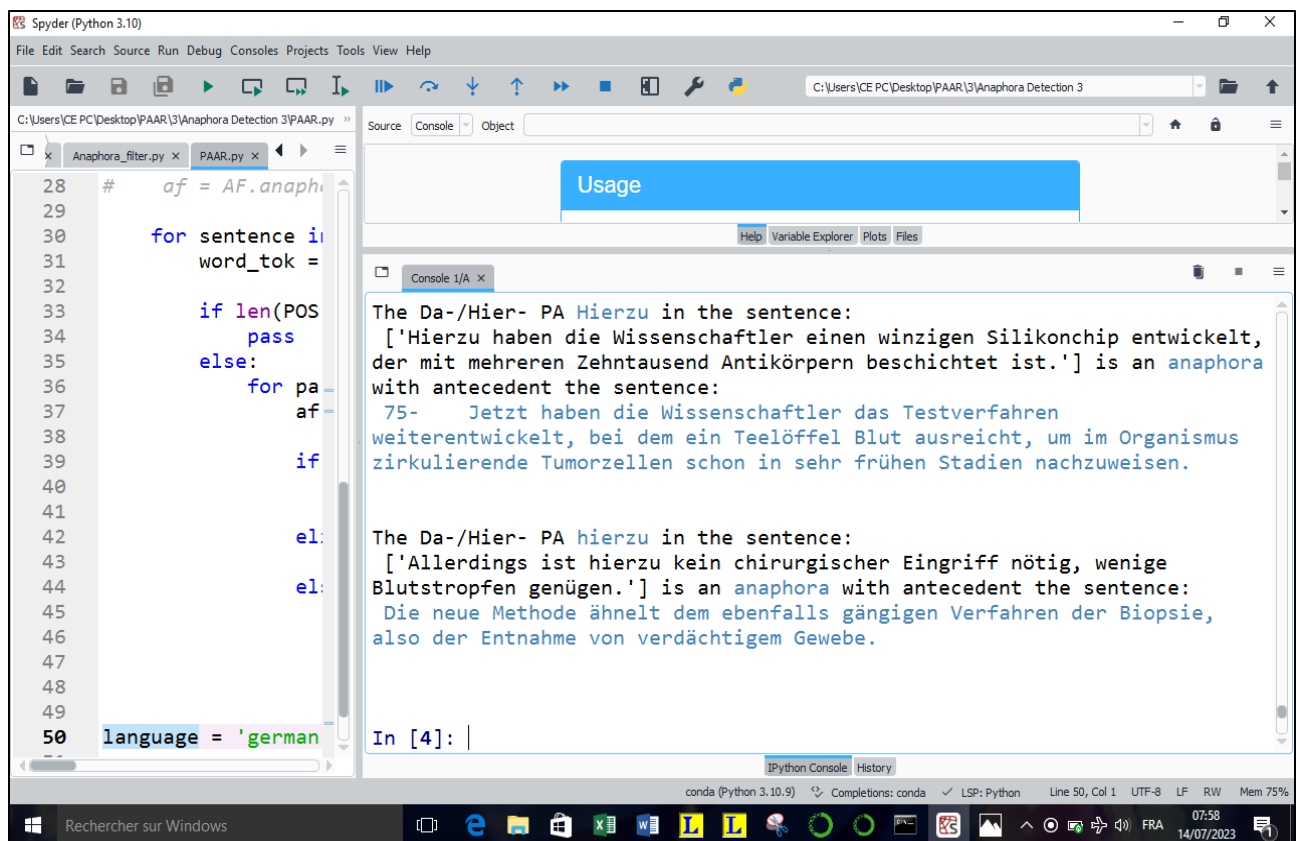


Abbildung 101: Einsatz des PAAR auf den Textabschnitt (N°51-75) (20)

In der Abbildung 94 steht es nämlich fest, dass PAAR zwei Erscheinungen von „Hierzu“ aufgefunden und sie als anaphorische PA klassifiziert hat. Etwas, das eigentlich nach Näherbetrachtung des Textes von einer menschlichen Instanz (dem Linguisten) auch nachvollziehbar ist. Weiterhin stellen wir fest, dass die Auflösung des ersten PA „Hierzu“ ihm (PAAR) gelungen ist. Was die Auflösung des zweiten aber angeht, ist das passende Antezedens nur die NP „Die neue Methode“ und nicht die Handlung, die durch das Verb „ähnelt“ realisiert wurde, wie PAAR es gemeint hat.

Darunter wird die Performanz vom PAAR bezüglich 3 Kriterien, wie etwa, die Erkennung von PA, deren jeweiligen Natur sowie auch deren Auflösung, wenn es sich um anaphorische Erscheinungen angeht, zusammengefasst.

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
40	42 (40PA+2Subjunktionen)	42/40 ??

Tabelle 67: Performanz vom PAAR bei Erkennung von PA im Text N°51-75

PA	Anzahl	Natur	Ergebnisse
Dazu	1	Katapher	0/1
Hierdurch	3	Anapher	3/3
davon	3	Anapher	3/3
Wodurch	1	Anapher	1/1
Hierein	3	Anapher	3/3
darauf	2	Anapher	2/2
davon	2	Katapher	2/2
Hierfür	2	Anapher	2/2
Hiergegen	1	Anapher	1/1
Hierhinter	1	Anapher	1/1
Hierin	2	Anapher	2/2

Damit	1	Katapher	0/1
Darüber	1	Anapher	1/1
Hiermit	1	Anapher	1/1
Dabei	2	Anapher	2/2
Hiernach	2	Anapher	2/2
Hierneben	1	Anapher	0/1
Hierneben	1	Katapher	0/1
Hierüber	1	Anapher	1/1
Hierum	1	Anapher	1/1
Hierunter	2	Anapher	2/2
Hierunter	1	Katapher	0/1
Hiervon	1	Anapher	1/1
Hierzu	3	Anapher	3/3
Danach	1	Anapher	1/1
TOTAL	40		35/40

Tabelle 68: Performanz vom PAAR bei Erkennung der Natur von PA im Text N°51-75

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Hierdurch	2/3
Davon	2/3
Wodurch	1/1
Hierein	2/3
Darauf	1/2
Hierfür	1,5/2
Hiergegen	1/1
Hierhinter	1/1
Hierin	0,5/2
Darüber	1/1
Hiermit	1/1
Hiernach	1,5/2
Dabei	1/2
Hierüber	1/1
Hierum	1/1
Hierunter	1/2
Hiervon	0,5/1
Hierzu	1/3
Danach	1/1
TOTAL	22/33

Tabelle 69: Performanz von PAAR bei der Auflösung von anaphorischen PA im Text N°51-75

4. Einsatz des PAAR auf den Textabschnitt (N°76-100)

3. ***Damit** nun das Inventar - Tische, Stühle, Schränke - angeschafft werden kann, musste zunächst die Genehmigung der Mitgliederversammlung eingeholt werden. Sie erteilte den Auftrag und gewährte ebenso einstimmig die Entlastung des Vorstands, die Kassenprüfer Albin Keßler und Joachim Voll erbeten hatten. Grundlage **hierzu** bildete der ausführliche Kassenbericht von Kassierer Christian Rost, der eine einwandfreie Bilanz vorlegen konnte. (SAZ11/JAN.00735 Saale-Zeitung, 10.01.2011, S. 5; Klein, aber unverzichtbar).*
4. *Wie der Fahrer eines Suzuki Vitara angab, befuhr er am Donnerstag zur Mittagszeit die Schlimpfhofer Straße und wollte an der Kissinger Straße nach rechts abbiegen. Eine 18-jährige Peugeot-Fahrerin, die auf der Kissinger Straße unterwegs war, wollte ihr Fahrzeug wenden und fuhr **hierzu** rückwärts in die Schlimpfhofer Straße ein. **Dabei** übersah die 18-jährige den herannahenden Suzuki, und es kam zum Zusammenstoß. Es entstand Sachschaden an den Fahrzeugen in Höhe von circa 1000 Euro. (SAZ11/JAN.00746 Saale-Zeitung, 10.01.2011, S. 7; ohne Titel)*
5. *Vorsitzender Leo Meder berichtete von über 30 Veranstaltungen. **Hierzu** gehörten Kameradschaftsabende, Wanderungen, kirchliche und gesellige Veranstaltungen, **wobei** sich die*

- Wehr um die Verkehrsregelung kümmerte oder mit Abordnungen beteiligte, Aktionen zum Ferienprogramm und Festbesuche bei benachbarten Vereinen. Für den Kameradschaftsabend am 22. Januar um 19 Uhr bat Leo Meder um Anmeldungen. (SAZ11/JAN.00748 Saale-Zeitung, 10.01.2011, S. 8; Viel Lob für engagierte Floriansjünger)
6. **Danach** landete Lepper als freier Mitarbeiter in Ammanns Museum für Moderne Kunst und der nicht minder bekannten Kunsthalle Schirn. Bis zu sieben Führungen leitete Lepper täglich, fühlte sich aber etwas zerrissen zwischen Vermittlungsfreude und "Kulturindustrie". In jener Zeit gab er auch ein Comeback an der Gießener Uni - als Dozent an den Instituten für Didaktik der politischen Bildung und Kunstgeschichte. **Hierzwischen** gebe es viel mehr Verbindung, als mancher denken würde, begründet Lepper die auf den ersten Blick ungewöhnliche Kombination. (GIZ17/FEB.00813 Gießener Anzeiger, 07.02.2017, S. 12; Ein anpackender Sozialromantiker...)
 7. Angeblich mault der Franke gern, er macht gerne nieder. Am liebsten sich selbst und das, was er hat. Der Kissinger mault über die Kur. "Die Kur ist tot", lautet ein oft gehörter Satz, **wobei** offenbleibt, was da mit Kur gemeint ist. Die private Vorsorge, die Stationäre Rehabilitation oder die Anschlussheilbehandlung? Schaut man auf die Statistik, hat die Übernachtungszahl bei den Versicherungsgästen sogar leicht zugenommen. Es kommt **darauf** an, was der Standort Bad Kissingen aus seiner Kompetenz macht. Zum Beispiel die Klinik am Kurpark mit ihrer Kira-Studie. Die "Kur" ist nicht tot. (SAZ11/JAN.00062 Saale-Zeitung, 03.01.2011, S. 1; Die Kur ist)
 8. Nach drei Siegen und einem Remis gegen den TSV Nordheim standen die Waldberger schon vor ihrem letzten Gruppenspiel als Halbfinalteilnehmer fest, doch die DJK peilte den Gruppensieg an, um in der Runde der letzten Vier dem SV Rödelmaier aus dem Weg zu gehen. Den nötigen Punkt **hierfür** holte Waldberg gegen den Kreisklassisten VfR Bischofsheim. Der musste — wie auch der TSV Uns leben — Lehrgeld zahlen. Unsleben holte den einzigen Sieg gegen den TSV Nordheim — und vermasselte dem **damit** den Einzug ins Halbfinale, **wobei** Benjamin Schirber mit der Schlusssirene einen Freistoß zum 2:1 ins Netz hämmerte. Nordheim hätte im letzten Spiel gegen Hohenroth einen Dreier benötigt, lag auch aussichtsreich 2:0 vorne, doch der FSV konterte, schaffte zehn Sekunden vor Ende durch Ralf Schwenkert das 2:2 — **damit** waren die Hohenröther aufgrund der besseren Tordifferenz — weiter und Nordheim draußen. Der FC Ober-/Mittelstreu wachte zu spät auf: Nach zwei Niederlagen holte die Truppe von Andreas Umla zwar noch sieben Punkte, doch mehr als der vierte Gruppenplatz war nicht **drin**. (SAZ11/JAN.00165 Saale-Zeitung, 04.01.2011, S. 19; Der Wahnsinn vom Punkt)
 9. In Gruppe B ließ der Bezirksoberligist SV Rödelmaier keinen Zweifel **daran**, dass der Weg zum Gruppensieg nur über ihn führen würde: Fünf Siege in fünf Spielen sprechen eine deutliche Sprache. **Dahinter** wurden dem FC Schönauf-Wegfurt die größten Chancen eingeräumt — und auch der Kreisligist wurde den Erwartungen gerecht. Er stand schon vor dem letzten Gruppenspiel — dank des Erfolgs im direkten Vergleich gegen den FC Hendungen — als Halbfinalteilnehmer fest. Die A-Klassisten SG Unterweißenbrunn/Frankenheim, TSV Stockheim und TSV Ober-Unterwaldbehrungen machten die Platzierungen in der zweiten Tabellenhälfte unter sich aus. Mit 18 Gegentreffern stellte der TSV Ober-Unterwaldbehrungen die anfälligste Hintermannschaft, **wobei** Aushilfsstorwart Steffen Malzer (BOL-Schiedsrichter) die geringste Schuld traf. (SAZ11/JAN.00165 Saale-Zeitung, 04.01.2011, S. 19; Der Wahnsinn vom Punkt)
 10. **Dazwischen** sang er ausgewählte, bekannte und weniger bekannte Lieder aus Dowlands "Books of Songs or Ayres" wie "Come away, sweet love", "Sorrow stay", "I saw my lady weep", "What poor astronomers!" oder, natürlich, "Fine knacks for ladies", aber auch einige Lieder von Claudio Saracini, Giulio Caccini, Claudio Monteverdi oder Pierre Guédron. **Dazu** gab's ein paar kleinere Instrumentalstücke, **wobei** John Bulls "In Nomine" für Cembalo solo — im Elfvierteltakt geschrieben, aber ein bisschen etüdenhaft einfallslos — aus der emotional geprägten, mal heiteren, mal melancholischen Linie des Abends deutlich herausfiel. (SAZ11/JAN.00597 Saale-Zeitung, 08.01.2011, S. 6; Eine musikalische Expedition, die nie zum Ziel führte)
 11. **Wobei** er die Anforderungen, die allein schon die Textmenge an die einzige Sängerin stellt — und auch an das Publikum — **dadurch** bewältigbar macht, dass der Gesangspart zwar durchaus emotional exaltiert geführt ist, dass er aber auf unsinnige Intervallquälereien um ihrer selbst willen verzichtet. Die Solistin hat immer die Möglichkeit, Freiräume für die Gestaltung zu nutzen, nicht von der Technik in den Würgegriff genommen zu werden, erzählend zu singen. (SAZ11/JAN.00598 Saale-Zeitung, 08.01.2011, S. 6; Vom großen Dämon der Rastlosigkeit)
 12. Vorsitzender Leo Meder berichtete von über 30 Veranstaltungen. **Hierzu** gehörten Kameradschaftsabende, Wanderungen, kirchliche und gesellige Veranstaltungen, **wobei** sich die Wehr um die Verkehrsregelung kümmerte oder mit Abordnungen beteiligte, Aktionen zum Ferienprogramm und Festbesuche bei benachbarten Vereinen. Für den Kameradschaftsabend am

22. Januar um 19 Uhr bat Leo Meder um Anmeldungen. (SAZ11/JAN.00748 Saale-Zeitung, 10.01.2011, S. 8; Viel Lob für engagierte Floriansjünger)
13. Bevor Landrat Thomas Bold auf das Thema "Generationenfreundlicher Landkreis" einging, streifte er zunächst die Kreistagspolitik. **Dabei** betonte Bold, dass 2010 auch im Landkreis ein wirtschaftlicher Aufschwung erkennbar sei, **wobei** aufgrund der besonderen Stellenwerte im Fremdenverkehr, der Kur und des Baugewerbes, die Beschäftigungszahlen schwanken. Dass die Krise aus 2009 überwunden sei, zeige auch, dass bei den Schulabgängern eine hervorragende Ausbildungsquote zu verzeichnen sei, die schon jahrelange nicht mehr so gut war. (SAZ11/JAN.00756 Saale-Zeitung, 10.01.2011, S. 12; Gemeinde würdigt die über 70-Jährigen)
 14. "Versteigerungen drohen" Doch **davon** sei der Arbeitsmarkt meilenweit entfernt. Es gebe immer mehr Menschen, die nicht angemessen bezahlt würden, zum Beispiel Verkäuferinnen oder Friseurinnen, und trotz Arbeit nicht genug Geld zum Leben hätten. Sie zahlten hierdurch auch nur geringe Beträge in die Rentenversicherung ein, **wodurch** ihnen auch im Alter Armut drohe. (SAZ11/JAN.00611 Saale-Zeitung, 08.01.2011, S. 9; Zunahme der Armut hat viele Gründe)
 15. Drei Verfahren möglich Schulte stellte drei mögliche Verfahren vor. Das erste ist die Selbstveranlagung der Bürger. **Hierbei** geben die Grundstücksbesitzer anhand eines Fragebogens die erforderlichen Angaben an die Abrechnungsstelle. Nachteil dieses Verfahrens sei die Überforderung mancher Betroffener und die Möglichkeit zu Falschangaben, **wodurch** ein erhöhter Beratungs- und Kontrollaufwand entstehe. (SAZ11/JAN.03359 Saale-Zeitung, 31.01.2011, S. 9; Abwassergebühren bleiben Thema in Ramsthal)
 16. Der Vermieter tauschte daraufhin die Schließanlage in dem Wohnhaus aus. Von dem Mieter verlangte er Schadensersatz. Ohne Erfolg: Ein Mieter sei zwar verpflichtet, alles zu unterlassen, **wodurch** an dem Mietobjekt Schaden entstehen könne, befanden die Richter. **Hierzu** zähle auch die sichere Aufbewahrung der Haus- und Wohnungsschlüssel. Der Mieter habe aber mit der Nutzung des Wertfachs seiner Verpflichtung Rechnung getragen. dpa (SAZ11/FEB.00750 Saale-Zeitung, 05.02.2011, S. 50; Schlüssel sind im Wertfach sicher genug)
 17. Bislang verwendete Knauf Gips aus Rauchgas-Entschwefelungsanlagen: Der fällt bei Kraftwerken an, wenn die Abgase entschwefelt werden. **Wofür** die Natur Million Jahre braucht, geschieht dann in wenigen Stunden. Doch immer mehr Kohlekraftwerke gehen vom Netz und weniger Abgase bedeuten weniger Gips. Knauf setzt Naturgips zu und weil die Iphöfer langfristig planen sucht der Bohrtrupp neue Vorkommen. (SAZ11/JAN.00310 Saale-Zeitung, 05.01.2011, S. 8; Auf)
 18. Während andernorts die Veranstalter mit dem Blitzeis und den **damit** verbundenen spiegelglatten Straßen haderten und Turniere absagen mussten, hatten sich im Römershager Schul- und Sportzentrum tatsächlich alle 16 Mannschaften eingefunden — und das trotz des frühen Turnierbeginns um 9.30 Uhr. "Ein Kompliment an die Mannschaften, die die Anreise gewagt haben. Das spricht auch für unser Turnier, wenn die Leute trotz solcher Bedingungen zu uns kommen", so Turnier-Organisator Jürgen Breitenbach vom ausrichtenden SV Römers hag über die treue Kundschaft. Nicht nur das Glatteis hatte den Veranstaltern im Vorfeld Kopfzerbrechen bereitet. "Die Turnhalle war ja aufgrund der Schneemassen zuvor gesperrt gewesen. Aber die Stadt hat das Hallendach dann räumen lassen, **wofür** wir uns an dieser Stelle sehr herzlich bedanken möchten", freute sich Breitenbach über den Sportsgeist der Entscheidungsträger. (SAZ11/JAN.00642 Saale-Zeitung, 08.01.2011, S. 20; Ein spezieller Treue-Beweis)
 19. IP-Adresse - **wofür** steht das eigentlich? IP (gesprochen: Ei Pi) steht für "Internet Protocol". Auf Deutsch: Internet-Protokoll. Das alte System aus Ziffern und Punkten wurde IPv4 abgekürzt. Das bedeutet Internet-Protokoll, Version 4. Eine Adresse kann so aussehen: 217.79.215.248. (SAZ11/FEB.00722 Saale-Zeitung, 05.02.2011, S. 34; ohne Titel)
 20. La Niña (spanisch für "das Mädchen") ist ein Wetterereignis, das häufig im Anschluss an El-Niño auftritt. Es ist eine außergewöhnlich kalte Strömung im äquatorialen Pazifik, sozusagen ein Anti-El-Niño. Durch die kalte Strömung entwickelt sich über Indonesien bei starken Passatwinden ein starkes Tiefdruckgebiet. **Dadurch** kühlt sich der östliche Pazifik weiter ab und es gibt in Indonesien viel Regen, **wogegen** in Peru kaum Regen fällt. Quelle: Wikipedia (SAZ11/JAN.01837 Saale-Zeitung, 19.01.2011, S. 34; Wettermacher)
 21. Immer wieder greift er mit skurriler Logik auf seine Protagonisten zurück: Mal stirbt Dörtes Katze eines nicht natürlichen Todes, **wobei** er die Ursache mit einem dezenten "Bobom — bobom" des Autos andeutete, mal lässt er Dörte und "dat Uschi" sich sieben Kilometer von der geplanten Windpark-Standorts an Bäume anketten, um den selten "Nußbaum-Zutzler" zu schützen, wobei die Hilferufe vom Landwirt auf dem Gülle verwirbelnden Bulldog nicht gehört werden. Dazwischen ein wenig Politik, wobei er mit dem "Hand-Dreieck" die Kanzlerin nur andeuten musste, **wogegen** "Plagiator" Gutenberg wenigstens einmal durch den Kakao gezogen wurde, und Lafontaine als "geiler Oscar

- mit dem heißesten Schnittchen des Kommunismus" für Heiterkeit sorgte. (SAZ12/MAR.00422 Saale-Zeitung, 05.03.2012, S. 6; "Dreggsagg")
22. Die Spannung steigerte sich, als die bisherige Rekordhöhe von 1,70 Meter von verbleibenden sechs Reiterinnen des Vereins übersprungen werden konnte. Da die vorhandene Mauer bislang lediglich bis 1,70 Meter gebaut werden konnte, hatte man entsprechende Vorsorge zum Höherbau getroffen. Bei der Mauerhöhe von 1,80 Metern, **wohinter** Parcourchef Klaus Conrad und seine Helfer nicht mehr wahrzunehmen waren, wuchs beim sachkundigen Publikum der Respekt vor dem Mut der im Rennen verbleibenden Damen, dieses Bauwerk mit ihren Pferden bewältigen zu wollen. (SBZ03/JAN.02430 Saarbrücker Zeitung, 08.01.2003; Die Jüngste sprang am höchsten)
 23. Müller, Ehrenpräsident der Narrengilde, kam als Sultan von Marrakesch, als der Herrscher im Orient auf die Bretter, die auch für Narren die Welt bedeuten. Er erzählte von Haremsdamen und Eunuchen und auch **davon**, wie er die zwischenmenschlichen Positionen in seinem orientalischen Palast umfunktionierte. Rudi Walz erntete tosenden Applaus in seiner Rolle als "Köllertaler Michel." Außer Rand und Band gerieten die Damen bei den Vorführungen des "Oberpolizisten Patrick" der mit einem Strip doch schon ziemlich gewagt zur Sache ging. "Die Glocken von Rom" ließen das Männerballett der Karnevalsgesellschaft "Die Eule" Rußhütte erklingen: eine umwerfende Show. Die Gildesänger hatten ihre Auftritte mit bekannten Melodien und leicht veränderten Texten, die Erotic lieferten. An den Zuckerhut der Copacabana entführte der Show-Tanz der Sa-Na-Gi-Dancers, **wohinter** sich die große Garde der Narrengilde verbirgt. Heribert Wagner und Willi Wagner übten sich als erotisches Gesangsduo, während Sarah Lee Bungert als "Es Quickie" letztes Eis schmelzen ließ. (SBZ03/FEB.06572 Saarbrücker Zeitung, 17.02.2003; Bei der Premiere knisterte es vor Erotik)
 24. Im Rund gestellte Tische und Stühle, ein allzeit bereiter Kühltisch, an dem sich jeder nach Lust und Durst bedienen darf, eine überdimensionale Tigerente in einer Ecke: nur wenig erinnert **daran**, dass man es hier mit einer Fahrschule zu tun hat. **Wohinter** durchaus System steckt: „Bei unseren Stunden geht es oft lustig zu. Da wird viel gelacht.“ Sturen Frontalunterricht hätte seine in der Regel junge Klientel lange genug gehabt. (SBZ04/AUG.04411 Saarbrücker Zeitung, 12.08.2004; Kein tröger Unterricht: Sven Kirsch betreibt die etwas andere Fahrschule)
 25. Entsprechend hoch durften die Erwartungen des Publikums sein. Und um es gleich vorweg zu sagen: Der Abend wurde zu einer tänzerischen Sensation. **Wobei** man ihn, begründet in den Choreographien, durchaus in zwei Teilen erleben konnte. Denn Mauro Bigonzetti hat sich zwischen 1994, als er erste Verbindungen zu Aterballetto knüpfte, und 2006 - aus dem Jahr stammt die jüngste Choreographie des Abends — nicht unbedingt grundlegend gewandelt, aber seine stilistischen Ausdrucksmittel erweitert. Oder um es anders auszudrücken: Je älter seine Arbeiten sind, desto athletischer sind sie. Je jünger, desto leichter, tänzerischer — **womit** er durchaus im Trend liegt. (SAZ11/JAN.01106 Saale-Zeitung, 13.01.2011, S. 4; Von der betörenden Leichtigkeit der Schwerstarbeit)
 26. In der Frauenklasse verpasste Julia Wenzke vom TSV Münnerstadt nur knapp den Medaillenrang. Sie lief die 400 Meter in 61,95 Sekunden und wurde **damit** Vierte. Im sehr stark besetzten Rennen über 200 Meter lief die Physiotherapeutin 27,41 Sekunden, **womit** sie aber letztlich das Finale verpasste. (SAZ11/JAN.02910 Saale-Zeitung, 27.01.2011, S. 21; Dreisprung fest in Kissinger Hand)
 27. Klaus Angerstein Der Schwarzspecht oder auch Dryocopus martius ist der Online-Enzyklopädie Wikipedia zufolge ein in Mitteleuropa weit verbreiteter Brutvogel, der selbst fragmentierte und kleinflächige Wälder besiedelt. Sogar in den Reichswaldfragmenten in Nähe des Nürnberger Flughafens scheint er sich wohl zu fühlen. Und zwar in möglichst hochstämmigen Bäumen. **Womit** die Probleme beginnen. Die Wipfel solch hochstämmiger Bäume übersteigen oft genug die Grenze dessen, was der Sicherheit auf bundesdeutschen Flughäfen zuträglich ist. 35 Meter hohe Bäume neben der Start- und Landebahn? Schwarzspechte mögen das vielleicht, für einen anderen Vogel — den Airbus beispielsweise — könnte das zu einem Sicherheitsrisiko werden. Weshalb die Vorschriften klar besagen: Die Bäume müssen weg. **Damit** der Airbus sicher starten und landen kann. Auch der Bund Naturschutz hat natürlich viel für Sicherheit übrig. Insbesondere, wenn es um den Schutz des Schwarzspechts geht. Der sei in seinem Bestand europaweit gefährdet. Die übliche Totschlagargumentation also. (SAZ11/JAN.03074 Saale-Zeitung, 28.01.2011, S. 23; Schwarzspecht contra Airbus)

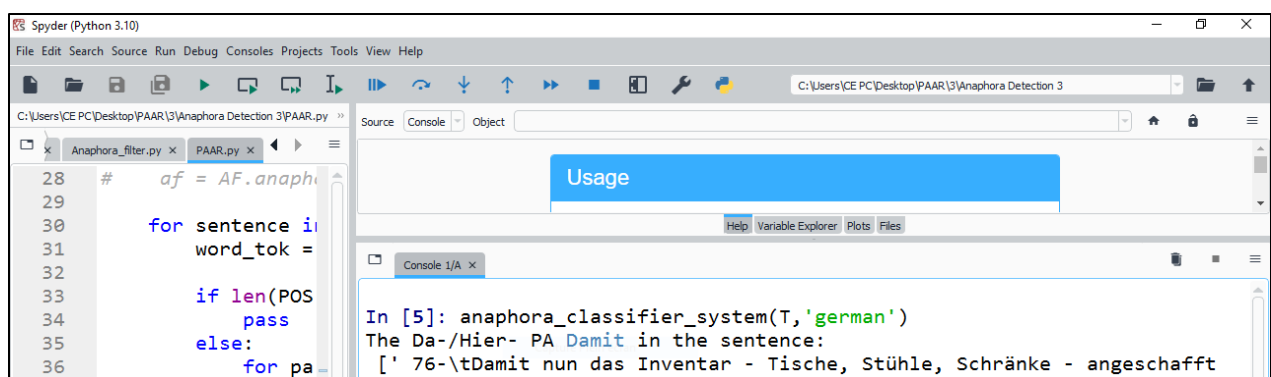


Abbildung 102: Einsatz des PAAR auf den Textabschnitt (N°75-100) (1)

Das Wort „Damit“ hier in der Abbildung 95 ist kein PA, sondern eher eine Subjunktion. Und was das Wort „hierzu“ angeht, handelt es sich zwar um ein PA aber mit dem Antezedens „Die Entlassung des Vorstandschefs“ und nicht der Textteil, den PAAR annotiert hat.

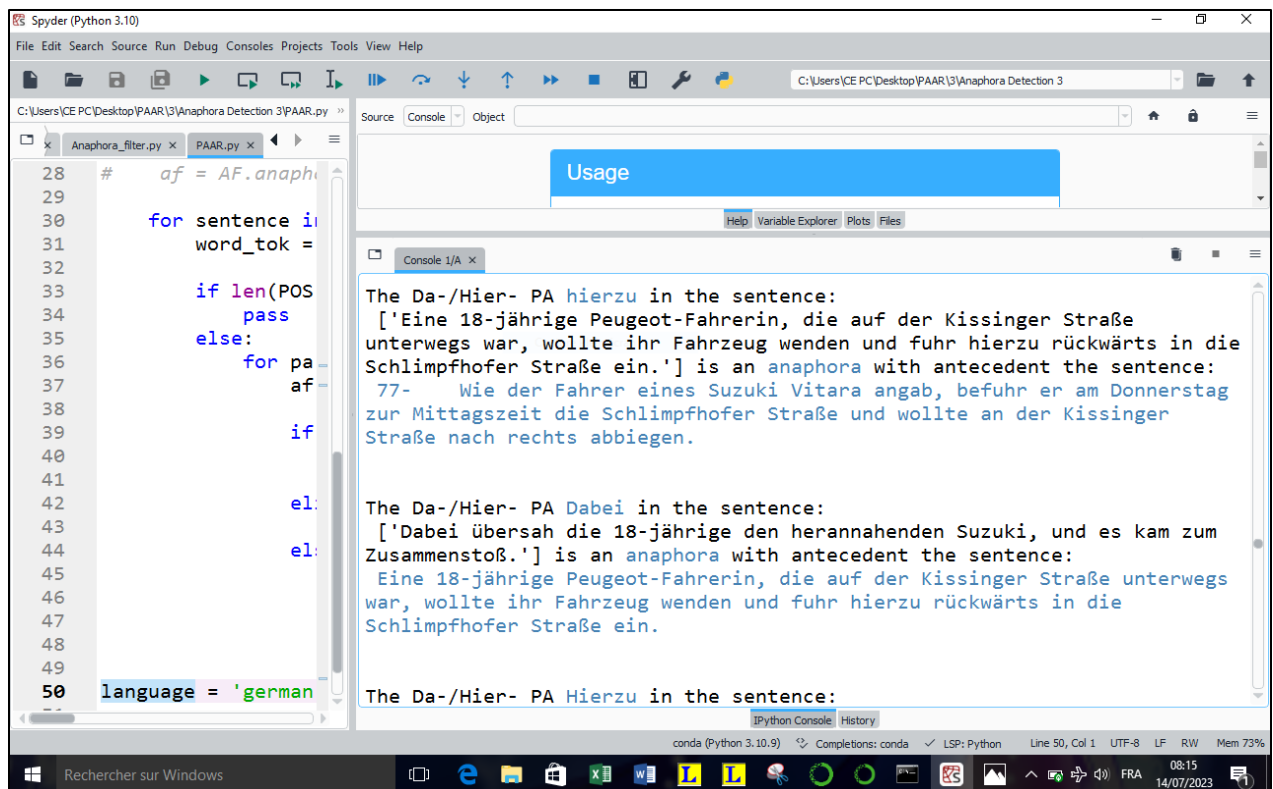


Abbildung 103: Einsatz des PAAR auf den Textabschnitt (N°75-100) (2)

Das PA „Hierzu“ in der Abbildung 96 ist zwar eine Anapher aber mit dem Antezedens „Eine 18-jährige Peugeot-Fahrerin, die auf der Kissinger Straße unterwegs war, wollte ihr Fahrzeug wenden“. Was das PA „Dabei“ angeht, wurde es richtig vom PAAR aufgelöst.

Das PA „Hierzu“ in der Abbildung 97 ist zwar eine Anapher, aber mit der PP („über 30 Veranstaltungen“) Antezedens. Was das PA „wobei“ angeht, wurde es vom PAAR richtig aufgelöst.

Das PA „Danach“ ist zwar eine Anapher, aber man kann es nicht auflösen, denn es steht am Anfang des Textes N°79. Für solche Vorfälle wird PAAR normalerweise das PA tatsächlich als Anapher mit unbestimmbarem

Antezedens verzeichnet. Aber, da wir ihm hier einen ganzen Textabschnitt (N°75-100) zur Verarbeitung geben haben, wird er das PA „danach“ verarbeiten als stünde es nicht am Anfang eines eigenen Textes. Was das PA „Hierzwischen“ angeht, wurde es richtig aufgelöst.

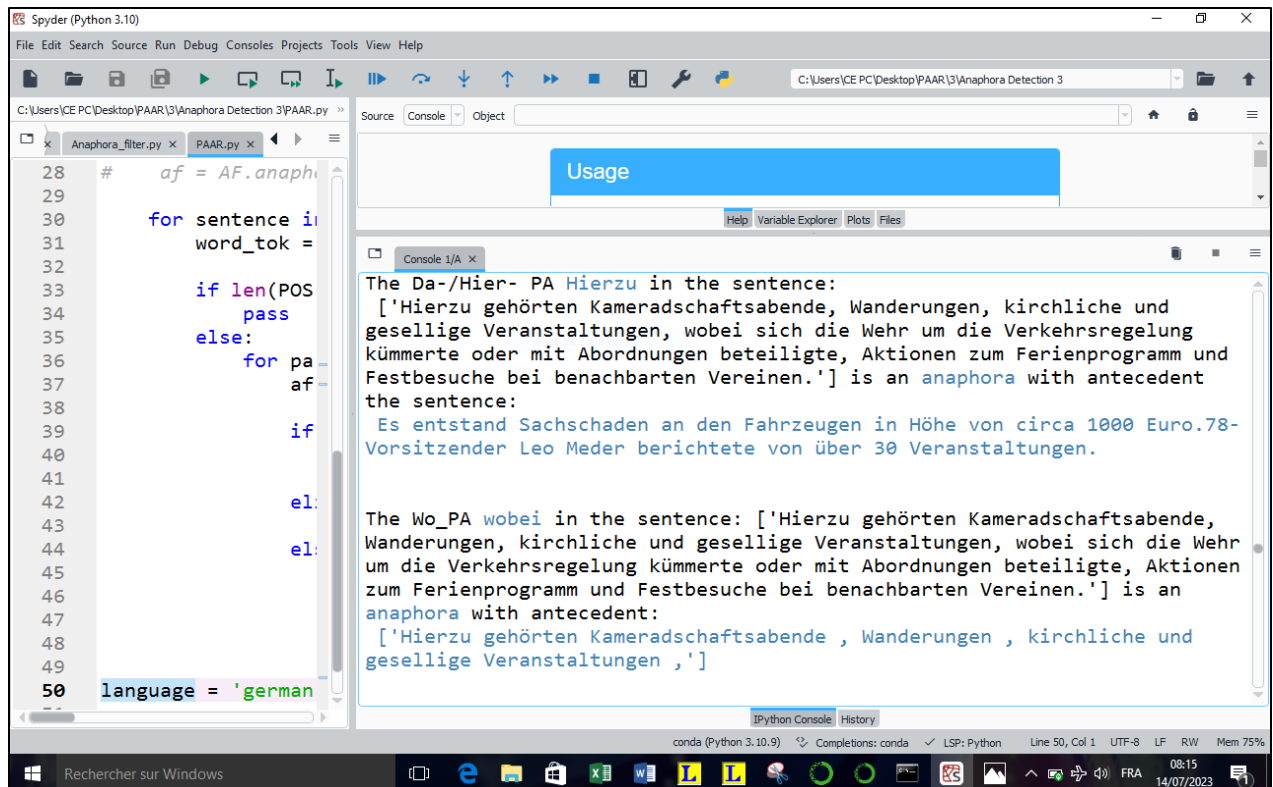


Abbildung 104: Einsatz des PAAR auf den Textabschnitt (N°75-100) (3)

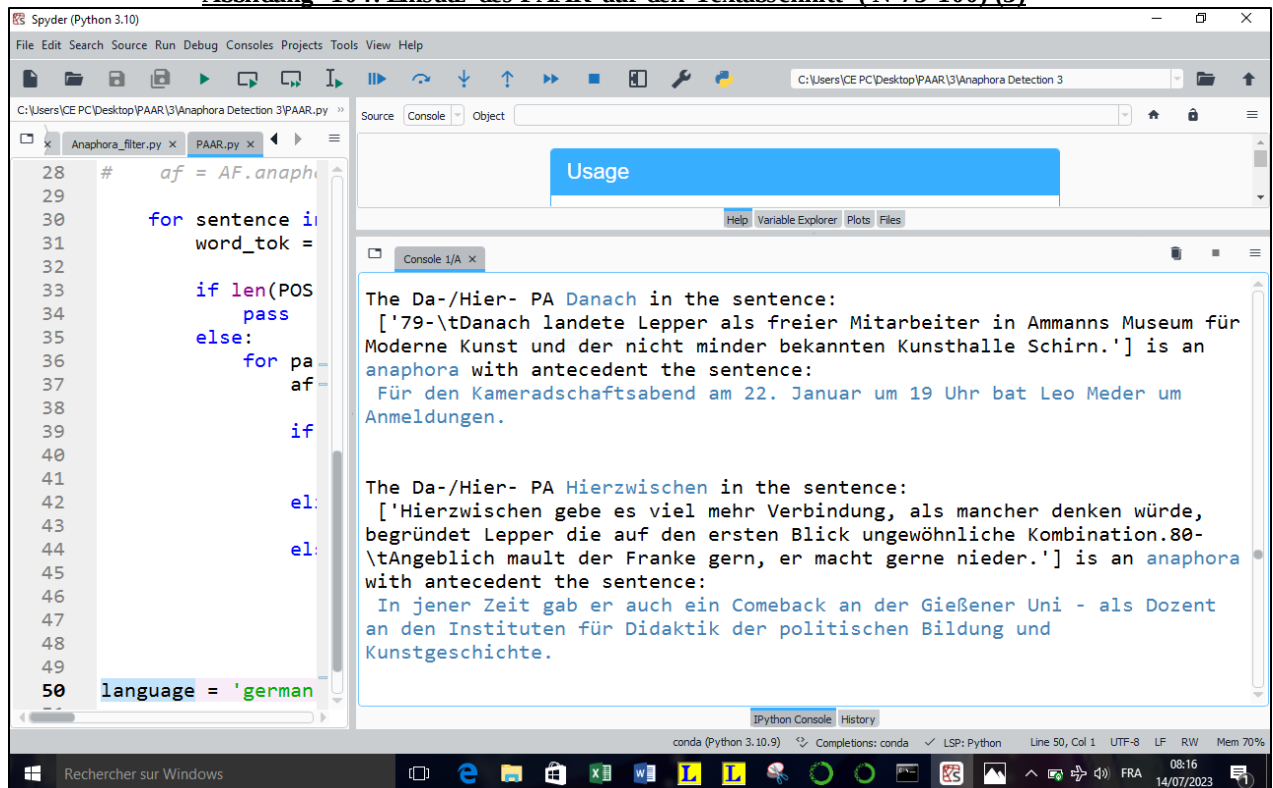


Abbildung 105: Einsatz des PAAR auf den Textabschnitt (N°75-100) (4)

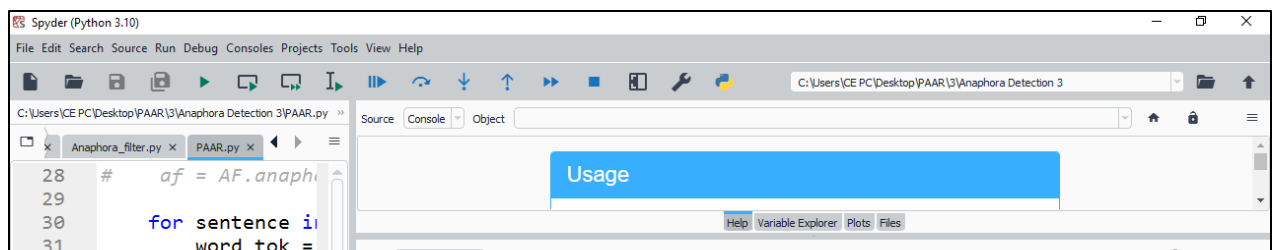


Abbildung 106: Einsatz des PAAR auf den Textabschnitt (N°75-100) (5)

Nach Betrachtung der Angaben aus der Abbildung 99 kann gesagt werden, dass PAAR das PA „wobei“ richtig aufgelöst hat. Im Gegensatz dazu wurde das PA „Hierfür“ leider nicht gut aufgelöst. Denn „Hierfür“ nimmt hier Bezug nur auf den Textteil „*doch die DJK peilte den Gruppensieg an, um in der Runde der letzten Vier dem SV Rödelsmaier aus dem Weg zu gehen*“. Was das PA „darauf“ angeht, handelt es sich tatsächlich um eine Katapher

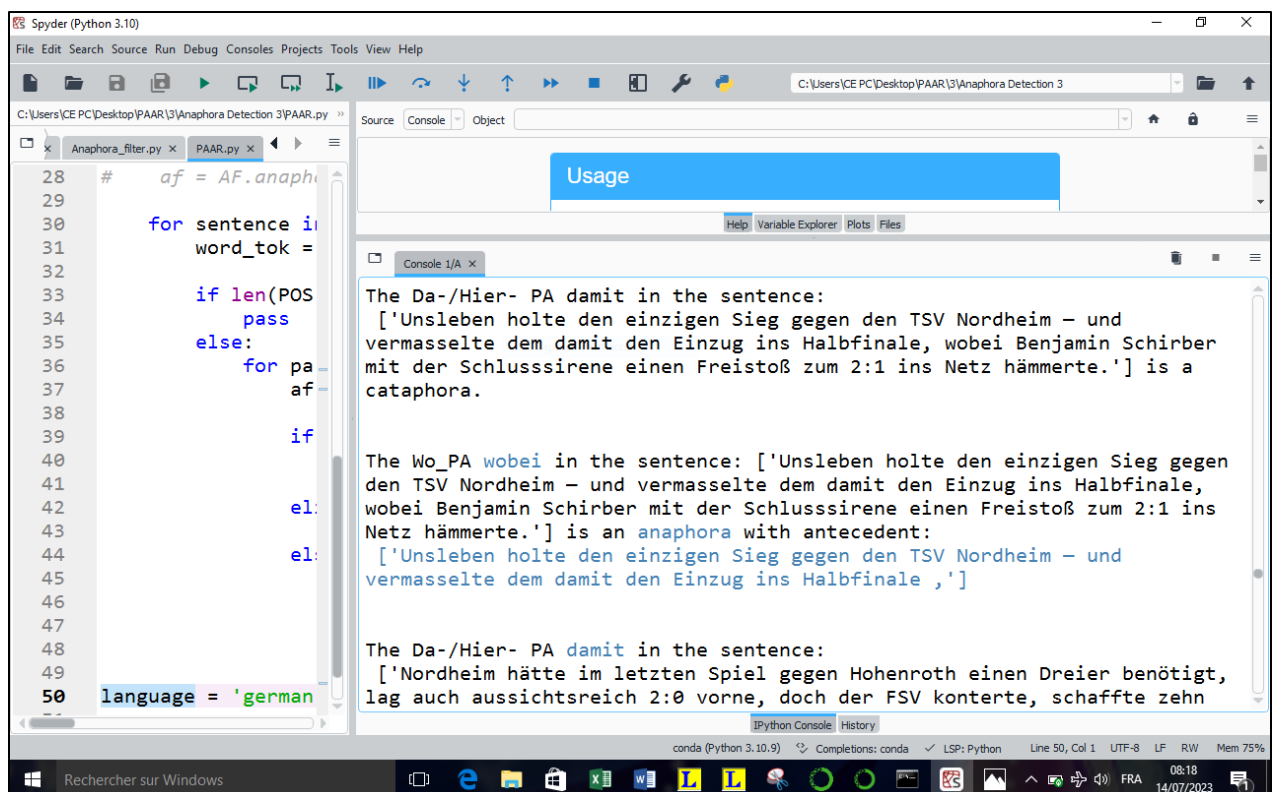


Abbildung 107: Einsatz des PAAR auf den Textabschnitt (N°75-100) (6)

Das PA „damit“ aus der Abbildung 100 ist keine Katapher, sondern eine Anapher mit dem Antezedens den Satzteil „Unsleben holte den einzigen Sieg gegen den TSV Nordheim“. Das PA „wobei“ ist zwar anaphorisch aber mit dem Antezedens „den einzigen Sieg“

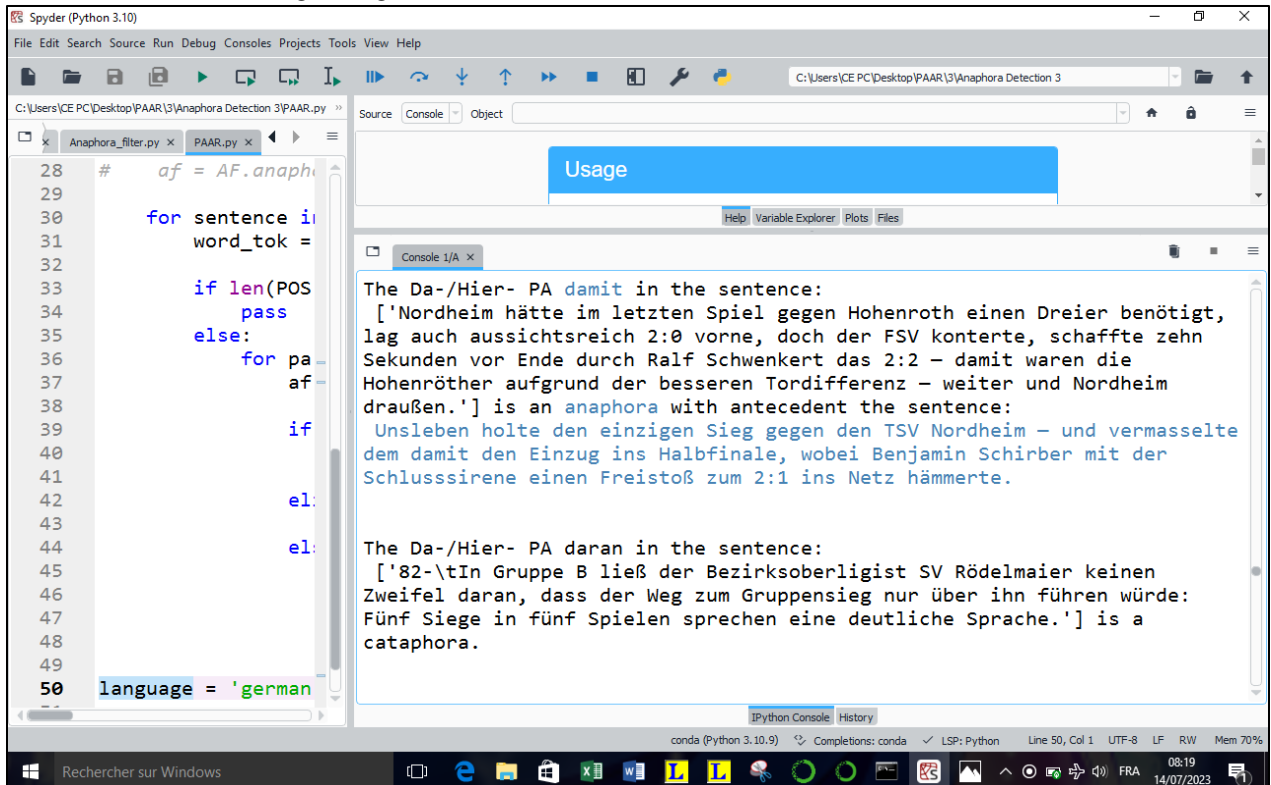


Abbildung 108: Einsatz des PAAR auf den Textabschnitt (N°75-100) (7)

Das PA „damit“ in der Abbildung 101 ist zwar anaphorisch aber mit dem Antezedens die NPP „das 2:2“. Was das PA „daran“ angeht handelt es sich tatsächlich um eine Katapher

Das PA „Dahinter“ aus der Abbildung 102 ist zwar eine Anapher aber mit dem Antezedens Die NP „der Bezirksoberligist SV Rödelmaier“. Was das PA „wobei“ angeht, wurde es richtig aufgelöst.

Die PA „Dazwischen“ in der Abbildung 103 ist keine Anapher. Was das PA „wobei“ angeht, wurde sie gut vom PAAR aufgelöst.

Das PA „Dazu“ in der Abbildung 104 wurde nur zum Teil richtig aufgelöst. Denn das richtige Antezedens fängt mit dem Wort „dazwischen“ an und endet mit dem Wort „Guédron“. Was das PA „dadurch“ aus der gleichen Abbildung angeht, handelt es sich tatsächlich um eine Katapher.

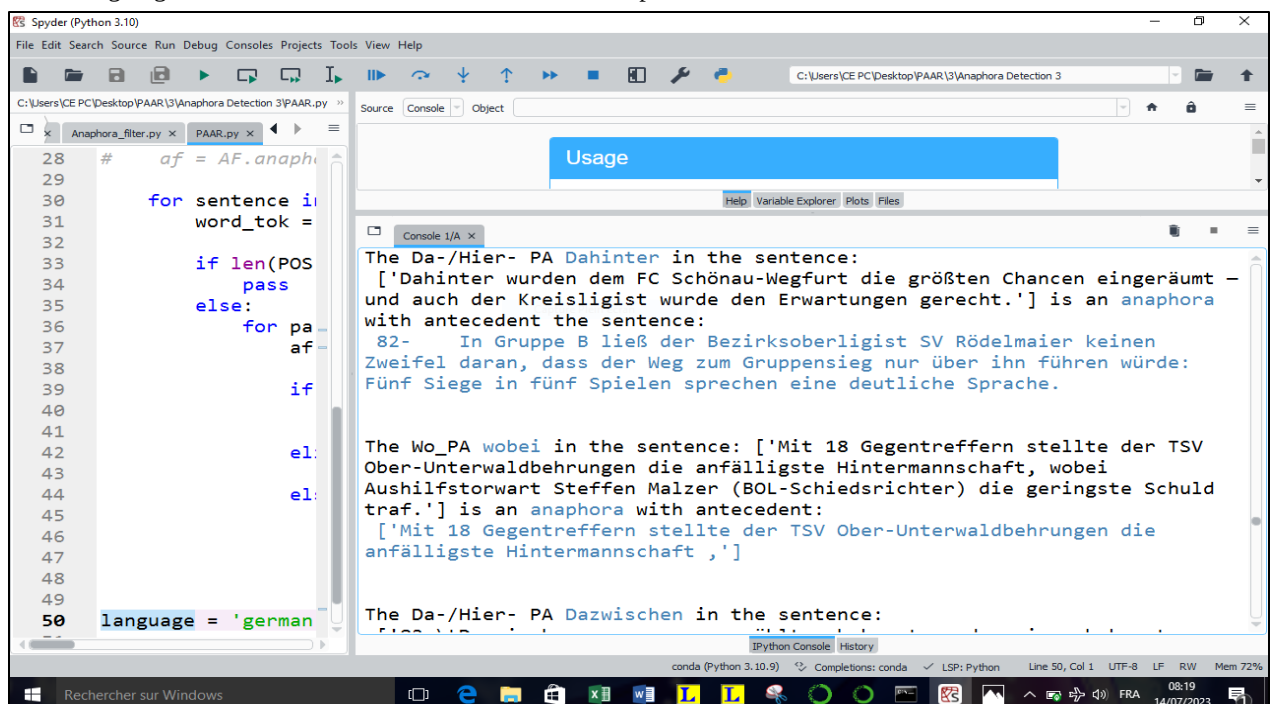


Abbildung 109: Einsatz des PAAR auf den Textabschnitt (N°75-100) (8)

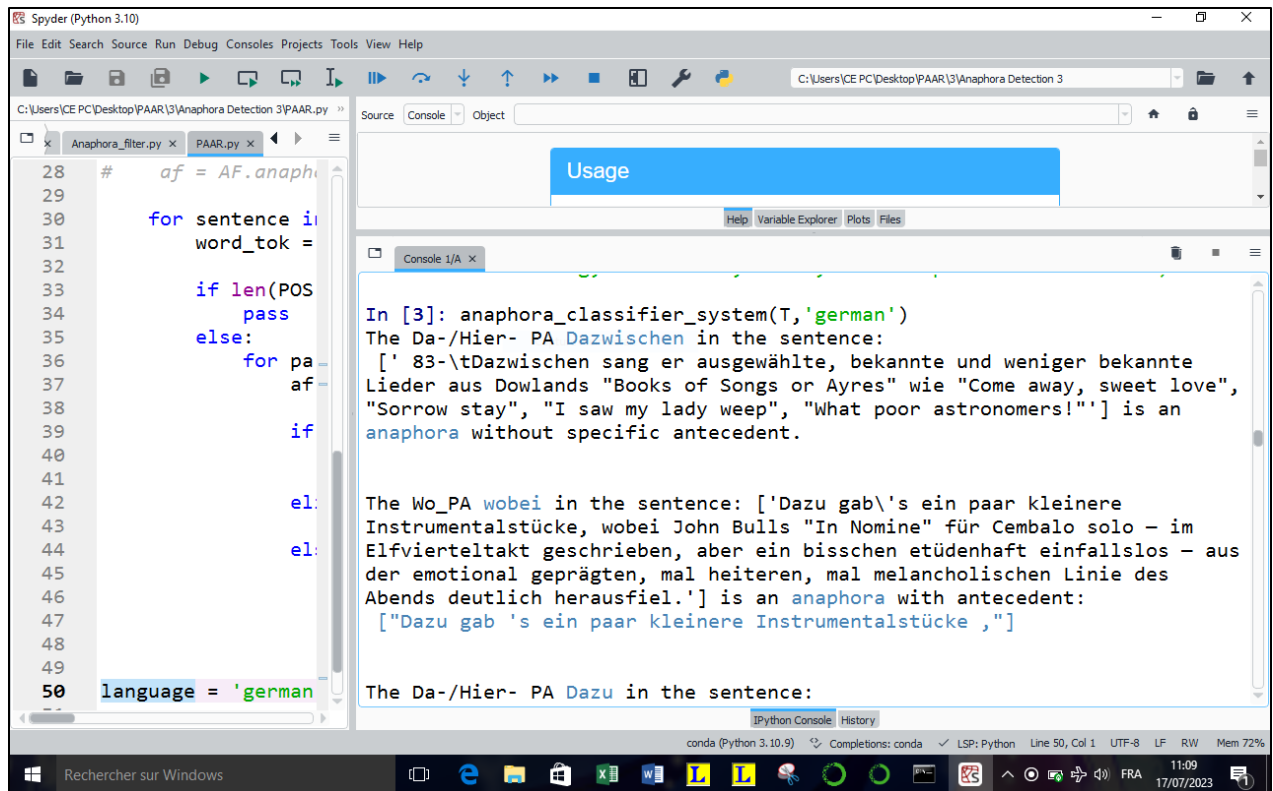


Abbildung 110: Einsatz des PAAR auf den Textabschnitt (N°75-100) (9)

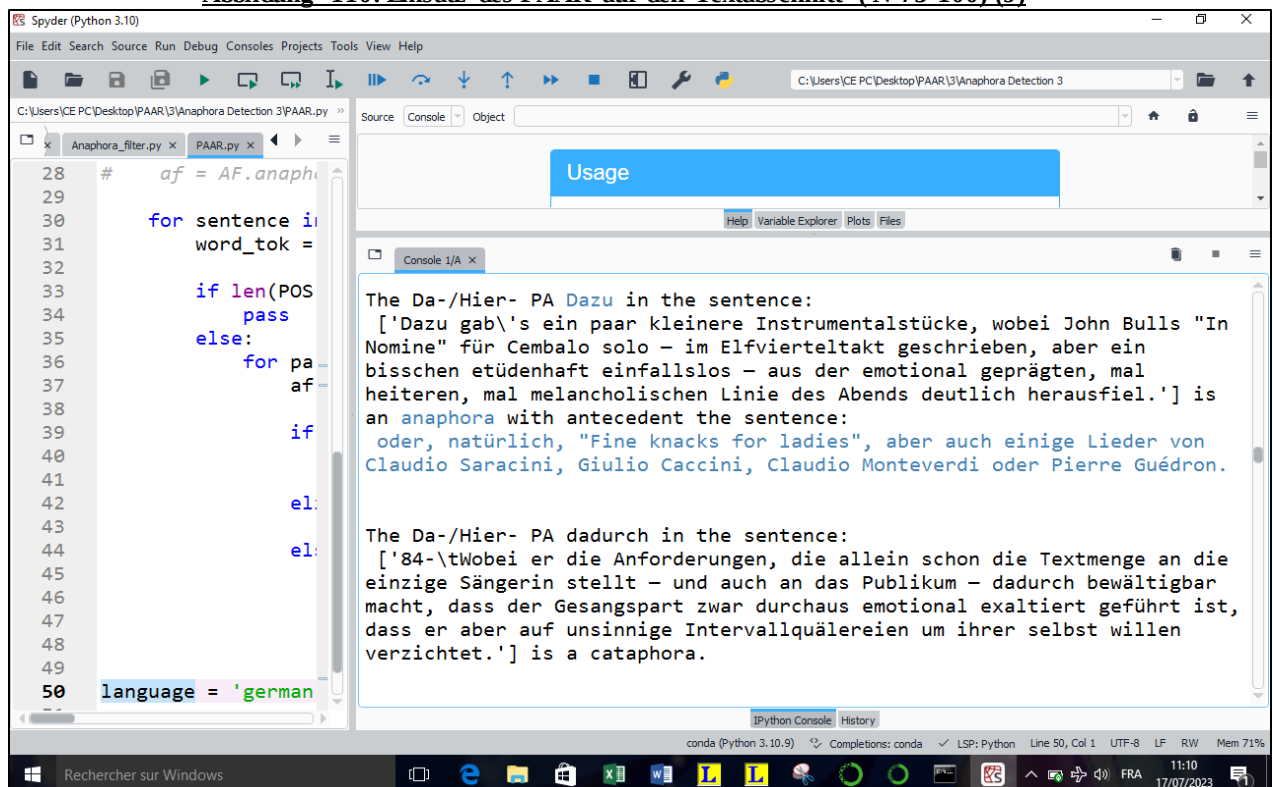


Abbildung 111: Einsatz des PAAR auf den Textabschnitt (N°75-100) (10)

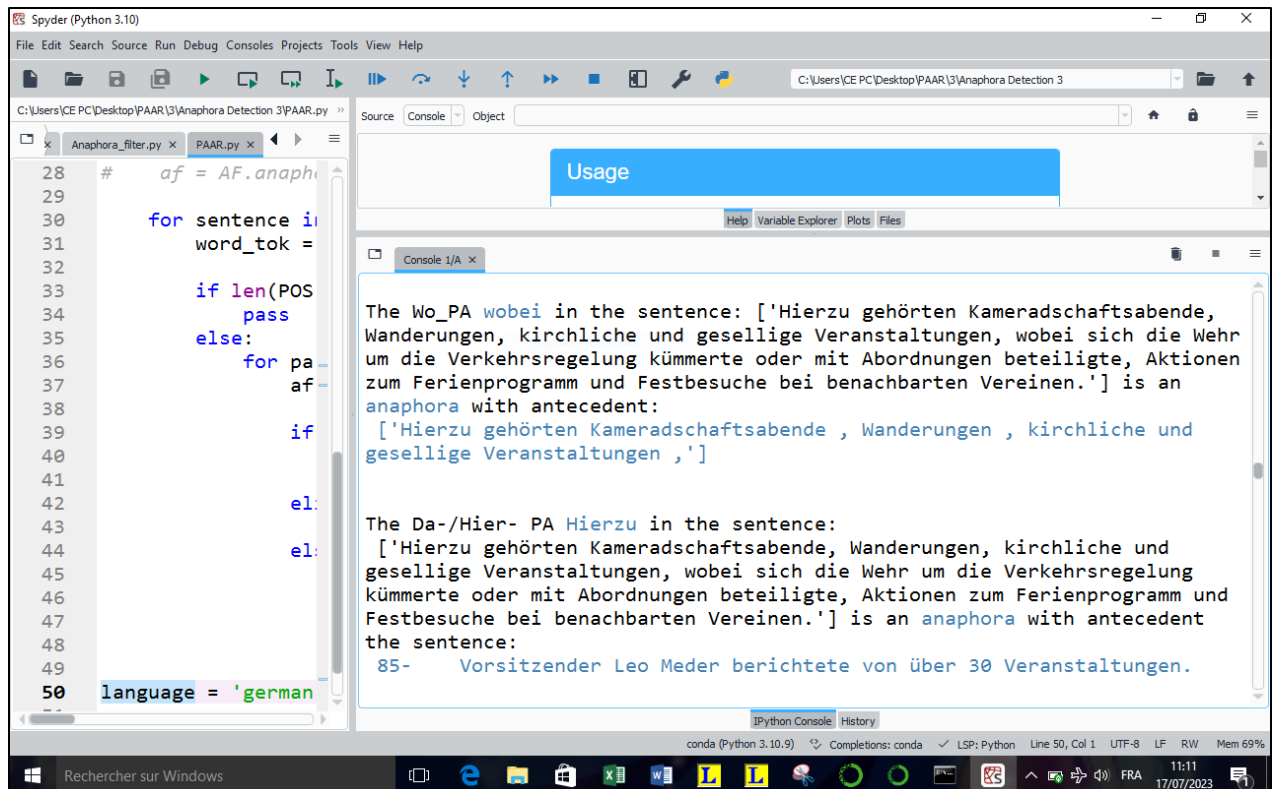


Abbildung 112: Einsatz des PAAR auf den Textabschnitt (N°75-100) (11)

Das PA „Hierzu“ in der Abbildung 105 ist zwar anaphorisch aber sein richtiges Antezedens ist „über 30 Veranstaltungen“. Was das PA „wobei“ angeht, wurde es richtig vom PAAR aufgelöst.

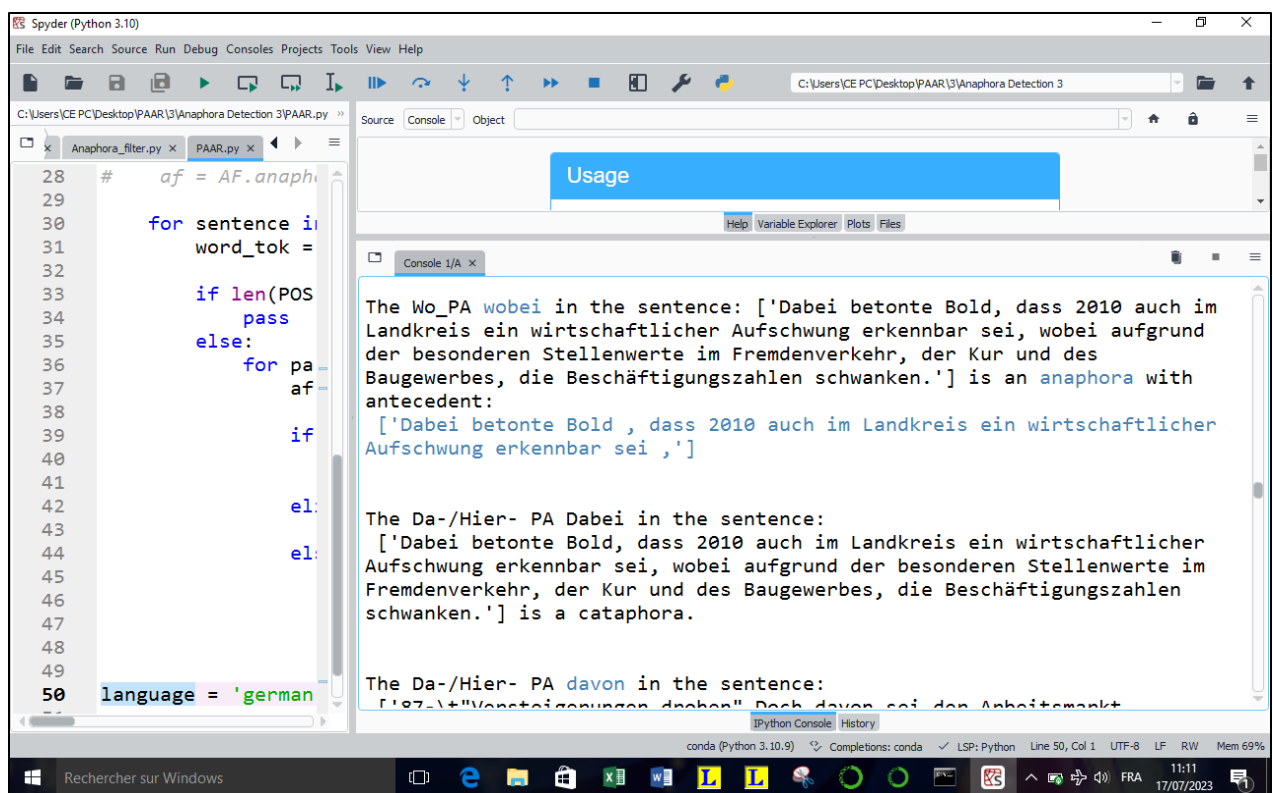


Abbildung 113: Einsatz des PAAR auf den Textabschnitt (N°75-100) (12)

Das PA „Dabei“ aus der Abbildung 106 ist keine Katapher, sondern eine Anapher mit Bezug auf den Textteil „Bevor Landrat Thomas Bold auf das Thema „Generationenfreundlicher Landkreis“ einging, streifte er zunächst die Kreistagspolitik“. Was das PA „Wobei“ angeht, wurde es richtig aufgelöst.

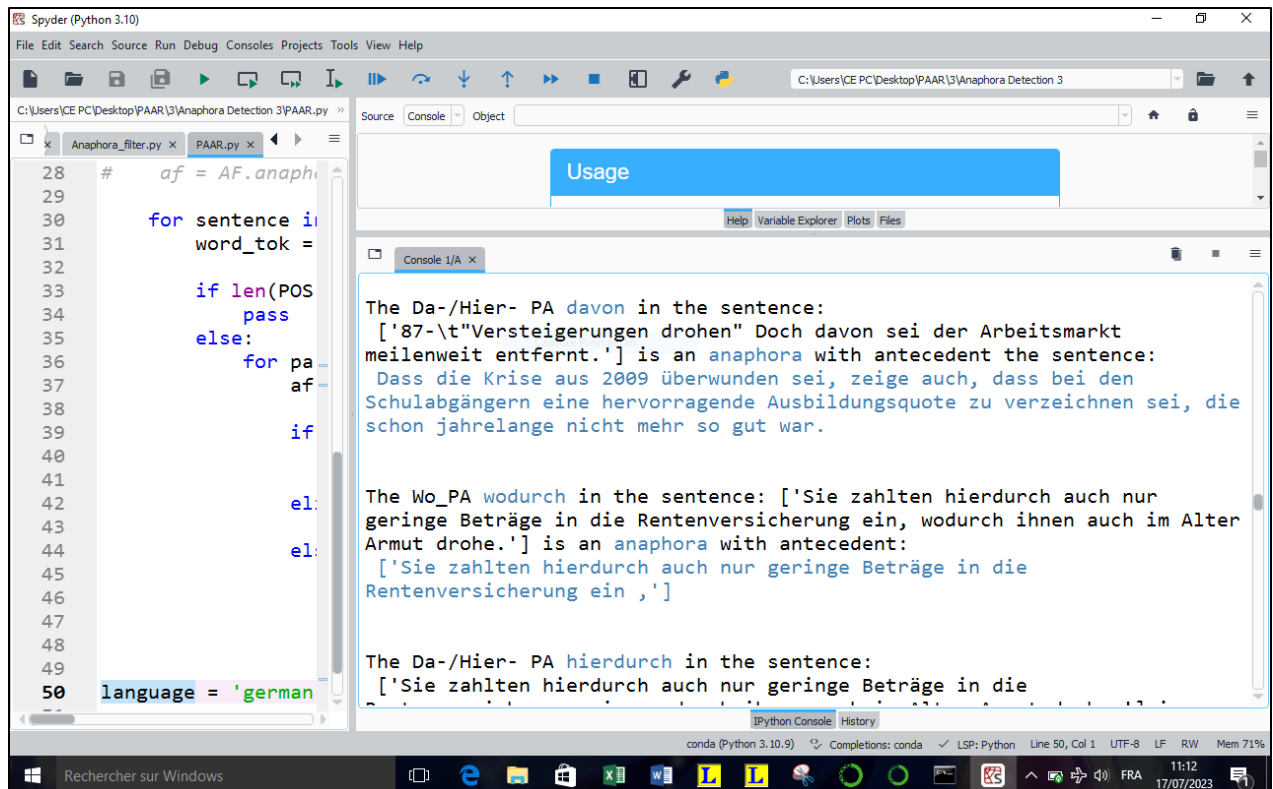


Abbildung 114: Einsatz des PAAR auf den Textabschnitt (N°75-100) (13)

Das PA „davon“ in der Abbildung 107 ist zwar anaphorisch aber mit dem Antezedens „Versteigerungen drohen“ und was die PA „wodurch“, „hierdurch“, „Hierbei“ & „wodurch“ au der Abbildung 108 angeht, wurden sie alle richtig aufgelöst

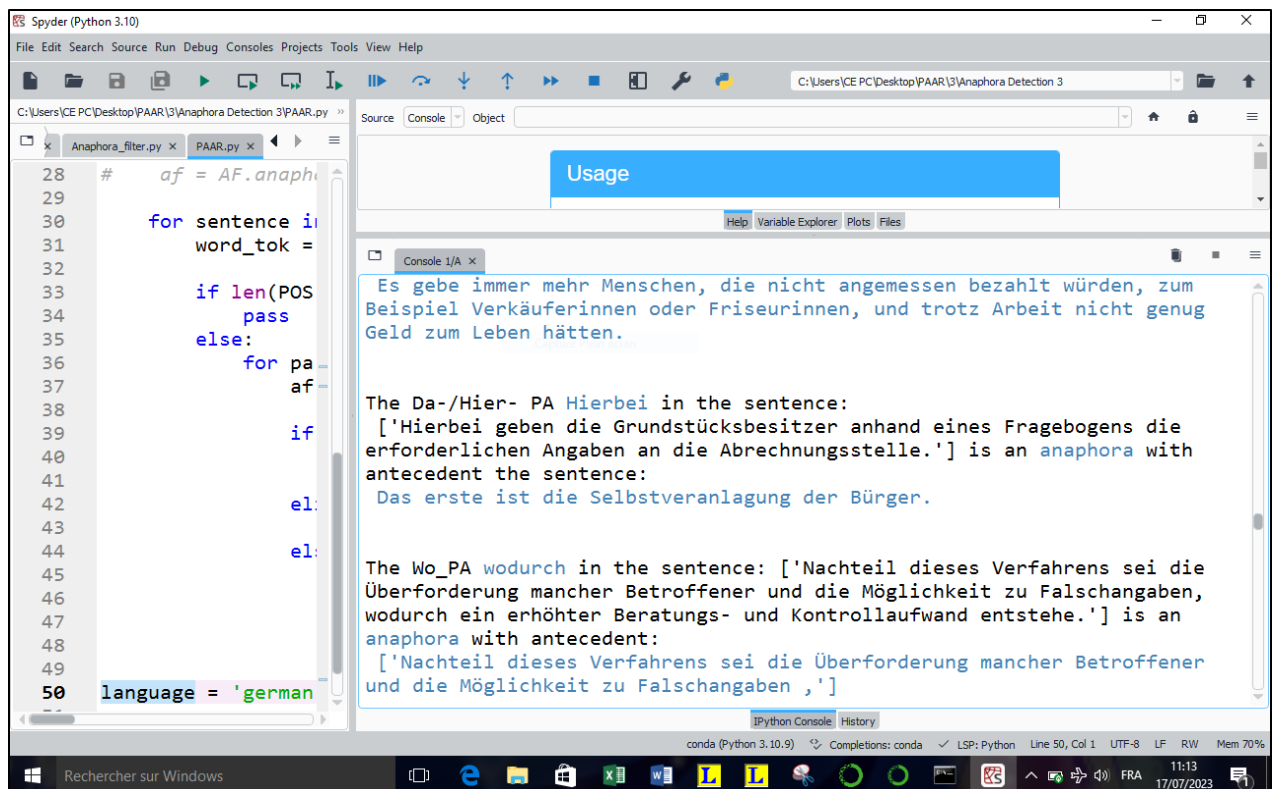


Abbildung 115: Einsatz des PAAR auf den Textabschnitt (N°75-100) (14)

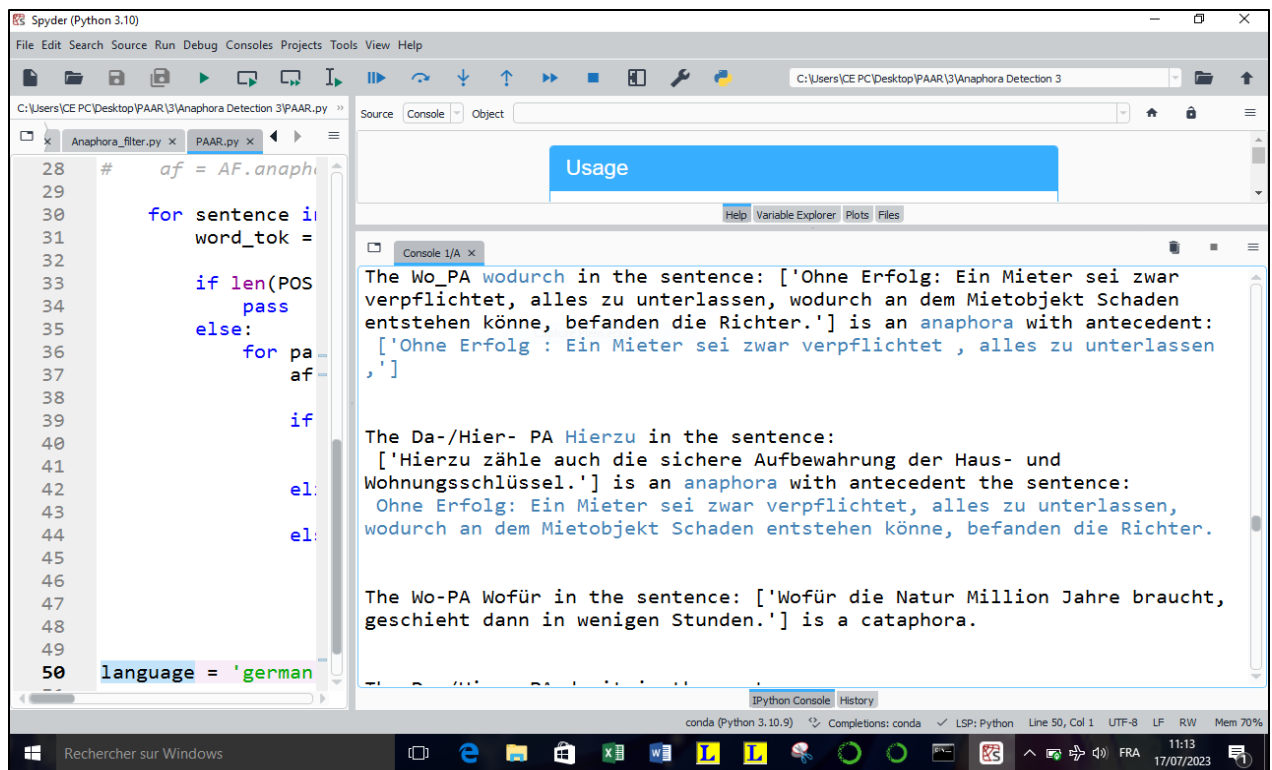


Abbildung 116: Einsatz des PAAR auf den Textabschnitt (N°75-100) (15)

Das PA "wodurch" aus der Abbildung 109 ist zwar anaphorisch aber mit dem Antezedens das Wort "alles". Die gleiche Überlegung gilt fürs PA "Hierzu". Es hat nicht den ganzen Satz fürs Antezedens, wie PAAR es meint, sondern eher nur das Wort "alles". Was das PA "Wofür" angeht, handelt es sich tatsächlich um eine Katapher

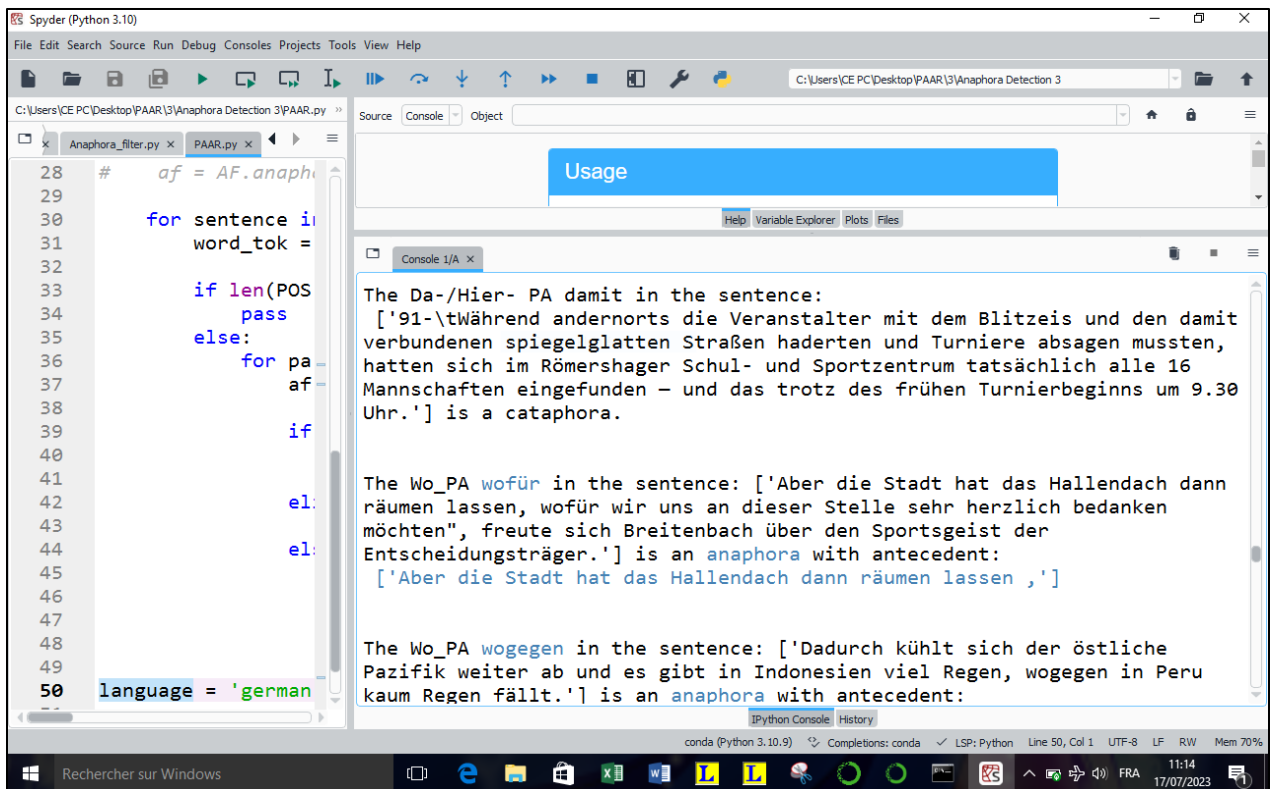


Abbildung 117: Einsatz des PAAR auf den Textabschnitt (N°75-100) (16)

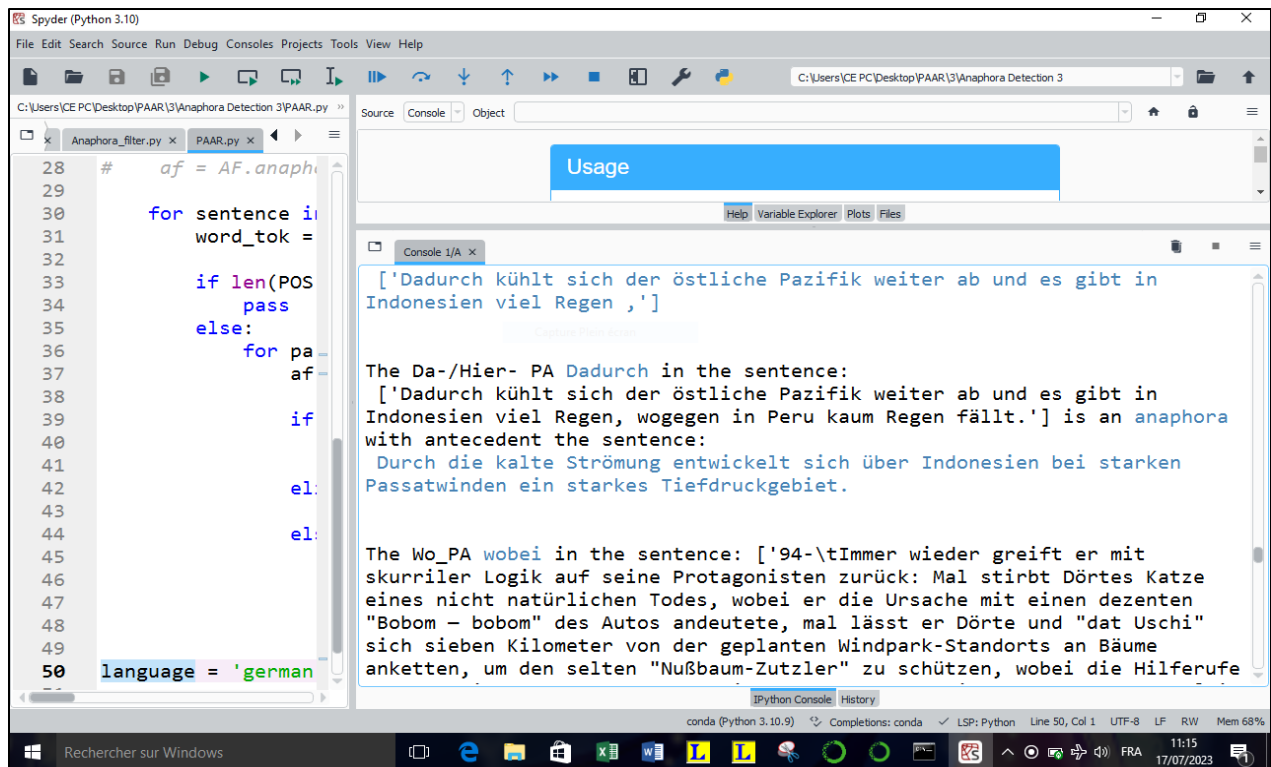


Abbildung 118: Einsatz des PAAR auf den Textabschnitt (N°75-100) (17)

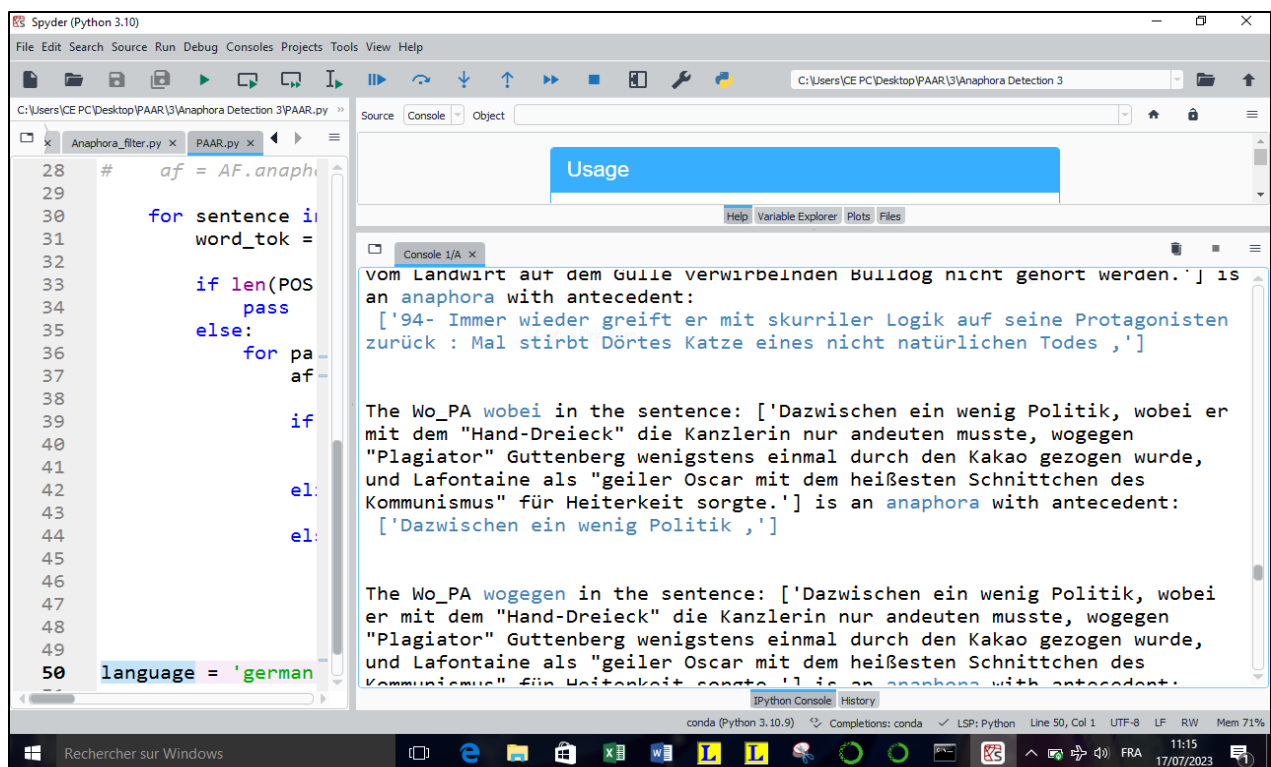


Abbildung 119: Einsatz des PAAR auf den Textabschnitt (N°75-100) (18)

In der Abbildung 110 ist das PA "Damit" keine Katapher, sondern eine Anapher mit Bezug auf die PP "mit dem Blitzeis". Das PA "wofür" hingegen ist anaphorisch und wurde richtig aufgelöst. Was das PA "Wogegen" angeht, wurde es nur zum Teil richtig aufgelöst. Denn das richtige Antezedens ist der Satzteil "und es gibt in Indonésien viel Regen" (Siehe Abbildung 111). Was das nachstehende PA "dadurch" angeht, handelt es sich umeine An apher und es wurde richtig aufgelöst. Weiterhin hat man in der Abbildung 111 das PA "wobei". Es ist zwar anaphorisch aber mit dem folgenden Textteil als Antezedens "Mal stirbt Dörtes Katze eines nicht natürlichen Todes". Die PA ("wobei"; "Wogegen" & "Dazwischen") aus den Abbildungen 112 & 113 sind alle richtig aufgelöst worden

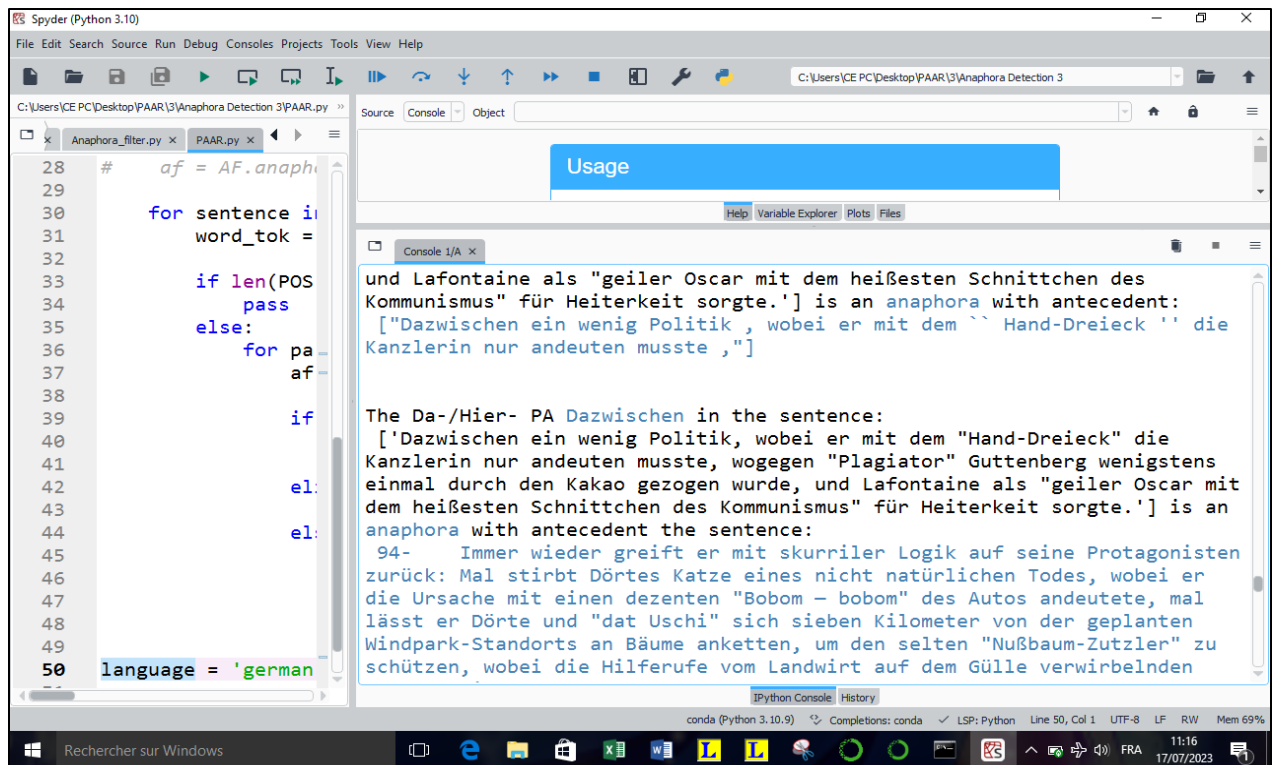


Abbildung 120: Einsatz des PAAR auf den Textabschnitt (N°75-100) (19)

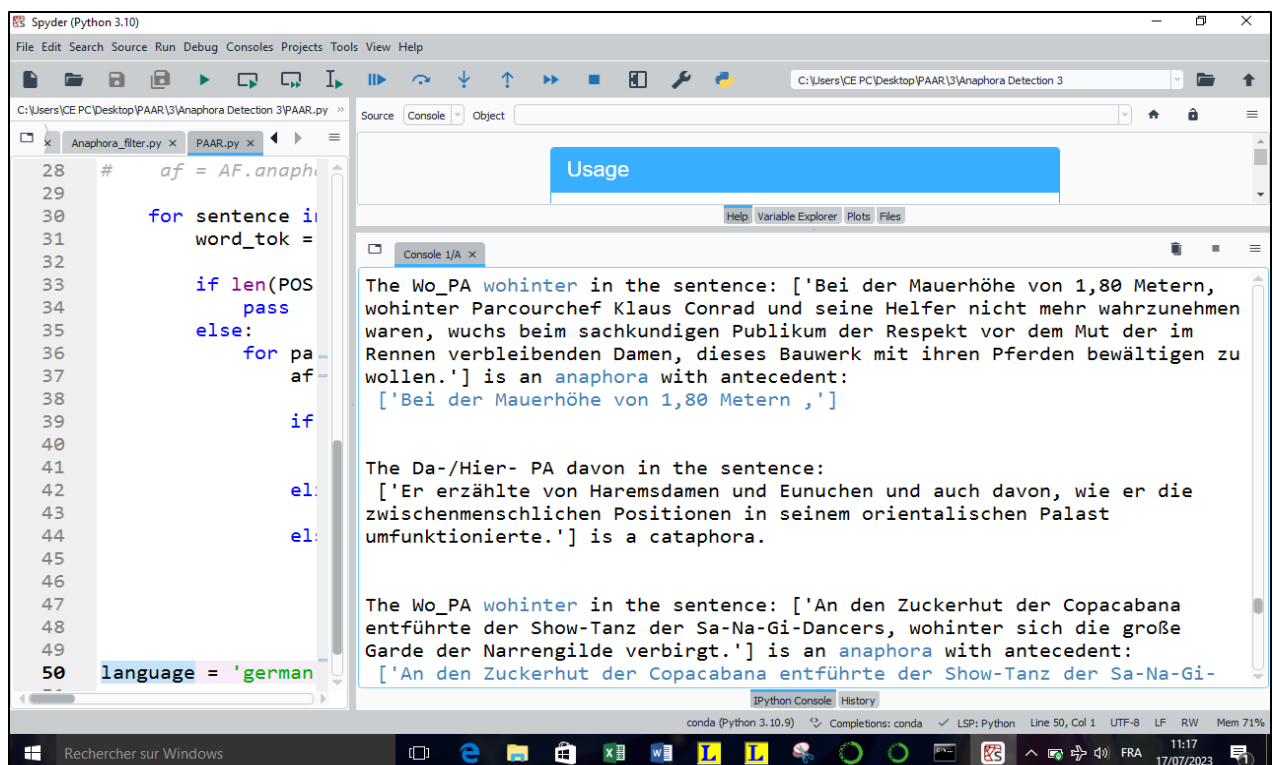


Abbildung 121: Einsatz des PAAR auf den Textabschnitt (N°75-100) (20)

Das PA "Wohinter" erscheint zweimal in der Abbildung 114. Was die erste Erscheinung angeht, wurde sie richtig vom PAAR aufgelöst. Was aber die zweite angeht, wurde diese nur zum Teil richtig aufgelöst. Denn das richtige Antezedens hierfür ist die PP "An den Zuckerhut der Copacabana". Was die PA "davon" und "daran" (Abbildung 115) angeht, handelt es sich tatsächlich um Kataphern. Dies hingegen sind die PA "Wohinter" & "Wobei" keine Kataphern, sondern Anaphern mit den jeweiligen Antezedenzen "mit einer Fahrschule" & "Der Abend wurde zu einer tänzerischen Sensation"

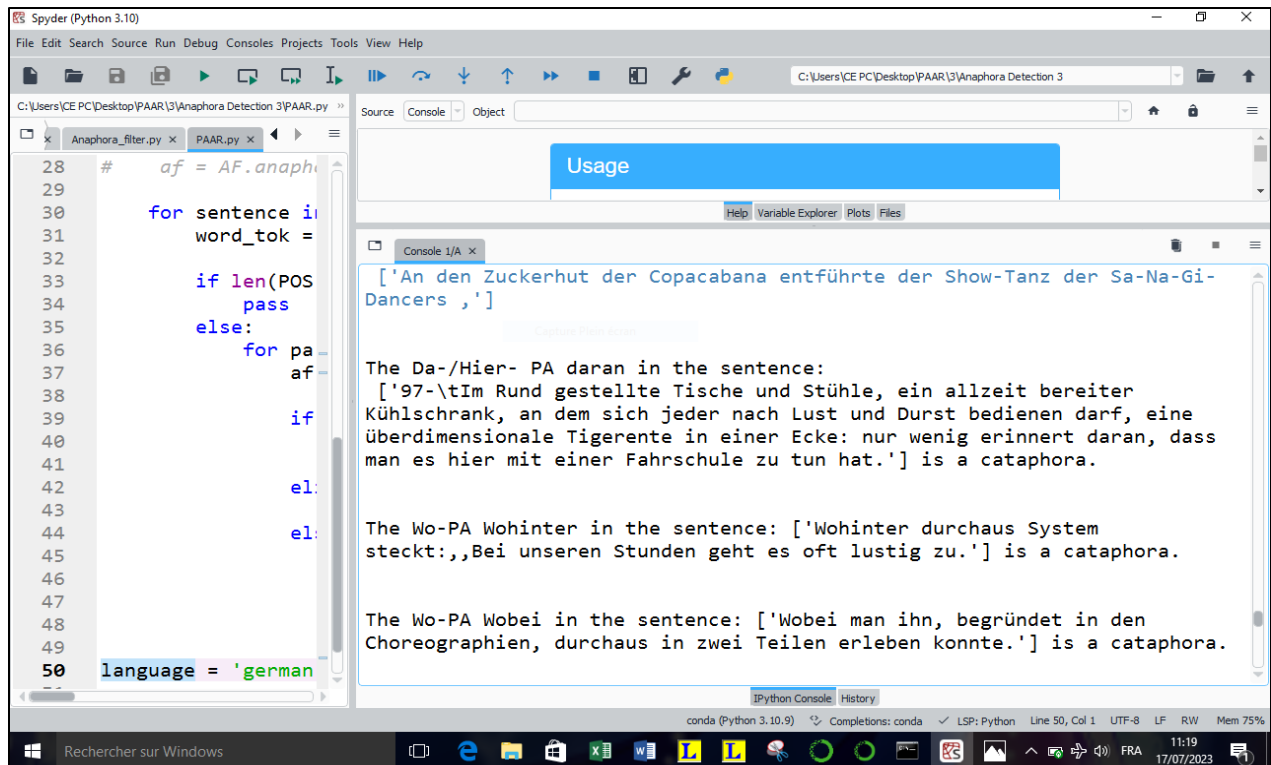


Abbildung 122: Einsatz des PAAR auf den Textabschnitt (N°75-100) (21)

Weiterhin stellen wir fest, dass das PA "damit" in der Abbildung 116 zwar anaphorisch ist aber mit Antezedens "Sie lief die 400 Meter in 61,95 Sekunden". Was die erste Erscheinung des PA "womit" angeht, wurde es richtig aufgelöst. Die zweite Erscheinung hingegen ist keine Katapher, sondern eine Anapher mit dem Textteil "Sogar.....hochstämmigen Bäumen" als Antezedens

Letztendlich stellen wir fest, dass die Erscheinung von "damit" in der Abbildung 117 zwar anaphorisch ist, wie PAAR es gemeint hat, aber mit dem Textteil "Die Bäume müssen weg." als Antezedens

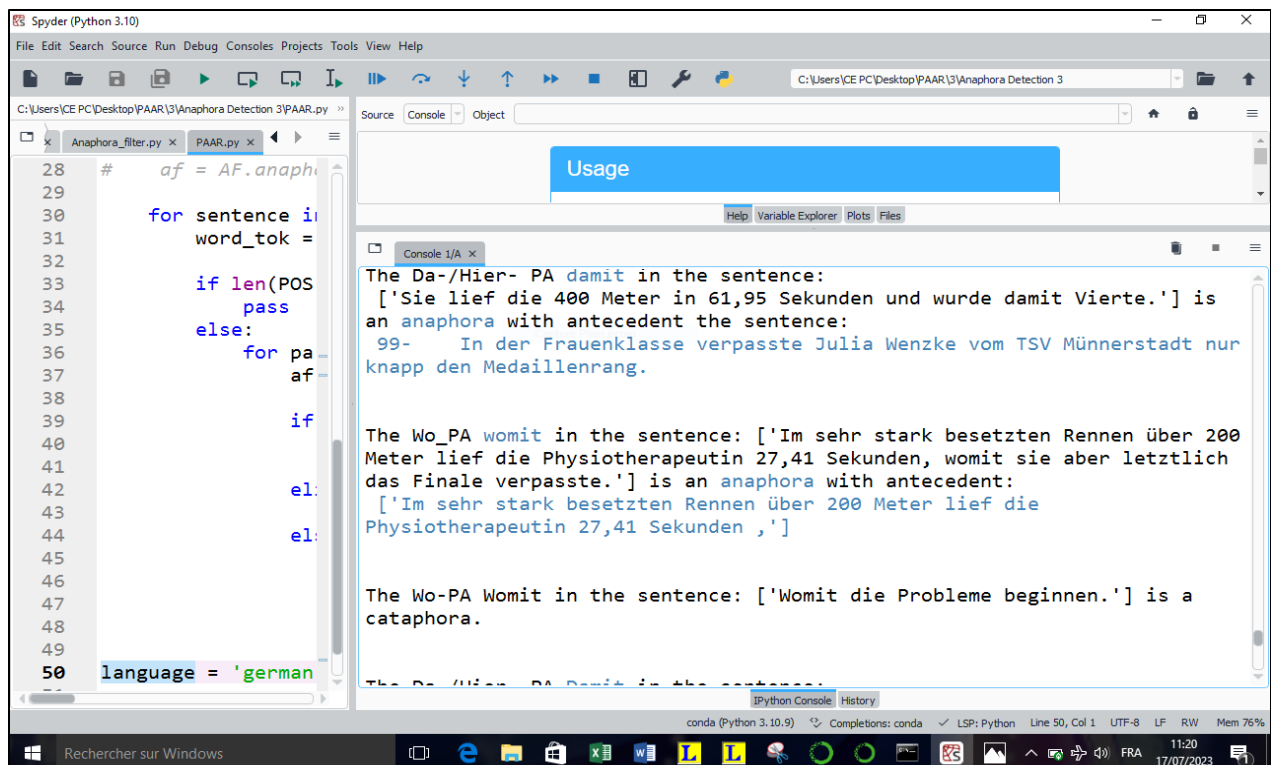


Abbildung 123: Einsatz des PAAR auf den Textabschnitt (N°75-100) (21)

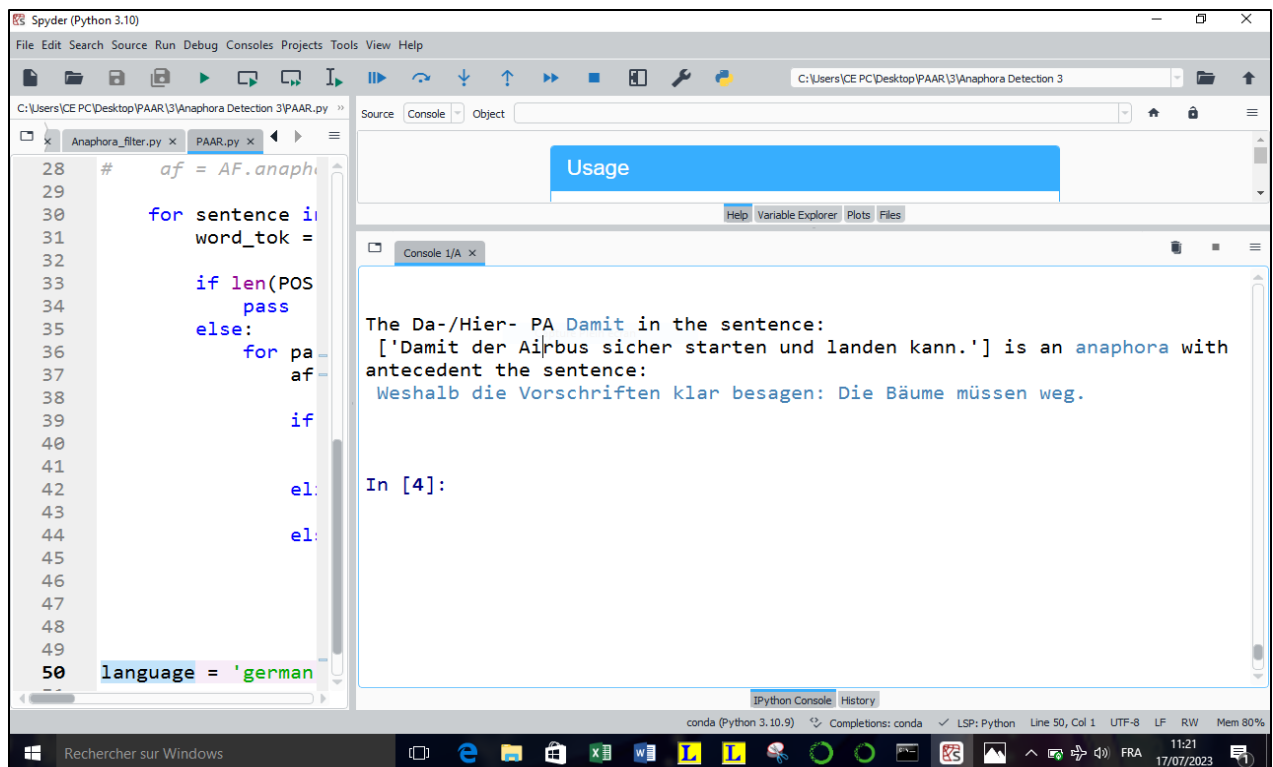


Abbildung 124: Einsatz des PAAR auf den Textabschnitt (N°75-100) (22)

ZUSAMMENFASSUNG UND ERGEBNISSE DES EINSATZES

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
50	51(50PA+ 1Subjunktion)	51/50???

Tabelle 70: Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°76-100)

PA	Anzahl	Natur	Ergebnisse
hierzu	5	Anapher	5/5
Dabei	1	Anapher	1/1
Wobei	9	Anapher	9/9
Danach	1	Anapher	1/1
Hierzwischen	1	Anapher	1/1
Hierfür	1	Anapher	1/1
Darauf	1	Katapher	1/1
Damit	2	Katapher	0/2
Damit	3	Anapher	3/3
Daran	2	Katapher	2/2
Dahinter	1	Anapher	1/1
Dadurch	1	Katapher	1/1
Dazwischen	2	Anapher	1/2
Dazu	1	Anapher	1/1
Dabei	1	Katapher	0/1
Davon	1	Anapher	1/1
wodurch	3	Anapher	3/3
Hierdurch	1	Anapher	1/1
Hierbei	1	Anapher	1/1
wofür	1	Katapher	1/1
wofür	1	Anapher	1/1
Wogegen	2	Anapher	2/2
Dadurch	1	Anapher	1/1
Wohinter	2	Anapher	2/2

Davon	1	Katapher	1/1
Wohinter	1	Katapher	0/1
Wobei	1	Katapher	0/1
Womit	1	Anapher	1/1
Womit	1	Katapher	0/1
TOTAL	50		44/50

Tabelle 71: Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°76-100)

PA	Noten der Auflösung von anaphorischen PA durch PAAR
hierzu	1/5
Dabei	1/1
Wobei	7,75/9
Hierzwischen	1/1
Hierfür	0/1
Damit	1,75/3
Dahinter	0,25/1
Dazu	0,5/1
Davon	0/1
Wodurch	2,5/3
Hierdurch	1/1
Hierbei	1/1
Wogegen	1,5/2
Wofür	1/1
Dadurch	1/1
Dazwischen	1/1
Wohinter	1,5/2
Womit	1/1
TOTAL	24, 75/33

Tabelle 72: Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°76-100)

5. Einsatz des PAAR auf den Textabschnitt (N°101-154)

- 101-** Ähnliches gilt für die Reserve des FC 06 Bad Kissingen angesichts der Verpflichtungen von Matthias Beck und Benedikt Wedlich für die Bezirksliga-Elf. **Womit** gewährleistet sein dürfte, dass die 06er auch in der Rückrunde eine gelungene Mischung aus Talenten und Routiniers aufs Grün schicken können. Den Verein verlassen hat Christoph Kohlhepp, der zum FC Westheim wechselt. (SAZ11/FEB.00498 SaaleZeitung, 04.02.2011, S. 12; Führungs-Quartett steht zu ehrgeizigen Zielen)
- 102-** Auf bessere Zeiten hofft auch Klaus Katzenberger, Vorsitzender des FC Reichenbach. Schließlich haben sich die Reihen der verletzten Spieler endlich gelichtet, **womit** unter dem neuen Trainer Marc Hartmann das Saisonziel "10. Platz" offensiv angegangen werden darf. Seinen Beitrag **dazu** leisten will künftig Keeper Steffen Kleinhenz, der von der DJK Schlimpfhof zu den Teutonen wechselte. (SAZ11/FEB.00498 Saale-Zeitung, 04.02.2011, S. 12; Führungs-Quartett steht zu ehrgeizigen Zielen)
- 103-** Betrifft: "Bloß nicht zurücktreten" vom 22. Februar:G" Ich rechne schon lange **damit**, dass seine Gegner so lange in seiner Lebensgeschichte herum wühlen, bis sie was finden, **womit** sie ihm ans Bein pinkeln können'. Dieser Satz fiel heute früh hier. Er ist nicht von mir, aber er könnte es sein. (SAZ11/FEB.02925 Saale-Zeitung, 23.02.2011, S. 19; Was zu erwarten war)
- 104-** Im vergangenen Dezember wurde das Gutachten eines Rechtsmediziners präsentiert, **wonach** die Frau frühestens um 23.45 Uhr gestorben ist. Dass beide Opfer um 23.30 Uhr getötet wurden, ist nach Angaben des Sachverständigen nicht unmöglich, aber sehr unwahrscheinlich. Wie Landgerichtspräsident Wich-Knoten dazu erklärte, ist der Angeklagte den Erkenntnissen zufolge in Tatortnähe gewesen, allerdings wohl nicht zu der Zeit, als das Liebespaar erschossen wurde. Für die Bluttat gibt es keine Zeugen, auch die Waffe fehlt bisher — eine schwierige Beweislage. "Das ist eine Indiziengeschichte", sagte Wich-Knoten. (SAZ11/JAN.00511 Saale-Zeitung, 06.01.2011, S. 25; Angeklagter auf freiem Fuß)

- 105- In keinem anderen Bundesstaat haben die politischen Emotionen im vergangenen Jahr derartige Wellen geschlagen wie in Arizona. Höhepunkt war der Versuch der Regierung in Phoenix, superstrenge Maßnahmen gegen Ausländer einzuführen. In letzter Minute konnte ein Bundesgericht die Absicht kippen, **wonach** die Polizei praktisch jeden Ausländer jederzeit kontrollieren kann, ob er illegal im Land ist. Zeitweise hatte der Streit das ganze Land aufgewühlt und die überhitzten Gemüter weiter angestachelt. Doch unter politischen Psychologen ist es längst gängige Erkenntnisse, dass ein solches Klima idealer Nährboden für politische Wirtköpfe und Gewalt ist. (SAZ11/JAN.00797 Saale-Zeitung, 10.01.2011, S. 20; ohne Titel)
- 106- Die ägyptische Führung unter Präsident Mubarak schweigt weiter zu den Protesten. Das einzige Zugeständnis war am Donnerstag eine Meldung der staatlichen Medien, **wonach** das Parlament am kommenden Sonntag über Maßnahmen zur Armutsbekämpfung, eine Anhebung des staatlichen Mindestlohnes und eine bessere Gesundheitsversorgung debattieren soll. Seit Beginn der Proteste wurden nach Darstellung ägyptischer Medien mindestens sechs Menschen getötet. (SAZ11/JAN.03071 Saale-Zeitung, 28.01.2011, S. 23; Proteste erreichen den Jemen)
- 107- Und haben Sie dann überhaupt noch Zeit, andere Dinge zu tun? Ich lese ja nicht alle Texte bis zum Ende. Bei manchen stelle ich schon nach fünf bis sechs Seiten fest: Das kommt nicht in Frage. **Woran** merken Sie das? Manchmal passt das Thema einfach nicht zu uns. Oder die Geschichte ist zu brutal. Wir geben in meinem Bereich ja vor allem Geschichten für Kinder und Jugendliche heraus. Manchmal ist es aber auch langweilig geschrieben. Oder die Idee ist alt. (SAZ11/JAN.02234 Saale-Zeitung, 21.01.2011, S. 30; "Leidenschaft für Sprache, Liebe zu Wörtern")
- 108- Ein überdurchschnittlich hoher Prozentsatz türkischstämmiger Jugendlicher beendet die Schule ohne Abschluss — **woran** das liegt, ist umstritten. Die Auszubildende Özen ist inzwischen im dritten Lehrjahr und möchte mit der neuen Zusatzqualifikation "Wirtschaftskennerin Türkei" ihre Chancen auf dem Arbeitsmarkt verbessern. **Dazu** wird sie neben ihrer Ausbildung von Februar an zusätzlich noch 130 Unterrichtsstunden etwa in Wirtschaftstürkisch und in Wirtschaftsgeschichte belegen. Bescheinigt wird die Qualifikation von der Industrie- und Handelskammer (IHK) Nürnberg. "Bei mir besteht die Möglichkeit, nach der Ausbildung in die Türkei zu gehen", sagt Özen. Sie sei neugierig auf das Leben in der Heimat ihrer Eltern. Die neue Qualifikation soll ihr **dabei** helfen, dort einen Job in einem deutschen Unternehmen zu finden. Die Firmen könnten nach Özens Ansicht in vielfacher Weise von ihrer Landeskunde und ihrem Know-how profitieren. (SAZ11/JAN.03291 Saale-Zeitung, 31.01.2011, S. 27; Cigdem Özen will Brücken bauen)
- 109- Ungeachtet dessen treten Mandatsträger und Verwaltung mit dem Anspruch an, den Mangel bestmöglich zu gestalten. **Dazu** hat es ein erstes Treffen gegeben, das hinter verschlossenen Türen stattgefunden hat. Über drei Stunden lang gab es Informationen von Seiten des Bürgermeisters Ernst Stross (SPD) und von Kämmerer Bernhard Blum. Ohne Details der nichtöffentlichen Beratungen bekannt zu geben, machte das Stadtoberhaupt im Gespräch mit der Redaktion deutlich: "Durch viele Vorgriffe ist der Haushalt 2011 bereits maßgeblich vorbelastet. Jeder weiß, **woran** wir sind." Bürgermeister Ernst Stross wollte auf Grund der prekären Finanzlage der Kommune damit einfach Verständnis dafür wecken, dass nicht alles Wünschenswerte leistbar sein werde. „Stross verwies auf die laufenden Millionenprojekte Dreifachturnhalle und Hallenbad. Die Einrichtungen müssten auf Grund ihres Alters mit hohem Aufwand generalsaniert oder gar neu gebaut werden. **Damit** sind bereits wesentliche Mittel gebunden", stellte Stross klar. (SAZ11/FEB.02128 Saale-Zeitung, 17.02.2011, S. 9; Mangel bestmöglich gestalten)
- 110- Erste Hilfe leisten heißt mit einfachsten Handgriffen, die jeder erlernen kann, Menschenleben retten können. **Dazu** gehört es, einen klaren Kopf zu behalten und zu wissen, **worauf** es ankommt. Wichtig ist eine gewisse Selbstsicherheit, und dass Laien, die gerade vor Ort sind, bis zum Eintreffen des Rettungsdienstes überbrücken können. Und dass sie gegebenenfalls die professionellen Retter unterstützen. (SAZ11/JAN.01594 Saale-Zeitung, 17.01.2011, S. 2; Was ist Erste Hilfe?)
- 111- Vielmehr können sie sie auch mit dem Hinweis, **worauf** sie setzen wollen, dem Croupier geben. Das hatte, wie berichtet, während der Ermittlungen der Leitende Oberstaatsanwalt Rainer Vogt, erläutert. Laut Landeskriminalamt aber soll der betagte Stammgast beim Setzen seiner Chips keine zu spielende Zahl genannt haben. Was es den beteiligten Croupiers leicht machte, ungeachtet des Ergebnisses des Spiels und **damit** unabhängig von der gefallenen Zahl Gewinne auszubezahlen. (SAZ11/JAN.02576 Saale-Zeitung, 25.01.2011, S. 1; Wenn die Kugel amerikanisch rollt)
- 112- Freizeit Was kann es im Winter Schöneres geben als zu rodeln? Zumindest wenn genug Schnee liegt und ein Rodelhang in der Nähe ist. **Worauf** gerutscht wird, ist eigentlich egal. (SAZ11/JAN.02647 Saale-Zeitung, 25.01.2011, S. 30; Rodeln ist ein toller Spaß)
- 113- Ungewöhnlich viele prominente Spieler für schwindelerregende Summen wechselten die Seiten. Als teuerster Winter-Transfer geht Vorjahres-Torschützenkönig Dzeko in die Historie ein. Den Abgang

des Angreifers zum englischen Premier-League-Club ließ sich Wolfsburg mit 35 Millionen Euro versüßen. Bayern München holte sich den Hoffenheimer Luiz Gustavo für 17 Millionen Euro, **worauf** sich Kapitän Mark van Bommel wegen der "besseren sportlichen Perspektive" für den AC Mailand entschied. dpa (SAZ11/FEB.00066 Saale Zeitung, 01.02.2011, S. 20; Helmes stürmt künftig für die Wölfe)

- 114- **Worauf** die Narren an den Tischen in der Schwedenberghalle gewartet hatten, das war ein Büttensbeitrag mit einer ordentlichen Portion Lokalkolorit. Sprechende Mülltonnen gaben mit ihrer Schilderung des Dorfgeschehens **hierauf** die passenden Antworten. Angefangen vom Fußball-Brachlandproblem, vergessenen Trikots und Reiseerlebnissen der Kicker, weiterhin über willkommene Reihernahrung durch eingesetzte Fischeier in der Saale bis hin zum Diebstahl von Weihnachtsplätzchen und Streugut betrachteten die Mülltonnen ihre Dorfwelt aus ihrer jeweiligen Sicht. (SAZ11/FEB.02600 Saale-Zeitung, 21.02.2011, S. 11; ohne Titel)
- 115- Die Mitglieder applaudierten, allerdings blieb unklar, **woraus** dieser Gewinn entstanden ist. Vielmehr wurden Nachfragen durch den Hinweis beantwortet, dass alles eingesehen werden kann und es früher auch nicht anders gewesen sei. Für den Haushalt 2011 kalkuliert Beck mit steigenden Kosten — allein für die Hochwasserschäden sind 4000 Euro eingeplant — und trotzdem mit einem Überschuss. Die Entlastung von Schatzmeister und Vorstand erfolgten einstimmig. (SAZ11/FEB.01649 Saale-Zeitung, 14.02.2011, S. 3; Der Ärger ist noch nicht verraucht)
- 116- Wenn die Schwaden von Pulverdampf draußen verraucht sind und der dicke Nebel im eigenen Kopf im Laufe des Neujahrstags sich allmählich lichtet, nagt gern ein Tierchen an der Seele: der Gewissenswurm. Das alte Jahr hat man sündig beendet und möchte ins neue ganz, ganz anders starten. Nun heißen die wenigsten von uns Saulus, und nach Damaskus will man sich in diesen Zeiten auch nicht so gern auf den Weg machen. **Woraus** folgt, dass der alte Adam auch 2012 ganz schön alt aussehen und bleiben wird. Hat der gute Vorsatz zum Jahresanfang nicht überhaupt etwas Puritanisches, zutiefst Genuss- und Lustfeindliches, ja Spießiges? Demnächst lebt man von Reformhauskost, geht im härenen Büßergewand und schaut im Fernsehen immer Reinhold Beckmann und Markus Lanz. So raunt das Teufelchen im Kopf bereits merklich lauter, als der G'wissenswurm nagt. Irgendwann in diesem Jahr wird er dennoch siegen. Er hat ja noch 365 Tage Zeit. Bis dahin halten wir's mit dem Philosophen Nietzsche: ". . . mit dem neuen Sündchen löscht man das alte aus"! (SAZ12/JAN.00081 Saale-Zeitung, 02.01.2012, S. 25; Ein Vorsatz:)
- 117- So verdeutliche die Nacktheit einiger der Dargestellten beispielsweise, dass der Mensch verletzlich sei. "Letztlich stellt auch dieses Werk existenzielle Fragen, die sich wie ein roter Faden durch die gesamte Ausstellung ziehen: Wer bin ich? **Woraus** lebe ich? Woraufhoffe ich?", sagte Lenssen. Wer den von Gierkes Werk gestalteten Raum betrete, werde hinein genommen in die Bühne des Lebens und müsse seinen Platz finden. "Abendmahl ist für mich überall da, wo Feier des Lebens geschieht." (SAZ12/MAR.00145 Saale-Zeitung, 02.03.2012, S. 33; Besucher als Teil des Kunstwerks)
- 118- Als ganz persönliche Zutat zum Programm improvisierte Stoiber über gegebene Themen, **darunter** das Kirchenlied "Was Gott tut, das ist wohlgetan" (**worein** sich auch das Männlein, das gewöhnlich im Walde steht, mogelte) in Verbindung mit "Wenn der weiße Flieder wieder blüht". Eine Sternstunde im Orgelzyklus. (SAZ12/SEP.01166 Saale-Zeitung, 12.09.2012, S. 2; Vom Beherrschen großer Räume)
- 119- Für ein attraktives Wohnumfeldes sei es oft nur eine Frage, **worein** man sein Engagement stecke, meinte Wolfgang Becker, Berater der Firma Argus Concept in Sachen Dorferneuerungen, und präsentierte Ideen zu Straßen, Gärten und Plätzen. Ein Tipp: Stauden statt Gras. "Die Zeit, um jede Woche den Rasen zu mähen, kann man in schönere Effekte stecken." (SBZ06/SEP.10248 Saarbrücker Zeitung, 21.09.2006; hauptzeile 01 Tag mit Dummy-Text Ortskerne zu neuem Leben erwecken)
- 120- Die Schobert und Winter GbR aus Strigleben hat zu diesem Zweck auf dem Feld gleich am Straßenrand einen Kasten installiert, **worein** Sonnenblumenfreunde eine Spende tun können. Ein Euro pro Blume erbitten die Feldbesitzer. Das Geld soll Milchbauern zu Gute kommen. Denn die haben gegenwärtig zu kämpfen, weil der Preis pro Liter derzeit mit 21,8 Cent so niedrig ist wie schon lange nicht mehr. (SVZ09/AUG.04015 Schweriner Volkszeitung, 08.08.2009, S. 9; Sonnenfeld für die)
- 121- Ich denke, es lohnt sich sehr, ein Kind in unserer Schule anzumelden. Unabhängig **davon**, dass das neue Gebäude und moderne Unterrichtsräume auch äußerlich beste Lernbedingungen bieten, ist das Schulkonzept darauf ausgerichtet, dass alle Kinder eine Chance haben. Die Betonung liegt auf alle, **worein** wir viel Kraft investieren, ohne dies ständig zu betonen, denn es ist selbstverständlich geworden. Die Mädchen und Jungen werden ebenso auf den Übergang zum Gymnasium nach der sechsten Klasse wie auf das Erreichen eines Schulabschlusses am Ende der neunten oder zehnten

- Klasse vorbereitet. (SVZ14/FEB.04593 Schweriner Volkszeitung, 17.02.2014, S. 9; Hauptbeteiligte nicht einbezogen)
- 122- **Worein damit?** Man könnte eine Obsttüte nehmen, wie sie auch das Werkbundarchiv nutzt. Oder eine Schachtel, die immer was von Schatzkiste hat. (TSP19/JAN.02550 Der Tagesspiegel, 27.01.2019, S. 208; Nummer 11 223 aus der Serie Glänzende Aussichten)
- 123- Bundeskanzlerin Angela Merkel (CDU) auf einer Pressekonferenz mit Frankreichs Präsident Nicolas Sarkozy in BrüsselG "Weder Deutschland noch Frankreich haben im Detail dargelegt, **worin** dieser Wettbewerbspakt besteht." Luxemburgs Ministerpräsident Jean-Claude Juncker zum von Deutschland und Frankreich vorgeschlagenen Pakt für eine verbesserte Wettbewerbsfähigkeit (SAZ11/FEB.00711 SaaleZeitung, 05.02.2011, S. 31; Gipfelzitate)
- 124- Seit den 1990er Jahren wird im Erlebniskindergarten Westheim, einer "Lernwerkstatt für Kinder", nach dem Prinzip der Reggio-Pädagogik gearbeitet. So kam im Jahr 2000 der Dachboden **dazu**, der das Atelier beherbergt, **worin** die Kinder malen, handwerklich arbeiten und sich mit naturwissenschaftlichen Themen befassen können. Der Um- und Erweiterungsbau, der nun kurz vor dem Abschluss steht, war notwendig, da der Kindergarten seit zwei Jahren eine Krippengruppe anbietet, die bisher räumlich nicht optimal untergebracht war. Möglich geworden ist er durch die Krippenförderung des Freistaats Bayern, der die Einrichtung von zwölf Krippenplätzen finanziert. (SAZ11/MAR.03900 Saale-Zeitung, 31.03.2011, S. 13; Kindergarten Westheim)
- 125- Die Situation entbehrt nicht einer gewissen Ironie. Als die Vorbereitungen für das Nato-Außenministertreffen — das erste in Deutschland seit 1996 — begannen, hatte niemand auch nur die leiseste Ahnung **davon**, dass sich das Militärbündnis in einem neuen Einsatz in Libyen befinden könnte. Jetzt überlagert die Operation "Vereinte Schutzmacht" alle anderen Debatten — egal ob Afghanistan-Einsatz oder das künftige Verhältnis zu Russland. Und ausgerechnet Westerwelle, der sein Land beim Kriegseinsatz in Libyen unter keinen Umständen **dabei** haben will, hat nun zur Aufgabe, die unterschiedlichen Positionen zusammenzuführen. Die Nato hat mit solchen Momenten schon Erfahrung. Also wird betont, **worin** man sich einig ist: dass der Diktator Muammar al-Gaddafi endlich abtreten müsse, dass das geschundene Land eine politische Lösung brauche, dass es Sache der Libyer sei, über ihre Zukunft zu entscheiden. (SAZ11/APR.01952 Saale-Zeitung, 15.04.2011, S. 30; Nato-Differenzen)
- 126- Schwierige Gespräche stehen seit Mittwoch zwischen zwei wichtigen Politikern an: dem amerikanischen Präsidenten Barack Obama und dem Staatschef von China, Hu Jintao. Hu Jintao war **dafür** nach Washington, der Hauptstadt der USA, gereist. Zwischen den USA und China gibt es oft Knatsch und Misstrauen. Politikexperten sagen **dazu**: Die Beziehungen zwischen den Ländern sind angespannt. Und **worüber** sind sich die beiden nicht einig? Zum Beispiel fordern die USA, dass China politische Gefangene freilässt. (SAZ11/JAN.02237 SaaleZeitung, 21.01.2011, S. 30; ohne Titel)
- 127- Diese Dokumente über die Verhandlungen zwischen Israel und der Palästinenserführung zu publizieren, sei "ein fehlerhaftes Verhalten". Er zweifelte die Echtheit der Dokumente jedoch nicht ausdrücklich an. "Wir verbergen nichts von dem, **worüber** wir verhandeln", sagte Abbas. "Wir haben alle Informationen mit den arabischen Staaten geteilt." Der Generalsekretär der Palästinensischen Befreiungsorganisation (PLO) Jassir Abed Rabbo, warf dem Golfemirat Katar, Sitz von Al-Dschasira, gezielte Sabotage vor. Ziel der Berichte sei es, die radikal-islamische Hamas auch im Westjordanland an die Macht zu bringen. (SAZ11/JAN.02674 Saale-Zeitung, 25.01.2011, S. 27; Enthüllungen setzen Präsident Abbas unter Druck)
- 128- Wie der Sender Al-Arabija meldete, rief die NDP Mubarak zum Einschreiten auf. Der staatliche Satellitenkanal Nile News zeigte erstmals Bilder von den andauernden Protesten. **Dagegen** funktionierten Internet und Handys — **worüber** sich Regierungsgegner und Oppositionelle verständigen — zeitweise nicht mehr. "Heute Morgen hat noch mein ägyptisches Handy funktioniert, das ist aber mittlerweile tot", berichtete die Kairoer Journalistin Ragah Kamel am Freitag. "Das gleiche gilt für mein deutsches Handy. Ich habe weder Internet noch Handy zurzeit." Ein Zeitungsverkäufer rät ihr: "Wenn du in Gefahr gerätst, dann gehe ins nächste Haus, da bist du geschützt." (SAZ11/JAN.03248 SaaleZeitung, 29.01.2011, S. 31; Feuer in Kairo)
- 129- "Die Vielseitigkeit macht das Bunte aus", sagen Annette und Elmar Wehner, die schon seit vielen Jahren ins Zahlbacher Schützenhaus kommen. "Die Mischung mit den Showtänzen macht das Ganze auch nicht so eintönig." Michael Frank ist schon seit zwölf Jahren als Akteur mit den Burkardröther Kindsköpf, als Büttenredner und als Regisseur der "Junge Grüner" **dabei**. Er sagt: "Es ist ein buntes Programm von und für Jung und Alt, mit Männlein und Weiblein." Alexandra Groten vom Organisationsteam der Burkedröther Norr'n ergänzt: "Mich überraschen jedes Jahr aufs Neue die verschiedenen und verborgenen Talente vieler Mitbürger. Außerdem kommt viel Neues zur Sprache,

- worüber man auch nachdenken muss." (SAZ11/MAR.00089 Saale-Zeitung, 01.03.2011, S. 7; Der letzte Auftritt der "Kindsköpfe")*
- 130-** *Tatsächlich hat Lammert zwar nur formuliert, worüber sich der Ältestenrat aufgeregt hat. Aber vor einer Woche erst ist er der Kanzlerin mit seinem "Sargnagel"-Vergleich schon einmal in die Parade gefahren. In der CDU wird geraunt, Lammert räche sich, weil die Kanzlerin ihn nicht gefragt habe, als sie vergangenes Jahr einen Bundespräsidenten suchen musste. (SAZ11/MAR.01552 Saale-Zeitung, 12.03.2011, S. 26; Die Ruhige)*
- 131-** *Die Kanzlerin verteidigte die Enthaltung Deutschlands bei der Entscheidung des UN-Sicherheitsrats. Sie glaube, dass eine Luftoperation über Libyen "nicht hundertprozentig durchdacht" sei, sagte Merkel am Freitag in einer Libyen-Sondersitzung der Unionsfraktion. Die Bundesregierung erwägt aber, die Nato an anderer Stelle zu entlasten. "Worüber wir mit der Nato im Gespräch sind, das ist die Frage, ob wir im Rahmen der Gesamtbelastung der Nato Aufgaben übernehmen könnten zum Beispiel im Bereich von Awacs in Afghanistan", sagte Merkel. (SAZ11/MAR.02464 Saale-Zeitung, 19.03.2011, S. 25; Nato entschlossen zu Einsatz in Libyen)*
- 132-** *Britisches Flair zum Nulltarif... gibt es im Old Bailey, dem historischen Strafgerichtshof, der schon in den Romanen von Charles Dickens mitspielt. Alles sieht dort heute noch genauso aus einschließlich der Richterperücken, die man unter anderem aus "Ein Fisch namens Wanda" mit John Cleese kennt. Einfach reingehen und in irgendeinen beliebigen Prozess reinsetzen. Worum es geht, steht angeschlagen. Nur Fotoapparate und Taschen darf man nicht mitnehmen, und es gibt auch keine Schließfächer. (SAZ11/FEB.02499 Saale-Zeitung, 19.02.2011, S. 47; London muss nicht viele Pfunde kosten)*
- 133-** *Mitte des zweiten Spielabschnitts riss der Faden bei der SG. Den Gastgeberinnen gelang es, zum 16:16 auszugleichen. Das Trainerduo Jordan/Wenzke reagierte mit einer Auszeit. "Wir haben unseren Spielerinnen klar gemacht, worum es hier geht und was wir als Mannschaft wollen", so Jordan. Anschließend gelang es den SGlerinnen, wieder Ruhe in die Partie zu bekommen. Routiniert spielten sie nun die letzten Minuten herunter und verließen als verdiente Siegerinnen das Parkett. (SAZ11/FEB.03508 Saale-Zeitung, 28.02.2011, S. 16; Nach der Auszeit wieder auf Kurs)*
- 134-** *Niemand glaubt Ihnen das Versteckspiel und Sie sollten endlich den Mut aufbringen, mit offenen Karten zu spielen. Es weiß ja sowieso schon jeder, worum es geht. (SAZ11/MAR.01188 Saale-Zeitung, 10.03.2011, S. 2; ohne Titel) Sie erhalten im Laufe des Tages eine seltsame Nachricht, die Sie nachhaltig irritiert. Worum es sich dabei handelt, wird sich zeigen. Sie dürfen gespannt sein! (SAZ11/APR.00074 Saale-Zeitung, 01.04.2011, S. 2; ohne Titel)*
- 135-** *Worum geht es bei diesem Gesetz über Bildungs- und Teilhabeleistungen? Die "Hartz IV"-Sätze für Kinder regeln zwar das tägliche Leben und Überleben, aber sie tragen nicht der Tatsache Rechnung, dass Kinder auch in eine Gesellschaft hinein wachsen müssen, dass sie den Kontakt zu Gleichaltrigen brauchen. Wo das nötige Geld dazu fehlt, springen jetzt die Kommunen mit Fördermitteln ein. (SAZ11/APR.01598 Saale-Zeitung, 13.04.2011, S. 3; Die Kinder können leichter Freunde finden)*
- 136-** *Was braucht man nicht alles für eine Erstkommunion: Festkleidung, gutes Essen, Geschenke. Dabei sind das nur Nebensachen. Die achtjährige Katharina erklärt, worum es eigentlich geht. (SAZ11/APR.02771 Saale-Zeitung, 21.04.2011, S. 34; Katharinas aufregender Sonntag)*
- 137-** *Nach einem Empfang durch den neuen Bürgermeister Stephane Perrin im Salle des Fetes (Festsaal) wurden die Münnerstädter Gäste, worunter sich auch elf junge Feuerwehrleute befanden, die guten Kontakt zu den "Pompieri" von Stenay pflegen, in ihre Gastfamilien verteilt. Dort gab es überall ein mehrgängiges Abendessen, bei dem die Kontakte wieder neu geknüpft wurden. (SAZ11/APR.02710 Saale-Zeitung, 21.04.2011, S. 15; Die Gastfreundschaft ist immer riesig)*
- 138-** *Bei der Ermittlung der Verkehrsfrequenz wurden an einem Sonntag 1740 Autos gezählt, die durch den Ort fuhren. Der Spitzenwert wurde an einem Mittwoch mit 2708 Fahrzeugen erreicht, worunter 173 Lkw waren. Die meisten Geschwindigkeitsüberschreitungen wurden zwischen 22 und 9 Uhr festgestellt. Am schnellsten waren die Raser aber tagsüber. (SAZ12/JUL.03281 Saale-Zeitung, 30.07.2012, S. 10; Raser im)*
- 139-** *All das findet sich in der Musik von "Pipit Farlouse". Schon 1998 waren die Elsaß-Lothringer zu Gast beim Liedermacherfestival in Riegelsberg. Ihr leider zu kurzer Auftritt schlug ein wie eine Bombe, was den Kulturverein Riegelsberg nicht ruhen ließ. Jetzt haben sie es geschafft, das Quartett erneut zu engagieren, und der Andrang in Jochems Kneipe war groß. Kein Wunder: Wenn "Pipit Farlouse" spielen, geht auch an einem total verregneten Samstagabend die Sonne auf. "Pipit Farlouse" sind Vertreter einer sehr modernen Richtung des französischen Chansons. Begonnen haben sie mit elektrisch verstärkten Instrumenten, seit zwei Jahren aber spielen sie rein akustisch: Stimme, Klavier, Gitarre, Bass, Schlagzeug, sonst nichts. Was ihrer Musik sehr zugute gekommen ist, sie ist noch einen Tick durchsichtiger geworden. Worunter Energie und Ausstrahlung aber*

- keineswegs gelitten haben, sie sind vielleicht sogar noch besser geworden als früher. Das fängt an bei der Frontfrau, Catherine Campanella, einer ebenso zierlichen wie kraftvollen Sängerin (und Pianistin), die unter anderem Alain Souchon zu ihren Vorbildern zählt. In den leisen Passagen der Balladen erinnert ihre glockenklare Stimme an Suzanne Vega, genauso kann sie aber jazzmäßig oder regelrecht rockig singen. Campanella schreibt auch die Texte der Band; außer vom zeitlosen Thema Liebe lässt sie sich **dabei** vor allem von einer längst vergangenen Zeit inspirieren: den Kelten und ihren Sagen. Von König Artus und seinen Rittern handelte etwa ein Lied. (SBZ00/JAN.10873 Saarbrücker Zeitung, 31.01.2000; Groove und Swing mit glockenklarer Stimme)
- 140- Alles ist super, nett, hochmotivierte Lehrer gestalten den Unterricht, persönliche Beratungsgespräche finden nach dem Unterricht statt. Könnte es sein, dass einige geschickt ausgewählte Schüler überhaupt keinen Realitätsbezug zu ihrer schulischen Umgebung mehr verspüren? Im Gegensatz zum KBBZ Wadern gibt es in Merzig noch genügend Missstände, die es zu beheben gilt. Wir beanstanden, dass in Merzig eine Klassenstärke von 30 Schülern als normal angesehen werden muss, **worunter** Schüler, aber auch Lehrer leiden. Auch scheint kein Geld vorhanden zu sein, die ohnehin schon völlig veralteten Schreibmaschinen zu reparieren, **damit** der Unterricht reibungslos ablaufen kann. Wer sich die Altersstruktur der Lehrer ansieht, wird feststellen müssen, dass vorwiegend ältere Lehrer, schon etwas müde geworden, ihren Unterrichtsstoff monoton im Unterricht umsetzen und **damit** keineswegs zu den Motivationskanonen zählen, wie es sie in Wadern ausschließlich gibt. (SBZ00/APR.03164 Saarbrücker Zeitung, 08.04.2000; ohne Titel)
- 141- Das GUS-Land liegt an der Grenze Europas zu Asien. Rund 5,3 Millionen Menschen leben dort, **darunter** Griechen, Türken, Armenier und Russen. In einem Bürgerkrieg haben sich die Abchasen abgespalten und über eine Million Flüchtlinge verteilt sich im ganzen Land. Alleine 490000 leben derzeit in der Hauptstadt Tbilissi. Vor allem wegen des Bürgerkrieges hat das rund 70000 Quadratkilometer große Land, das entspricht ziemlich genau der Fläche Bayerns, enorme wirtschaftliche Probleme, **worunter** besonders alte und kranke Menschen leiden. Im Westen grenzt es an das Schwarze Meer, im Osten an Aserbaidschan und Tschetschenien. Mit seinem mediterranen Klima ist Georgien ein recht fruchtbares Land. Der Verein "Saarländisch-georgische Freundschaft Jugend und Kultur" kümmert sich um junge Leute in Georgien und um einen Austausch zwischen Georgien und dem Saarland. Informationen gibt Wolfgang Wagner unter Tel. (06853) 922992. (SBZ00/APR.08965 Saarbrücker Zeitung, 26.04.2000; Saarländisch-georgische Freundschaft...)
- 142- Den Schülern wurde durch die "Tafel" ein Interview mit einer HartzIV-Empfängerin ermöglicht. Die 47-jährige, alleinerziehende Mutter dreier Kinder arbeitet als Sporttrainerin, muss aber auch als Zimmermädchen schuften und Zeitungen austragen, um über die Runden zu kommen. Sie verdient 300 Euro im Monat, **wovon** berufliche Kosten abgehen. Nachdem sie sechs Jahre versucht hat, alleine für ihre Familie zu sorgen, bezieht sie seit einem Jahr HartzIV. Vorher sei sie zu stolz gewesen, sich als "arm" zu bezeichnen. (SAZ11/JAN.00207 Saale-Zeitung, 04.01.2011, S. 5; Die Not wohnt gleich nebenan)
- 143- Mit der Vergrößerung des Gebietes wird es auch ein neues Faltblatt geben. "Wir werden versuchen, das etwas größer aufzuziehen, nicht nur für die Gäste, sondern auch für die Einwohner der Region." Mit dem Start der Fahrradsaison im April/Mai werden dann die Elektroräder auch wieder im Haus der Schwarzen Berge auszuleihen sein. Und dass das Fahren mit dem E-Bike so richtig Spaß macht, davon haben sich Geerd Müller und seine Mitstreiter schon selbst überzeugt. Schließlich wollen sie ja auch wissen, **wovon** sie reden, wenn die Gäste nachfragen. (SAZ11/JAN.02161 Saale-Zeitung, 21.01.2011, S. 15; Stolz auf Deutschlands schönsten Wanderweg 2010)
- 144- **Wovon** diese jungen Menschen träumen und wovon sie sich fürchten, wird vor allem durch Projektionen sichtbar, für die in unterschiedlicher Ästhetik sowohl Bühnenbildner Patrick Bannwart als auch Kostümbildner Falko Herold verantwortlich zeichnen. Mal wirken diese Bilder wunderbar poetisch und märchenhaft, dann wieder sind sie so aufdringlich und aufgebläht, dass sie die Protagonisten zu Statisten degradieren. (SAZ11/JUL.02754 Saale-Zeitung, 23.07.2011, S. 35; Pubertärer Geniestreich mit einigen Fragezeichen)
- 145- **Wozu** die DSV-Läufer dennoch imstande sind, zeigte der Frankenhainer Jens Filbrich. Auf dem klassischen Teilstück diktierte er das Rennen mit. "Ich habe mich richtig gut gefühlt. Doch dann kam ein Sturz **dazwischen**, und schon waren 20, 25 Plätze verloren. Zudem habe ich mir das Schienbein richtig wehgetan und konnte im Skating nicht so, wie ich wollte", sagte der Thüringer. (SAZ11/JAN.00148 Saale-Zeitung, 04.01.2011, S. 20; Teichmann steigt nach Rang 56 aus)
- 146- Eine Bürgerversammlung nur mit Frauen? Das gibt es nur in Fuchstadt und nur im Fasching. Da wurde der Bürgermeister alias Edgar Stockmann schon auch gefragt, **wozu** Sonnenkollektoren auf der Nordseite gut seien. Ob jeder dritte Fuchster hässlich ist, sei dahingestellt. Treffe dies auf sie

- zu, habe sie wenigstens die "Mittlere Reife", sagte Edith Stockmann. (SAZ11/FEB.02594 Saale-Zeitung, 21.02.2011, S. 9; ohne Titel)
- 147- Das Fernsehen lieferte die Vorlage für das Publikumsspiel "Die schnelle Minute", bei dem Frank Wenzel Saalkandidaten zur Gaudi des Publikums in verschiedenen Wettkämpfen gegeneinander antreten ließ. **Wozu** ein Männerballett in der Lage ist, zeigte die "Ramsthaler Puppenkiste". Seelefant Thorsten Kühnlein, Urmel Frank Herterich, eine Erdmännchenkolonie, die Blechbüchsenarmee oder die Piratenbande — alle Figuren überzeugten. (SAZ11/MAR.00191 Saale-Zeitung, 02.03.2011, S. 11; Von Erdmännchen bis Elfen)
- 148- Der CDU-Gesundheitsexperte Jens Spahn zeigte sich enttäuscht von der Stellungnahme des Rats aus Medizinern, Naturwissenschaftlern, Juristen, Philosophen und Theologen. "Es stellt sich einmal mehr die Frage, **wozu** es einen Ethikrat braucht", sagte der Anhänger eines kompletten Verbots der "Rheinischen Post". Seine FDP-Kollegin Ulrike Flach verteidigte den Rat. "Es gibt keine vereinheitlichte Ethik", sagte die Anhängerin einer Zulassung. Der Sprecher der Arbeitsgruppe Fortpflanzungsmedizin des Rats, Wolf-Michael Catenhusen, meinte, die Stellungnahme sei hilfreich - Positionen seien überprüft, Aspekte präzisiert worden. (SAZ11/MAR.01133 Saale-Zeitung, 09.03.2011, S. 24; Verstoß gegen das Recht auf Leben?)
- 149- Antrag auf Ratenzahlung "Sie sollten künftig besser **darauf** achten, **wozu** sie sich verpflichtet haben", sagte Dümig. Der Kläger folgte schließlich dem Rat des Gerichts, seine Klage zurückzunehmen und **dadurch** wenigstens einen Teil der Prozesskosten zu sparen. (SAZ11/MAR.02127 Saale-Zeitung, 17.03.2011, S. 16; Ein aussichtsloses Unterfangen)
- 150- Überhaupt kann man sich über den Sinn und Unsinn der Fragen streiten. Aber muss man denn bei "Wer wird Millionär?" wirklich alles wissen? "Power of Ten" ist zumindest insofern aufschlussreich, weil man erfährt, wie die Deutschen so ticken. Und weil man denken muss, was sie denken oder zumindest, was sie sagen, was sie denken würden, **wozwischen** manchmal ein kleiner, aber feiner Unterschied besteht. Soll man zum Beispiel tatsächlich glauben, dass 70 Prozent im Falle von Arbeitslosigkeit für fünf Euro die Stunde Spargel stechen würden? Oder dass nur sieben Prozent Versicherungsbetrug für ausgleichende Gerechtigkeit halten? Solche Fragen sind immerhin noch einigermaßen relevant, auch die **danach**, wie viel Prozent der Deutschen sich im Auslandsurlaub schon einmal geschämt haben, Deutscher zu sein. Aber will man unbedingt wissen, wie viele Menschen auf der Toilette telefonieren, "wo das Besetztsymbol eine ganz andere Bedeutung bekommt", wie Bach kalauert? Da würde man doch gern mal Mäuschen sein, wenn sich die Redaktion wo auch immer die Fragen ausdenkt. (STZ08/APR.05276 Stuttgarter Zeitung, 21.04.2008, S. 16; Telefonieren viele auf dem Klo?)
- 151- Ich sage das gerade vor dem Hintergrund des zwanzigsten Mauerfall-Jubiläums. In der DDR erlaubte sich das Politische Übergriffe in allzu viele Lebensbereiche. Die Machthaber hatten nicht nur den Willen, sondern auch den **dazu** nötigen Staat, Gesetze zu lenken und zu leiten, Lebensentwürfe aufgehen oder scheitern zu lassen. Die Bürger der DDR waren Staatsbürger - unabhängig **davon**, ob sie ihrem Staat zustimmend oder ablehnend gegenüber standen. Das Politische wehte einem jedem als eisiger Hauch im Nacken. Erschwerend kam hinzu, daß sich die Verhältnisse nicht einfach abwählen ließen; deshalb waren die Wahlen in der DDR nichts, was die Bezeichnung Wahlen verdient hätte. Es gab nichts, **wozwischen** man wählen konnte; das deutlichste Bekenntnis bestand darin, nicht zur Wahl zu gehen. (TSP09/AUG.03105 Der Tagesspiegel, 24.08.2009; Warum wählen? // Nicht der Nichtwähler ist eine Gefahr für die Demokratie, sondern die Wahlkämpfe sind es / Von Thomas Brüssig)
- 152- Die Commode war mit Röcken, Wäsche, Büchern pyramidenförmig aufgehäuft, alle Schubladen herausgezogen, in den Ecken waren Gewehre aller Art aufgefplant, die zwei vorhandenen Stühle waren besetzt mit Büchern, Briefschaften, Hausgerät, **wozwischen** rote Tücher als Friedensfahnen herahingen", beschrieb Wilhelm dem Bruder Jacob die spärlich möblierte Dreizimmerwohnung Arnims: "Hier schreibt er mitten in dieser Unordnung die herrlichsten und göttlichsten Dinge." (TSP12/DEZ.02025 Der Tagesspiegel, 16.12.2012; Es war einmal // ... eine Künstler-WG in der Mauerstraße. Dort lernt der junge Wilhelm viel über Berlin. Weihnachten 1812 veröffentlicht er mit seinem Bruder Jacob Grimm sein erstes Märchenbuch. Auch dank seiner neuen Berliner Freunde)
- 153- "Mit den Donauwellen ist das Leitmotiv vorgegeben: Es geht um flüssige Übergänge, egal **wozwischen**, wieso und wohin." (HAA10/DEZ.00902 Hamburger Abendblatt, 07.12.2010, S. 15; Querschläger)
- 154- Der Ort symbolisierte für beide ihre Zerrissenheit zwischen Berlin und Hamburg, **wozwischen** Hallaschka regelmäßig pendelte, um beim NDR das Verbrauchermagazin "Markt" und in Vertretung die "NDR Talk Show" zu moderieren. "Wir wollten unsere Freunde auf halber Wegstrecke treffen." Den Pastor vor Ort konnten sie allerdings nicht **davon** überzeugen, dass nur der Geistliche aus Anne-Katrins Heimatdorf die kirchliche Trauung übernehmen sollte. "Als wir in

die Kirche kamen, standen zwei Pastoren am Altar. Das fand ich erst befremdlich und dann ziemlich gut", erinnert Hallaschka sich. (HAA18/DEZ.02427 Hamburger Abendblatt, 11.12.2018, S. 6002; "Meine Prominenz ist nur geliehen")

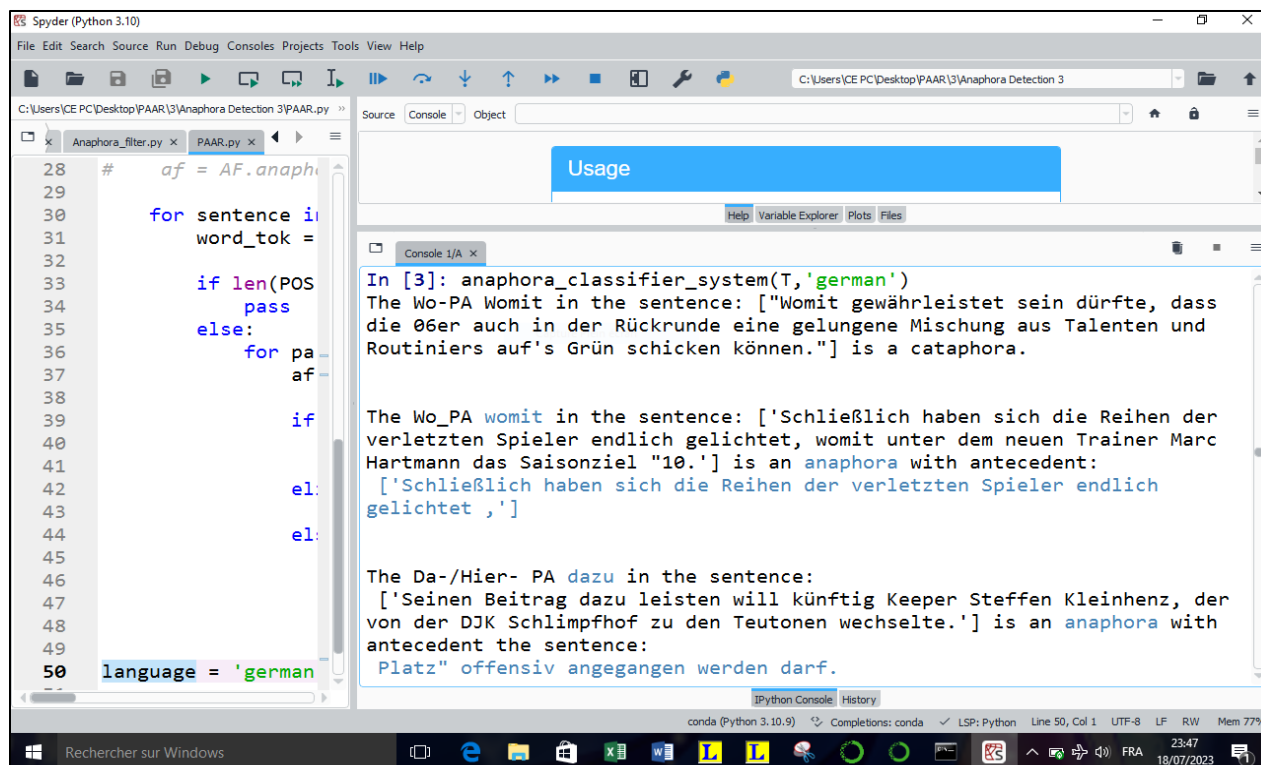
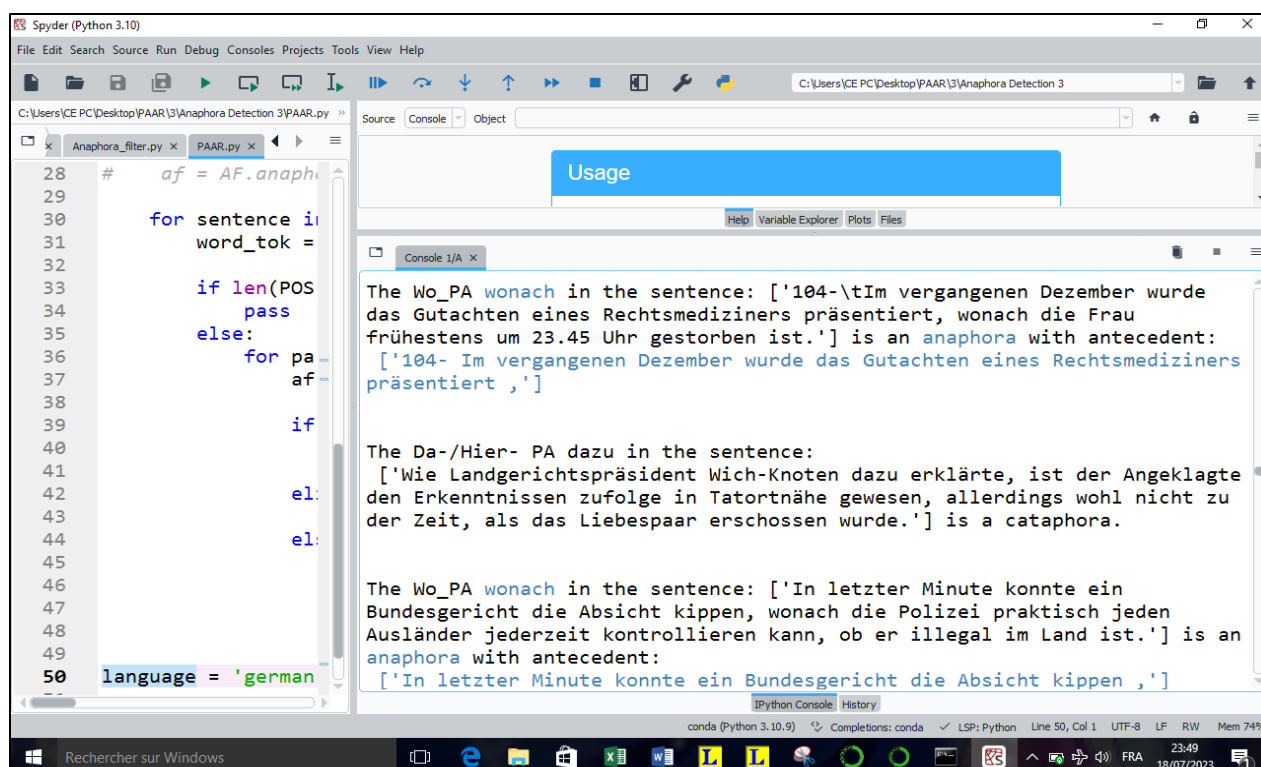


Abbildung 125: Einsatz des PAAR auf den Textabschnitt (N°101-154) (1)

Die erste Erscheinung des PA "Womit" in der Abbildung 118 ist keine Katapher, sondern eher eine Anapher mit dem Textteil "Ähnliches....Bezirksliga-Elf." Als Antezedens. Was die zweite Erscheinung angeht, wurde es vom PAAR richtig aufgelöst. Das dritte PA der Liste ("Dazu") ist zwar anaphorisch aber mit dem Textteil "Schliesslich.... werden darf" als Antezedens.



Das PA "Womit" in der Abbildung 119 ist zwar anaphorisch aber mit dem Textteil "bis sie was finden" als Antezedens. Was das PA "damit" angeht, handelt es sich tatsächlich um eine Katapher.

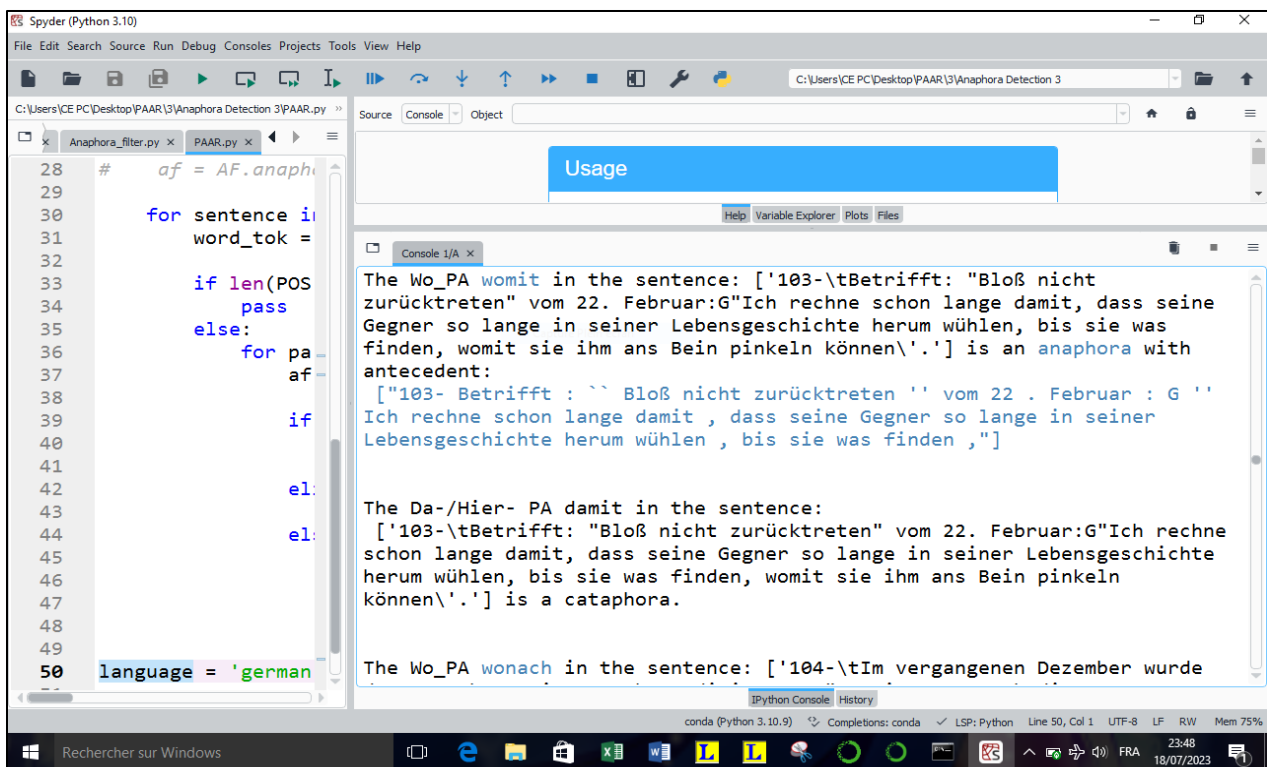


Abbildung 127: Einsatz des PAAR auf den Textabschnitt (N°101-154) (3)

Die beiden Erscheinungen des PA "wonach" in der Abbildung 120 sind zwar anaphorisch aber mit den Textteilen "Das Gutachtens eines Rechtsmedeziners" und "Die Absicht" als jeweiliges Antezedens. Was das PA "dazu" angeht, handelt es sich um keine Katapher sondern um eine Anapher mit dem Textteil "Dass beide Opfer um 23.30 Ihr...unwahrscheinlich" als Antezedens.

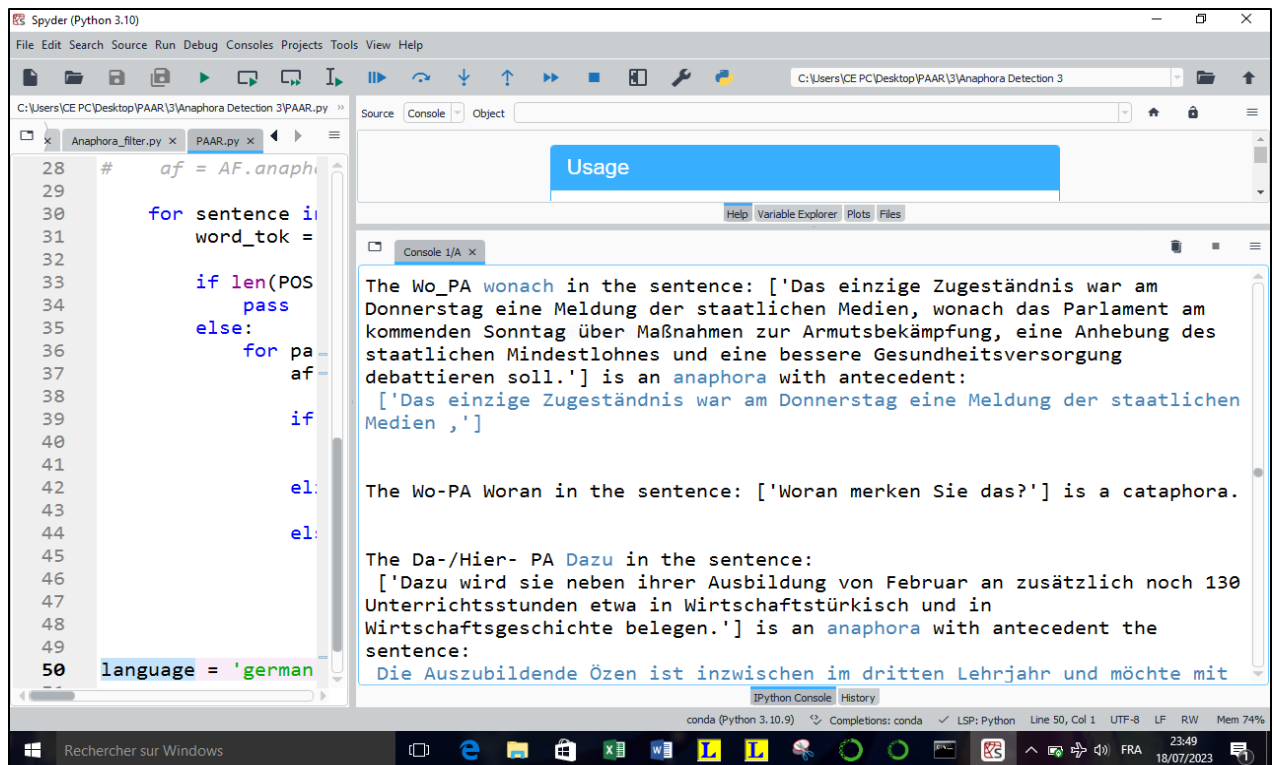


Abbildung 128: Einsatz des PAAR auf den Textabschnitt (N°101-154) (4)

Das PA "Wonach" in der Abbildung 121 ist zwar anaphorisch aber bezieht sich dabei auf die NP "eine Meldung der stattlichen Medien". Was das PA daran angeht, handelt es sich tatsächlich um keine Anapher. Weiterhin ist das PA "dazu" richtig vom PAAR aufgelöst worden.

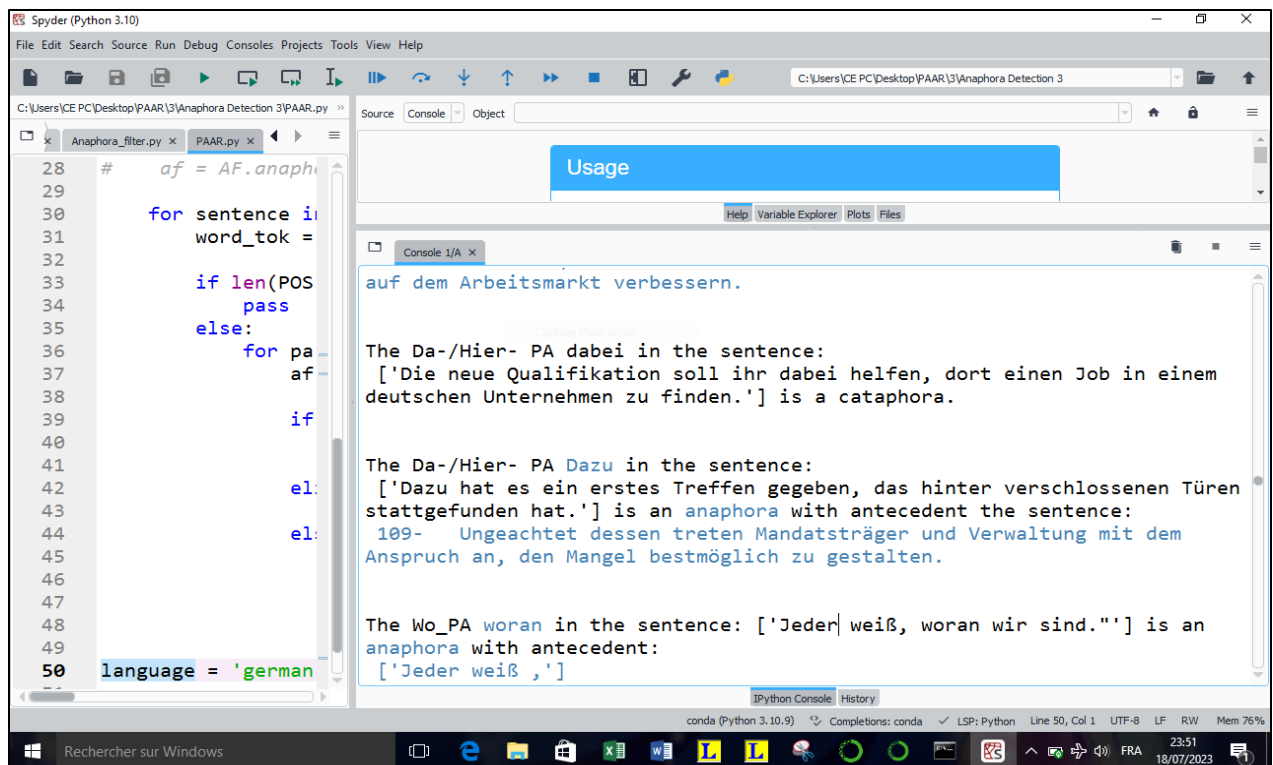
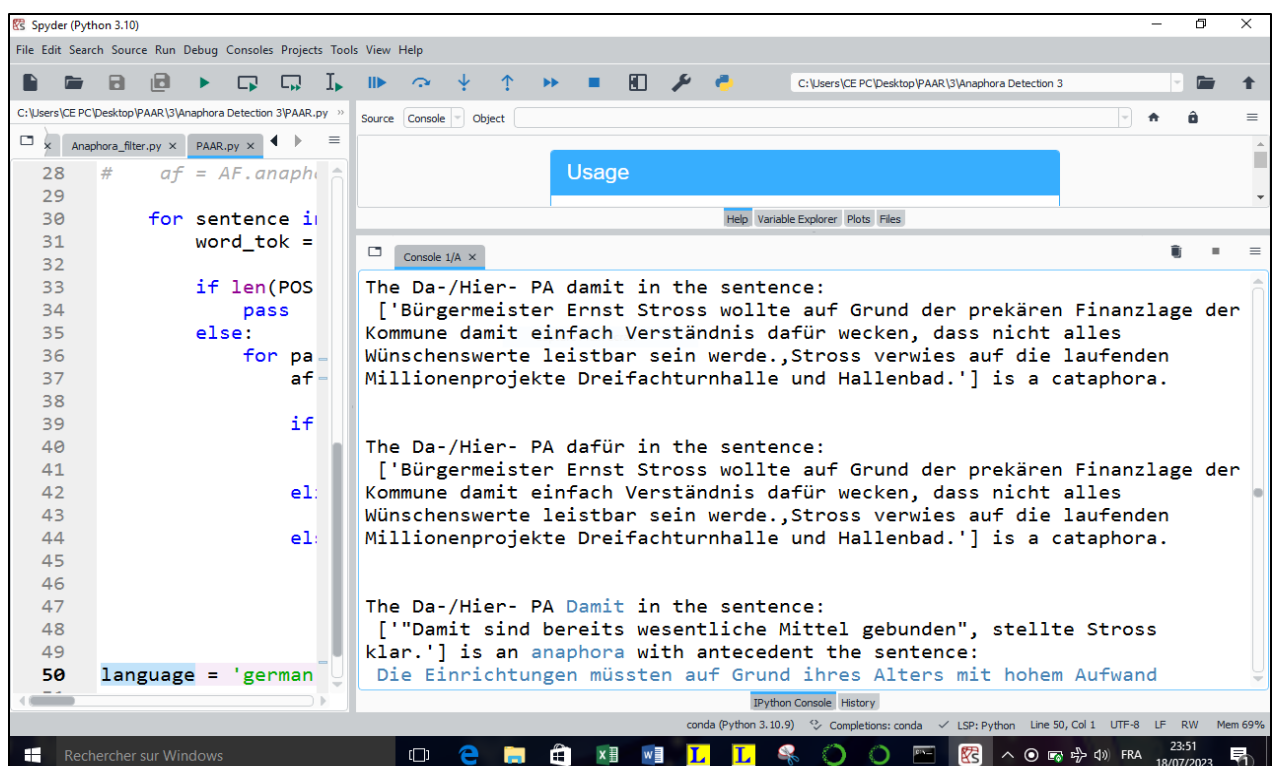


Abbildung 129: Einsatz des PAAR auf den Textabschnitt (N°101-154) (5)

Das PA "dabei" in der Abbildung 122 ist tatsächlich eine Katapher, denn es nimmt Bezug auf vorwärtsstehende sprachliche Einheiten. Was das PA "Dazu" angeht, wurde es nur zum Teil richtig aufgelöst und zwar, weil das richtige Antezedens ist nur der Textil "den Mangel bestmöglich zu gestalten". Weiterhin merkt man, dass das PA "woran" keine Anapher ist, sondern eher ein Deixis.



Das PA "damit" erscheint zweimal in der Abbildung 123 und wurde dabei richtig vom PAAR verarbeitet. Was das PA "darfür" angeht, handelt es sich tatsächlich um eine Katapher.

Das PA "dazu" in der Abbildung 124 ist richtig aufgelöst worden. Was die erste Erscheinung des PA "worauf" angeht, handelt es sich um ein Deixis ; was aber die zweite Erscheinung angeht, handelt es sich tatsächlich um eine Anapher aber mit der PP "mit dem Hinweis" als Antezedens (Abbildungen 125).

Das PA "damit" in der Abbildung 125 ist zwar anaphorisch, aber mit dem Textteil " Was es den beteiligten Croupiers leicht machte" als Antezedens. Was die erste Erscheinung vom PA "Worauf" angeht, handelt es sich tatsächlich um eine Katapher. Die zweite Erscheinung wurde ebenso richtig aufgelöst.

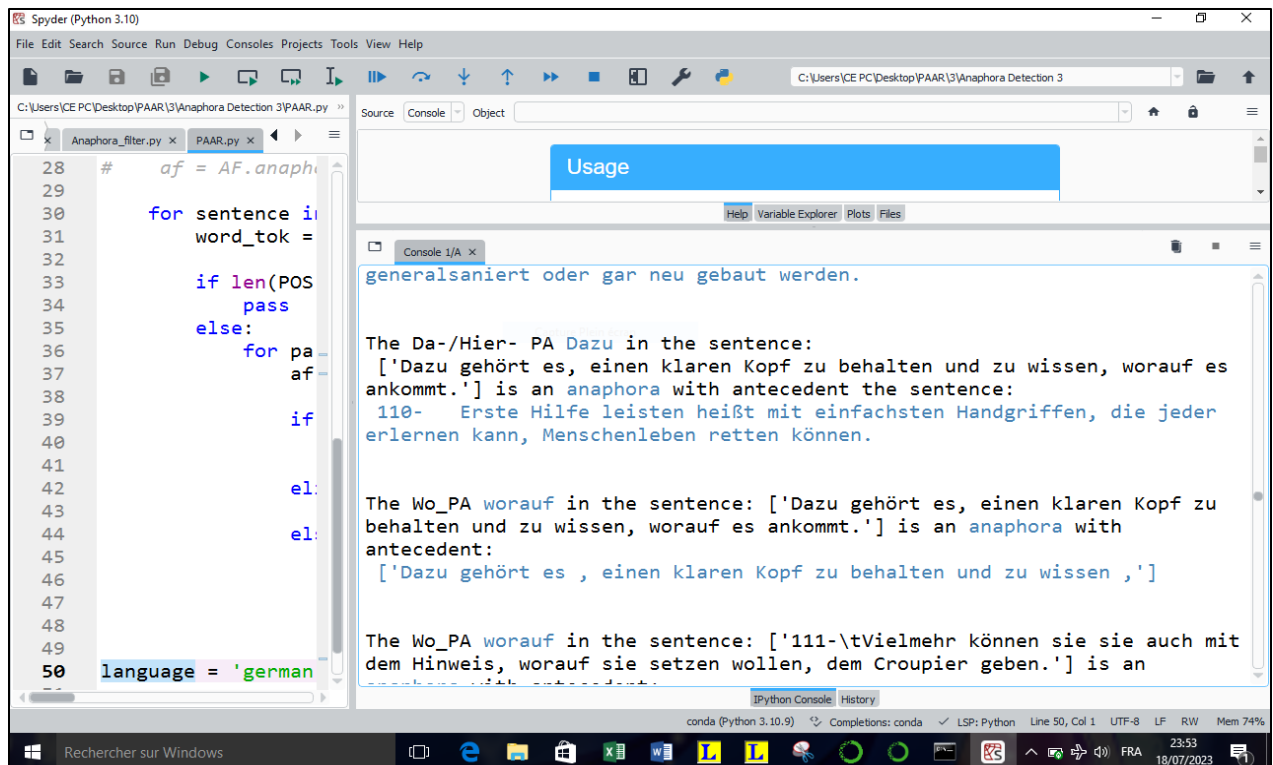


Abbildung 131: Einsatz des PAAR auf den Textabschnitt (N°101-154) (7)

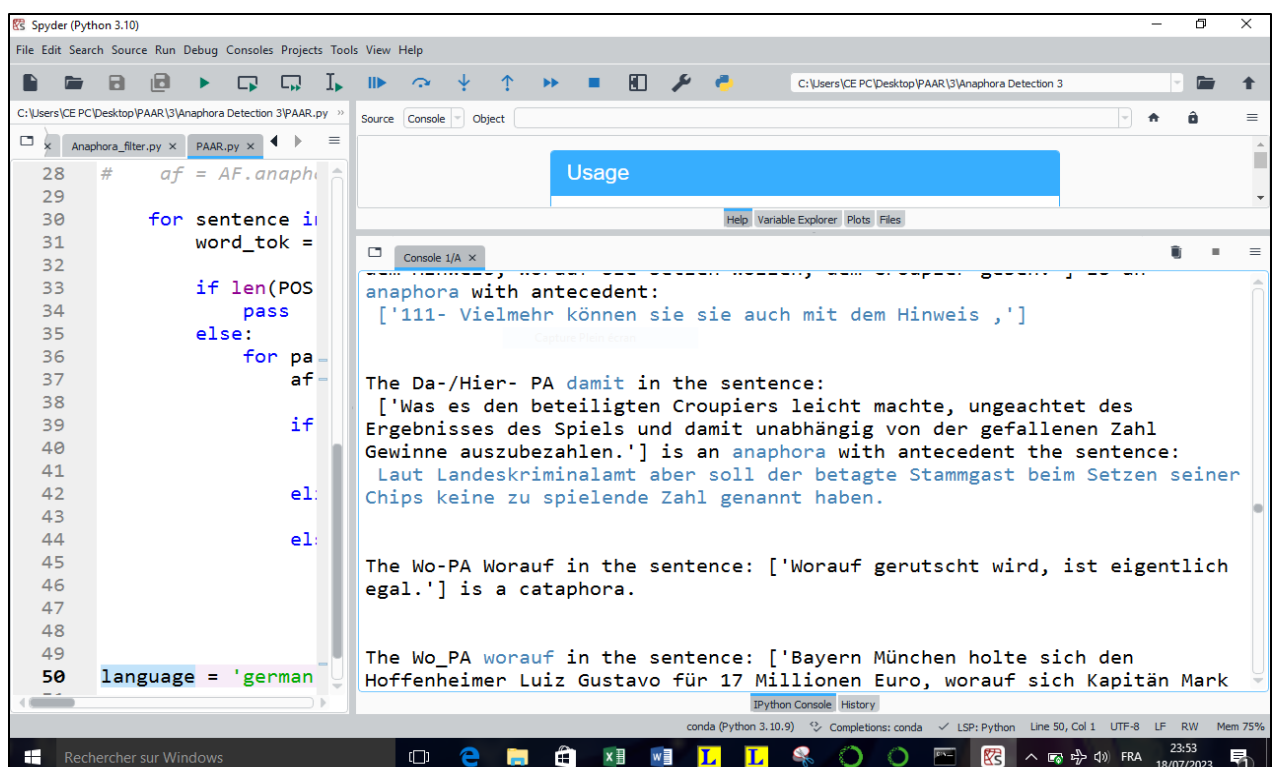


Abbildung 132: Einsatz des PAAR auf den Textabschnitt (N°101-154) (8).

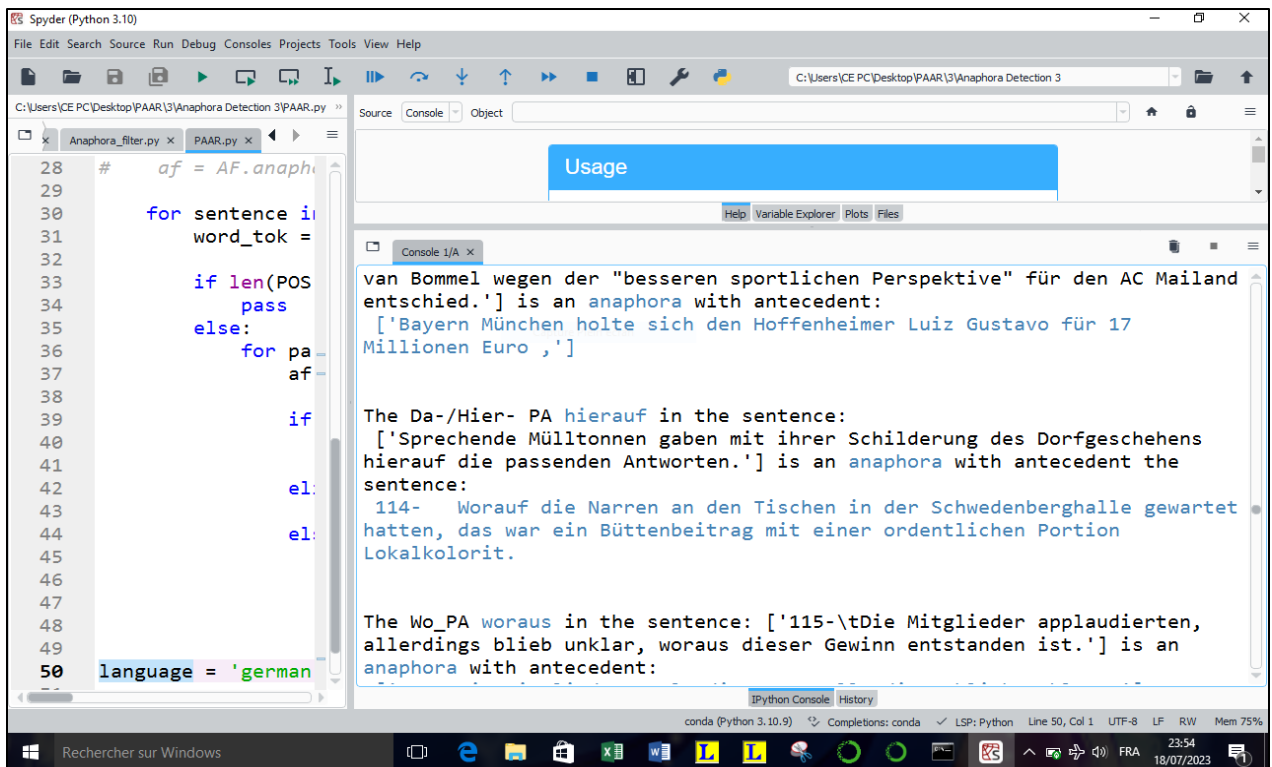


Abbildung 133: Einsatz des PAAR auf den Textabschnitt (N°101-154) (9)

Das PA "hierauf" in der Abbildung 126 ist richtig aufgelöst worden. Was das PA "woraus" angeht, handelt es sich eher um eine deiktische Erscheinung und um keine Anapher, wie PAAR es herausgefunden hat.

Das PA "Woraus" in der Abbildung 127 erscheint zweimal. Die erste Erscheinung ist aber keine Katapher sondern eine Anapher mit dem Textteil "wenn die Schwaden.....Weg machen". Was die zweite Erscheinung aber angeht, ist sie ebenso, wie die einzige Erscheinung des PA "worauf" deiktisch. Was das letzte PA der Liste ("Darunter") angeht, handelt es sich um keine Katapher, sondern eher um eine Anapher mit dem Textteil "über gegebene Themen" als Antezedens.

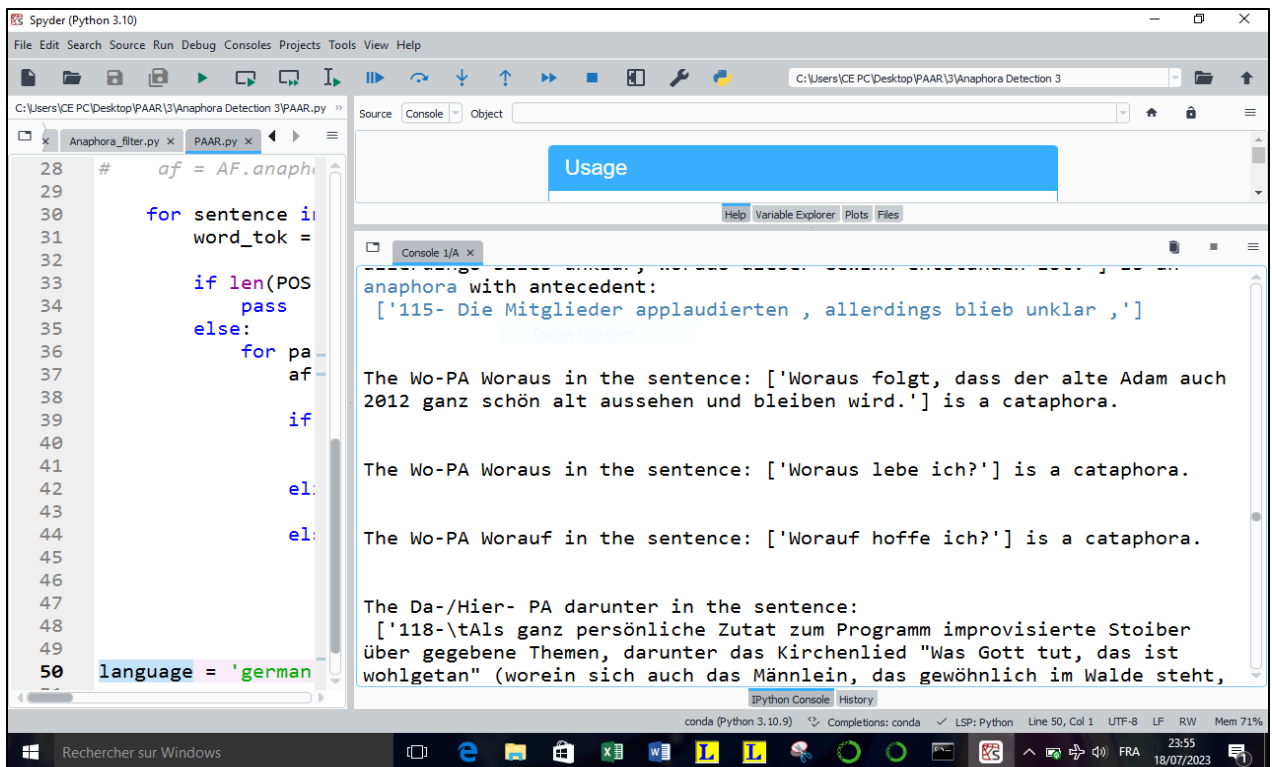


Abbildung 134: Einsatz des PAAR auf den Textabschnitt (N°101-154) (10)

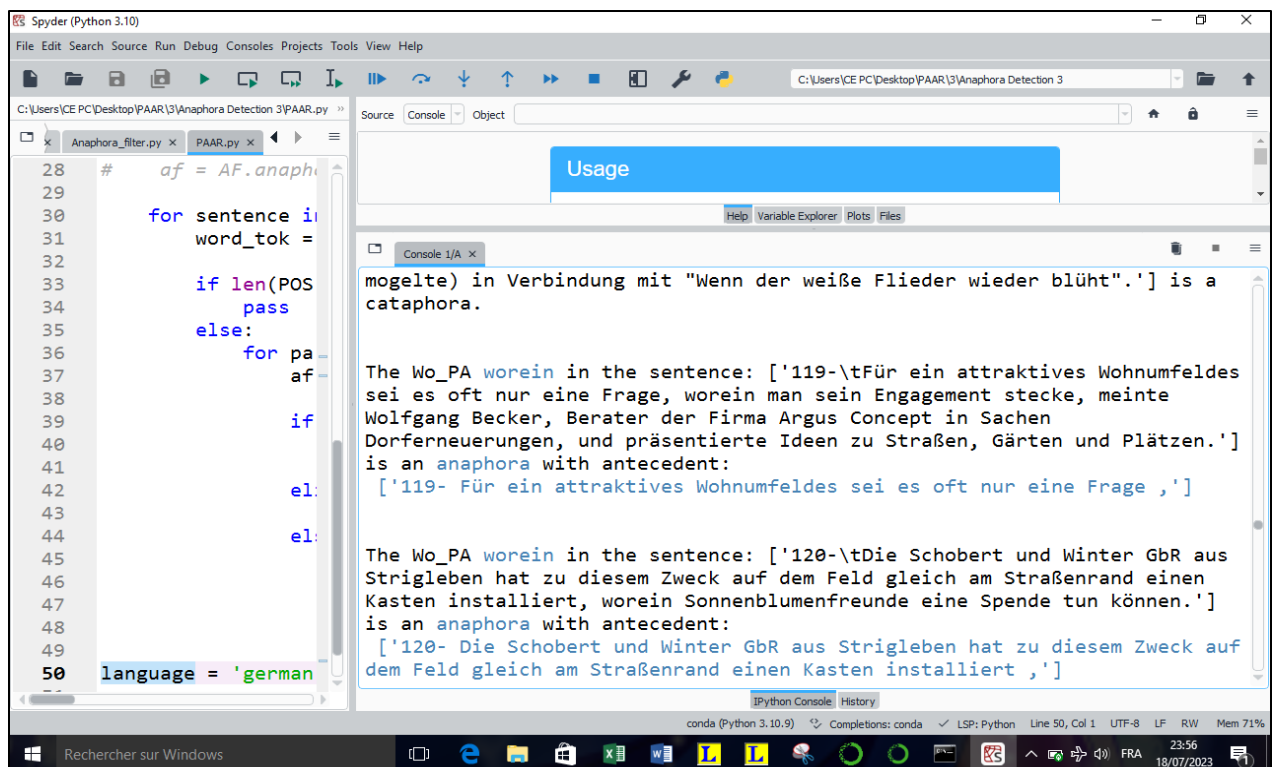


Abbildung 135: Einsatz des PAAR auf den Textabschnitt (N°101-154) (11)

Das PA "worein" in der Abbildung 128 ist zwar anaphorisch aber mit Bezug auf die NP "eine Frage". Die gleiche Überlegung gilt für die zweite Erscheinung von "worein" aber dieses Mal mit der NP "einen Kasten"

Die PA "darauf" & "davon" in der Abbildung 129 sind tatsächlich kataphorisch. Was das PA "worein" angeht, handelt es sich zwar um eine Anapher aber mit Bezug auf die PP "auf alle"

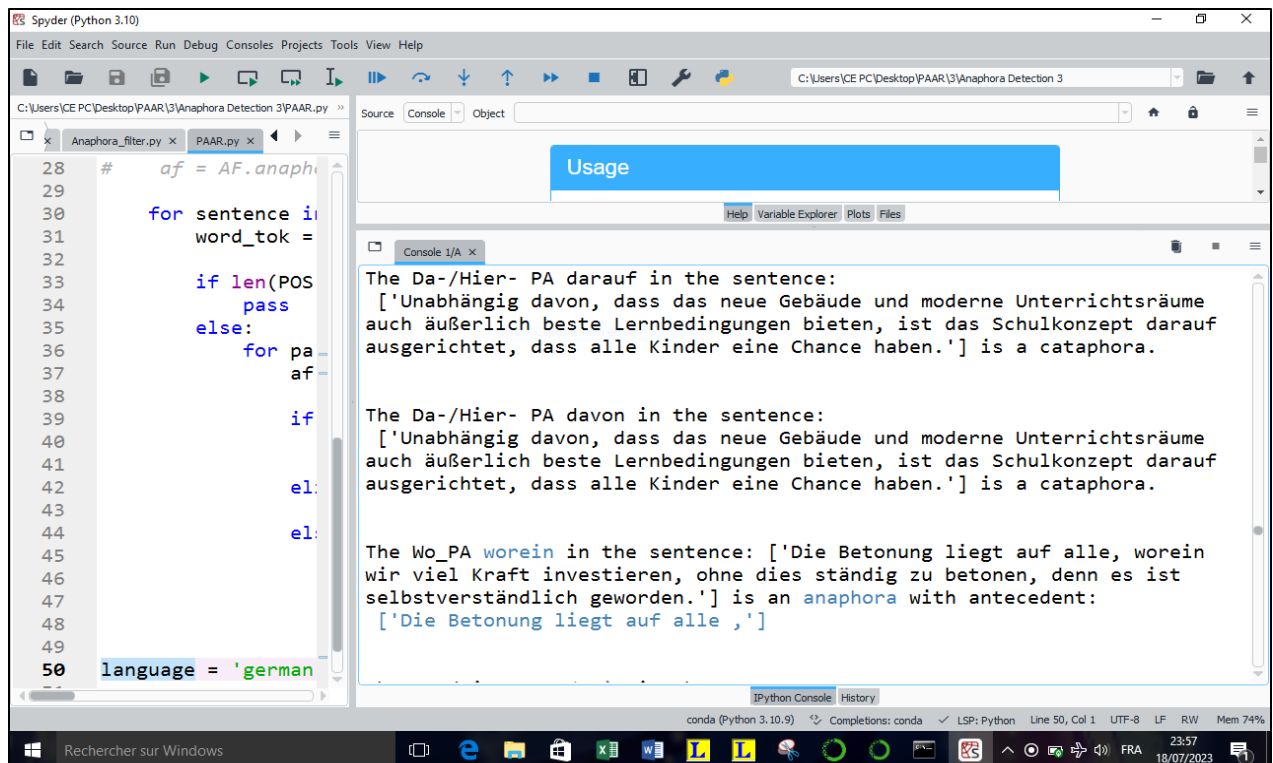


Abbildung 136: Einsatz des PAAR auf den Textabschnitt (N°101-154) (12)

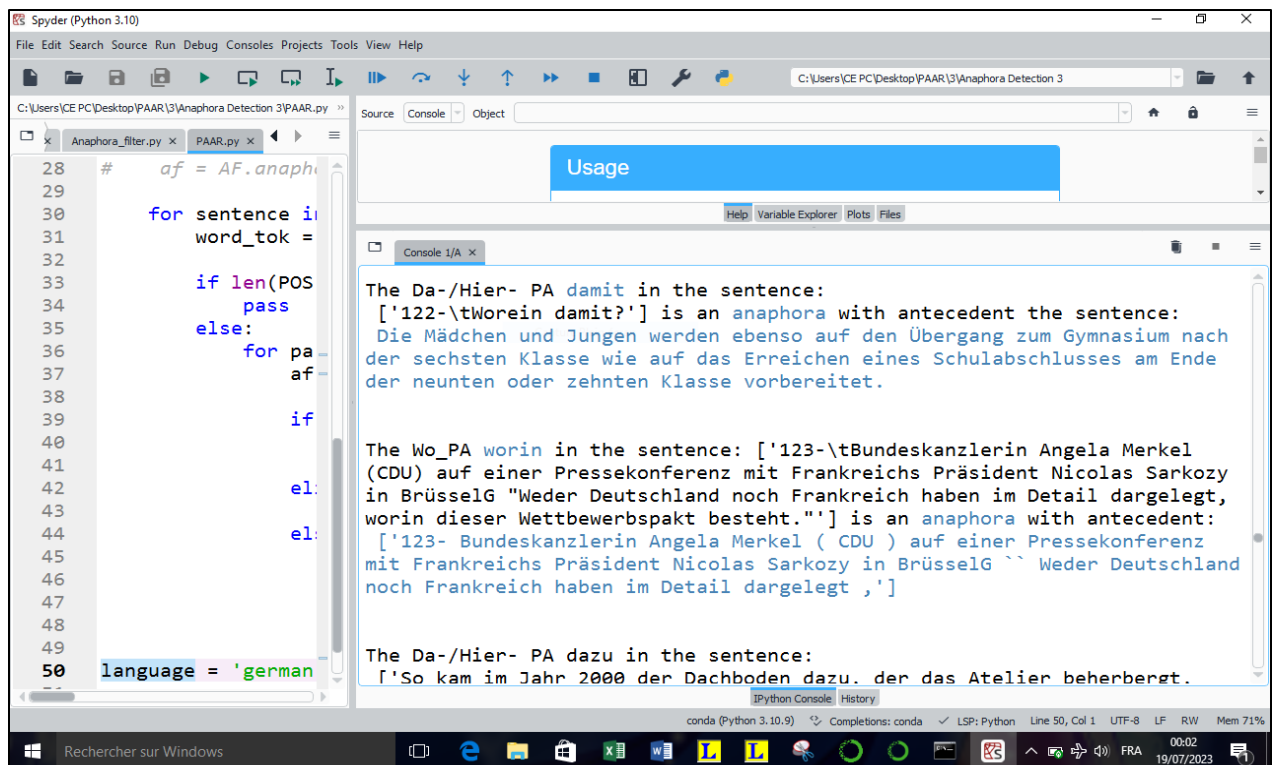


Abbildung 137: Einsatz des PAAR auf den Textabschnitt (N°101-154) (13)

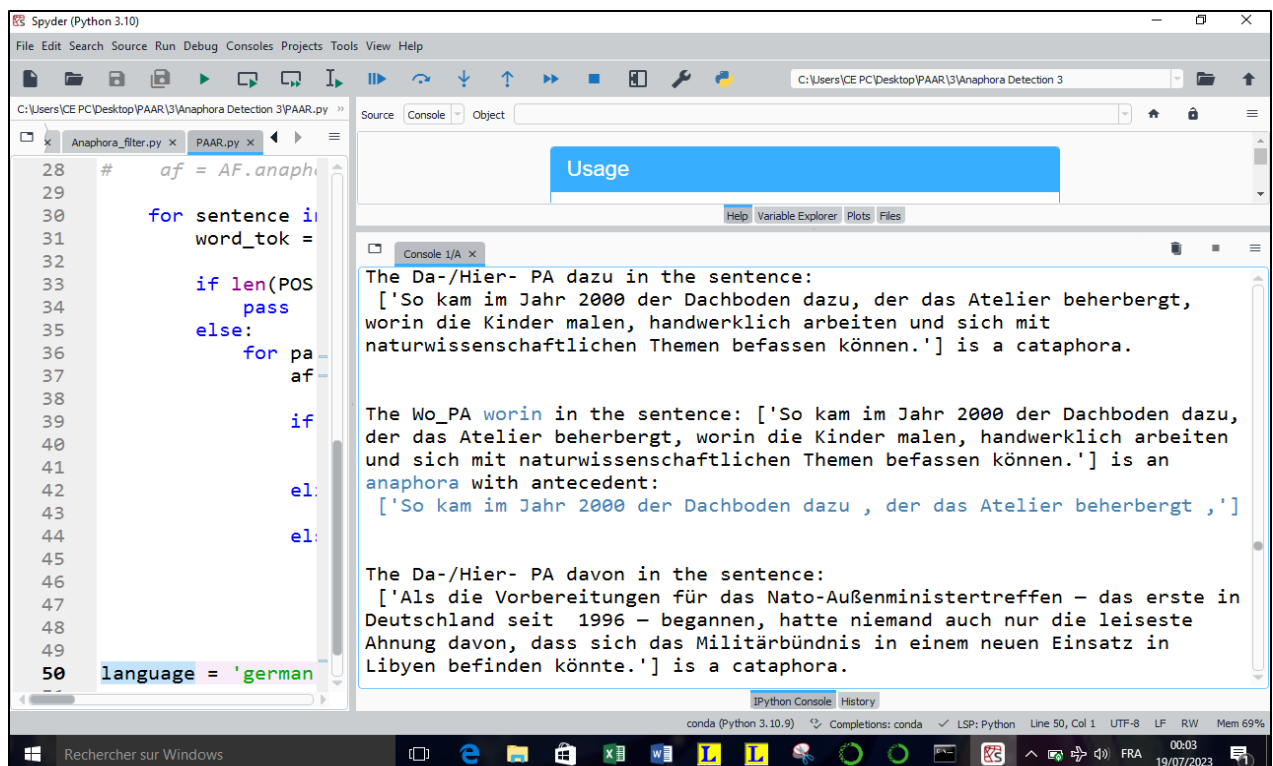


Abbildung 138: Einsatz des PAAR auf den Textabschnitt (N°101-154) (14)

Aus der Abbildung 131 ergibt sich, dass PAAR 2 PA gefunden hat und es zugleich als anaphorisch klassifiziert. Was das PA "damit" angeht, handelt es sich zwar um eine Anapher aber mit keinem bestimmtem Antezedens. Denn der Textausschnitt ist nicht in seiner Ganzheit abrufbar. Was das PA "Worin" angeht, handelt es sich um keine Anapher, sondern eher um eine Katapher.

Das PA "dazu" in der Abbildung 131" ist keine Katapher, sondern eine Anapher mit Bezug auf"

Lernwerkstatt für Kinder". Was das PA "worin" angeht, handelt es sich zwar um eine Anapher aber mit der NP "das Atelier" als Antezedens. Das PA "Davon" hingegen wurde gut kategorisiert.

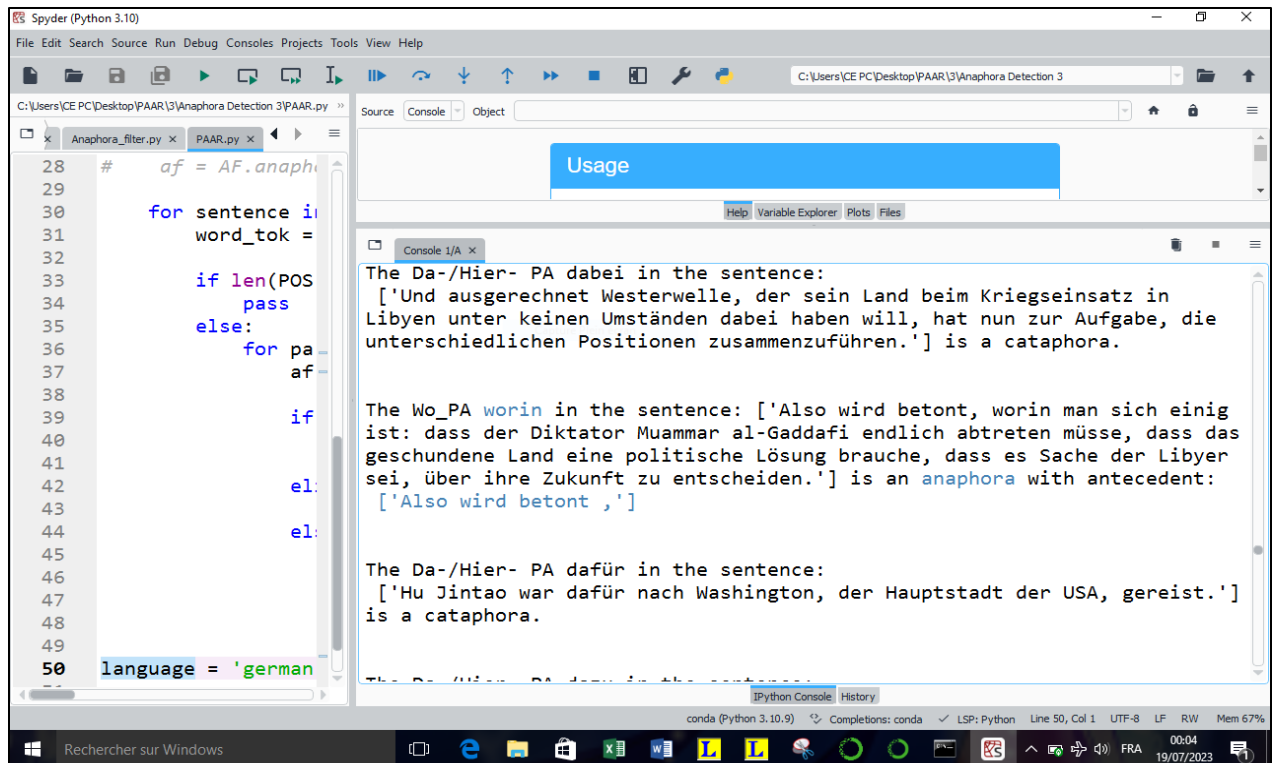


Abbildung 139: Einsatz des PAAR auf den Textabschnitt (N°101-154) (15)

Das "Dabei" aus der Abbildung 132 bezieht sich auf die PP "beim Kriegseinsatz". Das PA "worin" der gleichen Abbildung ist keine Anapher, sondern eine Deixis. Was das weitere PA ("dafür") angeht, handelt es sich um keine Katapher, wie PAAR es gemeint hat, sondern eher um eine Anapher mit Bezug auf den Textteil "Schwierige Gespräche stehen.....Hi Jintao"

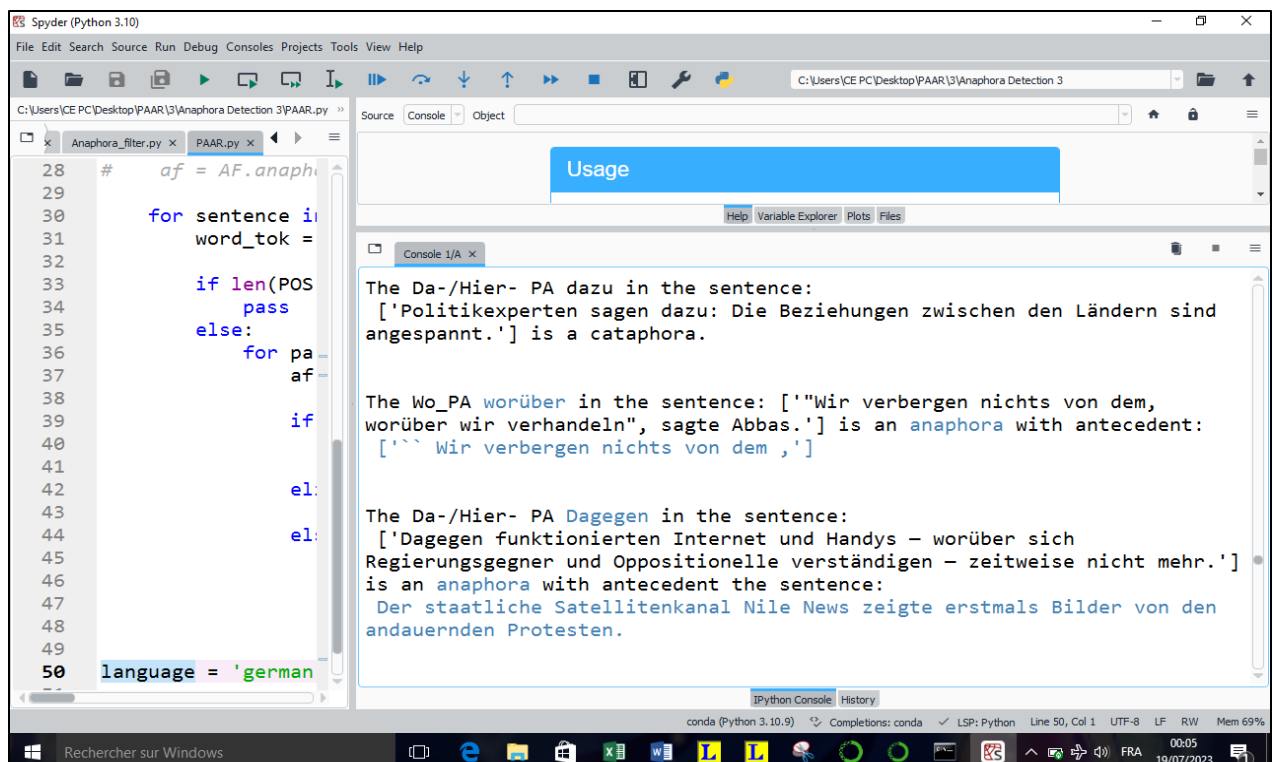


Abbildung 140: Einsatz des PAAR auf den Textabschnitt (N°101-154) (16)

Das PA "dazu" in der Abbildung 133 ist keine Katapher, sondern eine Anapher mit dem Textteil "Zwischen den USA und China gibt es oft Knasch und Misstrauen". Das PA "worüber" ist zwar anaphorisch aber nur mit Bezug aufs Pronomen "dem". Dies hingegen wurde das PA "dagegen" richtig aufgelöst.

Das PA "damit" in der Abbildung 134 ist zwar anaphorisch aber mit Bezug auf "als Akteur mit den Burkardröhter...der 'Junge grüner'". Im Gegensatz dazu wurden die beiden Erscheinungen des PA "worüber" richtig aufgelöst.

Das PA "worum" in der Abbildung 135 erscheint 4 Male. Und funktioniert bei allen diesen Erscheinungen immer als ein kataphorischer Ausdruck bzw. als keine Anapher. Etwas das PAAR nur zum Teil festgestellt hat.

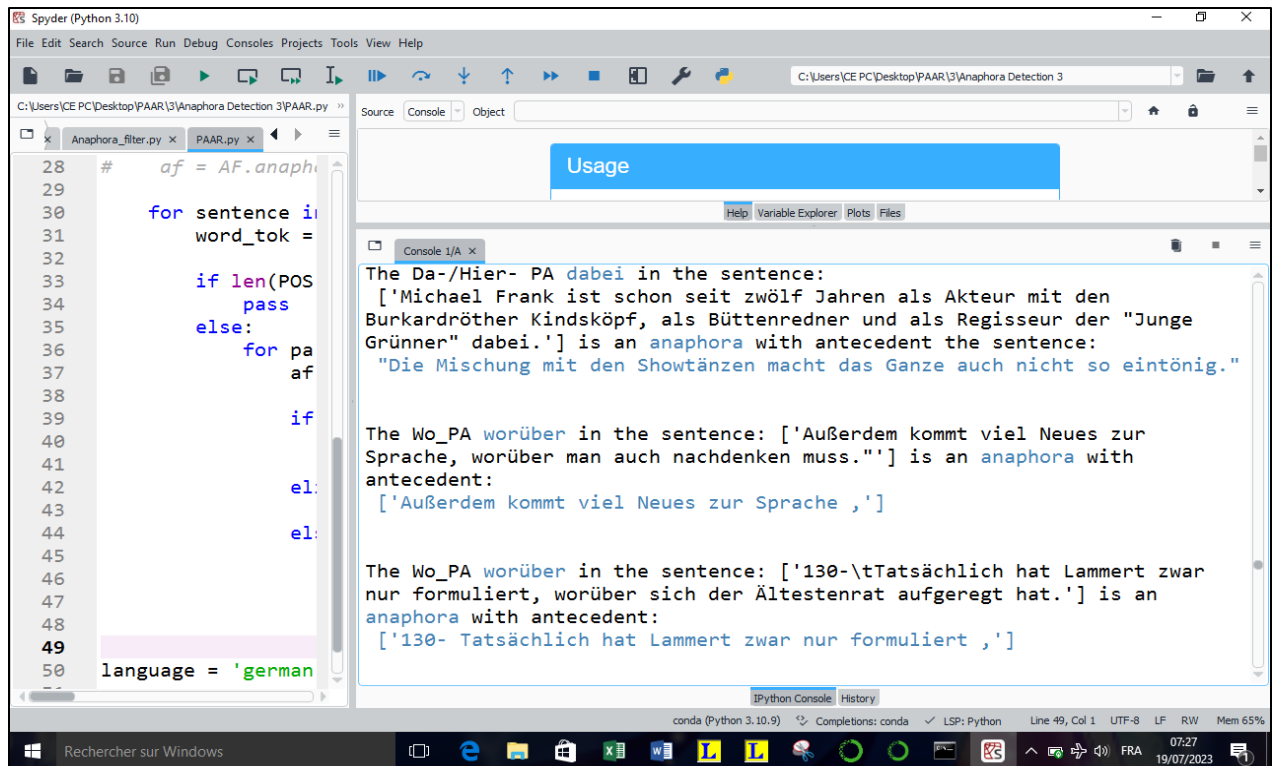


Abbildung 141: Einsatz des PAAR auf den Textabschnitt (N°101-154) (17)

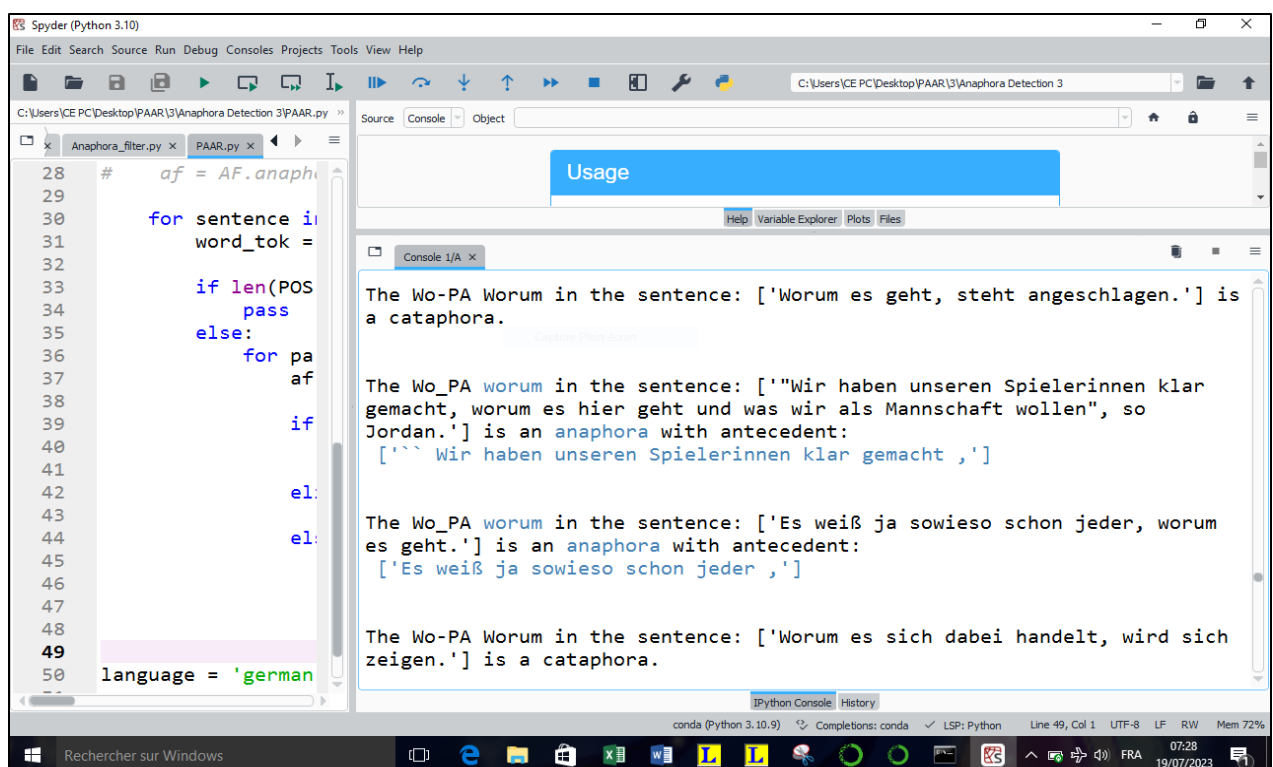


Abbildung 142: Einsatz des PAAR auf den Textabschnitt (N°101-154) (18)

Das PA "dabei" erscheint zwei Male in der Abbildung 136. Einmal als Katapher und ein anderes Mal als Anapher. Während es in der zweiten Erscheinung tatsächlich um eine Anapher geht, welche eben auch richtig aufgelöst worden ist, handelt es sich bei der ersten Erscheinung um keine Katapher, sondern um eine Anapher mit dem Textteil "....." Als Antezedens. Was die Erscheinungen der PA "dazu" und "worum" angeht, merkt man auch den gleichen Fehler vom PAAR. "Worum" ist keine Anapher, sondern entweder ein Deixis oder eine Katapher. Was "dazu" angeht ist es keine Katapher, sondern eine Anapher mit Bezug auf ", dass sie den Kontakt zu Gleichaltrigen brauchen."

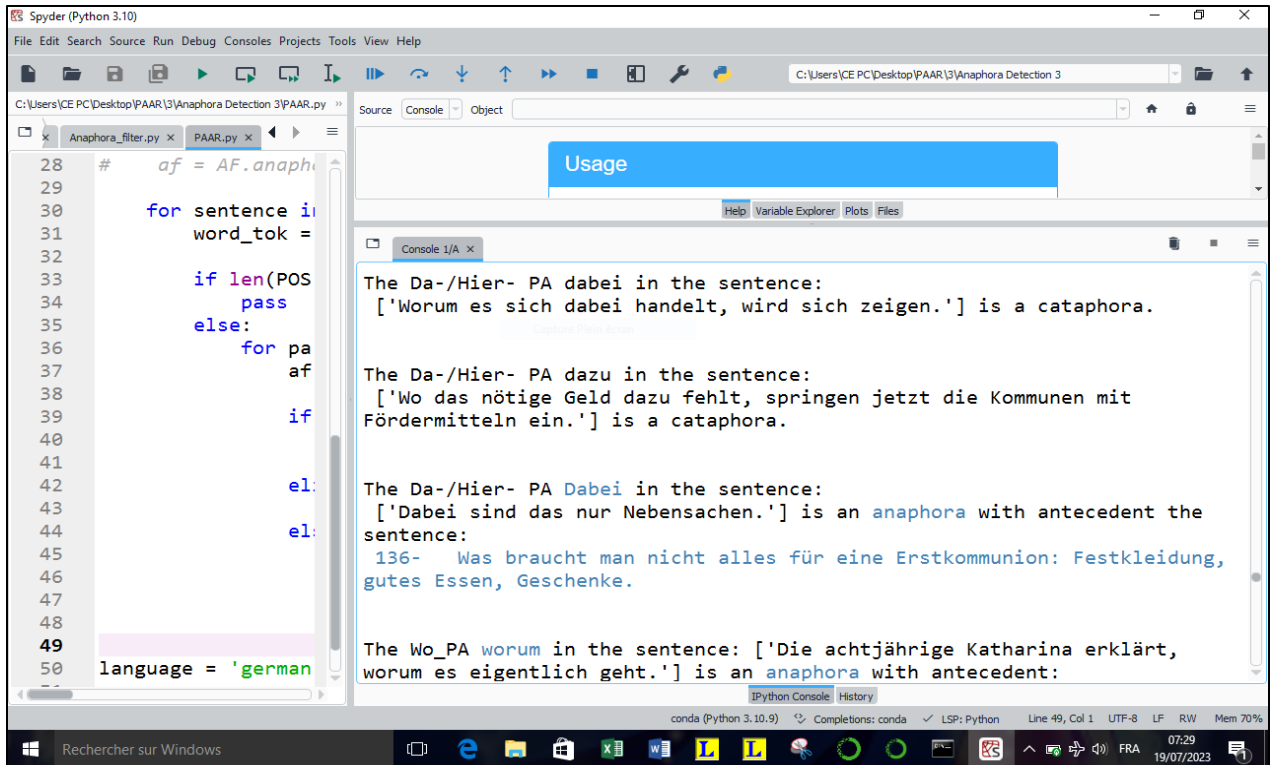


Abbildung 143: Einsatz des PAAR auf den Textabschnitt (N°101-154) (19)

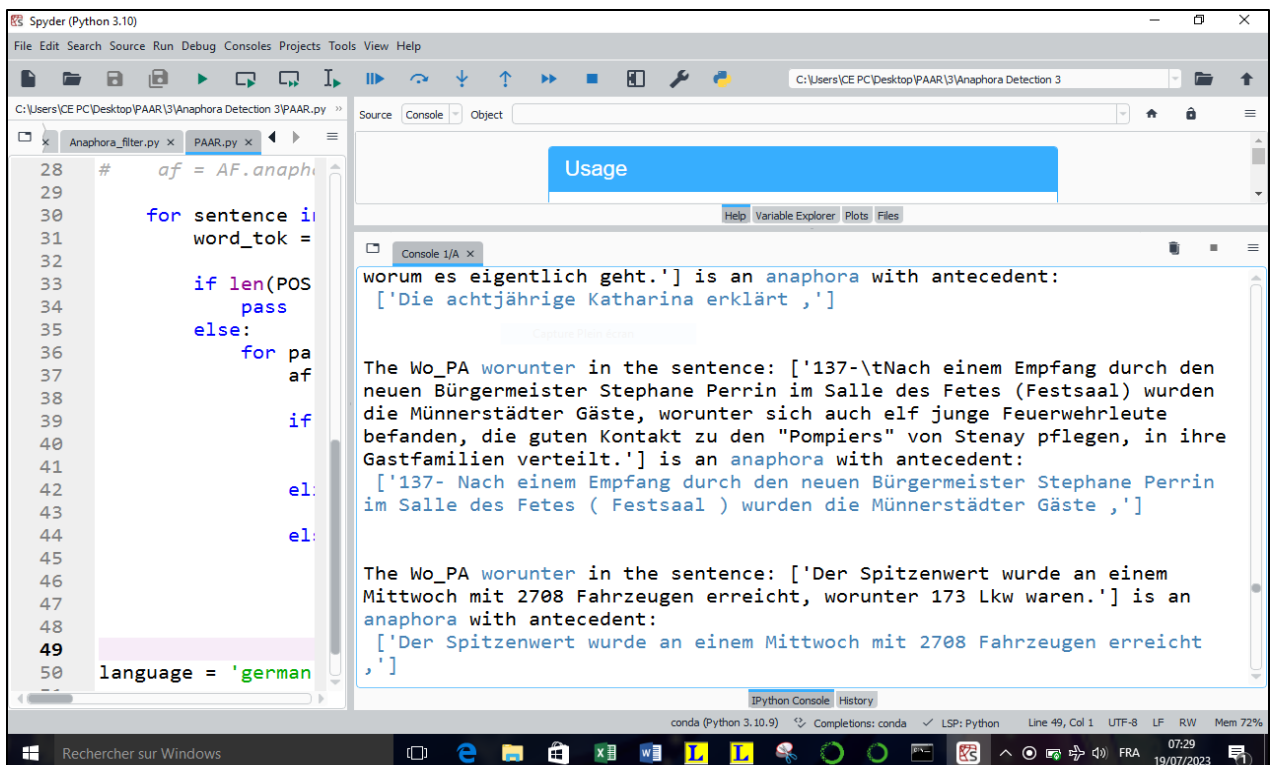


Abbildung 144: Einsatz des PAAR auf den Textabschnitt (N°101-154) (20)

Das PA "worunter" in der Abbildung 137 erscheint 2 Male und wurde jedes Mal nur zum Teil richtig aufgelöst. Die erste Erscheinung bezieht sich in der Tat nur auf die NP "die Mühnerstädter Gäste" und die zweite nur auf die NP "2708 Fahrzeuge"

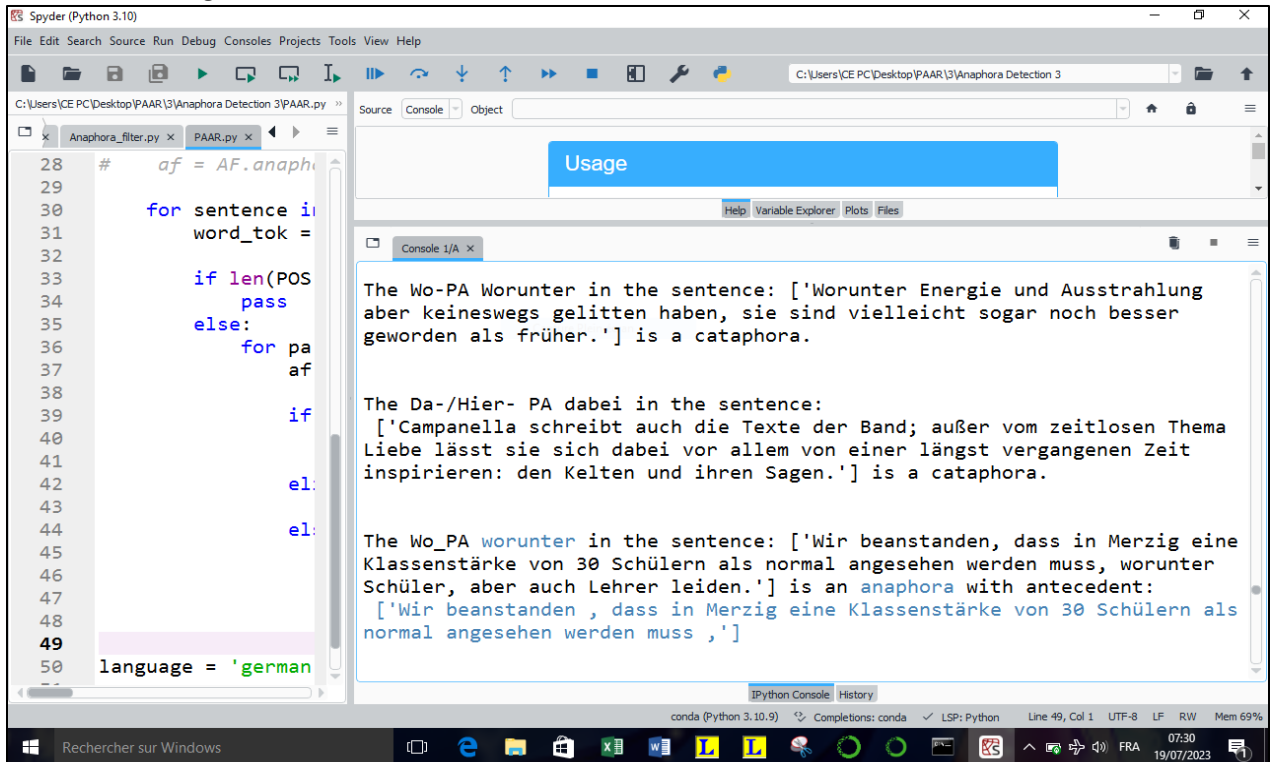


Abbildung 145: Einsatz des PAAR auf den Textabschnitt (N°101-154) (21)

Die PA "Worunter" und "dabei" in der Abbildung 138 sind keine Kataphern, sondern Anaphern mit jeweiligem Bezug auf "Was ihrer Musik sehr zugute gekommen ist, sie ist noch einen Tick" und "Campanella schreibt auch die Texte der Band". Das PA "worunter" ist zwar anaphorisch aber mit der NP "eine Klassenstärke von 30 Schülern" als Antezedens

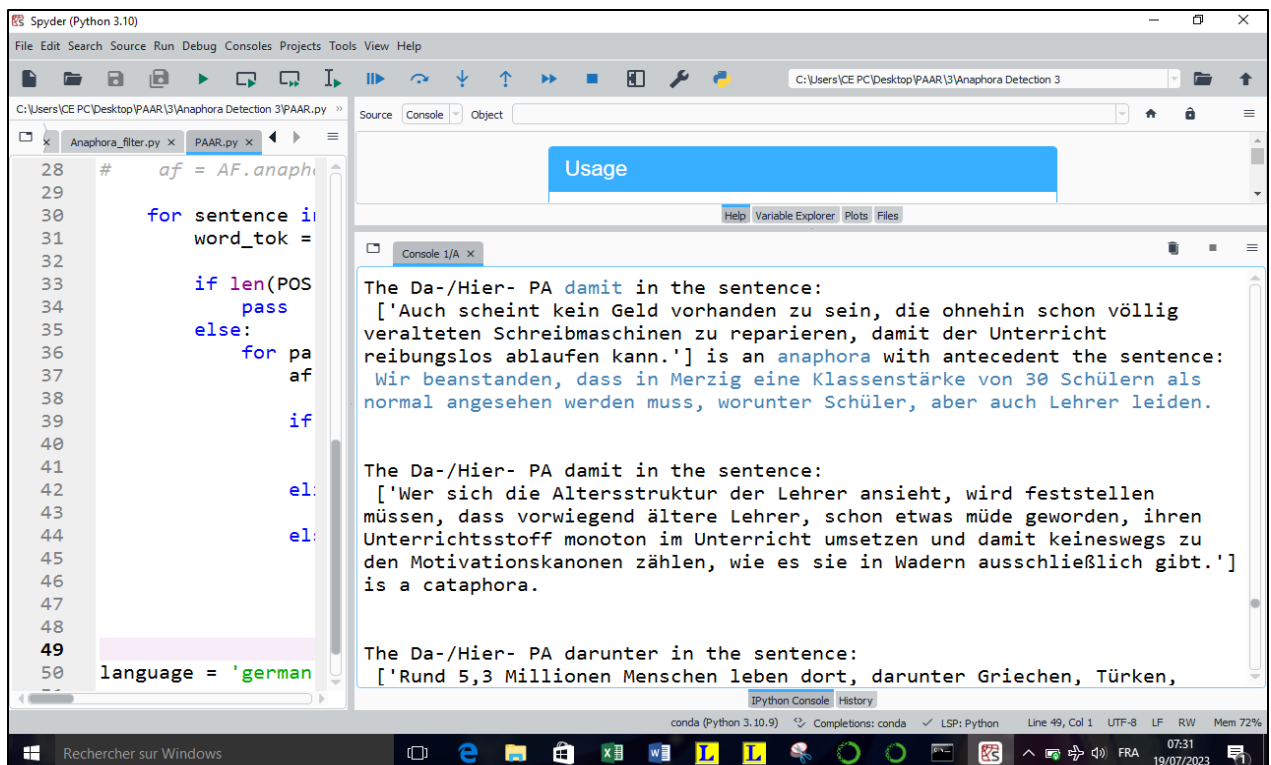


Abbildung 146: Einsatz des PAAR auf den Textabschnitt (N°101-154) (22)

Das Wort "Damit" aus der Abbildung 139 ist kein PA, sondern eine Subjunktion. Bei der zweiten Erscheinung des Worts "Damit" handelt es sich tatsächlich um ein PA aber um keine Katapher, sondern um eine Anapher mit Bezug auf "ihren Unterrichtsstoff monoton im Unterricht umsetzen". Die gleichen Überlegungen gelten fürs PA "Darunter". Es handelt sich in der Tat um keine Katapher, sondern um eine Anapher mit Bezug auf "Rund 5,3 Millionen Menschen leben dort".

Das PA "worunter" in der Abbildung 140 ist zwar anaphorisch aber bezieht sich auf "enorme wirtschaftliche Probleme". Was das PA "wovon" der gleichen Abbildung angeht, können wir sagen, es wurde richtig aufgelöst.

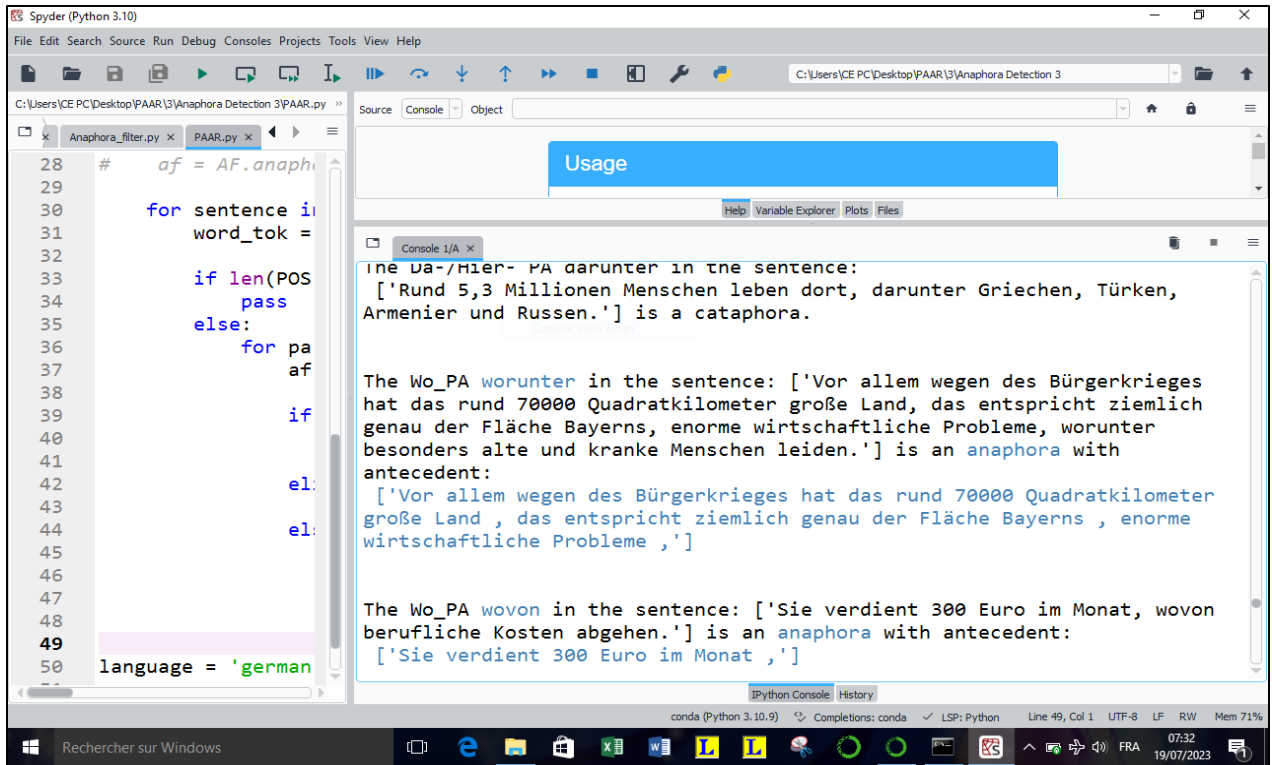


Abbildung 147 Einsatz des PAAR auf den Textabschnitt (N°101-154) (23)

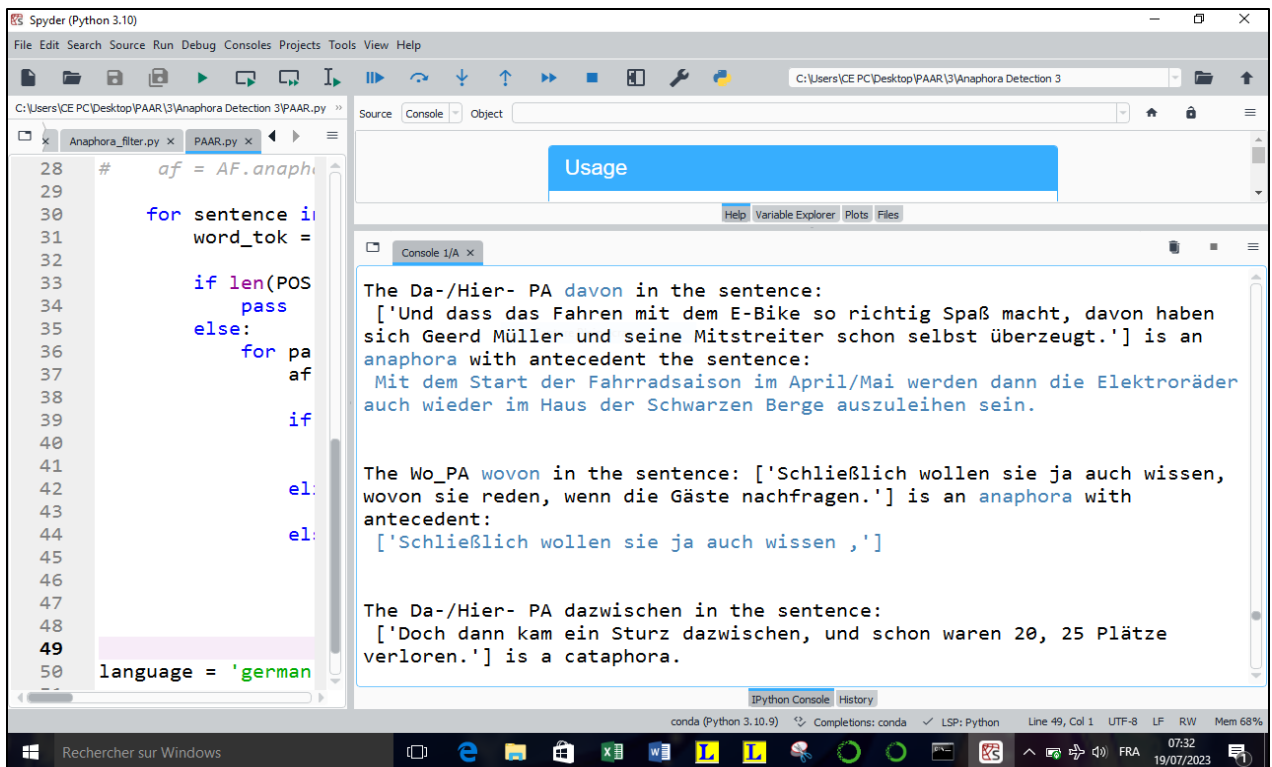
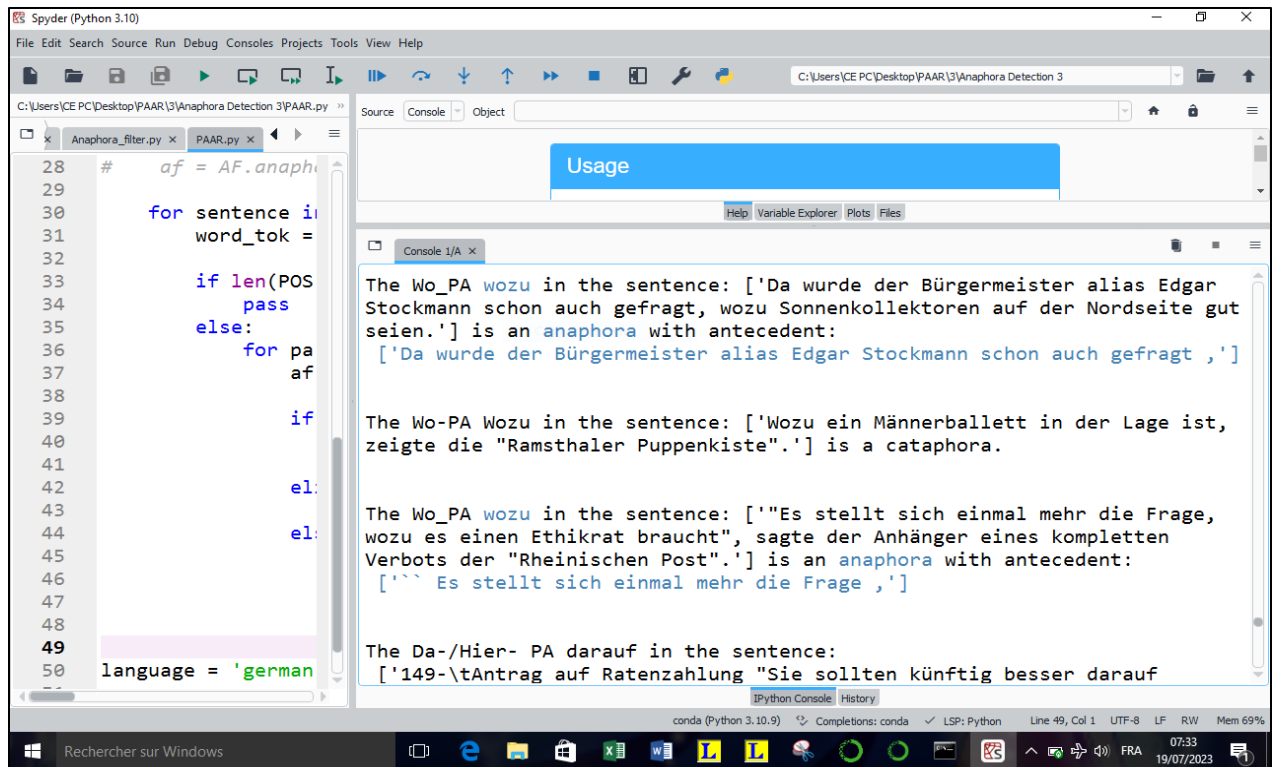


Abbildung 148: Einsatz des PAAR auf den Textabschnitt (N°101-154) (24)

Das PA "davon" in der Abbildung 141 ist zwar anaphorisch aber mit Bezug auf "Und dass das Fahren mit dem E-Bike so richtig Spaß macht,". Was das PA "Wovon" angeht, handelt es sich um keine Anapher, sondern um eine Katapher/Deixis. Das PA "Dazwischen" aus der gleichen Abbildung ist keine Katapher, sondern eine Anapher mit Bezug auf ""Ich habe mich richtig gut gefühlt."

Die 3 Erscheinungen des PA "Wozu" in der Abbildung 142 sind kataphorische bzw. dektische Erscheinungen und keine anaphorischen, wie PAAR es bei manchem Auftaucht klassifiziert hat. Doch was das PA darüber angeht, handelt es sich tatsächlich um eine Katapher.



```
28 # af = AF.anaphora
29
30 for sentence in
31     word_tok =
32
33     if len(POS)
34         pass
35     else:
36         for pa
37             af
38
39         if
40
41
42     el:
43
44     el:
45
46
47
48
49
50 language = 'german'
```

Usage

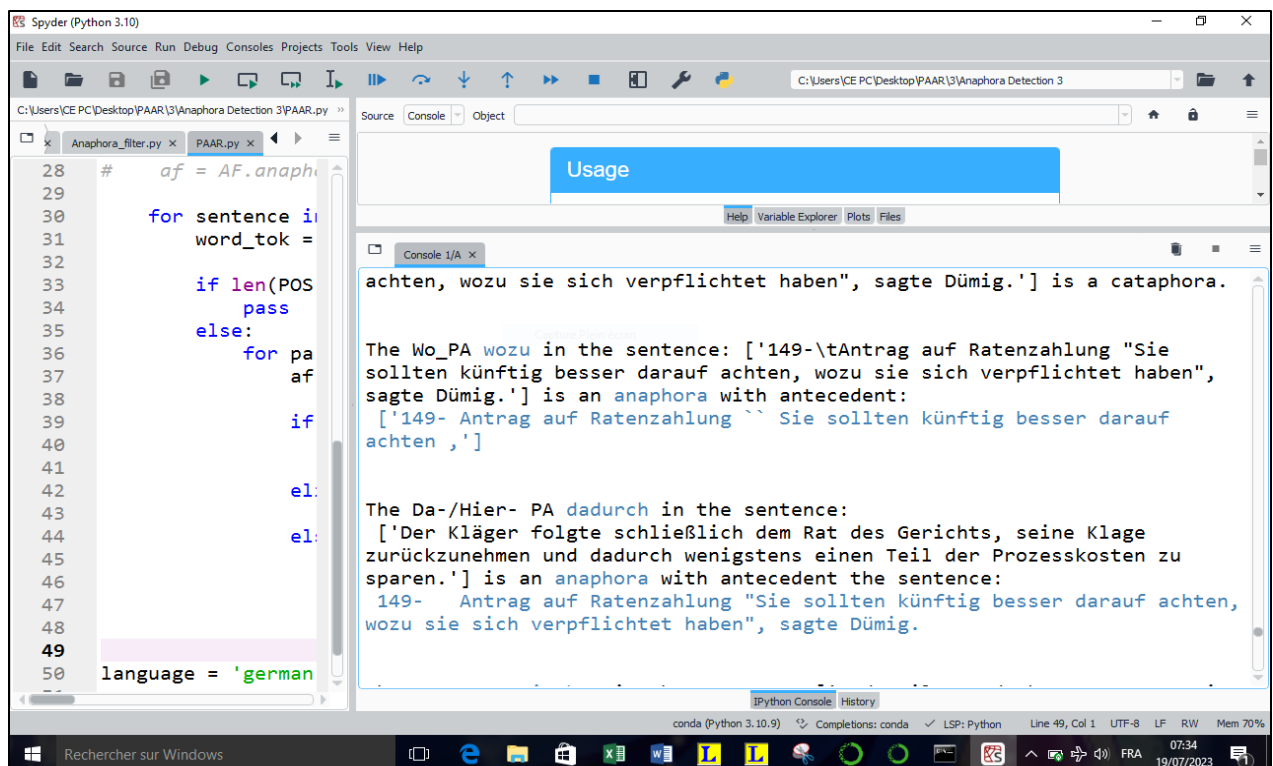
The Wo_PA wozu in the sentence: ['Da wurde der Bürgermeister alias Edgar Stockmann schon auch gefragt, wozu Sonnenkollektoren auf der Nordseite gut seien.']. is an anaphora with antecedent: ['Da wurde der Bürgermeister alias Edgar Stockmann schon auch gefragt ,']

The Wo-PA Wozu in the sentence: ['Wozu ein Männerballett in der Lage ist, zeigte die "Ramsthaler Puppenkiste."'] is a cataphora.

The Wo_PA wozu in the sentence: ['"Es stellt sich einmal mehr die Frage, wozu es einen Ethikrat braucht", sagte der Anhänger eines kompletten Verbots der "Rheinischen Post".'] is an anaphora with antecedent: ['" Es stellt sich einmal mehr die Frage ,']

The Da-/Hier- PA darauf in the sentence: ['149-\tAntrag auf Ratenzahlung "Sie sollten künftig besser darauf

Abbildung 149: Einsatz des PAAR auf den Textabschnitt (N°101-154) (25)



```
28 # af = AF.anaphora
29
30 for sentence in
31     word_tok =
32
33     if len(POS)
34         pass
35     else:
36         for pa
37             af
38
39         if
40
41
42     el:
43
44     el:
45
46
47
48
49
50 language = 'german'
```

Usage

achten, wozu sie sich verpflichtet haben", sagte Dümig.']. is a cataphora.

The Wo_PA wozu in the sentence: ['149-\tAntrag auf Ratenzahlung "Sie sollten künftig besser darauf achten, wozu sie sich verpflichtet haben", sagte Dümig.']. is an anaphora with antecedent: ['149- Antrag auf Ratenzahlung `` Sie sollten künftig besser darauf achten ,']

The Da-/Hier- PA dadurch in the sentence: ['Der Kläger folgte schließlich dem Rat des Gerichts, seine Klage zurückzunehmen und dadurch wenigstens einen Teil der Prozesskosten zu sparen.']. is an anaphora with antecedent the sentence: 149- Antrag auf Ratenzahlung "Sie sollten künftig besser darauf achten, wozu sie sich verpflichtet haben", sagte Dümig.

Das PA "Wozu" hier (Abld.143) ist ein Deixis und keine Anapher. Was das PA "dadurch" angeht, handelt es sich tatsächlich um eine Anapher aber mit Bezug auf "Der Kläger folgte schließlich dem Rat des Gerichts, seine Klage zurückzunehmen"



The screenshot displays the Spyder Python IDE interface. The left pane shows a Python script named `PAAR.py` with the following code:

```

28 # af = AF.anaphora
29
30 for sentence in sentences:
31     word_tok = tokenize_words(sentence)
32
33     if len(POS):
34         pass
35     else:
36         for pa in parse_anaphora(word_tok):
37             af.add(pa)
38
39 if len(af):
40     print('Anaphora detected in the sentence:')
41     for pa in af:
42         print(pa)
43 else:
44     print('No anaphora detected in the sentence.')
45
46 language = 'german'

```

The right pane shows the IPython console output, which is titled "Usage". The output text is as follows:

```

The Da-/Hier- PA davon in the sentence:
['Die Bürger der DDR waren Staatsbürger - unabhängig davon, ob sie ihrem
Staat zustimmend oder ablehnend gegenüber standen.'] is a cataphora.

The Da-/Hier- PA darin in the sentence:
['Es gab nichts, wozwischen man wählen konnte; das deutlichste Bekenntnis
bestand darin, nicht zur Wahl zu gehen.'] is a cataphora.

The Wo_PA wozwischen in the sentence: ['Es gab nichts, wozwischen man
wählen konnte; das deutlichste Bekenntnis bestand darin, nicht zur Wahl zu
gehen.'] is an anaphora with antecedent:
['Es gab nichts ,']

The Wo_PA wozwischen in the sentence: ['152-\\tDie Commode war mit Röcken,
Wäsche. Büchern pyramidenförmig aufgehäuft. alle Schubladen herausgezogen.

```

384

Die PA "davon"; darin" und "wozwischen" sind vom PAAR richtig verarbeitet worden. Was die zweite Erscheinung des PA "Wozwischen" angeht, wurde sie nur zum Teil richtig aufgelöst. Denn das richtige Antezedens ist "die zwei vorhandenen Stühle waren besetzt mit Büchern, Briefschaften, Hausgerät" (Siehe Abbildung 146.) Die gleiche Analyse gilt für die weitere Erscheinung des PA "wozwischen". Es ist zwar eine Anapher aber mit Bezug nur auf "'zwischen Berlin und Hambourg". Beim PA "davon" handelt es sich, wie PAAR es herausgefunden hat, um eine Katapher. (Abbildung 147)

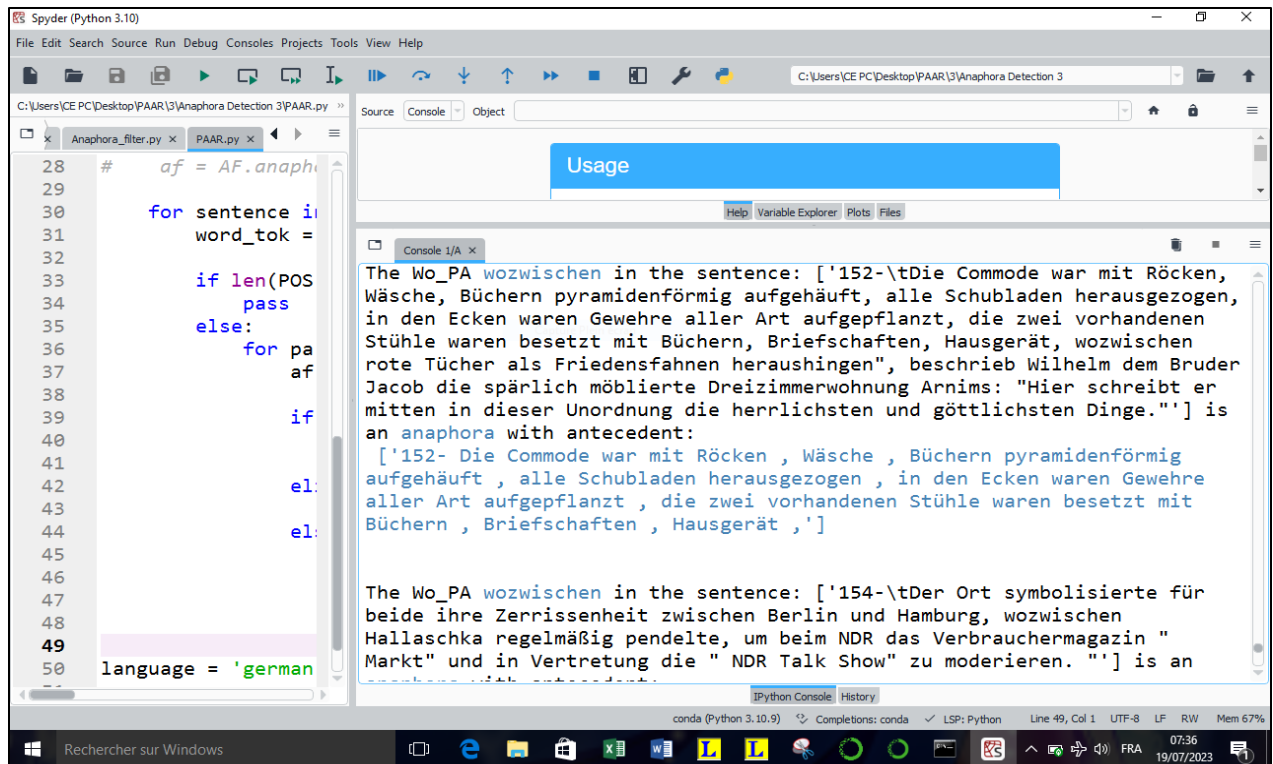


Abbildung 153: Einsatz des PAAR auf den Textabschnitt (N°101-154) (29)

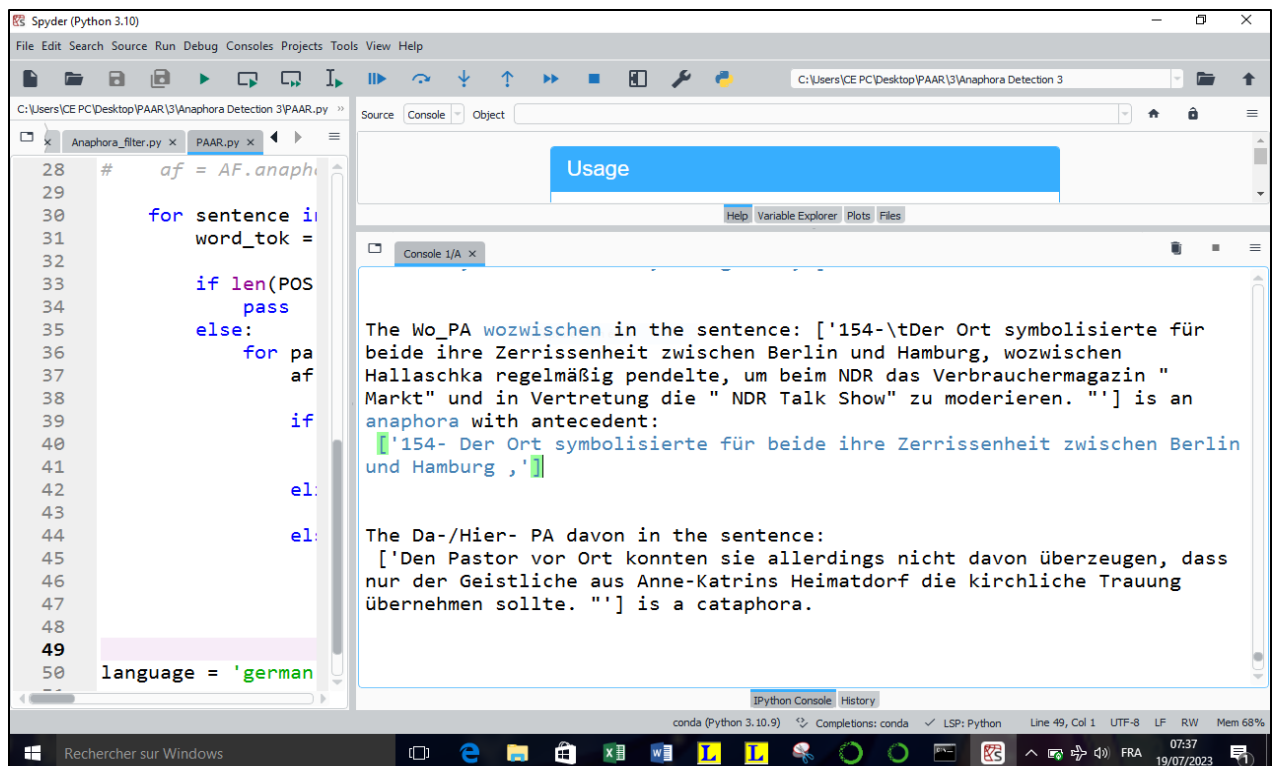


Abbildung 154: Einsatz des PAAR auf den Textabschnitt (N°101-154) (30)

ZUSAMMENFASSUNG UND ERGEBNISSE DES EINSATZES

Im Text enthaltene PA	Durch PAAR erkannte PA	Ergebnisse
85	84(83+1SUBJUNKTION)	84/85

Tabelle 73: Performanz vom PAAR bei Erkennung von PA im Textabschnitt (N°101-154)

PA	Anzahl	Natur	Ergebnisse
Womit	1	Katapher	0/1
Womit	2	Anapher	2/2
Dazu	4	Anapher	4/4
Damit	3	Katapher	2/3
Wonach	3	Anapher	3/3
Dazu	4	Katapher	0/4
woran	1	Deixis	1/1
Woran	1	Anapher	0/1
Dabei	4	Katapher	1/4
Damit	3	Anapher	3/3
Dafür	1	Katapher	1/1
Worauf	3	Anapher	2/3
Worauf	2	Katapher	2/2
Hierauf	1	Anapher	1/1
Woraus	1	Anapher	0/1
Woraus	2	Katapher	1/2
Darunter	2	Katapher	0/2
Worein	3	Anapher	3/3
Darauf	2	Katapher	2/2
Davon	4	Katapher	4/4
Worin	3	Anapher	1/3
Dafür	1	Katapher	0/1
Worüber	3	Anapher	3/3
Dagegen	1	Anapher	1/1
Dabei	2	Anapher	2/2
Worum	3	Anapher	0/3
Worum	2	Katapher	2/2
Worunter	4	Anapher	4/4
Worunter	1	Katapher	0/1
Wovon	2	Anapher	1/2
Davon	1	Anapher	1/1
Dazwischen	1	Katapher	0/1
Wozu	3	Anapher	0/3
Wozu	1	Katapher	1/1
Dadurch	1	Anapher	1/1
Wozwischen	4	Anapher	4/4
Danach	1	Katapher	1/1
Dazu	1	Katapher	1/1
Darin	1	Katapher	1/1
TOTAL	83		59/83

Tabelle 74: Performanz vom PAAR bei Erkennung der Natur von PA im Textabschnitt (N°101-154)

PA	Noten der Auflösung von anaphorischen PA durch PAAR
Womit	1,25/2
Dazu	3/4
Wonach	1,5/3
Damit	2/3
Worauf	1,75/2
Hierauf	1/1

Worein	1,5/3
Worin	0,5/1
Worüber	2,5/3
Dagegen	1/1
Dabei	1/2
Worunter	2/4
Wovon	1/1
Davon	0/1
Dadurch	0/1
Wozwischen	3,5/4
TOTAL	21,5/32

Tabelle 75: Performanz von PAAR bei der Auflösung von anaphorischen PA im Textabschnitt (N°101-154)

6. TDR: Termes de références

TDR

SYSTÈME DE RÉOLUTION DE L'ANAPHORE PRONOMINALE ADVERBIALE (PAAR)

Alex Hermann SOCKENG & Nelson Leroi SOFAC

June 12, 2023

1 Description

On souhaite lier deux parties d'un discours. L'une des parties est un mot (adverbe pronominal en abrégé PA) et l'autre partie est une phrase précédant le mot (l'antécédant).

Le PA peut être aussi bien relié à la phrase précédente (antécédent) qu'à la phrase qui le suit (postcédant). Lorsque le PA est relié à la phrase précédente, il est appelé **anaphore** et lorsqu'il est relié à la phrase suivante il est appelé **cataphore**.

Il est important de noter qu'il ne peut y avoir redondance d'un PA dans une phrase.

2 Objectif

L'objectif est pour un paragraphe donné, d'identifier automatiquement un PA dans chacune des phrase et de décélér s'il fait référence à un **antécédent** ou un **postcédent**.

3 Sources d'informations existantes

Le processus consistant à décélér si un PA est un antécédant ou un postcédent à été modéliser via un algorithme par le linguiste. On a entre autre:

1. Algorithme de résolution de l'anaphore pronominale adverbiale. C'est l'agorithme qui établi le lien définitif entre le PA et sont antécédent. Il requière l'utilisation des algorithmes suivant:
 - 2.1. Algorithme qui détecte si le PA a un postcédent i.e si le PA est une cataphore. Si oui on passe au prochain PA, sinon on considère ce PA comme ayant un antécédent (le PA est une anaphore). On distingue deux formes de cet algorithme: l'un pour les **Da-/Hier-PA** il identifie tout les PA qui sont suivi par maximum 3 mots + une virgule + et l'un des mots suivant: **dass, ob**, et les **W-Satz** (environ 40). Il identifie aussi les PA suivi d'au plus 1 mot et d'une virgule ou de deux points.
 - 2.2. **Wo-PA** il identifie les PA qui sont en début de phrases.

4 Aspects méthodologiques

Il faudra **tokeniser** le texte en phrases et mots permettant ainsi d'identifier les PA. Ceci nécessite l'usage d'un logiciel approprié notamment **python** qui est l'un des langages si ce n'est le langage de programmation par excellence de l'IA dont la linguistique computationnelle qui n'est que du NLP en fait parti.

5 Participation et responsabilité des parties prenantes

Le linguiste et le "computational engineer" travaillent en collaboration. Le linguiste a pour tache principlae la description et la présentation des algorithmes réalisés à partir de l'exploitation des principes linguistiques, notamment l'agorithme de filtrage de cataphores ou encore l'agorithme de résolution globale. Le "computational engineer" quant à lui appréhende les ces méthodes et les implemente en language informatique.

7. DAS PAAR

Anbei ein CD-ROM; eine SD-Karte oder ein USB-Stick, worin das Programm archiviert und zum sofortigen Einsatz bereits ist. Dafür müsste man aber den im Kapitel 4 angegebenen Vorschriften folgen.

.

.