

die Frau – sie – ø: Eine korpusbasierte multifaktorielle Analyse zur Wahl anaphorischer Formen im Deutschen

Abstract

Der Terminus *Anapher* bezeichnet den Verweis auf bereits erwähnte Entitäten im Text und dient überwiegend der thematischen Fortführung. Beim Wiederaufgreifen von im Text schon eingeführten Entitäten stehen unterschiedliche anaphorische Formen zur Verfügung. Als die wichtigsten Kategorien von Anaphern gelten lexikalische NP-Anapher (NA), pronominale Anapher (PA) und Zero-Anapher (ZA). Die Entscheidung für eine passende anaphorische Form, welche die Muttersprachler/-innen mittels ihres Sprachgefühls zumeist intuitiv treffen können, fällt den Nicht-Muttersprachler/-innen oftmals schwer. Der vorliegende Beitrag blickt auf die produzentenseitige Auswahl anaphorischer Formen aus mehreren Optionen im Deutschen. Das Ziel ist, den Entscheidungsmechanismus (NA vs. PA vs. ZA) auf der Grundlage authentischer Korpusdaten transparent zu beschreiben sowie die möglichen Einflussfaktoren auf die Entscheidung mithilfe der *multinomial logistic regression* herauszuarbeiten und in ihrer Zusammenwirkung zu gewichten. Die Untersuchung zeigt, dass neun kontextuelle Faktoren eine Schlüsselrolle bei der Wahl von anaphorischen Formen im Deutschen spielen. Mit ihnen lässt sich die Entscheidung von Deutsch-Muttersprachler/-innen mit einer hohen Genauigkeit (84,24%) modellieren.

The term *anaphora* denotes the reference to entities which have already been mentioned in the text and serves mainly to continue on a topic. There are different anaphoric devices available for referring back to entities already introduced in a text. The most important categories of anaphora are lexical NP anaphora (NA), pronominal anaphora (PA) and zero anaphora (ZA). Mostly native speakers make anaphoric choices intuitively, but it is not always easy for non-native speakers. This article focuses on the production-side choice of different anaphoric devices in German. It aims to describe the anaphoric choice mechanism (NA vs. PA vs. ZA) in a transparent way, based on authentic corpus data. By applying *multinomial logistic regression*, this study systematically explores and evaluates the conglomerate of factors influencing anaphoric choices. The study shows that nine contextual factors play a key role in the choice of anaphoric forms in German. They can be used to model the choice made by native German speakers with a high degree of accuracy (84.24%).

Keywords: anaphorische Formen, kontextuelle Faktoren, Verteilungsunterschiede, logistische Regression, Zugänglichkeit

1. Einleitung^{1, 2}

Die Frau, [...] tat es fast grob. Sie nahm meinen Arm und ø führte mich durch den dunklen Hausgang in den Hof. [...] Die Frau drehte den Hahn auf, ø wusch zuerst meine Hand und ø klatschte mir dann das Wasser, das sie in ihren hohlen Händen auffing, ins Gesicht.

(Bernhard Schlink, *Der Vorleser*)

¹ Korrespondenzautorin: Jing Li, Asso.-Professorin an der Beijing Foreign Studies University.

² Dieser Beitrag ist im Rahmen des von The National Social Science Fund of China geförderten Projekts (国家社科基金项目) „Constructing a Multilingual Foreign-Chinese Dictionary Database from the Perspective of Probabilistic Contextual Co-selection“ (概率语境共选视角下的多语外汉词典数据库建设与研究, Nr. 21BY021) entstanden.

Der Terminus *Anapher* bezeichnet den Verweis auf bereits erwähnte Entitäten im Text und dient überwiegend zur Fortführung von im Text schon eingeführten oder sonst präsenten Gegenständen oder Sachverhalten (vgl. Hoffmann 2000, S. 295). Als ein sprachliches Phänomen auf der Textebene trägt die Anapher erheblich zur Textkohärenz bei. Es findet sich kaum ein Text, der keine Anapher enthält. Allein in dem titelgebenden Beispielabschnitt kommt die Anapher beim Wiederaufgreifen von *Die Frau* siebenmal vor und wird durch unterschiedliche Kodierungsformen realisiert, nämlich durch Pronomen (*sie, ihren*), eine lexikalische NP-Phrase (*die Frau*) sowie die Nullform ($\emptyset$). Laut Bußmann (Hg.) (2002, S. 79) stellen diese Formen die drei wichtigsten Anapher-Kategorien dar: die pronominale Anapher (PA), die lexikalische NP-Anapher (NA) und die Zero-Anapher (ZA). Diese Studie untersucht die Wahl aus unterschiedlichen anaphorischen Formen im Deutschen, wobei es vor allem um die Entscheidung zwischen PA, NA und ZA geht.

Es ist bemerkenswert, dass die Stellungen von PA und NA recht flexibel sind, wohingegen die Stellung der ZA deutlich strikteren syntaktischen Regeln unterliegt. Dass diese drei Anapherotypen in der vorliegenden Studie gleichwertig berücksichtigt werden, rührt unter anderem daher, dass sie zu den wesentlichen Anapherotypen gehören und in zahlreichen natürlichen Sprachen zu finden sind (Huang 2000; Givón 2017). Unterschiedliche Sprachen machen jedoch in unterschiedlichem Maße Gebrauch von verschiedenen anaphorischen Formen. Beispielsweise sind einige ostasiatische Sprachen wie Chinesisch und Japanisch im Vergleich zu anderen stärker von Zero-Anaphern geprägt. Deswegen fällt den Fremdsprachen-Lernenden sowie den Übersetzenden die richtige Entscheidung für eine passende anaphorische Form nicht immer leicht, da sie sich der Beeinflussung ihrer Muttersprache entziehen müssen.

In der Sprachwissenschaft zählt die Anapher zu einem der beliebtesten Forschungsthemen, die bereits viel Aufmerksamkeit von Linguisten aus verschiedenen Bereichen auf sich gezogen hat. Grob gesehen kann man zwischen formalen und funktionalen Ansätzen unterscheiden (Jiang 2006). Formale Ansätze konzentrieren sich primär auf die Anapher innerhalb des Satzes, wobei syntaktische Regeln eine zentrale Rolle spielen. Dazu gehört vor allem Chomskys *Goverment & Binding Theory* (Chomsky 1981, 1982). Funktionale Ansätze erweitern die Untersuchung der Anapher auf die Textebene, wobei verschiedene Perspektiven aus der Pragmatik (Levinson 1987, 1991; Huang 1991, 2000), der Kognition (Ariel 1990) usw. einbezogen werden. Die vorliegende Studie schließt sich den funktionalen Ansätzen an und orientiert sich an Ariels *Accessibility Theory*, die besagt, dass die Anapher auf der Textebene wie auch auf der Satzebene letztendlich vom Grad der Zugänglichkeit (engl. *degree of accessibility*) des Antezedens in der mentalen Repräsentation abhängt. Für Anaphern auf der Textebene wie auch auf der Satzebene sind die exakt selben kontextuellen Faktoren wirksam, die syntaktische, semantische sowie pragmatische Faktoren umfassen (Ariel 1990, S.161f.). Zudem wird diese Studie auch durch die Konstruktionsgrammatik geprägt, die von einer *Usage-based*-Perspektive ausgeht und die Anapher (bzw. Alternation von PA/NA/ZA) als eine Konstruktion im weiteren Sinne betrachtet (vgl. Goldberg 2006, S. 5). Demzufolge unterscheiden sich die drei Anapherotypen – wie die konkurrierenden Konstruktionselemente – in Form und Funktion (einschließlich Semantik, Pragmatik usw.). Diese Unterschiede spiegeln sich in statistischen Wahrscheinlichkeitsverteilungen wider.

Die vorliegende Untersuchung setzt sich das Ziel, die Verteilungsunterschiede von PA, NA und ZA hinsichtlich der kontextuellen Faktoren transparent zu beschreiben. Aufgrund von authentischen, selbst annotierten Korpusdaten und mithilfe des statistischen Algorithmus *multinomial logistic regression* werden die möglichen kontextuellen Einflussfaktoren herausgearbeitet und in ihrer Zusammenwirkung gewichtet, so dass im Rahmen der *Accessibility Theory* allgemeinere Geltungsprinzipien zur Wahl anaphorischer Formen im Deutschen extrahiert werden können. Daraus ergeben sich die folgenden Fragestellungen:

- 1) Welche kontextuellen Einflussfaktoren erweisen sich empirisch unter mehreren möglichen Faktoren als Schlüsselfaktoren für die Wahl zwischen PA, NA und ZA im Deutschen?
- 2) Wie wirken sich die ermittelten Schlüsselfaktoren auf die Wahl zwischen PA, NA und ZA aus?
- 3) Wie sehen die typischen Verwendungskontexte für PA, NA und ZA jeweils aus?

Um diese Forschungsfragen zu beantworten, wird eine multifaktorielle Analyse (vgl. Gries 2013) durchgeführt. Zunächst wird ein ausgewogenes Korpus in Anlehnung an das Brown Corpus (1961) erstellt, aus dem insgesamt 3.000 Anapher-Fälle identifiziert und als unmittelbarer Untersuchungsgegenstand genutzt werden. Für jeden Anapher-Fall werden in dessen unmittelbaren Kontexten alle Einflussfaktoren manuell annotiert. Welche die Einflussfaktoren sind und wie sie annotiert werden können, ergibt sich aus der Forschungsliteratur, die im Kapitel 2 detailliert erläutert wird. Die annotierten Daten werden anschließend mit Hilfe der Software R (R Core Team 2021) einer statistischen Analyse unterzogen. Eingesetzt wird der Algorithmus *multinomial logistic regression*, welcher ermöglicht, die Schlüsselfaktoren sowie deren Auswirkungen vor allem in ihrer Zusammenwirkung bei der Entscheidung für eine bestimmte anaphorische Form im Deutschen zu ermitteln. Daraus können die typischen Kontexte für jede anaphorische Form (d. h. NA, PA und ZA) abgeleitet werden.

Der Aufbau des Beitrags ist wie folgt: Im zweiten Kapitel werden die theoretischen Grundlagen vorgelegt. Darauf basierend wird in Kapitel 3 die empirische Untersuchung vorgestellt. Die Forschungsergebnisse werden hinsichtlich 1) der Schlüsselfaktoren, 2) deren Auswirkungen sowie 3) der typischen Kontexte mit sprachlichen Belegen zusammenfassend präsentiert. Schließlich lassen sich in Kapitel 4 die wichtigsten Erkenntnisse resümieren und im Rahmen von Ariels *Accessibility Theory* interpretieren. Dem schließt sich ein kurzer Ausblick auf weitere Forschungsmöglichkeiten an.

2. Theoretische Grundlagen und bisherige Forschungen

2.1 Anapher im Deutschen

Der Begriff der Anapher ist von dem griechischen Wort „anaforá“ abgeleitet und bedeutet „das Hinauftragende“, „das Zurückweisende“. In linguistischen Kontexten werden Anaphern als sprachliche Ausdrücke bezeichnet, mit denen Sprachbenutzer auf einen bereits erwähnten Referenten im Text oder Diskurs erneut Bezug nehmen (Schwarz 1997, S. 445). Dieser Status wird ihnen nicht aufgrund inhärenter Merkmale zugesprochen, sondern sie sind durch die Angabe von textsemantischen und referenzpragmatischen Kriterien funkti-

onal definiert (Schwarz 2000, S. 52). Insofern ist eine weitere Abgrenzung hilfreich, so dass der unmittelbare Untersuchungsgegenstand der vorliegenden Untersuchung eindeutig festgelegt werden kann.

Zuerst wird zwischen der „direkten Anapher“ und der „indirekten Anapher“ unterschieden (Bußmann 2002, S. 79):

- (1) Als Anna_i das Klassenzimmer_j betrat, hatte der Unterricht_j längst begonnen. Sie_i ist zu spät gekommen.

Für direkte Anaphern ist eine im vorangehenden Kontext stehende sprachliche Einheit erforderlich. Im Beispielsatz (1) lässt sich erkennen, dass das Antezedens *Anna* und die pronominale Anapher *sie* eine anaphorische Koreferenz bilden. Bei indirekten Anaphern ist dies nicht der Fall. In (1) gibt es zwischen den beiden definitiven Nominalphrasen, nämlich *das Klassenzimmer* und *der Unterricht*, einen gewissen referenziellen Zusammenhang: Der Unterricht ist eine erwartbare Aktivität im Klassenzimmer. In diesem Sinne bilden die beiden Nominalphrasen ebenfalls eine Anapher, und zwar eine indirekte. Zum Zweck eines operationalisierbaren Aufgreifens werden in dieser Studie nur direkte Anaphern untersucht.

Die anaphorische Funktion lässt sich durch unterschiedliche Ausdruckstypen realisieren. Im Hinblick auf syntaktische Kategorien können Anaphern grob in zwei Kategorien unterteilt werden: NP- und VP-Anapher. Die vorliegende Studie konzentriert sich auf diejenigen NP-Anaphern, die eine komplette Nominalphrase – und nicht nur ein Kopfnomen³ – als Antezedens haben (vgl. Huang 2000, S. 2f.). Unter denen lassen sich wiederum drei lexikalische Realisierungsmittel unterscheiden: Pronomen, lexikalische Nominalphrasen sowie Nullformen. Im Folgenden wird der Untersuchungsgegenstand der vorliegenden Studie hinsichtlich dieser drei Kategorien näher bestimmt und mit Korpusbelegen ausgeführt.

Zur **pronominalen Anapher (PA)**: Aus forschungspraktischen Gründen werden in dieser Studie nur die Personalpronomen und Possessivpronomen der 3. Person behandelt. Die Fokussierung auf die Pronomen der 3. Person geht von der Annahme aus, dass die Pronomen der 1. und 2. Person deiktisch sind, da mit ihnen auf sprechende bzw. an der Kommunikation direkt beteiligte Personen verwiesen wird. Die Pronomen der 3. Person dagegen werden eher als Nomen-Ersatzformen begriffen (vgl. Graefen 2009, S. 661).

Korpus-Beispielinstanz:

- (2) Hilde Knef_i beschrieb später, warum sie_i den Aufruhr nicht hatte kommen sehen: Sie_i hatte die Säuglingsjahre der supersauberen BRD verpasst – zur selben Zeit scheiterte gerade ihr_i erster Versuch einer Hollywood-Karriere; und in den Jahren davor hatte sie_i Deutschland so sauber nicht finden können. (DE09C01A)

³ Die NP-Anapher nimmt eine komplette Nominalphrase als Antezedens, wohingegen die N-Anapher sich nur auf das Kopfnomen bezieht. Im Beispielsatz (i) verweist *er* auf die gesamte Nominalphrase *einen neuen Wagen* und bildet in diesem Sinne eine NP-Anapher. Die Nullanapher in (ii) referiert auf das Nomen *Wagen*, jedoch nicht auf die komplette Nominalphrase *einen neuen Wagen*, was sie als N-Anapher ausweist.

(i) *Anna hat einen neuen Wagen_i gekauft. Einige Monate später wurde er_i zur Reparatur gebracht.*

(ii) *Anna hat einen neuen Wagen_j gekauft. John hat einen gebrauchten Ø_j gekauft.*

Zur lexikalischen NP-Anapher (NA): Anders als pronominale Anaphern verfügen die lexikalischen NP-Anaphern über eine eigenständige Bedeutung. Sie werden vor allem durch bestimmte Nominalphrasen und Eigennamen realisiert. In Anlehnung an Mitkov (2002, S. 10 f.), Xu (2003, S. 142–155) und Qian (2006, S. 451–459) wird angenommen, dass zwischen den lexikalischen NP-Anaphern und ihren Antezedenzen vor allem zwei mögliche Relationen bestehen. Folgende Fälle (Tab. 1) werden in der vorliegenden Studie als NP-Anaphern berücksichtigt:

Relationen	Kategorie	Beispiele	
		Antezedens	Anapher
Lexikalische Wiederholung	Vollständige Wiederholung	<i>der 16-jährige Junge</i>	<i>der 16-jährige Junge</i>
	Wiederholung des Kopfnomens	<i>der 16-jährige Junge</i>	<i>der Junge</i>
Semantische Relation	Synonymie	<i>das Auto</i>	<i>der Wagen</i>
	Hyperonymie	<i>der Pudel</i>	<i>der Hund</i>
	Hyponymie	<i>der Baum</i>	<i>der Apfelbaum</i>
	Metapher	<i>die Debatte</i>	<i>der Krieg</i>

Tab. 1: Relationen zwischen NP-Anaphern und ihren Antezedenzen

Korpus-Beispielinstanz:

- (3) Da war beispielsweise das legendäre Adria_i, das Ende 2007 schließen musste. 35 Jahre lang versorgte das Schwabinger Lokal_i an der Giselastraße viele Nachtschwärmer bis drei Uhr morgens mit Pizza und Pasta. (DE09B03A)

Zur Zero-Anapher (ZA): Eine der wichtigsten Klassen anaphorischer Ausdrücke bildet die sogenannte „Zero-Anapher“ (auch „Null-Anapher“, „Ellipse“ oder „gap“ genannt). Hierbei handelt es sich um Anaphern, die nicht auf der Oberflächenstruktur des Satzes realisiert werden, wo ein Pronomen oder eine Vollform kanonischerweise stehen müsste (Frank 2019, S. 202). In der vorliegenden Studie wird diesbezüglich die methodische Herangehensweise von Givón (2017) übernommen, nach der ausschließlich syntaktisch nicht-regierte (ungoverned) bzw. diskursive (discourse) Zero-Anaphern berücksichtigt werden. Die syntaktisch regierten (governed) Zero-Anaphern fallen daher nicht in den Untersuchungsbereich dieser Arbeit.

Korpus-Beispielinstanz:

- (4) Seine Frau_i stand unschlüssig in der Küche und Ø_i betrachtete nachdenklich den weißen Umschlag. (DE09L02)

2.2 Anaphorische Formen und *Accessibility* des Referenten

Zur Wahl anaphorischer Formen wurden schon zahlreiche Untersuchungen durchgeführt, aus denen einzelne Erklärungsansätze wie *Accessibility Marking Scale* (Ariel 1990), *Givenness Hierarchy* (Gundel/Hedberg/Zacharski 1993), *Familiarity Scale* (Prince 1981), *Accessibility Hierarchy* (Pu 2011) usw. entwickelt wurden. Trotz begrifflicher Verschiedenheit wie *accessibility*, *givenness*, *salience* u. ä. liegt diesen Ansätzen letztlich folgende gemeinsame Annahme zugrunde: Die anaphorische Form hängt von dem Aktivierungsstatus des Referenten im Diskurs ab, wobei die kontextuellen Einflussfaktoren eine Schlüsselrolle spielen. Je aktivierter ein Referent in der mentalen Repräsentation ist, desto reduzierter wird die anaphorische Form.

Die vorliegende Studie stützt sich hauptsächlich auf Ariels (1990) *Accessibility Theory*, weil diese ein kohärentes und umfassendes Erklärungsmodell bereitstellt. Der Terminus *Accessibilty* (dt. *Zugänglichkeit*) benennt den Aktivierungsgrad eines Referenzobjektes, d. h. wie leicht es ist, auf das Referenzobjekt zuzugreifen oder dieses zu erinnern. Sofern ein Referenzobjekt in der mentalen Repräsentation einen höheren Grad an Zugänglichkeit aufweist, besteht die Tendenz zur Verwendung einer reduzierten anaphorischen Form. Ariel zufolge gelten vier zentrale Dimensionen zur Bestimmung der Zugänglichkeit (Ariel 1990, S. 28 f.): 1) Distanz zwischen der Anapher und dem Antezedens, 2) Konkurrenz mit anderen Referenten, 3) Salienz des Referenten und 4) Einheit (engl. *unity*) von der Anapher und dem Antezedens. Ariel betrachtet es als ein sprachübergreifend geltendes Prinzip, dass die Zugänglichkeit des Referenzobjekts zu einem erheblichen Grad die Form der Anapher determiniert. In verschiedenen Sprachen zeigt sich jedoch eine sprachspezifische Verteilung der anaphorischen Formen bezüglich unterschiedlicher Zugänglichkeit. Besonders in Bezug auf Anapher-Formen mit hoher Zugänglichkeit (ZA, PA) ist die Verteilung zwischen verschiedenen Sprachen teilweise äußerst unterschiedlich, so dass große Schwierigkeiten für Lernende und Übersetzende entstehen.

Vor diesem Hintergrund strebt die vorliegende Untersuchung vermittlels der *Accessibility Theory* nach einer transparenten Beschreibung der Verteilungsunterschiede der drei wesentlichen Anapherntypen (PA, NA und ZA) im Deutschen hinsichtlich der kontextuellen Einflussfaktoren, was sowohl für den Prozess des Fremdsprachenerwerbs als auch für Übersetzungsaufgaben sinnvoll sein sollte.

2.3 Einflussfaktoren in der bisherigen Forschung

In der bisherigen Forschung wurden verschiedene Einflussfaktoren empirisch ermittelt. Allerdings beschränkten sich die einzelnen Untersuchungen meistens auf eine begrenzte Anzahl möglicher Einflussfaktoren. Die vorliegende Studie hingegen versucht, alle ermittelten Einflussfaktoren aufeinander zu beziehen und die Schlüsselfaktoren bzw. ihre Funktionsweise mithilfe statistischer Algorithmen zu erhellen. Ziel ist, eine holistische Perspektive zu erarbeiten. In diesem Kapitel werden die bisherigen Untersuchungen zu Anaphern im Deutschen resümiert und alle erwähnten kontextuellen Einflussfaktoren systematisch ausgewertet. Diese Studie widmet sich der Anapher-Produktion, doch aufgrund der engen Verbindung zwischen Anapher-Produktion und der Anapher-Interpretation (vgl. Kehler/Rohde 2013) werden an dieser Stelle alle Einflussfaktoren – sowohl vonseiten der Produktion als auch vonseiten der Interpretation – berücksichtigt.

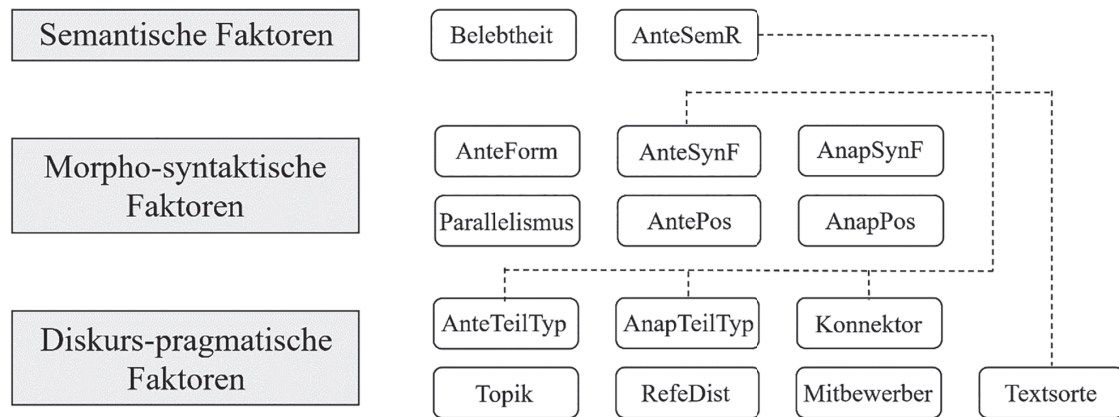


Abb. 1: Einflussfaktoren für die Auswahl anaphorischer Formen im Deutschen

Insgesamt lassen sich 15 mögliche Einflussfaktoren für die Auswahl von anaphorischen Formen im Deutschen feststellen, die in drei Dimensionen eingeteilt werden können (siehe Abb. 1). Auf der semantischen Ebene spielen die Belebtheit des Referenten (Belebtheit) und die semantische Rolle des Antezedens (AnteSemR) eine wichtige Rolle. Zu den morpho-syntaktischen Faktoren zählen die Form des Antezedens (AnteForm), die syntaktische Funktion des Antezedens (AnteSynF), die syntaktische Funktion der Anapher (AnapSynF), der syntaktische Parallelismus zwischen der Anapher und dem Antezedens (Parallelismus), die Position des Antezedens (AntePos) und die Position der Anapher (AnapPos). Bezüglich der diskurs-pragmatischen Dimension sind die folgenden Faktoren wichtig: der Teilsatztyp des Antezedens (AnteTeilTyp), der Teilsatztyp der Anapher (AnapTeilTyp), die Semantik des Konnektors (Konnektor), der Topikstatus des Referenten (Topik), die referenzielle Distanz zwischen der Anapher und dem Antezedens (RefeDist), der potenzielle Mitbewerber mit dem Antezedens (Mitbewerber) und die Textsorte (Textsorte). Darüber hinaus soll ebenso das Interaktionspotenzial zwischen einzelnen Faktoren (in Abb. 1 mit gestrichelten Linien markiert) beachtet werden.

2.3.1 Semantische Faktoren

Bei der **Belebtheit eines Referenten** handelt es sich um eine ontologische Kategorisierung, die nicht auf zufälliger grammatischer Funktionszuweisung, sondern auf einer text- und situationsunabhängigen semantischen Klassifizierung beruht (vgl. Frank 2019, S. 103). Unterschieden wird zwischen unbelebten und belebten Entitäten und innerhalb letzterer wiederum zwischen Menschen und Tieren. Es wird festgestellt, dass die belebten Entitäten eine Präferenz für reduzierte anaphorische Formen zeigen. Beispielsweise werden Personalpronomen im Vergleich zu Demonstrativpronomen (vgl. Portele/Bader 2016; Ellert/Holler 2011; Ellert 2010) und Vollformen (vgl. Frank 2019) wie auch Nullformen im Vergleich zu allen anderen Formen (vgl. Speyer 2016) präferiert, wenn der Diskursreferent eine belebte Entität ist. Zu bemerken ist, dass solch eine Tendenz sprachspezifisch sein könnte. In Ellert (2010) erweist sich die Belebtheit im Deutschen bei der Auflösung (d.h. Interpretation) von Personalpronomen als Einflussfaktor, was im Niederländischen aber nicht der Fall ist.

Der zweite semantische Faktor bezieht sich auf die **semantische Rolle des Antezedens**. Die semantische Rolle bzw. die Agentivität tragen zur Salienz eines Referenten im Diskurs bei. Wenn das Antezedens eine Agens-Rolle einnimmt, wird es mental auffälliger repräsentiert, weshalb eine reduzierte anaphorische Form gefordert wird. In manchen Untersuchungen wurde durch experimentelle Methoden herausgefunden, dass dieser Faktor im Vergleich zu anderen Einflussfaktoren (wie z. B. syntaktische Funktion, Position vom Antezedens usw.) einen stärkeren Einfluss haben könnte (Schumacher/Roberts/Järvikivi 2017; Schumacher/Dangl/Uzun 2016; Schumacher/Backhaus/Dangl 2015). In seiner Korpusanalyse stellte Speyer (2016) jedoch nur einen schwachen Einfluss der semantischen Rolle des Antezedens fest. Die Nullformen im Althochdeutschen weisen zwar häufig auf ein Agens hin, doch ist diese Tendenz gemäß der Untersuchung statistisch nicht signifikant. Die kontroversen Untersuchungsergebnisse lassen sich eher auf die unterschiedlichen Untersuchungsmethoden zurückführen.

2.3.2 Morphosyntaktische Faktoren

Der erste Einflussfaktor ist die **morphologische Form des Antezedens**, die etwas mit der Bestimmtheit eines Referenten zu tun hat. In der vorhandenen Literatur wurde durch Korpusanalysen (vgl. Portele/Bader 2016) sowie Experimente (vgl. Wilson 2009) gezeigt, dass die Bestimmtheit eines Antezedens sowohl die Produktion als auch die Auflösung von Anaphern beeinflussen könnte. Möglicherweise lässt sich die Wirkung der morphologischen Form des Antezedens auf den Diskursstatus des Referenten als diskurs-alt oder diskurs-neu erklären. Ein diskurs-älterer Referent könnte als auffälliger wahrgenommen werden, was zur Verwendung einer reduzierten anaphorischen Form führen könnte.

Einer der am häufigsten in der Literatur erforschten Faktoren ist die **syntaktische Funktion des Antezedens**. Zuerst wurde in einigen Untersuchungen eine vereinfachte Tendenz festgestellt, nämlich ein subjektives Antezedens eine morphologisch kürzere Anapher-Form bevorzugt. Beispielsweise wird ein Personalpronomen im Vergleich zu einem Demonstrativpronomen (vgl. Bosch/Katz/Umbach 2007; Bosch/Rozario/Zhao 2003; Kaiser 2011) oder eine Nullform im Vergleich zu anderen Formen bevorzugt (Speyer 2016), wenn das Antezedens Subjekt ist. Diese Tendenz wurde sodann in einigen Forschungsbeiträgen angezweifelt. Da das Subjekt in einem Satz normalerweise sowohl das Agens als auch die erste eingeführte Entität darstellt, sei es schwierig, eindeutig zu bestimmen, welcher der drei Faktoren – Subjekt, Agens oder die erste Entität – eine wichtigere Rolle spielt. Untersuchungen haben gezeigt, dass andere Einflussfaktoren, beispielsweise die Agentivität des Antezedens (Schumacher/Roberts/Järvikivi 2017; Schumacher/Dangl/Uzun 2016; Schumacher/Backhaus/Dangl 2015), der Topikstatus des Referenten (vgl. Bosch/Umbach 2007; Bosch/Hinterwimmer 2016) oder die Wechselwirkung mehrerer Faktoren (vgl. Bader/Portele 2019; Portele/Bader 2017; Portele/Bader 2016), ausschlaggebend sein können. Zudem könnte die Auswirkung sprachübergreifend variieren. Laut Colonna/Schimke/Hemforth (2014) hat die syntaktische Funktion des Antezedens im Deutschen eine andere Wirkung als beispielsweise im Französischen.

Der dritte Einflussfaktor betrifft die **syntaktische Funktion der Anapher**. Unterschiedliche anaphorische Formen werden von verschiedenen syntaktischen Funktionen bevorzugt. Zum Beispiel wurde in Bosch/Katz/Umbach (2007) festgestellt, dass die meisten Personalpronomen und Demonstrativpronomen in ihrem Korpus von Zeitungstexten die Funktion des Subjekts übernehmen. Dies erlaubt die Erklärung, dass sich eine Anapher,

die eine prominentere syntaktische Funktion z. B. als Subjekt in einem Satz einnimmt, tendenziell in einer kürzeren Form realisiert. In Sauermann/Gagarina (2017) und Streb/Rösler/Hennighausen (1999) wurde nicht nur die syntaktische Funktion der Anapher in den Vordergrund gerückt, sondern ebenso der sogenannte **syntaktische Parallelismus** zwischen Anapher und deren Antezedens. Experimentelle Methoden haben gezeigt, dass die Interpretation von Anaphern deutlich erleichtert wird, wenn die Anapher dieselbe syntaktische Funktion wie das Antezedens hat. Dieser Interpretationsvorteil begünstigt eine Realisierung der Anapher in kürzerer Form.

Die **Position des Antezedens** spielt ebenfalls eine wichtige Rolle. Aufgrund der relativ flexiblen Wortstellung in der deutschen Sprache wurden viele Untersuchungen durchgeführt, um zu ermitteln, welche Faktoren die Prominenz eines Diskursreferenten bestimmen, nämlich Subjektivität, Agentivität oder die Position als erste Entität. In einigen Untersuchungen wurde festgestellt, dass sich darauf die Position des Antezedens auswirken könnte – im Zusammenspiel mit anderen Einflussfaktoren wie der syntaktischen Funktion und der semantischen Rolle (vgl. Bosch/Umbach 2007; Bader/Portele 2019; Portele/Bader 2017; Portele/Bader 2016). Schumacher und seine Kollegen haben zudem deutlich gemacht, dass alle drei Faktoren, wenn auch in unterschiedlichem Ausmaß, gemeinsam zur Prominenz des Diskursreferenten beitragen (vgl. Schumacher/Roberts/Järvikivi 2017; Schumacher/Dangl/Uzun 2016; Schumacher/Backhaus/Dangl 2015). Es wird postuliert, dass eine Präsentation des Antezedens als initiale Entität in einem Satzgefüge zu einer verstärkten Salienz führt, was zu einer reduzierten anaphorischen Form führt.

In der Studie von Bosch/Katz/Umbach (2007) wurde **die Position der Anapher** in Erwägung gezogen. Sie haben die Positionen von Personalpronomen und Demonstrativpronomen (*vor* vs. *nach dem Verb*) in ihrem Korpus gezählt und kamen zu dem Schluss, dass die verschiedenen Typen von Anaphern nicht gleichmäßig in allen Positionen auftreten könnten. Die Analyse offenbarte eine ausgeprägte und statistisch signifikante Tendenz zur Verwendung des Personalpronomens in der postverbalen Satzposition. Die manifeste Bevorzugung des Demonstrativpronomens zur präverbalen Position wurde zwar suggeriert, erreichte jedoch keine statistische Signifikanz.

2.3.3 Diskurs-pragmatische Faktoren

Zunächst lässt sich feststellen, dass sich eine Reihe von Untersuchungen der Diskursrelation gewidmet haben. Das Hauptaugenmerk liegt dabei bisher auf dem **Konnektor**, weil dieser in der Lage ist, die Diskursrelation explizit zu kennzeichnen. Es wurde ermittelt, dass sowohl der **semantische Inhalt** (*adversativ* vs. *sonstig*) als auch die **strukturelle Eigenschaft** (*eingebettet* vs. *nicht-eingebettet*) des Konnektors wirksam sind, indem sie die Salienz des Referenten beeinflussen (vgl. Holler/Suckow 2016). Weiterhin wurde herausgearbeitet, dass sich der semantische Inhalt sowie die strukturelle Eigenschaft des Konnektors nicht direkt auswirken, sondern zuerst mit der semantischen Rolle interagieren (ebd.; Schumacher/Dangl/Uzun 2016). Die durch die semantische Rolle initiierte Präferenz bei der Pronomenauflösung kann aufgehoben oder sogar umgekehrt werden, wenn in einem koordinierten Nebensatz ein konzessiver Ausdruck präsentiert wird. Es ist erwähnenswert, dass Ellert/Holler (2011) und Speyer (2016) zu gegensätzlichen Ergebnissen gekommen sind: Weder der semantische Inhalt noch die strukturelle Eigenschaft des Kon-

nektors üben einen Einfluss aus. In beiden Untersuchungen wurde jedoch die Wechselwirkung der Faktoren nicht berücksichtigt.

Ein weiterer häufig erwähnter Einflussfaktor ist der **Topikstatus** des Referenten im Satz oder im Diskurs. Der Topikstatus verleiht einem Referenten Salienz, was zu einer Präferenz für reduzierte anaphorische Formen führen kann. Bosch/Hinterwimmer (2016) und Bosch/Umbach (2007) beschäftigten sich mit der Unterscheidung des Topikstatus von anderen Einflussfaktoren und betonten dessen Schlüsselrolle bei der Interpretation von Anaphern. Dabei zeigen Personalpronomen eine Vorliebe für ein Antezedens mit Topikstatus, während Demonstrativpronomen eine Dispräferenz gegenüber Topik aufweisen (vgl. Bosch/Umbach 2007; Ellert/Holler 2011; Ellert 2013; Bittner/Dery 2015; Bosch/Hinterwimmer 2016). Diese Präferenz zeigt sich nicht nur im Deutschen, sondern auch im Französischen (vgl. Colonna/Schimke/Hemforth 2012, 2014). Da der Topikstatus auf der Diskursebene nicht immer leicht zu bestimmen ist, wurde in Portele/Bader (2016) die Bekanntheit (*diskurs-alt* vs. *diskurs-neu*) eines Referenten als Ersatz herangezogen, deren Korpusstudie eine ähnliche Tendenz nachweist.

Der dritte Einflussfaktor ist die **Distanz zwischen Antezedens und Anapher**. Dieser Faktor scheint hauptsächlich verantwortlich zu sein für den Aktivierungsgrad einer Entität in der kognitiven Repräsentation und damit für die Prominenz eines Diskursreferenten zu einem gewissen Zeitpunkt im Textverarbeitungsprozess. Unterschiedliche Untersuchungen haben gezeigt, dass mit zunehmender Distanz zum Antezedens informations- und struktureichere Ausdrücke als Anapher zum Einsatz kommen (vgl. Frank 2019; Bosch/Rozario/Zhao 2003; Portele/Bader 2016).

Es ist offenkundig, dass das Vorhandensein **potenzieller Antezedens-Konkurrenten** die Identifikation des passenden Antezedens erschweren kann, so dass die Verwendung von einer komplizierteren anaphorischen Form erforderlich wird (vgl. Frank 2019, S. 113–115). In Portele und Bader (2016) wird der potenzielle Mitbewerber eines Antezedens berücksichtigt. Laut ihnen gilt eine nicht koreferierende Entität, welches dasselbe Genus und denselben Numerus mit dem Antezedens aufweist, als ein potenzieller Mitbewerber des Antezedens. Wenn ein potenzieller Mitbewerber im unmittelbar vorangegangenen Satz steht, ist ein Einsatz von Personalpronomen im Vergleich zu Demonstrativpronomen weniger wahrscheinlich.

Das Register gilt auch als ein wichtiger Einflussfaktor auf die anaphorische Form. In dieser Korpusstudie wird ein besonderes Augenmerk auf die Varianz verschiedener **Textsorten** gerichtet. In bisherigen Untersuchungen konnten differenzierte Verteilungen von Personalpronomen und Demonstrativpronomen beobachtet werden, zum Beispiel im schriftlichen vs. gesprochenen Deutschen (vgl. Bosch/Katz/Umbach 2007) oder in formellen vs. informellen Texten (vgl. Patil/Bosch/Hinterwimmer 2020). Der Einsatz von *der*-Demonstrativa ist lediglich auf das gesprochene Register beschränkt. Demgegenüber tendieren *diese*-Demonstrativa zu einer Dominanz im formellen Sprachregister. Des Weiteren zeigt sich in Patil/Bosch/Hinterwimmer (ebd.) eine interagierende Auswirkung zwischen Textsorte und der syntaktischen Funktion des Antezedens. Es lässt sich eine verstärkte Dispräferenzierung des *diese*-Demonstrativpronomens gegenüber einem Subjekt in formellen Texten feststellen.

Die in diesem Unterkapitel vorgestellten Einflussfaktoren werden allesamt in der empirischen Untersuchung berücksichtigt und ins Annotationssystem (Kapitel 3.1) integriert.

3. Korpusstudie

3.1 Datengrundlage und Vorgehensweise

Aus der umfassenden Darlegung der bisherigen Untersuchungen und der erwähnten Einflussfaktoren wird ersichtlich, dass die meisten Untersuchungen entweder auf einzelne oder auf nur wenige Einflussfaktoren ausgerichtet sind. In der vorliegenden Studie wird versucht, alle zuvor erläuterten Einflussfaktoren aufeinander zu beziehen und in ihrer Zusammenwirkung zu gewichten, um eine holistische Perspektive zu gewinnen und damit der Komplexität bezüglich der Auswirkung von Einflussfaktoren gerecht zu werden. Dafür wird eine korpusbasierte multifaktorielle Analyse durchgeführt. Methodisch gesehen besteht eine solche Analyse aus drei Schritten: i) Erstellung eines Korpus für ein bestimmtes Sprachphänomen, ii) Annotation der zu untersuchenden sprachlichen Formen und ihrer kontextuellen Einflussfaktoren, iii) Anwendung statistischer Verfahren auf umfangreiche Korpusdaten, um die Zusammenwirkung einzelner Faktoren zu ermitteln.

Für die vorliegende Untersuchung wurde in Anlehnung an das *Brown Corpus* (1961) ein ausgewogenes Korpus aufgebaut, das zu 25% aus Zeitungstexten, zu 25% aus Gebrauchsliteratur, zu 25% aus wissenschaftlichen Texten und zu 25% aus belletristischen Texten besteht. Insgesamt sind 60 Schriftstücke mit jeweils etwa 2.000 Wörtern gesammelt worden. Jedem Text wurde ein Code zugewiesen, und zwar nach dem Muster „DE09 + Subkategorie + Nummer“.

Textsorten	Subkategorien			Anzahl der Texte
Zeitungstexte	A	Reportage	Politik (1), Wirtschaft (1), Sport (1), Gesellschaft (1), Kultur (1), Technik (1), Umwelt (1), Reise (1)	15
	B	Editorial	Leitartikel (1), Kommentar (1), Kolumne (1), Leserbrief (1)	
	C	Rezension	Theaterkritik (1), Buchkritik (1), Musikkritik (1)	
Gebrauchsliteratur	D	Religiöser Text (1)		15
	E	Anweisung (3)		
	F	Folklore (4)		
	G	Biografie (5)		
	H	Bericht und offizielles Dokument (2)		
Wissenschaftliche Texte	J	Wissenschaft	Naturwissenschaft (2), Medizin (1), Mathematik (1), Sozialwissenschaft (3), Politikwissenschaft (1), Jura (1), Pädagogik (1), Geisteswissenschaften (3), Ingenieurwissenschaft (2)	15
Belletristik	K	Allgemeine Fiktion (4)		15
	L	Krimi und Thriller (3)		
	M	Science-Fiction (1)		
	N	Abenteuerroman (3)		
	P	Liebesroman (3)		
	R	Humor (1)		
Gesamtzahl				60

Tab. 2: Ein Überblick über das ausgewogene Korpus

Insgesamt wurden in dem Korpus 3.000 Anapher-Vorkommnisse identifiziert. Jedes Anapher-Vorkommnis (auch „Instanz“ genannt) ist mit einer Antwortvariable (d.h. die Anaphertypen NA, PA oder ZA) und 15 erklärenden Variablen (d.h. die kontextuellen Einflussfaktoren) annotiert worden. Die folgende Tabelle listet alle Faktoren mitsamt ihren Werten auf.

Faktoren		Werte	Skalenniveau
Semantische Faktoren	Belebtheit	belebt, unbelebt	kategorial
	AnteSemR	Agens, nicht-Agens	kategorial
Morpho-syntaktische Faktoren	AnteForm	definit, indefinit	kategorial
	AnteSynF	Subjekt, nicht-Subjekt	kategorial
	AnapSynF	Subjekt, nicht-Subjekt	kategorial
	Parallelismus	parallel, nicht-parallel	kategorial
	AntePos	final, nicht-final	kategorial
	AnapPos	Vorvorfeld, Vorfeld, Mittelfeld, Nachfeld	kategorial
Diskurs-pragmatische Faktoren	AnteTeilTyp	eingebettet, nicht-eingebettet	kategorial
	AnapTeilTyp	eingebettet, nicht-eingebettet	kategorial
	Konnektor	kausalgegenläufig, sonstige, ohne	kategorial
	Topik	innerhalb von 5 Sätzen [1, 2, 3...]	numerisch
	RefeDist	Absatz [1, 2, 3...], Satz [1, 2, 3...], Teilsatz [1, 2, 3...]	numerisch
	Mitbewerber	innerhalb von 1 Teilsatz [ja, nein]	kategorial
	Textsorte	Zeitungstexte, Gebrauchsliteratur, Wissenschaftliche Texte, Belletristik	kategorial

Tab. 3: Gesamtüberblick über alle zu annotierenden Einflussfaktoren und ihre Werte

Korpus-Beispielinstanz:

(5) Im Sommer 1812 war Faraday einundzwanzig. Er war geküsst worden. (DE09K03)

Beispielsweise lässt sich die Anapher-Instanz aus DE09K03 wie folgt annotieren: *Belebtheit* = *belebt*, *AnteSemR* = *Agens*, *AnteForm* = *definit*, *AnteSynF* = *Subjekt*, *AnapSynF* = *Subjekt*, *Parallelismus* = *parallel*, *AntePos* = *final*, *AnapPos* = *Vorfeld*⁴, *AnteTeilTyp* = *nicht-eingebettet*, *AnapTeilTyp* = *nicht-eingebettet*, *Konnektor* = *ohne*, *Topik* = *1*⁵, *Refe*

⁴ In der vorliegenden Studie besetzen nur 7 Anaphern das Vorvorfeld und 31 Anaphern das Nachfeld in ihrem Teilsatz. Um die Datenknappheit zu vermeiden, werden die beiden Werte konsolidiert. Diese methodische Konsolidierung wird durch ein baumstrukturelles Modell realisiert, basierend auf dem Algorithmus *conditional inference tree*. Nach dem Modell beeinflusst die Position der Anapher, ob sie im Vorvorfeld oder im Nachfeld steht, die anaphorische Form nicht signifikant. Daher wird in der weiterführenden Analyse lediglich zwischen Vorfeld, Mittelfeld und einer aggregierten restlichen Kategorie unterschieden.

⁵ Ebenso wird mittels eines Conditional-Inference-Tree-Modells die Variable Topik konsolidiert. Mit dieser Variable wird eigentlich die Erwähnungshäufigkeit eines Referenten quantifiziert, so dass der Topikstatus des Referenten bewertet werden kann. Indem die ursprünglich numerische Variable in eine kate-

*Dist = Absatz0/Satz1/Teilsatz1*⁶, *Mitbewerber = ja*, *Textsorte = Belletristik*. Die annotierten Daten werden in Excel tabellarisch dargestellt, und zwar mit dem Format „Instanz-Variable“ (vgl. Gries 2013, S. 167). Das heißt, jede Zeile bildet eine Instanz von einer Anapher, jede Spalte stellt eine Variable dar.

Um die Annotationsqualität zu sichern, wurde das allgemeinere Maß des *Inter-Annotator-Agreements* mittels *Cohens Kappa* (d. h. der κ -Koeffizient) (vgl. Perkuhn/Keibel/Kupietz 2012, S. 60) eingesetzt. Acht aus insgesamt 60 Texten – jeweils 2 Texte aus den vier Textsorten – wurden stichprobenweise ausgewählt und kreuzannotiert. Für jede der manuell annotierten Variablen wurde die Inter-Annotator-Übereinstimmung zwischen beiden Annotatorinnen mittels des κ -Koeffizienten gemessen. Alle manuell annotierten Variablen weisen einen relativ hohen κ -Koeffizient ($\kappa > 0,7$) sowie eine statistische Signifikanz ($p < 0,001$) auf. Deswegen wird davon ausgegangen, dass die Annotation in der vorliegenden Studie zuverlässig ist.

Bevor man mittels statistischer Modellbildung die multifaktorielle Analyse durchführt, müssen zunächst die annotierten Daten so bearbeitet werden, dass sie den jeweiligen statistischen Voraussetzungen von unterschiedlichen Algorithmen entsprechen. Behandelt wurden mittels des Algorithmus *conditional inference tree* die beiden intervallskalierten Variablen – RefeDist und Topik – und auch die kategoriale Variable mit vier Stufen – AnapPos.

Die bereits annotierten Daten wurden in einem weiteren Schritt mittels der Software R (R Core Team 2021) statistisch ausgewertet. Um die Forschungsfragen dieser Studie zu beantworten, wird in R ein multinomiales logistisches Regressionsmodell unter Verwendung des Pakets {polytomous} (Arppe 2013) implementiert. Nach dem Zufallsprinzip wurden zuerst 70% der Korpusdaten ausgewählt, die als Trainingsdaten der Modellbildung dienen. Die verbleibenden 30% dienen als Testdaten. Das heißt: Zuerst wurde ein Regressionsmodell anhand von 70% der Korpusdaten erstellt, sodann wurde das Modell auf die verbleibenden 30% der Daten angewendet. Das Ziel des Verfahrens ist, die Gültigkeit des Modells zu bewerten.

goriale Form überführt wird, erlaubt es eine differenziertere Erfassung und Einordnung des Topikstatus. In der empirischen Analyse wird demzufolge nur zwischen Tpk0 (nicht einmal erwähnt), Tpk1–2 (ein- bis zweimal erwähnt) und Tpk2+ (mehr als zweimal erwähnt) unterschieden.

⁶ Um die Multikollinearität zu umgehen, werden die drei Maßeinheiten für die referenzielle Distanz vereint. Durch die Anwendung eines Conditional-Inference-Tree-Modells wird eine neue operationalisierte Maßeinheit zur Erfassung der referenziellen Distanz bestimmt. Diese Variable wird von numerischen Variablen zu einer kategorialen Variable mit acht Werten transformiert. Detaillierte Informationen zu dieser Variable sind im Abschnitt 3.2.1 zu finden.

3.2 Ergebnisse

Das durch die Analyse ermittelte optimale Modell enthält neun signifikante Variablen, nämlich AnteSynF, AnapSynF, AnapPos, AnapTeilTyp, RefeDist, Belebtheit, Mitbewerber, Parallelismus und Textsorte. Die Prädiktionsgenauigkeit beträgt für die Trainingsdaten 84,24% und 81,33% für die Testdaten. Die Prädiktionsgenauigkeit des Modells mit neun erklärenden Variablen ist deutlich höher als die Baseline₁ (51,90%), wenn man stets die anaphorische Form mit der höchsten Häufigkeit (NA) auswählt; diese ist auch deutlich höher als die Baseline₂ (33,33%), wenn man immer zufällig aus den drei Alternativen (NA, PA, ZA) eine Auswahl trifft. Zudem weist das Modell ein Nagelkerke R^2 ⁷ von 0,78 und eine AIC⁸ von 1770 auf.

3.2.1 Schlüsselfaktoren und ihre Auswirkungen

Das ermittelte Regressionsmodell mit neun signifikanten Schlüsselfaktoren lässt sich in der folgenden Tabelle darstellen. In der ersten Spalte werden die neun Schlüsselfaktoren (erklärende Variablen) aufgelistet. In den nachfolgenden drei Spalten werden die jeweiligen Koeffizienten von NA, PA und ZA angegeben, mit denen die Auswirkung jeder einzelnen erklärenden Variable auf die Antwortvariable dargestellt werden soll. Die Koeffizienten zeigen an, um wie viel sich die Wahrscheinlichkeit der Entscheidung für eine bestimmte anaphorische Form (im logit-Raum) erhöht oder verringert, wenn sich der Wert der erklärenden Variablen ändert. Ein positiver Koeffizient besagt, dass ein bestimmter Wert der erklärenden Variable im Vergleich zu dem Referenzwert die Wahrscheinlichkeit der entsprechenden anaphorischen Form erhöht. Ein negativer Koeffizient hingegen bedeutet, dass die erklärende Variable mit einem bestimmten Wert die Wahrscheinlichkeit der entsprechenden anaphorischen Form vermindert. Der Absolutwert eines Koeffizienten spiegelt das Maß der Auswirkungen wider. Zudem ist anzumerken, dass die Koeffizienten auf dem Niveau von 0,05 nicht statistisch signifikant sind, wenn sie in Klammern stehen.

⁷ Der Nagelkerke R^2 ist ein von Nagelkerke verbessertes Maß des generalisierten Determinationskoeffizienten. Es wird insbesondere im Kontext von logistischen Regressionen verwendet, um die Güte der Modellanpassung im Vergleich zum maximal möglichen Log-Likelihood-Wert zu bewerten, wobei der Wertebereich zwischen 0 und 1 liegt. Ein Wert über 0,2 bedeutet, dass die Anpassung des Modells akzeptabel ist, während ein Wert über 0,5 auf eine sehr gute Modellanpassung hinweist.

⁸ Das Akaike Information Criterion (AIC) dient dem Vergleich verschiedener Modelle. Der Index berücksichtigt sowohl die statistische Anpassung eines Modells als auch die Anzahl der Parameter, die benötigt werden, um diese Anpassung zu erreichen. Modelle mit kleineren AIC-Werten – was auf eine angemessene Anpassung mit weniger Parametern hinweist – werden bevorzugt.

Erklärende Variablen	NA	PA	ZA
<i>Intercept</i>	2,441	-1,587	(-22,37)
1 belebt = unbelebt	1,173	-0,9907	(0,1551)
2 AnteSynF = Subj	-1,245	0,9473	(-1,599)
3 AnapSynF = Subj	-0,5207	(0,1829)	1,563
4 Parallelismus = unpara	(0,0707)	0,6765	-3,883
5 AnapPos = sons	2,554	-1,954	(-15,65)
AnapPos = Vor	-0,3925	-0,3645	1,439
6 AnapTeilTyp = neinb	(-0,04917)	-0,7234	1,375
7 RefeDist = S0T0	-6,563	5,154	(3,175)
RefeDist = S0T1	-4,885	1,518	(21,6)
RefeDist = S0T1+	-3,187	1,578	(20,29)
RefeDist = A0S0+T1	-2,642	2,404	(-0,1873)
RefeDist = A0S0+T2	-1,592	1,3	(0,1577)
RefeDist = S0+T3-4	(-0,4642)	(0,224)	(0,3498)
RefeDist = S0+T4+	1,47	-1,547	(15,08)
8 Mitbewerber = nein	-0,8793	0,5661	(0,3243)
9 Textsorte = Gebr	0,8195	-0,3471	(-0,5861)
Textsorte = Wiss	1,485	-0,8623	(-0,7217)
Textsorte = Zeit	(0,3756)	(0,1674)	-1,559

Tab. 4: Koeffizienten von den neun Schlüsselfaktoren in dem ermittelten optimalen Regressionsmodell

Belebtheit des Referenten: Wenn sich eine Anapher auf einen unbelebten Referenten im Diskurs bezieht, ist der Koeffizient für NA positiv, während der Koeffizient für PA negativ ist. Das heißt, im Vergleich zu dem Referenzwert (d. h. Belebtheit = belebt) erhöht sich die Chance für den Einsatz von NA, doch die Chance für den Einsatz von PA verringert sich. Der Koeffizient für ZA ist positiv, steht aber in Klammern. Das bedeutet, dass diese Tendenz statistisch nicht signifikant ist, obwohl ZA eine Tendenz zum unbelebten Referenten aufweist. Im linguistischen Sinne lässt sich feststellen, dass **beim unbelebten Referenten NA präferiert wird, während beim belebten Referenten PA bevorzugt wird.**

Syntaktische Funktion des Antezedens: In dieser Studie wird zwischen subjektivem Antezedens und nicht-subjektivem Antezedens unterschieden. Wenn das Antezedens von einer Anapher als Subjekt im Satz fungiert, ist der Koeffizient für NA negativ und für PA positiv. Dies kann so interpretiert werden, dass im Vergleich zu dem Referenzwert (AnteSynF = nSubj) die Chance für den Einsatz von NA kleiner und die Chance für PA größer ist. Der Koeffizient für ZA ist ebenfalls ein negativer Wert, steht aber in Klammern. Das heißt: **Beim subjektiven Antezedens wird PA bevorzugt, während beim nicht-subjektiven Antezedens NA präferiert wird. Obwohl das subjektive Antezedens auch eine Dispräferenz gegen ZA aufweist, ist diese Tendenz statistisch nicht signifikant.**

Syntaktische Funktion der Anapher: Bei der Annotation wird ebenso zwischen subjektiven Anaphern und nicht-subjektiven Anaphern unterschieden. Der Koeffizient für NA ist negativ: Im Vergleich zu dem Referenzwert ($\text{AnapSynF} = \text{nSubj}$) verringert sich die Chance für den Einsatz von NA, wenn die Anapher ein Subjekt ist. Im Gegensatz dazu weisen die Koeffizienten für PA und NA jeweils einen positiven Wert auf. Der erste steht aber in Klammern. Das heißt: Bei einer subjektiven Anapher steigt die Chance für den Einsatz von einem Pronomen oder einer Nullform, wobei die Bevorzugung von Pronomen statistisch nicht signifikant ist. Mit anderen Worten: **Im Fall einer subjektiven Anapher wird eine Nullform präferiert, während bei einer nicht-subjektiven Anapher eine lexikalische NP-Phrase bevorzugt wird.**

Parallelismus: Mit Parallelismus wird annotiert, ob eine Anapher über dieselbe syntaktische Funktion mit ihrem Antezedens verfügt. Wenn die beiden dieselbe syntaktische Funktion aufweisen, werden sie als „(syntaktisch) parallel“ angesehen, ansonsten als „(syntaktisch) unparallel“. Der Koeffizient für NA hat einen kleinen Absolutwert und erscheint in Klammern. Das heißt, der Einflussfaktor Parallelismus hat fast keinen signifikanten Einfluss auf NA. Für PA ist der Koeffizient positiv: Im Vergleich zu dem Referenzwert ($\text{Parallelismus} = \text{para}$) ist die Wahrscheinlichkeit für den Einsatz von PA bei unparallelen syntaktischen Funktionen größer. Im Gegensatz dazu weist der Koeffizient für ZA einen negativen Wert auf, und zwar mit einem relativ großen Absolutwert. Das bedeutet, dass die Wahrscheinlichkeit für den Einsatz von ZA bei unparallelen syntaktischen Funktionen deutlich niedriger ist. Insgesamt zeigt sich, dass **im Fall der parallelen syntaktischen Funktion zwischen Anapher und ihrem Antezedens eine starke Präferenz für ZA vorliegt.**

Korpus-Beispielinstanz:

(6) Ein Paternoster_i hat keine Türen und Ø_j fährt ganz langsam. (DE09N01)

Position der Anapher: Dieser Einflussfaktor umfasst drei Werte, nämlich Vorfeld, Mittelfeld und sonstige Felder. Mittelfeld ergibt sich als Referenzwert und wird nicht in der obigen Tabelle angegeben. Zuerst lassen sich die Koeffizienten für NA wie folgt interpretieren: Im Vergleich zu dem Referenzwert ($\text{AnapPos} = \text{Mittel}$) verringert sich die Chance für den Einsatz von NA, wenn die Anapher im Vorfeld des Satzes auftritt. Wenn die Anapher jedoch in sonstigen Feldern (d. h. Nachfeld und Vorvorfeld) auftritt, erhöht sich die Chance für NA. Die Koeffizienten für PA sind negativ. Das heißt: Im Vergleich zum Mittelfeld ist die Chance für den Einsatz von PA geringer, wenn sich die Anapher im Vorfeld oder sonstigen Feldern des Satzes befindet. Zudem weist der Koeffizient für den Wert $\text{AnapPos} = \text{sons}$ einen größeren Absolutwert, was wie folgt interpretiert werden kann: Im Vergleich zum Vorfeld wird PA in sonstigen Feldern noch stärker dispräferiert. ZA hat einen negativen Koeffizienten für den Wert $\text{AnapPos} = \text{sons}$ und einen positiven Koeffizienten für den Wert $\text{AnapPos} = \text{Vor}$. Trotz des großen Absolutwertes ist der erste Koeffizient statistisch nicht signifikant. Der letztere Koeffizient verdeutlicht, dass im Vergleich zum Mittelfeld die Chance für den Einsatz von ZA im Vorfeld höher ist. Zusammenfassend lässt sich feststellen, dass **im Vorfeld ZA, im Mittelfeld PA und in sonstigen Feldern NA präferiert wird.**

Korpus-Beispielinstanz:

- (7) Er_i schrieb ihn 1966 in einem Aufsatz mit dem Titel: „Gold und wirtschaftliche Freiheit“ und Ø_i outete sich als Fan eines gedeckten Währungssystems. (DE09B02B)
- (8) Auch wenn die alte Koalition_i im Amt bliebe, müsste sie_i einige ihrer wichtigsten Politikfelder neu entdecken. (DE09A01A)
- (9) Es gibt in diesem Roman eine Figur seines Namens_i, [...] Er ist mindestens ebenso ein Alter Ego des Autors wie die Figur gleichen Namens_i, umso gewagter, [...] (DE09C02A)

Teilsatztyp der Anapher: In den Korpusdaten wird annotiert, ob der Teilsatz, in dem die Anapher auftritt, eingebettet oder nicht-eingebettet ist. NA hat einen negativen Koeffizienten mit einem sehr geringen Absolutwert und weist keine statistische Signifikanz auf. Für PA ist der Koeffizient auch negativ. Das heißt, im Vergleich zum Referenzwert (AnapTeilTyp = einb) ist die Wahrscheinlichkeit für den Einsatz für PA in einem nicht-eingebetteten Teilsatz niedriger. Im Gegensatz dazu hat ZA einen positiven Koeffizienten. Insgesamt bedeutet es: **In einem nicht eingebetteten Teilsatz wird ZA bevorzugt, während in einem eingebetteten Teilsatz PA präferiert wird.**

Korpus-Beispielinstanz:

- 10) Petersdorf_i war groß gewachsen, gut aussehend und Ø_i hatte einen freundlichen, sympathischen Gesichtsausdruck. Es zeigte sich dann auch, dass man mit ihm_i zurechtkommen konnte. (DE09G04)

Referenzielle Distanz zwischen der Anapher und ihrem Antezedens: In dieser empirischen Studie wurden ursprünglich drei intervallskalierte Maßeinheiten (Teilsatz, Satz, Absatz) für die referenzielle Distanz genutzt. Aus statischen Gründen wurde eine Transformation mithilfe des Modells *Conditional-Inference-Tree* vorgenommen. Die konsolidierte neue kategoriale Variable mit acht Werten wird in der folgenden Tabelle dargestellt:

Wert	Beschreibung
S0T0	Die Anapher und ihr Antezedens sind in demselben Satz (S0) bzw. demselben Teilsatz (T0).
S0T1	Die Anapher und ihr Antezedens sind in demselben Satz (S0), aber durch einen Teilsatz getrennt (T1).
S0T1+	Die Anapher und ihr Antezedens sind in demselben Satz (S0), aber durch mehr als einen Teilsatz getrennt (T1+).
A0S0+T1	Die Anapher und ihr Antezedens sind in demselben Absatz (A0), aber durch mindestens einen Satz (S0+) und von einem Teilsatz (T1) getrennt.
A0S0+T2	Die Anapher und ihr Antezedens sind in demselben Absatz (A0), aber durch mindestens einen Satz (S0+) und durch zwei Teilsätze (T2) getrennt.
A0+S0+T1–2	Die Anapher und ihr Antezedens sind durch mindestens einen Absatz (A0+) und einen Satz (S0+) und durch einen oder zwei Teilsätzen (T1–2) getrennt.
S0+T3–4	Die Anapher und ihr Antezedens sind durch mindestens einen Satz (S0+) und durch drei oder vier Teilsätzen (T3–4) getrennt.
S0+T4+	Die Anapher und ihr Antezedens sind durch mindestens einen Satz (S0+) und fünf Teilsätze (T4+) getrennt.

Tab. 5: Operationalisierbare Einheit zur Messung der referenziellen Distanz

Bei der geringsten referenziellen Distanz – eine Anapher und ihr Antezedens liegen in demselben Teilsatz (d. h. RefeDist = S0T0) – ist der Koeffizient für NA negativ und zwar mit einem relativ großen Absolutwert. Das heißt, im Vergleich zum Referenzwert (RefeDist = A0+S0+T1–2) ist die Chance für den Einsatz von NA deutlich niedriger. Im Gegenteil dazu ist der Koeffizient für PA positiv und weist auch einen relativ großen Absolutwert auf. Demzufolge ist die Chance für PA deutlich höher. Obwohl ZA einen positiven Koeffizienten hat, gilt dieser jedoch als statistisch nicht signifikant. Wenn die Anapher und ihr Antezedens in demselben Satz, aber nicht mehr in demselben Teilsatz liegen (d. h. RefeDist = S0T1, RefeDist = S0T1+), weisen die drei Anapher-Typen eine sehr ähnliche Tendenz auf. Die Chance für NA ist kleiner (jeweils mit einem Koeffizienten von –4,885 und –3,187), die für PA ist ein minimal größer (jeweils mit einem Koeffizienten von 1,518 und 1,578). Die Chance für ZA ist erwähnenswert deutlich höher (jeweils mit einem Koeffizienten von 21,6 und 20,29), was aber statistisch nicht signifikant ist. Wenn die Anapher und ihr Antezedens nicht mehr in demselben Satz auftreten und gleichzeitig durch nicht zu viele Teilsätze getrennt sind (d. h. RefeDist = A0S0+T1, RefeDist = A0S0+T2, RefeDist = S0+T3–4), ist die Chance für NA niedriger und die Chance für PA höher. Zudem weist der jeweilige Koeffizient für ZA einen niedrigen Absolutwert auf, der statistisch nicht signifikant ist. Bei einer größeren referenziellen Distanz (d. h. RefeDist = S0+T4+) ist die Situation eine völlig andere: Im Vergleich zum Referenzwert steigt die Chance für NA und für PA sinkt sie. Der Koeffizient für ZA hat einen großen Absolutwert, ist aber statistisch nicht signifikant. Demnach kann man grob zwischen vier Kategorien von referenziellen Distanzen unterscheiden: Null < Kurz < Mitten < Lang. Je größer die referenzielle Distanz ist, desto größer ist die Chance für den Einsatz von NA und desto kleiner ist die Chance für

den Einsatz von PA. Das heißt, **mit steigender referenzieller Distanz verstärken sich ebenfalls die Präferenz für NA sowie die Dispräferenz gegen PA.**

Potenzielle Mitbewerber vom Antezedens: In der vorliegenden Studie gilt eine nicht koreferierende Entität, die dasselbe Genus und denselben Numerus mit dem Antezedens aufweist, als dessen potenzieller Mitbewerber. Wenn ein potenzieller Mitbewerber in demselben oder dem unmittelbar vorangegangenen Teilsatz auftritt, wird die Variable mit „ja“ annotiert, sonst mit „nein“. Wie oben in Tabelle 4 dargestellt wird, ist die Chance für NA im Vergleich zum Referenzwert (Mitbewerber = ja) kleiner und die Chance für PA größer, wenn es keine Mitbewerber gibt. Mit anderen Worten: **Wenn keine Ambiguität im Kontext vorliegt, wird ein Pronomen vor einer lexikalischen Nominalphrase als Anapher bevorzugt.**

Textsorte: Die vorliegende Studie basiert auf einem ausgewogenen Korpus, das aus Zeitungstexten, Gebrauchsliteratur, wissenschaftlichen Texten und Belletristik besteht. Im Vergleich zum Referenzwert (Textsorte = Bell) ist die Chance für den Einsatz von NA in der Gebrauchsliteratur und in wissenschaftlichen Texten größer und für PA kleiner. Zudem ist es bemerkenswert, dass die Chance für den Einsatz von ZA in Zeitungstexten im Vergleich zu Belletristik deutlich kleiner ist. Es kann angenommen werden, dass **bei steigender Fachlichkeit, die ein erhöhtes Maß an Fachkenntnissen oder Vorwissen des Lesers voraussetzt, komplexere anaphorische Formen bevorzugt werden.**

3.2.2 Typische Kontexte jeweils für NA, PA und ZA

Neben einer tabellarischen Auflistung von Schlüsselfaktoren (Variablen) mitsamt entsprechenden Koeffizienten erlaubt die Analyse mit dem Regressionsmodell außerdem die Ermittlung von typischen Kontexten (d. h. hochfrequente Kombinationen von erklärenden Variablen) für jeden Anapherotyp. Im Folgenden werden – jeweils für NA, PA und ZA – die 10 am häufigsten vorkommenden Kontexte aufgelistet und mit Beispielen veranschaulicht.

Die folgende Tabelle gibt einen Überblick über 10 verschiedene Kombinationen von Kontextvariablen, die vor allem für NA typisch sind. Es sind die wiederkehrenden Kontexte (d. h. mit relativ hoher Häufigkeit), in denen man sich fast ohne Zweifel für eine NA entscheidet (d. h. mit relativ hoher Wahrscheinlichkeit).

Häufig- keit	Belebt- heit	Ante SynF	Anap SynF	Parallel	Anap Pos	Anap TeilTyp	RefeDist	Mith.	Text- sorte	Wahrsch. von NA
20	unbelebt	nSubj	nSubj	unpara	Mittel	neinb	S0+T4+	ja	Wiss	0,993
18	unbelebt	nSubj	nSubj	para	Mittel	einb	S0+T4+	ja	Wiss	0,993
11	unbelebt	nSubj	Subj	unpara	Vor	neinb	S0+T4+	ja	Wiss	0,993
11	unbelebt	nSubj	nSubj	unpara	Vor	neinb	S0+T4+	ja	Gebr	0,991
17	unbelebt	nSubj	nSubj	unpara	Mittel	neinb	S0+T4+	ja	Gebr	0,988
14	unbelebt	nSubj	nSubj	unpara	Mittel	einb	S0+T4+	ja	Wiss	0,987
13	unbelebt	nSubj	nSubj	unpara	Mittel	neinb	S0+T4+	nein	Wiss	0,987
15	unbelebt	nSubj	nSubj	unpara	Mittel	neinb	S0+T4+	ja	Zeit	0,982
15	unbelebt	Subj	nSubj	unpara	Mittel	neinb	S0+T4+	ja	Wiss	0,981
12	unbelebt	nSubj	nSubj	unpara	Mittel	einb	S0+T4+	ja	Gebr	0,977

Tab. 6: Typische Kontexte (Kombinationen von erklärenden Variablen) für NA

Wie die obige Tabelle zeigt, wird NA in Kontexten mit folgenden Eigenschaften bevorzugt: 1. Der Referent ist unbelebt. 2. Weder die Anapher noch ihr Antezedens fungieren im Satz als Subjekt. 3. Die Anapher übernimmt nicht die identische syntaktische Funktion wie ihr Antezedens. Bezüglich der Position im Teilsatz wird die Anapher vor allem im Mittelfeld auftreten. Zwischen der Anapher und ihrem Antezedens gibt es normalerweise eine relativ größere referenzielle Distanz (d. h. RefeDist = S0+T4+). In demselben oder dem unmittelbar vorangegangenen Teilsatz der Anapher ist mindestens ein Mitbewerber zu finden. Texte, in denen NA mit einer höheren Wahrscheinlichkeit auftritt, weisen auch einen höheren Fachlichkeitsgrad auf – NA wird vor allem in wissenschaftlichen Texten und in der Gebrauchsliteratur bevorzugt.

Korpus-Beispielinstanz:

- (11a) Eine innovative Neuerung hinsichtlich neuartiger Konfiguration, stellt die Batch Rektifikation mit Mittelbehälter dar,
- (11b) die erstmalig von Robinson und Gilliland erwähnt,
- (11c) jedoch nicht weiter untersucht wurde.
- (12a) Die ersten Untersuchungen fanden unabhängig voneinander durch Bortolini und Guarise sowie Treybal statt.
- (12b) Sie demonstrieren aber lediglich die Durchführbarkeit des Prozesses.
- (12c) Hasebe et al. untersuchten diese Konfiguration, unter dem Begriff Multi-Effect Batch Distillation, [...] (DE09J14)

In DE09J14 bezieht sich die Anapher *diese Konfiguration* in (12c) auf einen unbelebten Diskursreferenten. Sie und ihr Antezedens *neuartiger Konfiguration* in (11a) fungieren im Satz nicht als Subjekt und weisen keine parallele syntaktische Relation auf. Die Anapher befindet sich im Mittelfeld eines nicht-eingebetteten Teilsatzes und ist von ihrem Antezedens weit entfernt. In dem unmittelbar vorangegangenen Teilsatz besteht ein Mitbewerber des Antezedens, nämlich *die Durchführbarkeit*. In diesem Kontext besteht eine äußerst hohe Wahrscheinlichkeit für einen Einsatz von NA (0,993).

Die folgende Tabelle bietet eine Zusammenfassung der zehn verschiedenen Kombinationen der Kontextvariablen, die für **PA** als typisch gelten.

Häufig- keit	Belebt- heit	Ante SynF	Anap SynF	Parallel	Anap Pos	Anap TeilTyp	RefeDist	Mit.	Text- sorte	Wahrsch. von PA
21	belebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	nein	Bell	0,997
10	belebt	Subj	nSubj	unpara	Mittel	einb	SOT0	nein	Bell	0,997
18	belebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	nein	Gebr	0,994
9	belebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	ja	Bell	0,993
11	belebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	ja	Zeit	0,990
12	belebt	Subj	nSubj	unpara	Mittel	einb	SOT1	nein	Bell	0,987
10	belebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	ja	Gebr	0,987
8	belebt	Subj	nSubj	unpara	Mittel	einb	SOT1	nein	Gebr	0,977
9	belebt	Subj	nSubj	unpara	Mittel	einb	SOT1	ja	Bell	0,972
8	unbelebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	nein	Wiss	0,961
8	unbelebt	Subj	nSubj	unpara	Mittel	neinb	SOT0	ja	Gebr	0,960

Tab. 7: Typische Kontexte (Kombinationen von erklärenden Variablen) für PA

Es lässt sich sofort erkennen, dass zwischen den typischen Kontexten von PA und den typischen Kontexten von NA wesentliche Unterschiede bestehen. PA zeigt eine deutliche Präferenz für Kontexte mit folgenden Eigenschaften: 1. Der Referent ist eine belebte Entität im Diskurs. 2. Im Gegensatz zu NA fungiert das Antezedens im Satz als Subjekt. 3. Die Anapher selbst ist nicht das Subjekt und befindet sich meistens im Mittelfeld eines Satzes. Bezüglich der referenziellen Distanz weist PA charakteristische Merkmale auf: Das Antezedens befindet sich entweder in demselben Teilsatz wie die Anapher (d.h. RefeDist = S0T0) oder im unmittelbar vorangegangenen Teilsatz (d.h. RefeDist = S0T1). Darüber hinaus wird PA in Texten mit einem niedrigeren Fachlichkeitsgrad bevorzugt.

Korpus-Beispielinstanz:

- (13a) Da er_i nie besonders ehrgeizig oder fleißig gewesen war,
- (13b) hatte er_i es auch nicht so weit wie sein_i Bruder gebracht. (DE09L02)
- (14a) Der Junge_i hob einen Stein auf,
- (14b) der vor ihm_i auf dem Bergpfad lag,
- (14c) [...] (DE09N02)

Die obigen Beispielsätze stellen die typischen Kontexte von PA dar. In (13b) ist die Anapher das Possessivpronomen *sein* und in (14b) ist die Anapher das Personalpronomen *ihm*. Beide Anaphern beziehen sich auf belebte Diskursreferenten und weisen relativ kürzere referenzielle Distanzen auf: *sein* in (13b) befindet sich in demselben Teilsatz ihres Antezedens und *ihm* im nächsten Teilsatz.

Schließlich lassen sich die zehn typischen Kontexte für ZA in Tabelle 8 demonstrieren. Zu bemerken ist, dass die von dem Regressionsmodell vorhergesagten Wahrscheinlichkeiten von ZA deutlich niedriger sind als die von NA und PA. Trotz geringerer Genauigkeit bei der Prädiktion zeigt ZA auch eine klare Bevorzugung von bestimmten Kontextvariablen gegenüber den anderen.

Häufig- keit	Belebt- heit	Ante SynF	Anap SynF	Parallel	Anap Pos	Anap TeilTyp	RefeDist	Mitb.	Text- sorte	Wahrsch. von ZA
11	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	ja	Wiss	0,740
18	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	nein	Wiss	0,737
15	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	ja	Gebr	0,732
17	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	nein	Gebr	0,684
7	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	nein	Bell	0,671
6	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	ja	Zeit	0,667
12	belebt	Subj	Subj	para	Vor	neinb	SOT1	ja	Gebr	0,639
12	belebt	Subj	Subj	para	Vor	neinb	SOT1	ja	Bell	0,631
17	unbelebt	Subj	Subj	para	Vor	neinb	SOT1	nein	Zeit	0,612
12	unbelebt	Subj	Subj	para	Mittel	neinb	SOT1	nein	Wiss	0,586

Tab. 8: Typische Kontexte (Kombinationen von erklärenden Variablen) für ZA

Die meisten typischen Kontexte von ZA weisen die folgenden kontextuellen Eigenschaften auf: AnteSynF = Subjekt, AnapSynF = Subjekt, Parallelismus = parallel, AnapPos = Vor, AnapTeilTyp = nicht-eingebettet, RefeDist = S0T1. Es lässt sich sofort erkennen, dass ZA vor allem bei einer sogenannten „Kette“ desselben Themas eingesetzt wird.

Korpus-Beispielinstanz:

- (15a) Das heißt,
- (15b) sie_i können die gefäß- und zellschädigenden Radikale im Körper ausschalten
- (15c) und Ø_i schützen auf diese Weise vor Arteriosklerose und Krebs. (DE09E01)
- (16a) Die Frauen_i hören konzentriert zu,
- (16b) Ø_i wägen ab,
- (16c) Ø_i antworten bedächtig,
- (16d) Ø_i danken Deutschland. (DE09A03A)

Wie in den beiden Beispielsätzen wird zunächst ein Referent in dem ersten Teilsatz eingeführt und als Thema festgelegt. Mit der Entfaltung des Themas wird weder ein Pronomen noch eine lexikalische Nominalphrase in den unmittelbar folgenden Teilsätzen eingesetzt, sondern eine Null-Form.

3.3 Zusammenfassung der Ergebnisse

In dieser Studie wurde mittels eines multinomialen logistischen Regressionsmodells die Auswahl spezifischer anaphorischer Formen aus NA, PA und ZA modelliert. Eine Reihe von relevanten Kontextvariablen, die signifikanten Einfluss auf die Wahl ausüben, wurden mithilfe des Modells identifiziert. Im Folgenden wird versucht, die neun ermittelten Schlüsselfaktoren in Bezug auf die vier von Ariel (1990) vorgeschlagenen Dimensionen zu verorten und im Rahmen der *Accessibility Theory* zu interpretieren.

Dimension 1: Distanz zwischen der Anapher und dem Antezedens

Die **referenzielle Distanz** zwischen der Anapher und dem Antezedens zeigt eine inverse Korrelation mit der kognitiven Zugänglichkeit des Referenten. Das bedeutet, dass der Aktivierungsgrad der referierten Entität in der mentalen Repräsentation mit zunehmender Distanz tendenziell abnimmt. Innerhalb der durchgeführten Untersuchung dieser Arbeit konnte empirisch bestätigt werden, dass die morphologisch komplexere und semantisch reichere NA – im Vergleich zu PA und ZA – eine signifikante Präferenz für größere referenzielle Distanzen aufweist.

Dimension 2: Konkurrierende Referenten

Die Variable **Mitbewerber** ist ebenfalls für den Aktivierungsgrad eines Diskursreferenten im Textverarbeitungsprozess verantwortlich. Die Präsenz potenzieller Konkurrenten zwischen der Anapher und dem Antezedens erschwert den kognitiven Zugriff auf den Referenten. In der vorliegenden Analyse konnte nachgewiesen werden, dass die morphologisch komplexere und semantisch nuanciertere NA – im Gegensatz zu PA und ZA – in einem Kontext mit stärkerer Interferenz bevorzugt wird.

Dimension 3: Salienz des Referenten

Die **Belebtheit** eines Diskursreferenten beeinflusst dessen kognitive Salienz unabhängig vom Kontext. Belebte Referenten sind in der menschlichen Kognition auffällender verankert als unbelebte (vgl. Kibrik 2011, S. 406). Ein belebter Referent nimmt folglich eine

herausragendere Rolle im Diskurs ein und wird eher durch PA als durch NA wieder aufgenommen, da die PA weniger Informationen explizit im Text kodiert, was die Korpusanalyse belegt.

Referenten, die mit der **syntaktischen Funktion** des Subjekts kodiert werden, erweisen sich als herausragende Entitäten im Diskurs (vgl. Ariel 1990, S. 45). Mit der vorliegenden Untersuchung lässt sich empirisch belegen, dass die syntaktische Rolle sowohl des Antezedens als auch der Anapher von signifikanter Relevanz ist. Ein Antezedens oder eine Anapher in subjektiver Funktion zeichnen sich durch eine erhöhte Zugänglichkeit in der mentalen Repräsentation aus. Bei dieser Konstellation wird eine kürzere anaphorische Form präferiert.

Bei der Variablen **Parallelismus** wird geprüft, ob die Anapher dieselbe syntaktische Funktion wie das Antezedens hat. In syntaktischen Strukturen, in denen die Entitäten infolge ihrer Subjektposition herausragender wirken oder in denen die Re-Aktivierung wegen des Parallelismus kognitiv erleichtert wird, wird NA dispräferiert. Und die morphologisch kürzeste und semantisch ärmste ZA bevorzugt Kontexte, in denen die Anapher und ihr Antezedens gleichzeitig als Subjekt fungieren.

In Bezug auf die **Position der Anapher** ist das Vorfeld gewöhnlich für ein thematisches Element (vgl. Duden 2009, S.1123) reserviert. Das Vorfeld dient nämlich zur Platzierung von sprachlichen Elementen, die als gegebene Information aus dem Kontext zu erschließen sind, oder von Elementen, die als Kontrast hervorgehoben werden sollen. Die Korpusdaten zeigen, dass ZA eine starke Präferenz für das Vorfeld im Satz hat. Das steht im Einklang mit der Hypothese: Je herausragender der Referent ist, desto weniger ist Information an der Kodierung auf der Textoberfläche erforderlich.

Dimension 4: Einheitlichkeit von Anapher und Antezedens

Der Faktor Einheitlichkeit betrifft die Frage, ob sich das Antezedens innerhalb desselben Rahmens/derselben Welt/des gleichen Standpunkts/des gleichen Abschnitts oder Absatzes wie die Anapher befindet (Ariel 1990, S. 29). **AnapTeilTyp** bezieht sich auf die strukturelle Eigenschaft (eingebettet vs. nicht-eingebettet) des Teilsatzes, in dem die Anapher auftritt. Das Auftreten der Anapher in einem eingebetteten Teilsatz, im Kontrast zu einem nicht-eingebetteten, mindert die Einheitlichkeit zwischen Anapher und Antezedens und kann den Aktivierungsgrad des Referenten beeinträchtigen. Es ist in der Korpusstudie nachweisbar, dass ZA – im Vergleich zu NA und PA – den nicht-eingebetteten Teilsatz bevorzugt. Dies scheint mit der Hypothese von Holler und Suckow (vgl. Holler/Suckow 2016, S. 82) übereinzustimmen, dass die Prominenz einer Diskursentität von der strukturellen Eigenschaft des Teilsatzes beeinflusst wird.

Zusätzlich zu den Einflussfaktoren, die in die obigen vier Dimensionen eingeordnet werden können, zeigt die vorliegende Studie, dass die Textsorte ebenfalls eine zentrale Rolle beim anaphorischen Verweis spielt. Aus den empirischen Ergebnissen geht hervor, dass abhängig von der jeweiligen Textsortenkategorie differenzierte Präferenzen für anaphorische Formen existieren. In der Gebrauchsliteratur und in wissenschaftlichen Texten ist die Wahrscheinlichkeit des Einsatzes von NA höher, während die Wahrscheinlichkeit des Einsatzes von PA reduziert ist. In Zeitungstexten ist die Verwendung von ZA verglichen mit literarischen Texten signifikant seltener. Dies lässt sich auf die Fachlichkeit eines Textes zurückführen. Mit dem zunehmenden Fachlichkeitsgrad (wissenschaftliche

Texte > Gebrauchsliteratur > Zeitungstexte > literarische Texte) werden mehr Vorkenntnisse, Vorwissen usw. vorausgesetzt, was die kognitive Aktivierung des betreffenden Referenten erschwert und insofern die Zugänglichkeit negativ beeinflusst. Das führt dazu, dass in solchen Fällen morphologisch komplexere und semantisch reichere NA präferiert werden, so dass die Identifikation reibungsloser zustande kommt.

Die vorliegende Studie hat Ariels *Accessibility Theory*, die mit vier eher abstrakten Dimensionen arbeitet, mit Hilfe von neun Schlüsselfaktoren konkretisiert. Vermittels Koeffizienten wird die Gewichtung der Einflussübung einzelner Schlüsselfaktoren ersichtlich. Besonders der Einflussfaktor der Textsorte, der weder in früheren entsprechenden Untersuchungen noch in Ariels Theorie genügend berücksichtigt worden ist, wird nun erneut ins Feld geführt.

Die ermittelten neun Schlüsselfaktoren können heuristisch in zwei Gruppen eingeordnet werden:

Intrinsische Faktoren: Diese Faktoren sind unmittelbar mit den Eigenschaften des Referenzobjekts verknüpft und tragen zur Steigerung seines Salienz- bzw. Aktivierungsgrads bei. Dazu gehören Einflussfaktoren wie Belebtheit, syntaktische Funktion und Position, die primär in Ariels Dimension 3 thematisiert werden.

Extrinsische Faktoren: Diese Faktoren fokussieren sich auf den Kontext, in dem das Referenzobjekt eingebettet ist. Diese Faktoren beleuchten, in welchem Maße externe Kontextbedingungen den Aktivierungsstatus des Referenzobjekts im Arbeitsgedächtnis beeinflussen. Dazu gehören Einflussfaktoren wie RefeDist, Mitbewerber, AnapTeilTyp, die vor allem in Ariels Dimension 1, 2 und 4 behandelt werden, sowie der durch die vorliegende Studie erneut ins Feld geführte Faktor der Textsorte.

4. Schlussbemerkungen und Ausblick

Dieser Beitrag geht davon aus, dass die Entscheidung von deutschen Muttersprachler/-innen für eine bestimmte anaphorische Form aus NA, PA und ZA im bestimmten Kontext nicht zufällig getroffen wird, sondern von vielen kontextuellen Faktoren abhängt. Das Ziel besteht darin, die Verteilungsunterschiede von NA, PA und ZA hinsichtlich der kontextuellen Faktoren mithilfe statistischer Methoden zu ermitteln. Basierend auf den vorangegangenen Untersuchungen wurden mehrere Einflussfaktoren erarbeitet und für die empirische Untersuchung bereitgestellt. Um der Komplexität der Zusammenwirkung einzelner Faktoren gerecht zu werden, wurde eine multinomiale logistische Regression durchgeführt.

Die Untersuchung hat ergeben, dass neun kontextuelle Faktoren eine Schlüsselrolle bei der Wahl anaphorischer Formen im Deutschen spielen: die Belebtheit vom Referenten (Belebtheit), die syntaktische Funktion des Antezedens (AnteSynF), die syntaktische Funktion der Anapher (AnapSynF), der syntaktische Parallelismus zwischen dem Antezedens und der Anapher (Parallelismus), die Position der Anapher (AnapPos), der Teilsatztyp der Anapher (AnapTeilTyp), die referenzielle Distanz zwischen dem Antezedens und der Anapher (RefeDist), potenzielle Mitbewerber des Antezedens (Mitbewerber) und die Textsorte (Textsorte). Mit ihnen lässt sich die Entscheidung für eine bestimmte anaphorische Form mit einer hohen Wahrscheinlichkeit (84,24%) vorhersagen.

Die erzielten Ergebnisse korrespondieren mit den vier Dimensionen von Ariels (1990) *Accessibility Theory* und können wiederum in zwei Kategorien (intrinsisch vs. extrinsisch) zusammengefasst werden. Die intrinsische Kategorie, repräsentiert durch die Faktoren Belebtheit, AnteSynF, AnapSynF, AnapPos und Parallelismus, betrifft direkt die ontologische Salienz des Referenzobjekts an sich. Ein belebtes Referenzobjekt, das syntaktisch als Subjekt agiert, wobei die Anapher ebenfalls in einer subjektähnlichen Position vor dem Verb steht, weist eher einen erhöhten Aktivierungsgrad auf, was eine gesteigerte Zugänglichkeit ermöglicht. Demgegenüber umfasst die extrinsische Kategorie die Faktoren AnapTeilTyp, RefeDist, Mitbewerber und Textsorte. Es kommt darauf an, wie diskursimmanente Rahmenbedingungen den Aktivierungsgrad des Referenzobjekts beeinflussen können. Wenn das Referenzobjekt in einem eingebetteten Nebensatz auftritt, die Distanz zwischen Anapher und Antezedens erheblich ist, potenzielle konkurrierende Referenten vorhanden sind und der Text fachlich ist, wird die Aktivierung des Referenzobjekts eher erschwert, was die Zugänglichkeit verringert. Vor diesem Hintergrund lässt sich die Auswahl einer anaphorischen Form als ein Abwägungsprozess zwischen diesen Faktoren verstehen. Bei einem erhöhten Aktivierungsgrad neigt man zur Wahl einer semantisch und morphologisch weniger komplexen anaphorischen Form wie ZA oder PA. Bei einem reduzierten Aktivierungsgrad tendiert die Präferenz hingegen zu morphologisch komplexeren und semantisch reicheren Formen wie NA.

Nicht zuletzt soll noch auf mögliche Einschränkungen der vorliegenden Untersuchung hingewiesen werden. Wie Ariel (ebd., S. 56) bereits formuliert hat, ist die Zugänglichkeit (d. h., *accessibility*) oft die einzige Bedingung für die angemessene Verwendung von Ausdrücken mit niedrigerer Zugänglichkeit (vor allem NA). Doch haben Sprachen normalerweise zusätzliche syntaktische Beschränkungen für Ausdrücke mit hoher Zugänglichkeit (vor allem ZA). Bei Ausdrücken mit hoher Zugänglichkeit wird zwar das Zugänglichkeitsprinzip ebenfalls beachtet, aber es stellt eher eine essenzielle und keine hinreichende Voraussetzung dar. Deutlich stärker wirken sich dabei die syntaktischen Regeln aus. Insofern will die vorliegende Studie keineswegs die traditionellen syntaktischen Theorien bezüglich der Anapher in Frage stellen, sondern sie versucht vielmehr, aus einer funktionalen Perspektive ein neues interpretatives Licht auf das Thema zu werfen. Zudem lässt sich methodisch anmerken, dass die in dieser Studie angewandte Modellierungstechnik darauf ausgerichtet ist, die Schlüsselfaktoren zu identifizieren. Tatsächlich handelt es sich jedoch um das einfachste Modell, das gemäß dem Prinzip des „Ockhamschen Rasiermessers“ (auch: Prinzip der Sparsamkeit), wonach Wesenheiten nicht über das Notwendige hinaus zu vermehren sind (lat. „*Entia non sunt multiplicanda praeter necessitatem*“), ausgewählt wurde und das in den Daten die größte Erklärungskraft aufweist.

Ungeachtet der erwähnten Einschränkungen ist dieser Beitrag sowohl für die linguistische Forschung als auch für die Fremdsprachendidaktik von Relevanz. Durch die Ermittlung von Schlüsselfaktoren sowie deren Wirkungsweise und durch die Beschreibung von typischen Kontexten lässt sich der Entscheidungsmechanismus, der bei Muttersprachler/-innen fast intuitiv abläuft, transparenter beschreiben. Ferner könnte diese Untersuchung methodische Anregungen für künftige kontrastive Sprachanalysen liefern, indem man Verteilungsunterschiede hinsichtlich der Einflussfaktoren (Prädiktorvariablen) ganzheitlich zu berechnen versucht.

Literatur

- Ariel, Mira (1990): *Accessing noun-phrase antecedents*. London: Routledge.
- Arppe, Antti (2013): *Polytomous: Polytomous logistic regression for fixed and mixed effects*. R package version 0.1.6. <https://CRAN.R-project.org/package=polytomous> (Stand: 16.4.2022).
- Bader, Markus/Portele, Yvonne (2019): The interpretation of German personal pronouns and d-pronouns. In: *Zeitschrift für Sprachwissenschaft* 38, 2, S. 155–190.
- Bittner, Dagmar/Dery, Jeruen E. (2015): The information structural effects of German P-and D-pro-nouns in discourse. In: *ZAS Papers in Linguistics* 58, S. 49–71.
- Bosch, Peter/Hinterwimmer, Stefan (2016): Anaphoric reference by demonstrative pronouns in German. In search of the relevant parameters. In: Holler/Suckow (Hg.), S. 193–212.
- Bosch, Peter/Katz, Graham/Umbach, Carla (2007): The non-subject bias of German demonstrative pronouns. In: Schwarz-Friesel, Monika/Consten, Manfred/Knees, Mareile (Hg.): *Anaphors in Text: Cognitive, formal and applied approaches to anaphoric reference*. (= *Studies in Language Companion Series* 86). Amsterdam/Philadelphia: Benjamins, S. 145–164.
- Bosch, Peter/Rozario, Tom/Zhao, Yufan (2003): Demonstrative pronouns and personal pronouns: German *der* vs *er*. In: *Proceedings of the EACL 2003 Workshop on the Computational Treatment of Anaphora*, S. 61–68.
- Bosch, Peter/Umbach, Carla (2007): Reference determination for demonstrative pronouns. In: *ZAS Papers in Linguistics* 48, S. 39–51.
- Bußmann, Hadumod (Hg.) (2002): *Lexikon der Sprachwissenschaft*. Stuttgart: Kröner.
- Chomsky, Noam (1981): *Lectures on government and binding*. Dordrecht: Foris.
- Chomsky, Noam (1982): *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT Press.
- Colonna, Saveria/Schimke, Sarah/Hemforth, Barbara (2012): Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. In: *Linguistics* 50, 5, S. 901–1073.
- Colonna, Saveria/Schimke, Sarah/Hemforth, Barbara (2014): Information structure and pronoun resolution in German and French: Evidence from the visual-world paradigm. In: Hemforth, Barbara/Mertins, Barbara/Fabricsius-Hansen, Cathrine (Hg.): *Psycholinguistic approaches to meaning and understanding across languages*. (= *Studies in Theoretical Psycholinguistics* 44). Cham: Springer International Publishing, S. 175–195.
- Duden (2009): *Der Duden in zwölf Bänden. Bd. 4: Die Grammatik. Unentbehrlich für richtiges Deutsch. 8., überarb. Aufl.* Mannheim/Wien/Zürich: Dudenverlag.
- Ellert, Miriam (2010): *Ambiguous pronoun resolution in L1 and L2 German and Dutch*. PhD Thesis. Nijmegen: Radboud University. Wageningen: Ponsen & Looijen.
- Ellert, Miriam (2013): Information structure affects the resolution of the subject pronouns *er* and *der* in spoken German discourse. In: *Discours* 12, S. 3–24.
- Ellert, Miriam/Holler, Anke (2011): Semantic and structural constraints on the resolution of ambiguous personal pronouns – a psycholinguistic study. In: Hendrickx, Iris/Lalitha Devi, Sobha/Branco, António/Mitkov, Ruslan (Hg.): *Anaphora Processing and Applications. DAARC 2011*. (= *Lecture Notes in Computer Science* 7099). Berlin/Heidelberg: Springer, S. 157–170.
- Frank, Markus (2019): *Phorische Verkettung im Deutschen – Eine exemplarische Untersuchung anhand von Diskursrelationen der kausalen Gruppe*. Berlin: De Gruyter.
- Givón, Talmy (2017): *The story of zero*. Amsterdam/Philadelphia: Benjamins.

- Goldberg, Adele E. (2006): *Constructions at Work. The nature of generalization in language*. Oxford: Oxford University Press.
- Graefen, Gabriele (2009): Pronomen. In: Hoffmann, Ludger (Hg.): *Handbuch der deutschen Wortarten*. Berlin/New York: De Gruyter, S. 657–705.
- Gries, Stefan Th. (2013): *Statistics for linguistics with R*. 2. Aufl. Berlin/Boston: De Gruyter.
- Gundel, Jeanette K./Hedberg, Nancy/Zacharski, Ron (1993): Cognitive status and the form of referring expressions in discourse. In: *Language* 69, 2, S. 274–307.
- Hoffmann, Ludger (2000): Anapher im Text. In: Brinker, Klaus/Antos, Gerd/Heinemann, Wolfgang/Sager, Sven F. (Hg.): *Text- und Gesprächslinguistik. Ein internationales Handbuch zeitgenössischer Forschung*. (= Handbücher für Sprach- und Kommunikationswissenschaft (HSK) 16.1). Berlin/New York: De Gruyter, S. 295–304.
- Holler, Anke/Suckow, Katja (Hg.) (2016): *Empirical perspectives on anaphora resolution*. (= *Linguistische Arbeiten* 563). Berlin/Boston: De Gruyter.
- Holler, Anke/Suckow, Katja (2016): How clausal linking affects noun phrase salience in pronoun resolution. In: Holler/Suckow(Hg.), S. 61–85.
- Huang, Yan (1991): A neo-Gricean pragmatic theory of anaphora. In: *Journal of Linguistics* 27, 2, S. 301–335.
- Huang, Yan (2000): *Anaphora: a cross-linguistic approach*. Oxford: Oxford University Press.
- Jiang, Wangqi (2006): 篇章与回指 (Diskurs und Anapher). In: *外语学刊* (Foreign Language Research) 131, 4, S. 33–40.
- Kaiser, Elsi (2011): On the relation between coherence relations and anaphoric demonstratives in German. In: Reich, Ingo/Horch, Eva/Pauly, Dennis (Hg.): *Sinn und Bedeutung 15. Proceedings of the 2010 Annual Conference of the Gesellschaft für Semantik*. Saarbrücken: universaar, S. 337–352.
- Kehler, Andrew/Rohde, Hannah (2013): A probabilistic reconciliation of coherence-driven and centering-driven theories of pronoun interpretation. In: *Theoretical Linguistics* 39, 1–2, S. 1–37.
- Kibrik, Andrej A. (2011): *Reference in discourse*. Oxford: Oxford University Press.
- Levinson, Stephen C. (1987): Pragmatics and the grammar of anaphora: a partial pragmatic reduction of binding and control phenomena. In: *Journal of Linguistics* 23, 2, S. 379–434.
- Levinson, Stephen C. (1991): Pragmatic reduction of the binding conditions revisited. In: *Journal of Linguistics* 27, 1, S. 107–161.
- Mitkov, Ruslan (2002): *Anaphora resolution*. London/New York: Routledge.
- Patil, Umesh/Bosch, Peter/Hinterwimmer, Stefan (2020): Constraints on German demonstratives: language formality and subject-avoidance. In: *Glossa: a Journal of General Linguistics* 5, 1, S. 1–14.
- Perkuhn, Reihner/Keibel, Holger/Kupietz, Marc (2012): *Korpuslinguistik*. Paderborn: UTB.
- Portele, Yvonne/Bader, Markus (2016): Accessibility and referential choice: personal pronouns and d-pronouns in written German. In: *Discours* 18, S. 3–38.
- Portele, Yvonne/Bader, Markus (2017): Personal pronouns and d-pronouns in German: Connecting comprehension to production. In: Featherston, Sam/Hörnig, Robin/Steinberg, Reinhild/Umbreit, Birgit/Wallis, Jennifer (Hg.): *Proceedings of Linguistic Evidence 2016 – Empirical, Theoretical, and Computational Perspectives*. <https://publikationen.uni-tuebingen.de/xmlui/handle/10900/77066> (Stand: 9.6.2022).
- Prince, Ellen F. (1981): Toward a taxonomy of given – new information. In: Cole, Peter (Hg.): *Radical pragmatics*. New York: Academic Press, S. 223–256.
- Pu, Ming M. (2011): *Discourse anaphora: a cognitive-functional approach*. München: Lincom.

- Qian, Wencai (2006): 汉德语言实用对比研究 (Praktische vergleichende Studie der chinesischen und deutschen Sprache). Beijing: 外语教学与研究出版社 (Foreign Language Teaching and Research Press).
- R Core Team (2021): R: A language and environment for statistical computing. R foundation for statistical computing. www.R-project.org/ (Stand: 3.9.2022).
- Sauermann, Antje/Gagarina, Natalia (2017): Grammatical role parallelism influences ambiguous pronoun resolution in German. In: *Frontiers in Psychology* 8, Art. 1205, S. 1–8.
- Schlink, Bernhard (1995): *Der Vorleser*. Zürich: Diogenes.
- Schumacher, Petra B./Backhaus, Jana/Dangl, Manuel (2015): Backward-and forward-looking potential of anaphors. In: *Frontiers in Psychology* 6, Art. 1746, S. 1–14.
- Schumacher, Petra B./Dangl, Manuel/Uzun, Elyesa (2016): Thematic role as prominence cue during pronoun resolution in German. In: Holler/Suckow (Hg.), S. 213–240.
- Schumacher, Petra B./Roberts, Leah/Järvikivi, Juhani (2017): Agentivity drives real-time pronoun resolution: evidence from German *er* and *der*. In: *Lingua* 185, S. 25–41.
- Schwarz, Monika (1997): Anaphern und ihre diversen Antezedenten – das Koreferenz und Konsorten. In: Dürscheid, Christa/Ramers, Karl Heinz/Schwarz, Monika (Hg.): *Sprache im Fokus: Festschrift für Heinz Vater zum 65. Geburtstag*. Tübingen: Niemeyer, S. 445–455.
- Schwarz, Monika (2000): *Indirekte Anaphern in Texten: Studien zur domänen gebundenen Referenz und Kohärenz im Deutschen*. Tübingen: Niemeyer.
- Speyer, Augustin (2016): Factors determining the choice of anaphora in Old High German: a survey of zero and personal pronoun usage in Otfrid. In: Holler/Suckow (Hg.), S. 113–142.
- Streb, Judith/Rösler, Frank/Hennighausen, Erwin (1999): Event-related responses to pronoun and proper name anaphors in parallel and nonparallel discourse structures. In: *Brain and Language* 70, 2, S. 273–286.
- Wilson, Frances (2009): *Processing at the syntax-discourse interface in second language acquisition*. PhD Thesis. Edinburgh: University of Edinburgh.
- Xu, Jiujiu (2003): 现代汉语篇章回指研究 (Eine Studie über Anaphora in modernen chinesischen Texten). Beijing: 中国社会科学出版社 (China Social Sciences Press).

Guying Zhou
Beijing Foreign Studies University
National Research Centre for Foreign Language Education
Xisanhuan North Road 19
100089 Beijing
VR China
E-Mail: guying.zhou@bfsu.edu.cn

Asso.-Prof. Dr. Jing Li (Korrespondenzautorin)
Beijing Foreign Studies University
School of German Studies
Xisanhuan North Road 19
100089 Beijing
VR China
E-Mail: lijing@bfsu.edu.cn