

**Referenzbezogene Registervariation zwischen konzeptuell öffentlichem
und nicht-öffentlichem Gebrauch gesprochener Sprache**

**Bachelorarbeit
zur Erlangung des akademischen Grades
Bachelor of Arts (B.A.)
im Fach Germanistische Linguistik**

**Humboldt-Universität zu Berlin
Philosophische Fakultät II
Institut für deutsche Sprache und Linguistik**

eingereicht von Nico Lehmann

**1. Gutachterin: Prof. Dr. Anke Lüdeling
2. Gutachterin Prof. Dr. Elisabeth Verhoeven**

Berlin, den 03. 09. 2015

Inhaltsverzeichnis

1. Einleitung	3
2. Diskurs, Referenz und Registervariation	7
2.1 Diskurs.....	7
2.1.1 Mentale Repräsentation	8
2.1.2 Referenz im Diskurs	10
2.2 Referenz und Referenzstrategien.....	12
2.2.1 Referenz.....	13
2.2.2 Koreferenz im Vergleich zu anderen Referenzrelationen	16
2.3 Register.....	20
2.3.1 Registervariation.....	21
2.3.2 Kriterien der Registerdifferenzierung.....	22
2.4 Ableitung spezifischer Hypothesen.....	24
3. Korpusbasierte Studie	28
3.1 Daten	28
3.2 Annotationsrichtlinien	30
3.3 Annotation	32
3.4 Ergebnisse und Auswertung	36
4. Zusammenfassung	41
Literaturverzeichnis	44
Anhang	48

1. Einleitung

Sprechakte beurteilt Hymes (1972:56) als „activities [...] that are directly governed by rules or norms for the use of speech“. Diese Normen und Regeln werden schon in den ersten Jahren gelernt. Eine Studie von Wagner u. a. hat gezeigt, dass Kinder ihre Sprache bereits nach einem Jahr an ihr soziales Umfeld anpassen können (2010:1679), im Alter von vier Jahren entwickeln sie schließlich ein Gespür für Register (2010:1985), der differenzierten Verwendung von Sprache bedingt durch situative Faktoren wie interpersonale Beziehungen, situative Umstände, Ziele des Sprechers etc. (Biber 2008:823) – siehe Abschnitt 2.3. Vierjährige verstehen, dass Sprecher sich abhängig vom Konversationspartner anderer sprachlicher Mittel bedienen und damit sogar soziale Beziehungen anzeigen (Wagner u. a. 2010:1685).

Da die Verwendung von Registern also eine fundamentale Fähigkeit der Kommunikation ist, die sich schon früh herausbildet, werden in zahlreichen Studien diese situativen Faktoren untersucht, um Rückschlüsse auf die Gründe und kulturellen Konventionen ziehen zu können (siehe Besnier 1988; Schikorsky 1990; Biber 2008; Kunz 2010; Wagner u. a. 2010; Neumann 2014; Kunz & Lapshinova-Koltunski 2015).

Aufgrund der Unmengen an verschiedenen Registern, deren wir uns tagtäglich bedienen, und der unscharfen Abstufungen zwischen diesen ist die Zuordnung von Texten¹ zu spezifischen Registern keineswegs eindeutig. Besnier (1988) und Biber (1988; 2008) haben daher einen multidimensionalen Ansatz gewählt und ausgewählte Textsorten für verschiedene Dimensionen entlang eines Kontinuums eingeteilt. Biber (2008:826) wählte für seine Studie von 1988 zum Unterschied zwischen gesprochener und geschriebener Sprache des Englischen 67 linguistische Features, nach denen er die Textsorten² charakterisierte. Die erste Dimension beschrieb ein Kontinuum mit dem Status ‚involved‘ am einen Ende, der interaktive Charakteristika hervorhebt, und ‚information production‘ mit Fokus auf den informativen, weniger interaktiven Charakter am anderen. Die Ergebnisse zu dieser Dimension zeigten, dass informelle Konversationen sich stark dem ‚involved‘ Ende annähern und hochgradig interaktiv sind, wobei öffentliche Konversationen eher eine interme-

¹ Der Begriff ‚Text‘ wird hier im Sinne von Halliday und Matthiessen (2013:3) verwendet, die sowohl mündliche als auch geschriebene Sprache einschließen: „When people speak or write, they produce text; and text is what listeners and readers engage with and interpret.“

² Textsorten (‚text types‘) unterscheiden sich laut Biber (2008:846) von Registern insofern, dass Textsorten primär linguistisch definiert werden und Register vorrangig anhand ihrer situativen Charakteristika.

diäre Position bestreiten, mit höherem informativen Charakter und mittlerer Interaktivität; am anderen Ende stehen akademische, geschriebene Texte mit sehr hoher Informativität und minimalen interaktiven Mitteln (Biber 2008:834). Daraus geht hervor, dass sich bestimmte mündliche Texte stark von schriftlichen unterscheiden, aber auch unter den mündlichen Texten unterschiedliche Ausprägungen erkennbar sind.

Die Studie von Biber deutet an, dass es Faktoren gibt, nach denen Textproduzenten³ informelle Sprache und formelle Sprache voneinander abgrenzen. Neben der kulturellen Abhängigkeit könnte ein Grund in der zuvor beobachteten differenzierten Verwendung von Registern abhängig von den Kommunikationspartnern liegen. Nach Ventola (1979:270) gehört die Kennzeichnung von sozialen und persönlichen Beziehungen zu den Makrofunktionen der Sprache. Dies geht einher mit der Beobachtung von Wagner u. a. (2010:1678), dass mit Sprache die Kollektivzugehörigkeit markiert wird. Maas (2010:38) nimmt eine Einteilung in drei Grundregister nach der sozialen Dimension vor und unterscheidet intime, informell-öffentliche und formell-öffentliche Register, wobei in letzterem die Personenbindung verschwinden soll. All das lässt vermuten, dass die Wahl des Registers von der Beziehung der Kommunikationspartner abhängt. Das soll anhand einer korpusbasierten Studie überprüft werden.

Für eine solche Untersuchung bedarf es eines kontrastiven Vergleichs von Texten, bei denen sich das Verhältnis der Gesprächspartner zueinander auf einer Skala der Vertrautheit bewegt. Für die Untersuchung werden nicht-öffentliche, private Gespräche mit öffentlichen Gesprächen verglichen, da sich diese der Annahme nach in der Vertrautheit der Gesprächspartner größtmöglich unterscheiden und sie der medial mündlichen, oraten⁴ Textsorte angehören, die eine Basis im Sinne der Interaktivität sicherstellt. Da die Wahl des Registers, wie zuvor festgestellt, von vielen Faktoren abhängt und sehr graduelle Unterschiede bestehen, müssen die Texte in der Dimension der Vertrautheit einen möglichst großen Abstand aufweisen – zur Auswahl der Texte siehe Abschnitt 3.1. Um kulturelle Faktoren auszuschließen, beschränkt sich diese Untersuchung auf Texte der deutschen Sprache.

³ Um keine Unterscheidung zwischen Sprecher, Schreiber etc. bzw. Hörer, Leser etc. zu machen, werden hier die Termini ‚Textproduzent‘ und ‚Textrezipient‘ verwendet.

⁴ Orat sind nach Maas (2010:29) interaktive Texte unabhängig vom Medium, bei denen es ein Gegenüber gibt – analog zu konzeptueller Mündlichkeit ohne intentionaler Konnotation.

Texte können anhand zahlreicher Variablen analysiert werden, darunter syntaktische Komplexität, Register und rhetorische Mittel (Prince 1992:296). Steiner (2013:15) untersuchte bereits die Hypothese, dass Referenz ein „distinguishing feature for register variation“ sei. Auch hier wird die Annahme gemacht, dass Referenzstrategien dazu verwendet werden, den Diskurs zu leiten und Beziehungen zwischen Kommunikationspartnern anzuzeigen. Nach Pu (2011:1) sind Untersuchungen zu Referenzstrategien sogar „fundamental to our understanding of the relationship among cognitive processes, discourse coherence, and information distribution.“ Referenzen sind Relationen zwischen zwei sprachlichen Ausdrücken, allerdings ‚referieren‘ nicht die Wörter selbst, sondern Textproduzenten benutzen sie, um Referenzen herzustellen (Schwarz-Friesel & Chur 2014:89). Mittels referierenden Ausdrücken werden Diskursreferenten (siehe Abschnitt 2.1.2) eingeführt, aufgegriffen und verändert; außerdem tragen sie erheblich zur Kohärenz von Diskursen bei (Kunz 2010:23). In dem Beispiel in (1) bezeichnen die unterstrichenen Ausdrücke mit dem Index 1 dieselbe Person und mit dem Index 2 denselben Gegenstand. Nach der ausführlichen Beschreibung des Gegenstands kann mit einem reduzierten Ausdruck erneut auf diesen Gegenstand verwiesen werden wie bei *diesen Bericht*, wodurch referierende Ausdrücke den mentalen Aufwand erleichtern. Jedoch unterliegen Referenzstrategien auch kommunikativen Bedürfnissen. Der Hörer muss verstehen können, wer was (mit wem, für wen ...) tut (Dooley & Levinsohn 2001:56).

- (1) S5 wir₁ erwarten als sachliche Grundlage unserer₁ Diskussion den jetzt vorhin schon angeführten Umweltschutzbericht der Landesregierung den₂ wir₁ zum ersten April erwarten und werden sicher diesen Bericht₂ zum Anlaß nehmen hier einiges ganz erhebliches weiter zu bearbeiten . (IDS 2015c)

Dooley & Levinsohn (2001:63) folgend werden referenzstiftende Mittel in Sprachen unterschiedlich verwendet. Steiner (2013:25) und Kunz & Lapshinova-Koltunski (2015:283) zeigen weitergehend, dass Unterschiede bei den Referenzstrategien in Registern zu beobachten sind. Steiner (2013:25) findet zum Beispiel heraus, dass gesprochene Texte mehr Pronomen zu enthalten scheinen als geschriebene Texte. Ein großer Mangel ist jedoch, dass bisherige Studien vor allem literate Register in den Vordergrund stellen. Eggins & Slade (2004:23) stellen fest, dass „casual conversation[s]“ (nicht-öffentlichen Gesprächen) generell bisher nur wenig Aufmerksamkeit in der linguistischen Forschung geschenkt wurde. Auch bei Kunz

& Lapshinova-Koltunski werden orate Register nur auf der medialen Ebene mit literaten Texten verglichen, nicht aber innerhalb der oraten Ebene.

Daraus ergibt sich folgende These, die hier mit einer korpusbasierten Studie untersucht wird (siehe Abschnitt 3): Das Verhalten deutscher Sprecher unterscheidet sich in der öffentlichen und nicht-öffentlichen (konzeptuell privaten) Verwendung gesprochener Sprache bezüglich der Nutzung von Referenzstrategien. Dazu wird in Abschnitt 2 ein Überblick über aktuelle Diskurstheorien, Referenzstrategien sowie Registervariation dargestellt. Aus dem theoretischen Hintergrund ergeben sich Kriterien für die Textauswahl (Abschnitt 2.3.2). Auf dieser allgemeinen Grundlage werden in Abschnitt 2.4 konkrete Hypothesen für die Korpusstudie abgeleitet. Die Durchführung der Korpusstudie wird in Abschnitt 3 beschrieben. Es werden Texte nach den Kriterien ausgewählt, Annotationsrichtlinien erläutert und die Schritte der Annotation dargelegt sowie die Ergebnisse präsentiert. In Abschnitt 4 folgt eine Zusammenfassung mit Rückblick auf die theoretischen Ansätze.

2. Diskurs, Referenz und Registervariation

Die Verwendung verschiedener Referenzstrategien wird durch die vom Textproduzenten gewählte Diskursstruktur bedingt (Grosz u. a. 1995:204) und diese unterliegt unter anderem der Sprechakttheorie von Grice:

We might then formulate a rough general principle which participants will be expected (*ceteris paribus*) to observe, namely: Make your conversational contributions such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. (1975:45)

Diskurse werden nach diesem Prinzip von Intentionen und Zielen beeinflusst und sind daher unerlässlich für die Analyse von Referenzen und Registereigenschaften. Daher müssen zunächst die aktuellen Annahmen über die Diskursstruktur erläutert werden, bevor die Formen der Registervariation und Zusammenhänge mit einzelnen Referenzstrategien betrachtet werden können.

2.1 Diskurs

Für Grosz u. a. (1995:206) besteht ein Diskurs⁵ dann, wenn in einem Text Anzeichen von Kohärenz zu verzeichnen sind, also ein „semantisch-kognitiver Sinnzusammenhang eines Textes“ (Vater 2005:157). Dieser inhaltliche Zusammenhang entsteht durch kohärenzstiftende sprachliche Mittel, die eine Verknüpfung zwischen den einzelnen Komponenten auf der Oberfläche des Textes herstellen (Vater 2005:156; Averintseva-Klisch 2013:2). Die Art und Weise, wie die semantischen Links innerhalb und über Satzgrenzen hinweg benutzt werden, um eine ‚Sinnkontinuität‘ aufzubauen, wird auch Kohäsion bezeichnet (Vater 2005:156f.; Halliday & Matthiessen 2013:114). Zu diesen sprachlichen Signalen gehören phonologische und morphologische Mittel, Referenzen, Rekurrenz, Substitutionen, Ellipsen, Konnektoren und lexikalische Kohäsion (Averintseva-Klisch 2013:11; Sanders & Pander Maat 2006:591; Halliday & Matthiessen 2013:603).

Zu unterscheiden ist hier zwischen lokaler und globaler Kohärenz, wobei erstere sich auf einen Diskursabschnitt beschränkt und letztere den Zusammenhang des ganzen Diskurses anzeigt (Averintseva-Klisch 2013:6; Grosz u. a. 1995:204). Dabei wird in neueren Ansätzen davon ausgegangen, dass Kohärenz weniger durch Verbindungen im Text, sondern vielmehr durch Zusammenhänge in der mentalen

⁵ In einigen Ausführungen werden ‚Text‘ und ‚Diskurs‘ synonym verwendet (Averintseva-Klisch 2013:3); für die Betrachtung von Diskursmodellen ist es jedoch sinnvoll, die sinnzusammenhängenden Eigenschaften dem ‚Diskurs‘ zugrunde zu legen.

Repräsentation des Diskurses hergestellt wird (Sanders & Pander Maat 2006:592). Textproduzenten benutzen sprachliche Signale nämlich als „clues to assist the hearers in coming up with an adequate mental representation“ (Dooley & Levinsohn 2001:13).

Die gelisteten kohärenzstiftenden Mittel tragen zur Kohärenz auf zwei Ebenen bei: referenzielle Kohärenz und relationale Kohärenz (Sanders & Pander Maat 2006:592), wobei vor allem referenzielle Kohärenz durch den engen Bezug zur mentalen Diskursrepräsentation sowohl auf lokaler als auch globaler Ebene Einfluss nimmt. Referenzstrategien bestimmen, welche Konzepte für den Diskurs relevant sind und damit im Fokus der mentalen Repräsentation stehen; mit anderen Worten, sie geben Rückschluss darauf, wovon der Diskurs handelt (Grosz u. a. 1995:205f.). Durch Einheitlichkeit der verwendeten Konzepte wird eine thematische Kontinuität erstellt (Grosz u. a. 1995:206), welche die Kohärenz des Diskurses fördert und kognitive Ressourcen sparsam verbraucht, wohingegen ein disruptiver Informationsfluss dazu führt, dass zwei kognitive Prozesse benötigt werden: einmal für das Hinzufügen neuer Referenten zur mentalen Repräsentation sowie für das Schließen von Inferenzen zur Überbrückung von Informationslücken (Pu 2011:2). Durch den einheitlichen Kontext der Medizin in Beispiel (2) können Textproduzenten direkt mit den unterstrichenen definiten Phrasen neue Konzepte im Diskurs aktivieren, ohne sie durch spezielle Mittel explizit einführen zu müssen.

(2) S3 ich mein die an die Privatpatienten wird er nich rankommen . S3 das macht meistens der Chefarzt . S3 aber er könnte gut und gerne so für die die Kassenpatienten . (IDS 2015b)

Referenzen sind also grundlegend für die Textproduktion und –rezeption und lenken den Diskurs auf verschiedenen Ebenen. Was sprachlichen Referenzen zugrunde liegt und wie sie kognitive Prozesse beeinflussen, wird im Folgenden dargestellt.

2.1.1 Mentale Repräsentation

Wenn Kommunikation zur Folge hat, dass eine mentale Repräsentation des Diskurses (auch ‚Diskurswelt‘ oder ‚Textwelt‘) erstellt wird (Kunz 2010:31; siehe auch Vater 2005; Gavins 2007), dann wird die Textrezeption vom ständigen Aufbauen, Anpassen, Aktualisieren und Verändern der Diskurswelt begleitet (Kunz 2010:32). Textproduktion und –rezeption bedingen also sprachunabhängige mentale Prozesse

(2010:23), die es uns erst ermöglichen, sprachliche Einheiten zu konzeptualisieren, i. e. mit Bedeutungen zu verbinden (Gavins 2007:2).

Gavins (2007:3ff.) stellt drei aktuelle Ansätze vor, die das Konstrukt der mentalen Repräsentation zu erklären versuchen. Im Sinne der Schema-Theorie werden sogenannte ‚Skripte‘ erstellt, die die wichtigsten Informationen über bekannte Ereignisse und Situationen wie in Handbüchern gespeichert haben – ein bekanntes Beispiel ist das ‚restaurant script‘⁶. Unbekannte Ereignisse und Situationen könnten so durch Rückgriff auf bereits vorhandene Skripte verstanden und eingeordnet werden (2007:3). Die kognitive Psychologie legt dagegen bei der mentalen Repräsentation ein mentales Modell der realen Welt zugrunde; statt eines separat gespeicherten Skripts, auf das zurückgegriffen werden kann, sind Abbilder der Erlebnisse vorhanden, die sowohl abstrakt als auch höchst detailliert sein können und übergreifende Rückschlüsse erlauben (2007:4). Ein ähnlicher, dritter Ansatz wird von Gavins im Rahmen der Text-World-Theorie postuliert, bei der Texte in ‚conceptual levels‘ unterteilt werden, wobei das erste Level situatives, persönliches und kulturelles Wissen einschließt (hier wird dieses Level als ‚kontextuelle Diskurswelt‘ bezeichnet). Das zweite Level betrifft die zuvor erwähnte Konzeptualisierung der sprachlichen Einheiten in mentale Repräsentationen (hier ‚textuelle Diskurswelt‘), die wie bei den mentalen Modellen der kognitiven Psychologie höchst detailliert sein können (2007:8ff.).

Gemein haben diese Theorien, dass sprachliche Einheiten mental verarbeitet werden müssen und in einer Weise für den Diskurs ‚gespeichert‘ werden. Diskurse weisen dabei ein interessantes Phänomen auf: sie sind meist hochgradig unterspezifiziert, das heißt, vieles, das verstanden wird, wird nicht von expliziten sprachlichen Einheiten ausgedrückt (Kunz 2010:34).

Der Sinn eines Textes ergibt sich [...] oft nicht nur aus der Summe der Satzbedeutungen, sondern in vielen Fällen erst aus der Aktivierung zusätzlichen Weltwissens sowie der Anwendung kognitiver Schlussfolgerungen. (Schwarz 2000:20)

Zahlreiche Linguisten gehen davon aus, dass zu viele explizite sprachliche Informationen das Arbeitsgedächtnis des Textrezipienten überfordern würden (siehe dazu Vater 2005; Schwarz 2000; Sperber & Wilson 1995), was sich auch in den Prinzipien der Sprachökonomie und Grices Kooperationsprinzipien wiederfindet

⁶ In dem Restaurantskript sollen spezifische Situationsmerkmale eines Restaurantbesuchs gespeichert sein, so dass Elemente wie Kellner, Bestellvorgang, Rechnung etc. nicht expliziert werden müssen.

(Kunz 2010:34). Chafe (1987:22) geht davon aus, dass nur geringe Informationsmengen im Fokus bzw. „active“ sein können (siehe Abschnitt 2.1.2).

Kunz (2010:36) deutet an, dass mithilfe von Referenzstrategien untersucht werden kann, wie Textproduzenten einschätzen, welche Dinge explizit ausgedrückt werden und welche implizit gelassen werden können. Referenzen werden als direkte Verbindung zur mentalen Diskurswelt angesehen (2010:35), weshalb mit einigen Referenzstrategien angezeigt werden kann, dass bestimmte Informationen bereits in ihr vorhanden sind. Welche Verbindung zwischen Diskurswelt und der Verwendung von Referenzen besteht, soll durch den nächsten Abschnitt deutlich werden.

2.1.2 Referenz im Diskurs

Um auf die in der Diskurswelt verfügbaren Informationen zurückzugreifen, werden mit referenziellen Ausdrücken (‘referring expressions’) Referenzen vollzogen (Schwarz-Friesel & Chur 2014:89). Das bedeutet, durch die Äußerung eines referenziellen Ausdrucks wird dem Textrezipienten angezeigt, wo und wie er eine bestimmte Information in der Diskursrepräsentation finden kann. Daneben gibt es Ausdrücke, die eine Information in der Diskurswelt überhaupt erst etablieren. Die Information, auf die mit solchen Ausdrücken referiert wird, ist bekannt als ‚Referent‘, auch ‚Diskursreferent‘ in Bezug auf den aktuellen Diskurs (2014:90) – siehe Abschnitt 2.2.1 zu Referenten und referenziellen Ausdrücken.

Im Laufe eines Diskurses wird die Diskurswelt angereichert mit Diskursreferenten, denn wie in Abschnitt 2.1.1 erwähnt werden Diskurswelten kontinuierlich erweitert und verändert. Je nach Status eines Referenten werden andere Referenzstrategien benutzt, um im Diskurs darauf zu verweisen. Für die zugrundeliegenden Prozesse und Motivationen gibt es verschiedene Theorien, die zu erklären versuchen, mit welchen referenziellen Mitteln welcher Referentenstatus angezeigt wird; hier soll auf vier hervorstechende eingegangen werden: die Accessibility-Theorie (Ariel 2001), die Givenness-Theorie (Gundel u. a. 1993), die Centering-Theorie (Grosz u. a. 1995) und die Theorie der ‚assumed familiarity‘ (Prince 1981).

Ariel (2001:29) schlägt vor, dass sich referenzielle Ausdrücke auf einer Skala positionieren, die den Grad der Zugänglichkeit für den Textrezipienten anzeigt. Die Grade sind dabei als allmählicher Übergang zu verstehen. Gundel u. a. (1993:275)

schlagen ganz ähnlich eine hierarchische Einordnung vor, bei der sprachliche Ausdrücke nach ihrem ‚Givenness-Status‘ bemessen werden. Die sprachliche Form soll demnach einen inhärenten kognitiven Status ausdrücken, der in der Hierarchie einer von sechs Kategorien zugeordnet wird. Bei dem schwächsten Grad lässt der Ausdruck nur die Identifikation des Referententyps zu und bei dem höchsten Grad ist der Referent im Fokus und damit in der Diskurswelt aktiv und eindeutig identifizierbar (Kunz 2010:42; Averintseva-Klisch 2013:73f.). Beide Annahmen sind insofern problematisch, als dass es keine eindeutige Beziehung zwischen der Form eines referenziellen Ausdrucks und seiner pragmatischen Funktion zu geben scheint (Kunz 2010:39) – siehe Abschnitt 2.2.1. Es spielen außerdem viele weitere Aspekte des Diskursmodells eine wesentliche Rolle bei der Verwendung von Referenzen.

Die Centering-Theorie beschränkt sich dagegen auf die lokale Kohärenz (Averintseva-Klisch 2013:76). Grosz u. a. (1995:208) entwickelten ein Modell, das sowohl sprachliche Form, syntaktische Funktion als auch syntaktische Position betrachtet, um eine Einordnung nach der Erschließbarkeit der Referenz vorzunehmen. Die Grundannahme ist, dass je mehr ein Referent im Zentrum der Aufmerksamkeit bleibt, desto weniger kognitiver Aufwand muss bei der Erschließung aufgewendet werden und desto kohärenter ist der Diskurs. Referenzielle Ausdrücke werden dafür in vorwärtsbezogene Zentren (CF) und rückwärtsbezogene Zentren (CB) eingeteilt und deren Vorkommen und Beziehung in benachbarten Äußerungen betrachtet (Grosz u. a. 1995:210f.; Averintseva-Klisch 2013:77f.). Kann das Zentrum der Aufmerksamkeit von Äußerung zu Äußerung fortgesetzt werden, so bleibt das CB gleich. Es ändert sich, indem es durch ein vorerwähntes CF (weicher Übergang) oder durch ein ganz neues (harter Übergang) ersetzt wird (Averintseva-Klisch 2013:78). Beispiel (3) zeigt einen weichen Übergang, wo das eingeführte CF *Schwefelabgase* im zweiten Satz, welcher durch die Sprecher-ID *S6* angezeigt wird, zum CB wird. Durch den Fokus auf der lokalen Kohärenz werden zahlreiche Phänomene mit diesem Modell allerdings außen vor gelassen.

- (3) *S6* der Stein₁ wird zerstört durch die Luftverunreinigung zu einem ganz erheblichen Teil durch Schwefelabgase₂. *S6* Schwefelabgase₂ entstehen nicht in der Innenstadt von Kraftfahrzeugen. (IDS 2015c)

Ebenfalls in Hinblick auf ‚Givenness‘ betrachtet Prince (1981:226ff.) referenzielle Ausdrücke, jedoch schlägt sie den Term ‚assumed familiarity‘ vor, um das Konzept von der Konnotation der Vorerwähnung abzugrenzen. Außerdem lässt der

Begriff auf einen wichtigen Punkt schließen, den Prince hervorhebt: Der Textproduzent macht durch die Verwendung von referenziellen Ausdrücken eine Annahme über die mentale Repräsentation des Textrezipienten (1981:228) sowie über dessen Fähigkeit, einen Referenten zu identifizieren (1981:230). Für referenzielle Ausdrücke schlägt sie drei Parameter vor, die den Bezug zur Diskurswelt wiedergeben. Als ‚new‘ (auch ‚hearer-new‘) werden referenzielle Ausdrücke klassifiziert, deren zugehörige Referenten nach der Annahme des Textproduzenten in der Diskurswelt des Textrezipienten kreiert werden müssen (‚brand-new‘) oder zuvor nicht benutzt wurden (‚unused‘) und auch nicht aus dem Kontext geschlossen werden können (1981:235f.). Als ‚evoked‘ gelten Ausdrücke mit Referenten, die im Diskurs bereits etabliert sind, entweder aufgrund sprachlicher Vorerwähnung (‚textually evoked‘) oder situativer Gegebenheit (‚situationally evoked‘), z. B. anwesende Personen. Zur Kategorie ‚inferrables‘ gehören Ausdrücke, deren Referenten durch Rückschlüsse vom Textrezipienten identifiziert werden sollten; das heißt, sie werden zwar zum ersten Mal verwendet und sind demnach quasi ‚new‘, allerdings gibt es kontextuelle oder textuelle Hinweise, mit denen der Referent erschlossen werden kann (1981:236). In Beispiel (3) ermöglicht so das Thema der Umweltschutzdebatte, den Referenten von *die Luftverunreinigung* trotz der erstmaligen Erwähnung kontextuell zu inferieren.

Diese Einteilung beschränkt sich zwar auf die interne Form referenzieller Ausdrücke, schließt dafür allerdings das Wissen der kontextuellen Diskurswelt ein. Außerdem wird das Konzept von ‚shared knowledge‘ oder ‚common ground‘ abgelöst und durch Annahmen des Textproduzenten über das Wissen und die mentale Repräsentation des Textrezipienten ersetzt, was für das Diskursmodell eine wichtige Anpassung ist (Kunz 2010:36). Wenn Textproduzenten generell Annahmen über das verfügbare Wissen bzw. die Diskurswelt machen, sollte sich das auch in den Referenzstrategien der zu untersuchenden Register widerspiegeln, da miteinander vertraute Diskursteilnehmer ganz andere Annahmen übereinander machen können.

2.2 Referenz und Referenzstrategien

Mit dem vorgestellten Diskursmodell erhält die folgende Ausführung zur Referenz eine wichtige Grundlage. Denn darin liegt die Motivation dafür, wie referenzielle Ausdrücke eingesetzt werden (Kunz 2010:36), und das nicht zuletzt um kognitive

Prozesse im Diskurs möglichst ökonomisch zu verteilen. Darauf aufbauend werden nun die sprachlichen Möglichkeiten zur Bildung von Referenzen erläutert.

2.2.1 Referenz

Bisher wurde Referenz lediglich als Möglichkeit gesehen, auf Informationen in der Diskurswelt zuzugreifen, einen Referenten zu spezifizieren. Dies ist jedoch eine vereinfachte Darstellung, die noch weiter ausgeführt werden muss.

Schwarz-Friesel & Chur (2014:89) unterscheiden zwei Formen sprachlicher Ausdrücke: die referenziellen Ausdrücke haben nur die Fähigkeit, Referenzen anzuzeigen, in einer konkreten Situation werden diese Ausdrücke jedoch als referierende Ausdrücke verwendet, denn hier signalisieren Textproduzenten die Erforderlichkeit eines Referenzprozesses. Dieser Prozess des Referierens ist die Herstellung einer Relation zwischen dem sprachlichen Ausdruck und einem Referenten. Referenten, also die Dinge, auf die referiert oder auch Bezug genommen werden kann, sind vielfältig und werden unterschiedlich klassifiziert. Lambrecht (1994:39) teilt Referenten in Entitäten und Propositionen (inklusive Ereignisse, Situationen und Zustände); Averintseva-Klisch (2013:30) benutzt Entitäten als Oberbegriff, worunter Objekte, Personen und Ereignisse fallen; auch Prince (1981:235) ordnet Referenten allgemein den Entitäten zu, die weitere Unterteilungen zulassen; Vater (2005:71f.) nimmt eine feingliedrigere Einteilung der Referenten nach Situationen, Objekten, Zeit und Ort und marginal auch nach Attributen, Modalitäten und Quantitäten vor. Hier soll der Oberbegriff ‚Entitäten‘ übernommen und wenn nötig weiter spezifiziert werden.

(4) S5 ich wäre gerne bereit , mir diese konkreten Fragen zu notieren und sie an das zuständige Ressort weiterzugeben . (IDS 2015c)

Diese Klassifizierungen haben die Gemeinsamkeit, dass Referenten ein Teil der mentalen Repräsentation sind und nicht etwa als Teil der realen Welt zu verstehen sind. Die in (4) unterstrichenen referierenden Ausdrücke haben z. B. ganz unterschiedliche Referenten, allerdings wird es schwer, *diese konkreten Fragen* einer Entität in der realen Welt zuzuordnen. Das Konzept einer Frage, egal wie ‚konkret‘ es auch sein mag, besteht nur in der textuellen Diskurswelt (Schwarz-Friesel & Chur 2014:94). Weiterhin ist der Referent von *diese konkreten Fragen* situationsgebunden, i. e. der Referent hat seine Gültigkeit nur in der aktiven Diskurswelt des Textes und der gleiche referenzielle Ausdruck kann sich in einem anderen Diskurs

auf einen anderen Referenten beziehen, so dass die Referenz auf andere Fragen abzielt. Aus diesem Grund ist die Referenz von der ‚Bedeutung‘ eines Ausdrucks zu unterscheiden (Averintseva-Klisch 2013:30).

Neben der Eigenschaft von Referenzausdrücken, einen Referenten zu spezifizieren, zeigt das Beispiel in (4) eine zweite Funktion von referierenden Ausdrücken. Der Ausdruck *sie* stellt zusätzlich zu dem Bezug zum Referenten auch eine Relation zu einem anderen referierenden Ausdruck her, durch den der Referentenbezug erst deutlich wird. Ohne die textuelle Vorerwähnung mit *diese konkreten Fragen* wäre der Referent von *sie* für den Textrezipienten nicht erschließbar.

Die Identifikation von Referenten aus sprachlichen Ausdrücken ist ein zentraler Aspekt von Referenzstrategien. Kunz vergleicht sie mit einer mentalen Suche in verschiedenen Bereichen (2010:37). Halliday & Matthiessen (2013:624f.) charakterisieren zwei Formen der Referenz, ‚exophora‘ und ‚endophora‘, wobei der Referent bei der exophorischen Referenz aus der außersprachlichen Welt, dem kontextuellen Umfeld des Textes, und bei der endophorischen Referenz aus dem Text selbst entnommen werden kann. Das korrespondiert mit den in Abschnitt 2.1.1 erläuterten konzeptuellen Levels der mentalen Repräsentation nach Gavin (2007:8ff.). Kunz (2010:68) und Pu (2011:2) merken an, dass Endophora nach verbreiteter Auffassung (siehe Schwarz 2000; Lambrecht 1994) sowohl einen linearen und hierarchischen Bezug zu einem anderen sprachlichen Ausdruck als auch zu einem außersprachlichen Referenten herstellen. Schwarz (2000:100) differenziert drei ‚search areas‘, ähnlich den Kontexten von Ariel (2001:30f.), aus denen Referenten entnommen werden können: das ‚world knowledge system‘, bei dem der Textrezipient im Langzeitgedächtnis zu suchen hat (entspricht Ariels ‚encyclopedic knowledge context‘); das ‚perceptive environment‘, also dem situativen Umfeld oder der mentalen Repräsentation der Umgebung (Ariels ‚physical context‘); und der ‚textual world‘, die aus den sprachlichen Äußerungen des Diskurses besteht (‚linguistic context‘ bei Ariel).

Obwohl die Form eines referenziellen Ausdrucks die Referenzidentifikation beeinflusst, kann keine eindeutige Beziehung zwischen Form und Funktion hergestellt werden, da mehrere Ausdrücke für dieselbe Strategie verwendet werden können, siehe für einige Tendenzen Tabelle 5 im Anhang, die keinen Anspruch auf Vollständigkeit hat. Daraus geht unter anderem hervor, dass für Referenten ohne Repräsentation in der Diskurswelt ein neues Konzept mithilfe von z. B. indefiniten

Phrasen und komplexen Nominalphrasen erstellt werden muss (Kunz 2010:48). Ähnlich werden auch Referenten markiert, „if the addressee can be assumed to know the meaning of the word ... and can therefore understand what type of thing a phrase ... describes“ (Gundel u. a. 1993:276). ‚Unused‘ sind sogenannte ‚unique referents‘, auf die ohne vorherige Einführung direkt zugegriffen werden kann, z. B. Eigennamen. Für generische Identifikation wird eine ganze Klasse von Referenten betrachtet, die in den Diskurs mithilfe von definiten und indefiniten Phrasen eingeführt werden kann (Kunz 2010:58), also aus der kontextuellen Diskurswelt erschließbar ist. ‚Situationally evoked‘ sind Referenten, die aus dem ‚perceptive environment‘ erschlossen werden können: personaldeiktische⁷ Begriffe (*ich, du*), Demonstrativa, Adverbien für örtliche und zeitliche Bezüge (2010:63ff.).

Referenzstrategien, die auf Außersprachliches verweisen, zählen zu den exophorischen Referenzen. Mit endophorischen Referenzen wird dagegen meist auf Referenten referiert, die ‚textually evoked‘ sind, also aus der ‚textual world‘ hervorgehen. Diese Strategie tritt in allen Texten sehr häufig auf und zeichnet sich besonders durch die Verwendung von Pronomen aus (Kunz 2010:68). Neue Referenten können jedoch auch von Textrezipienten endophorisch ‚inferiert‘, i. e. durch Rückschluss identifiziert, werden, indem sie in einer semantisch-pragmatischen Relation zu einem anderen Referenten in der bisherigen textuellen Diskurswelt stehen. Diese ‚inferrable‘ Referenten werden meist mit definiten und komplexen Nominalphrasen in den Diskurs eingeführt.

Referenten, die mit diesen Strategien in den Diskurs eingeführt wurden, bleiben jedoch nicht unweigerlich präsent. Im Sinne der Centering-Theorie (vgl. Abschnitt 2.1.2) werden eingeführte Referenten nur dann ‚aktiv‘ bleiben, wenn sie im Fokus des Diskurses stehen. Durch die Einführung eines neuen Referenten, der in den Fokus tritt, werden andere Referenten ‚deaktiviert‘ (Schwarz-Friesel 2007:6). Anders als die Centering-Theorie vermuten lässt, passieren diese Prozesse auch auf der globalen Diskursebene, d. h., je größer der Abstand zwischen einem referierenden Ausdruck und einer folgenden Referenz auf den gleichen Referenten ist, desto weniger ist der Referent greifbar in der Diskurswelt des Textrezipienten (Kunz 2010:41), so dass dieser erst wieder ‚reaktiviert‘ werden muss (Schwarz-Friesel 2007:6). Daraus folgt, dass verschiedene Referenten in Diskursen kontinuierlich

⁷ Unter dem Begriff ‚Deixis‘ fallen generell Bezüge zu der Sprechsituation, entweder in temporaler, lokaler oder personaler Hinsicht (Schwarz-Friesel & Chur 2014:101).

um die Aufmerksamkeit konkurrieren und Textproduzenten mit geeigneten Referenzstrategien deshalb zwischen Aktivierung und Reaktivierung wechseln müssen, um die Identifizierbarkeit der Referenten zu garantieren.

In diesem Abschnitt wurden verschiedene Möglichkeiten betrachtet, wie Referenzen hergestellt werden können und die Identität des Referenten anhand der textuellen und kontextuellen Diskurswelt kommuniziert wird. Im nächsten Abschnitt soll aufgegriffen werden, wie referenzielle Ausdrücke nicht nur in Beziehung zu einem Referenten, sondern auch in Relation zu anderen sprachlichen Ausdrücken stehen.

2.2.2 Koreferenz im Vergleich zu anderen Referenzrelationen

Nachdem nun dargelegt wurde, wie Referenten im Einzelnen anhand von referierenden Ausdrücken im Diskurs etabliert werden, soll das Augenmerk auf Relationen zwischen referierenden Ausdrücken gelegt werden.

Eine dominante Strategie des Diskurses ist, einen Referenten in der Diskurswelt zu etablieren und anschließend mithilfe von reduzierten referenziellen Mitteln erneut auf ihn zu verweisen (Halliday & Matthiessen 2013:370). Dafür sind zwei Relationen notwendig: einerseits wird ein zweites Mal auf denselben Referenten referiert, er wird also in der Diskurswelt reaktiviert, andererseits kann eine endophorische Relation zu dem vorangehenden referierenden Ausdruck hergestellt werden, wodurch erst deutlich wird, dass beide Ausdrücke denselben Referenten denotieren. Der Text in (5) expliziert, dass alle unterstrichenen referierenden Ausdrücke auf denselben Diskursreferenten verweisen, den die indefinite Nominalphrase *ein Arzt* einführt, indem ein unspezifisches Element der Menge der Ärzte vom Textproduzenten anvisiert wird und in der Diskurswelt vom Textrezipienten etabliert werden soll. Im Nachfolgenden reichen reduzierte sprachliche Ausdrücke in Form von Pronomen, damit der Rezipient die Referenz auf denselben Referenten herstellen kann; ohne den Bezug zur textuellen Diskurswelt würde der Rezipient bei der Identifizierung des Referenten jedoch scheitern.

- (5) S2 ... so ein Arzt , wenn er fertig ist , wie hier daß er heiratet und , ich weiß nicht , und richtet sich irgendwo seine Praxis ein und jetzt nämlich auf grund seiner Ausbildung macht er was er was er will , ne , also das kommt nicht in Frage . (IDS 2015b)

Diese erste Relation wird als Koreferenzrelation bezeichnet. Zwei referierende Ausdrücke stehen in einer Koreferenzrelation zueinander, sie ‚koreferieren‘, wenn sie denselben Referenten markieren (van Deemter & Kibble 2000:1; Poesio u. a. 1999:65; Kunz 2010:18; Azzam u. a. 1998:5; Halliday & Matthiessen 2013:626; Averintseva-Klisch 2013:43). Die zweite, anaphorische Relation wird häufig benutzt, um Koreferenz zweier Elemente anzuzeigen (Hawkins 1984:656). Nach Schwarz-Friesel & Consten sind Anaphern:

Ausdrücke, die innerhalb eines Textes einen anderen Ausdruck (das Antezedens oder der Antezedent) wieder aufgreifen und mit denen ein Sprecher somit auf einen im Text oder Diskurs bereits erwähnten Referenten erneut Bezug nimmt. (2007:271)

Demnach sind Anaphern Anweisungen, in der textuellen Diskurswelt nach einem korrespondierenden referierenden Ausdruck samt Referenten zu suchen; sie sind gerichtet, irreflexiv, unsymmetrisch und ‚nontransitiv‘ und ‚cannot be interpreted without information about where it occurs‘ (van Deemter & Kibble 2000:2). Der Begriff ‚Anapher‘ wird jedoch in der Literatur unterschiedlich verstanden, besonders mit Blick auf die verwendete Definition für Koreferenz. Für Schwarz-Friesel (2007:4) bedingen sich anaphorische Relationen und die Koreferenzrelation der zwei Ausdrücke. Für van Deemter & Kibble (2000:2) und Poesio u. a. (1999:65) sind anaphorische Relationen beschränkt auf die Wiederaufnahmefunktion. Anaphern sollen sich demnach auch auf nicht-referierende Ausdrücke beziehen können, wodurch beide Ausdrücke keinem Referenten zugeordnet werden können und eine Koreferenzrelation nach der strikten Definition von Referenzidentität nicht gegeben ist.

(6) Whenever a few solutions emerged, we embraced them. (van Deemter & Kibble 2000:92)

Die unterstrichenen Ausdrücke in (6) zeigen nach dieser Auffassung eine anaphorische Relation, da *them* sich auf *a few solutions* bezieht, allerdings keine Koreferenzrelation, da *a few solutions* nicht zu den referenziellen Ausdrücken gezählt wird. Solche quantifizierten Nominalphrasen, zusammen mit vielen indefiniten Nominalphrasen und prädikativen Nominalphrasen, markieren nach van Deemter & Kibble (2000:2) keinen spezifischen Referenten (siehe auch die Verwendung der indefiniten Nominalphrase *ein Arzt* in Beispiel (5)), da sie keine definite Menge markieren.

Allerdings scheinen solche Phrasen wie *a few solutions* durchaus eine Repräsentation in der mentalen Diskurswelt zu etablieren, welche als Referent fungieren könnte, sei dieser auch unspezifisch. Durch anaphorische Relationen wird diese Repräsentation wieder aufgegriffen und hat demnach in gewisser Weise denselben Referenten. Auch bei Komplex-Anaphern, dessen Antezedenten abstrakte Entitäten wie Phrasen oder Textabschnitte sind, wird eine abstrakte Repräsentation vom Textproduzenten anvisiert, die der Textrezipient erst durch den komplex-anaphorischen Ausdruck als Diskursreferenten etabliert (Marx 2011:10; Schwarz-Friesel & Chur 2014:108). Ähnliches könnte man bei quantifizierenden Nominalphrasen annehmen, die eventuell erst durch den anaphorischen Ausdruck einen Referenten etablieren. Für die Analyse von Koreferenzrelationen bleiben solche Fälle jedoch ein Problem, da keine eindeutige Zuordnung des Referenten vorgenommen werden kann. Hier wird daher eine Markierung verwendet, um diese Fälle gesondert betrachten zu können (siehe Abschnitt 3.2).

Auf der anderen Seite gibt es Koreferenzrelationen, deren Ausdrücke nicht zwangsweise in anaphorischer Relation zueinander stehen. Wiederholungen und die mehrfache Verwendung von Eigennamen denotieren jeweils dieselbe Entität, der Referent kann vom Textrezipienten aber individuell erschlossen werden (Kunz 2010:69). In (7) sind die referierenden Ausdrücke koreferent und getrennt voneinander aus der kontextuellen Diskurswelt erschließbar.

- (7) S5 und hier sind in der Tat intensive Beratungen mit dem Bund erforderlich schon allein deshalb weil wir entweder die Kompetenz auch hinsichtlich der Investitionen dem Bund werden geben müssen .
(IDS 2015c)

Kunz (2010:75) weitet die Verwendung von ‚Anapher‘ auf alle referierenden Ausdrücke aus, deren Referenten im Diskurs bereits vorerwähnt wurden, so dass als Antezedent nur die allererste Nennung einer Entität zählt. Jede Anapher steht jedoch nicht nur in Relation zu dem Referenten, sondern auch zu dem zuletzt im Diskurs erwähnten Ausdruck derselben Koreferenzkette⁸, denn „in on-line processing the reader must decide in each case of definite reference whether there is referential identity or not“ (Schwarz-Friesel 2007:4). In dieser Arbeit zählen als Anaphern daher diejenigen referierenden Ausdrücke, deren Referenzen nicht ohne

⁸ ‚Koreferenzketten‘ bezeichnen mehrere referierende Ausdrücke, die denselben Referenten markieren (Kunz 2010:24).

andere referierende Ausdrücke, i. e. ohne einen Bezug zur textuellen Diskurswelt, interpretiert werden können. Zu unterscheiden ist hier zusätzlich zwischen vorwärtsgerichteter und rückwärtsgerichteter endophorischer Referenz (Halliday & Matthiessen 2013:625). Erstere wird Katapher genannt und funktioniert anders als die rückwärtsgerichtete Anapher, schließlich muss der Rezipient die Referenz zeitweilig offen lassen und eine Leerstelle für den Referenten in der Diskurswelt speichern (Weinrich 2003:386); außerdem verfolgt ihre Verwendung sehr spezifische Ziele, wie zum Beispiel die Erzeugung von Spannung oder Interesse (Kunz 2010:69).

Ob deiktische Begriffe zu den Anaphern gezählt werden sollten, ist ebenfalls umstritten. Levinson (1983:86) differenziert die Anapher mit der textuellen Referenz von deiktischen Ausdrücken mit einer außertextuellen Referenz. Bei deiktischen Ausdrücken wird sowohl der Referent als auch die Bedeutung von der Kommunikationssituation bestimmt (Averintseva-Klisch 2013:39), allerdings gleichen die sprachlichen Mittel denen der anaphorischen Referenz.

(8) SB als Exponenten des Umweltschutzes als Vertreter seiner Partei möchte ich ihn fragen muß er da noch lang auf Bürgerinitiative warten ? (IDS 2015c)

In (8) entscheidet allein der Kontext, inwiefern die unterstrichenen Ausdrücke den deiktischen Ausdrücken zugeordnet werden können. Ist eine Person anwesend, die als Referent im ‚perceptive environment‘ gelten kann, sind sie deiktisch, andernfalls rein anaphorisch. Besonders häufig sind Referenzen zum Textproduzenten und Textrezipienten deiktisch (Kunz 2010:62). Da auch deiktische Ausdrücke in einer Koreferenzrelation auf der textuellen Ebene stehen, werden nachgehende Erwähnungen von Entitäten, die in der Kommunikationssituation anwesend sind, auch zusammengefasst unter den anaphorischen Relationen behandelt.

Nicht zu den Koreferenzrelationen zählen ‚indirekte Anaphern‘, deren Referenten durch sogenannte ‚bridging relations‘ identifiziert werden (Kunz 2010:71). Bei indirekten Anaphern wird zwar ebenfalls eine Verbindung zu einem zuvor erwähnten Ausdruck hergestellt, allerdings nicht im Sinne der Identität, sondern vielmehr auf der Basis von semantischer Assoziation oder konzeptueller Verwandtschaft (Halliday & Matthiessen 2013:61). Der Antezedent der indirekten Anapher fungiert als Hinweis oder ‚Anker‘, mit dem eine Verbindung zwischen der kontextuellen

und textuellen Diskurswelt hergestellt wird, so dass Textrezipienten den neuen Referenten inferieren können (Schwarz-Friesel 2007:8; Schwarz-Friesel & Chur 2014:107). ‚Co-classification‘ oder auch ‚comparative reference‘ ist ein Subtyp der indirekten Anapher, bei dem der Referent zur gleichen Klasse wie ein anderer Referent gehört, angezeigt unter anderem durch die Wiederholung des nominalen Kopfes wie bei *dieser Junge* und *der andere große Junge* oder Ellipsen und Substitution (Kunz 2010:33; Halliday & Matthiessen 2013:626). Im Gegensatz zur Koreferenz, bei der es um Äquivalenz geht, wird ein Kontrast angezeigt (Halliday & Matthiessen 2013:632).

Eine weitere Strategie zur Referenteneinführung auf Basis von indirekten Anaphern ist ‚co-extension‘, bei welcher der Referent mithilfe von semantischen Eigenschaften des Antezedenten erschlossen wird. Hierbei wird davon ausgegangen, dass sprachliche Ausdrücke einen mentalen Rahmen mit sich bringen, auf den in der Diskurswelt zurückgegriffen werden kann (Schwarz-Friesel & Chur 2014:107). Hierfür wird unter anderem lexikalische Kohäsion benutzt, aber auch das ‚world knowledge system‘. Beispiele dafür sind Meronymie, Hyperonymie, Antonymie, Frames und weitere (Kunz 2010:33f.; Halliday & Matthiessen 2013:61).

Nach Schwarz-Friesel & Chur (2014:103) ist Koreferenz eine der wichtigsten expliziten Kohärenzrelationen. Wegen ihres impliziten Charakters werden indirekte Anaphern hier nicht weiter betrachtet, jedoch stellen sie für weitere Untersuchungen ein sehr interessantes Feld dar. Im Fokus dieser Arbeit stehen explizite Referenzrelationen, vor allem Koreferenz, da die Eigenschaften von Koreferenzketten den Diskurs wesentlich beeinflussen und so Aufschluss über die Registervariation in Hinblick auf die Beziehungen zwischen Sprechern in öffentlichen und nicht-öffentlichen Diskursen geben sollten.

2.3 Register

Der Registerbegriff ist formal vom Stil eines Textproduzenten zu unterscheiden, denn Stile betreffen die Gewohnheiten und Präferenzen eines Individuums, wohingegen Register individuenübergreifend sind (Dooley & Levinsohn 2001:6). Halliday & Matthiessen (2013:593) verstehen Register als ein System mit bestimmten Mustern und Neumann (2014:14f.) sieht es als Manifestation des Sprachpotentials in bestimmten Kommunikationssituationen. Halliday & Hasan geben folgende Definition:

A register we can define as a variety according to use. In other words, the register is what you are speaking at the time, depending on what you are doing and the nature of the activity in which language is functioning. (1985:41)

Für einen kontrastiven Vergleich von Referenzstrategien werden Texte aus unterschiedlichen Registern benötigt. In der Einleitung wurde Register als kontextabhängige Sprachverwendung beschrieben, doch blieben dort die Faktoren der Abhängigkeit offen. Zuerst wird nun auf die Faktoren der Registervariation eingegangen, um anschließend individuelle Kriterien für die Differenzierung von Registern ableiten zu können.

2.3.1 *Registervariation*

Sprache variiert auf mehreren Ebenen, d. h. unter regionalen, historischen und soziologischen Aspekten (Neumann 2014:1). Register variieren, indem Textproduzenten systematische Entscheidungen in Anbetracht der situativen Faktoren bei der Aussprache und Wortwahl sowie des morphologischen und grammatischen Aufbaus treffen (Biber 2008:823). Neumann (2014:2) geht dabei von der Annahme aus, dass „people are never totally free in the way they use language, but rather that recurring situations predetermine their use of language“. Biber (2008:823) argumentiert dafür, dass bestimmte linguistische Merkmale nur in einem Register vorkommen, die er ‚register markers‘ nennt. Die Mehrzahl der Merkmale variiert dagegen lediglich in der Frequenz je nach Register, bekannt als ‚register features‘ (Registermerkmale). Die Frequenz wird bestimmt durch Präferenzen, die von der Sprachgemeinschaft festgelegt werden und sich wie alle Sprachvarietäten kontinuierlich verändern (Neumann 2014:2).

Nach diesen Präferenzen beeinflussen die situativen Faktoren also die Registermerkmale. Register können demnach anhand der situativen Faktoren oder der Merkmale unterschieden werden. Da hier die Registermerkmale mit Sicht auf die Referenzstrategien untersucht werden sollen, müssen die zu untersuchenden Texte zunächst nach den situativen Faktoren klassifiziert werden. Als Indikatoren für Register werden mehrheitlich drei Situationskontexte betrachtet, anhand derer die Art der Situation beschrieben werden kann: ‚field of discourse‘, ‚tenor of discourse‘ und ‚mode of discourse‘ (Eggins & Slade 2004:51; Neumann 2014:2; Halliday & Matthiessen 2013:33; Kunz 2010:191).

Unter ‚field of discourse‘ fallen inhaltliche Kriterien des Textes wie Themen und Ziele. Koreferenzrelationen sind wichtige Merkmale dieser Variable, da dominante Referenten Aufschluss über fokussierte Themen geben und weil die Distribution von Referenten anzeigt, wie die Entwicklung der Themenfelder im Text ist (Kunz 2010:195f.). Beziehungen zwischen Diskursteilnehmern sind im ‚tenor of discourse‘ verankert. Zwischenmenschliche Beziehung haben großen Einfluss darauf, wie sich Diskursteilnehmer während der Interaktion verhalten (Neumann 2014:35). Ein Grund dafür liegt in der Art, wie Diskursteilnehmer von bestimmten Personengruppen wahrgenommen werden wollen: ob sie sich mit den anderen Diskursteilnehmern identifizieren, sich integrieren oder einen Standpunkt außerhalb der Gruppe einnehmen wollen (Eggins & Slade 2004:52f.). Nach der in der Einleitung formulierten These sollten Referenzstrategien sich gerade hinsichtlich dieser Variable unterscheiden (siehe Abschnitt 2.4). Als ‚mode of discourse‘ werden Kriterien der Kommunikationsform bezeichnet. Diese bestimmen unter anderem, inwieweit Diskursteilnehmer interagieren können (Eggins & Slade 2004:51; Neumann 2014:16). Hierunter fallen Kommunikationswege, wie elektronische, graphische und phonetische Kommunikation, auch differenziert nach dem Medium, also geschrieben oder gesprochen, aber auch die Rolle der sprachlichen Kommunikation bei der Interaktion, ob sie nebensächlich oder wesentlicher Teil der Interaktion ist (Kunz 2010:251ff.) und dem Grad der Interaktion, zum Beispiel dialogisch oder monologisch (Halliday & Matthiessen 2013:37). Wie bereits in der Einleitung angemerkt, wurden bereits wesentliche Unterschiede bei dem Einsatz von Referenzstrategien auf dieser Ebene festgestellt, weshalb die Register der Untersuchung sich in dieser Variable möglichst nicht unterscheiden sollen.

Diese Variablen dienen als Hilfestellung bei der Klassifikation der Kommunikationssituationen, haben jedoch nicht den Anspruch, jeden Text eindeutig einem Register zuzuordnen. Die Unterschiede sind dafür zu graduell. Im folgenden Abschnitt werden aus den Unterkategorien dieser Variablen Kriterien für die Registerunterscheidung in der korpusbasierten Studie abgeleitet.

2.3.2 Kriterien der Registerdifferenzierung

Als Indikatoren für Register werden in diesem Abschnitt die von Neumann (2014) präsentierten Unterkategorien für ‚field of discourse‘, ‚field of tenor‘ sowie ‚field of mode‘ vorgestellt.

„Field of discourse“ teilt sich in Erfahrungsdomäne und Zielorientierung (2014:48). Die Erfahrungsdomäne beinhaltet die thematische Ebene des Textes und kann mithilfe des Vokabulars und der Frequenzen lexikalischer Felder, d. h. verwandter lexikalischer Begriffe, eingeschätzt werden (2014:49ff.). Bei der Zielorientierung können mehrere Typen gleichzeitig relevant sein, jedoch sollte eines dominant sein. Kategorisiert werden Texte hierbei in expositive, instruktive, argumentative, narrative und überzeugende Texte (2014:55). Nach Steiner (2004:16) gibt es zahlreiche Indikatoren für diese Kategorien, so mitunter Ellipsen, Zeitformen, Konjunktionen, Topikkonstruktionen und Paragrapheneinteilung für geschriebene Texte.

Soziale Position und Distanz sowie agentive Rolle gehören zum „tenor of discourse“. Unter agentiver Rolle wird verstanden, wie aktiv ein Diskursteilnehmer bei der Interaktion ist (Neumann 2014:63), gemessen unter anderem an der Anzahl der Äußerungen (Kunz 2010:233). Diskursteilnehmer werden anhand ihrer sozialen Position eingeteilt, indem sie in die gesellschaftliche Hierarchie eingeordnet werden. Nach Ventola (1979:269) können Diskursteilnehmer sowohl auf unterschiedlichen Ebenen stehen, wie Arbeitgeber und Arbeitnehmer, als auch auf gleicher Ebene, i. e. befreundete Personen. Außerdem wird das Machtverhältnis mit Blick auf Autorität, Expertise, Ausbildung und Statussymbole eingeschätzt (Neumann 2014:63; Eggins & Slade 2004:52). Die soziale Distanz gibt das Verhältnis der Sprecher in einem Kontinuum wieder. Nach Neumann (2014:67) und House (1997:41f.) lässt sich dieses in folgende Kategorien einteilen: formal, konsultativ, ungezwungen und intim. Je formaler das Verhältnis, desto eher seien explizierte Hintergrundinformationen und lexikalische statt pronominale Referenzausdrücke zu finden und dafür weniger Anzeichen für gemeinsames Hintergrundwissen. Die entgegenliegende Richtung verhalte sich dementsprechend umgekehrt mit vermehrt impliziten Äußerungen und Bezügen zum Textrezipienten (Neumann 2014:70). Eggins & Slade (2004:52f.) führen zusätzlich den Grad des Zuneigungsverhältnisses an, wonach Diskursteilnehmer im Verhältnis ungebunden, etwas gebunden wie Kollegen und Schulfreunde und hochgradig verbunden wie enge Freunde und Lebenspartner sein können. Eine wesentliche Rolle bei der sozialen Distanz spielt auch die Interaktionsgeschichte (Neumann 2014:67). Je häufiger sie interagieren, desto eher schwindet die soziale Distanz. Eine grobe Einteilung in regulären, diskontinu-

ierlichen und einmaligen Kontakt schlagen Eggins und Slade (2004:52f.) vor. Zusätzlich beeinflusst wird die Distanz von der Form des Kontakts, d. h. ob freiwillig und zu welchem Zweck. Falls der Kontakt lediglich einem pragmatischen Zweck dient, verringert sich die soziale Distanz weniger als wenn es primär um die Beziehung zueinander geht (2004:24).

„Mode of discourse“ wird unterteilt in Rolle der Sprache, Medium und Kanal. Beim Kanal wird die physikalische Beschaffenheit der Kommunikation gemessen und zwischen phonisch, graphisch und im Sinne dieser Arbeit auch elektronisch unterschieden (2014:75). Das Medium beinhaltet Aspekte des Mündlichen und Schriftlichen (Neumann 2014:67; Maas 2010:26), angezeigt unter anderem durch eine erhöhte Anzahl an Nominalisierungen in der Schriftlichkeit und pronominalen Referenzausdrücken in der Mündlichkeit (Neumann 2014:77f.). Bei der untergeordneten Rolle der Sprache seien Ellipsen und pronominale Referenz vordergründig und bei der dominanten Rolle der Sprache dafür „wohlgeformte“ Sätze und lexikalische Referenz (2014:74f.).

2.4 Ableitung spezifischer Hypothesen

Um die allgemeine These zu überprüfen, dass Sprecher abhängig von situativen Faktoren der Öffentlichkeit und Vertrautheit der Sprecher andere Referenzstrategien benutzen, bedarf es spezifischer Hypothesen, die in der folgenden korpusbasierten Studie untersucht werden können. Als abhängige Variablen stehen insbesondere die Eigenschaften der Koreferenzketten, die Rolle der syntaktischen Funktion, die Distanz referierender Ausdrücke und die unterschiedlichen Möglichkeiten der Aufnahme der Referenzen im Fokus.

Wie Biber (1988:47) beobachtet, sind literate Texte häufig strukturell komplexer als orale Texte und weisen eine geringere persönliche Zuneigung sowie eine größere Distanz und Abstraktion auf. In der Einleitung wurde bereits beschrieben, dass bestimmte Register eher in einem Kontinuum anzusetzen sind, das sich zum Beispiel von spontaner Konversation bis zu formalen akademischen Texten strecken kann (Ting 2009:51). Das legt die Annahme nahe, dass Sprecher in öffentlichen Situationen eher auf die sprachlichen Mittel der literaten Sprachverwendung zurückgreifen, da sie diesem Ende des Kontinuums wesentlich näher sind als Sprecher nicht-öffentlicher Texte. Unter Berücksichtigung dieser Annahme lassen sich nachstehende Hypothesen aufstellen.

Hypothese 1:

Sprache in nicht-öffentlichen Registern ist bei gleicher Anzahl referierender Ausdrücke in öffentlichen Registern markiert durch längere Koreferenzketten und eine geringere Anzahl an Koreferenzketten.

Diese Hypothese folgt der Auffassung von Kunz (2010:31), nach der Unterschiede „impinge on the frequency of nominal coreferring expressions“. Sie (2010:258) merkt an, dass literate Texte im Vergleich zu oraten Texten eine höhere Anzahl an referierenden Ausdrücken aufweisen, so dass hier die gleiche Anzahl an Markables betrachtet werden muss. Nach Schwarz-Friesel (2007:4) zeugen Koreferenzketten von Mittelbarkeit und Unterspezifizierung – Eigenschaften, die, wie in Abschnitt 2.1.1 erläutert, eher für vertraute Personen sprechen. In nicht-öffentlichen Diskursen sollten Textproduzenten leichter auf längere Koreferenzketten zurückgreifen können. Koreferenzketten korrelieren außerdem mit dem Stellenwert eines Referenten (Kunz 2010:17), wie Abschnitt 2.1.2 zu entnehmen ist. Öffentliche Diskurse sind mehr auf die inhaltliche Ebene fokussiert, so dass hier mehr Koreferenzketten zu erwarten sind.

Anzeichen für diese Hypothese zeigen die Ergebnisse der Studie von Kunz (2010:197), bei der 32,76 % der Nominalphrasen in literaten Aufsätzen Teil von Koreferenzketten sind im Vergleich zu 29,49 % bei den populärwissenschaftlichen Texten, die auf der Skala eher in Richtung Vertrautheit tendieren. Allerdings ist der Unterschied sehr gering und es wird nicht dargelegt, wie sich die Anzahl aller Nominalphrasen zusammensetzt.

Hypothese 2:

In nicht-öffentlichen Registern werden häufiger pronominale statt lexikalische referierende Ausdrücke verwendet als in öffentlichen Registern.

Nach Neumann (2014:78) werden in gesprochenen Registern mehr pronominale referierende Ausdrücke benutzt als in geschriebenen, unter anderem weil es leichter sein soll, sich auf das ‚perceptive environment‘ mit anderen als sprachlichen Mitteln zu beziehen, und weil Ambiguitäten leichter aufgelöst werden können (2014:163). Diese Tendenz sollte sich auch zwischen öffentlichen und nicht-öffentlichen Registern zeigen, da miteinander vertraute Sprecher sich einvernehmlicher in der Kommunikationssituation bewegen können und unbefangener Problemstrategien benutzen als weniger vertraute Sprecher in öffentlichen Sprechsituationen.

Weniger vertraute Sprecher sollten potentiell zweideutige Formulierungen eher vermeiden und stattdessen lexikalische Referenzmittel benutzen (Neumann 2014:163). Diese Vermutung wird unterstützt durch Steiners Studie (2013), dessen geschriebene Texte mehr lexikalische statt pronominale Kohäsion aufzeigen. Anhand der Markierung der Wortart soll daher zusätzlich die Häufigkeit verschiedener Pronominavorkommen aufgeschlüsselt und verglichen werden. Nach Besnier (1988:707) referieren vertraute Sprecher eher mithilfe von Personalpronomen, da sie in Diskursen mehr interagieren. Außerdem haben Bosch u. a. (2007:147f.) in ihrer Studie festgestellt, dass in ihren gesprochenen Texten 80 % der maskulinen Pronomen demonstrativ sind und in den geschriebenen Texten nur 7 %. Das spricht für einen allgemeinen Satzfokus bei nicht-öffentlichen Texten und einen nominalen Fokus in öffentlichen, so dass dieser Unterschied auch in Hinblick auf öffentliche und nicht-öffentliche Register zu vermuten ist.

Hypothese 3:

In öffentlichen Registern treten lexikalischen referierenden Ausdrücken häufiger in Subjektposition auf als in nicht-öffentlichen Registern.

Referierende Ausdrücke sind nach dem Diskursmodell aus Abschnitt 2.1 mitverantwortlich für die Strukturierung von Informationen. Nach Grosz u. a. (1995:212) hat die Subjektposition eine wichtige Funktion bei der Aufrechterhaltung eines Diskursreferenten. Um das zu untersuchen, müssen die syntaktischen Eigenschaften von referierenden Ausdrücken betrachtet werden, da sie die Ziele des Textproduzenten indizieren (Kunz 2010:39). Durch syntaktisch gleichmäßige Realisierung versucht der Sprecher, dem Referenten eine Relevanz zuzuschreiben, wohingegen wechselnde Positionen innerhalb einer Koreferenzkette schwankende Wichtigkeit anzeigen (2010:17). Prince (1992:298) bemerkt, dass Subjekte eher vorhandene Informationen aufgreifen. So beobachtet Kaiser (2005:263) in ihrer Studie, dass Diskursthemen stark von Subjekten bestimmt werden und daher eher Antezedenten von Pronomen sind. Daraus ergibt sich, dass bei öffentlichen Texten mit ihrem vermeintlich inhaltlichen Fokus mehr referierende Ausdrücke in Subjektposition vorzufinden sein sollten. Literate Texte aus der Studie von Steiner (2013:27) unterschieden sich bei den Pronomen in Subjektposition (86,08 %) nur minimal von den oraten. Bei den lexikalischen sind allerdings nur 24,46 % in Objektposition und 34,41 % der Referenzausdrücke Modifikatoren. Bei den oraten

Texten wurden Pronomina hauptsächlich in Subjektposition (97,19 %) realisiert, lexikalische Referenzausdrücke mit 53,90 % eher in Objektposition, aber auch zu 31,75 % in Subjektposition. Das deutet neben der allgemeinen Subjektpräferenz in öffentlichen Texten zusätzlich auf eine Variation bei der Form der referierenden Ausdrücke je nach syntaktischer Position hin. Bosch u. a. zeigen in ihrer Studie auch, dass Personalpronomen und Demonstrativpronomen generell in Subjektpositionen präferiert werden (2007:149), die laut Hypothese 2 häufiger bei nicht-öffentlichen Texten vorkommen, und dass nominale Antezedenten von Personalpronomen zwar eher Subjekte, nominale Antezedenten von Demonstrativpronomen dafür aber eher Objekte sind (2007:151).

Hypothese 4:

In nicht-öffentlichen Registern werden kürzere Distanzen zwischen referierenden Ausdrücken benutzt als in öffentlichen Registern.

Die Distanz determiniert, wie aktiv ein Referent in der Diskurswelt ist (von Heusinger 2007:134). Wenn Referenten länger nicht erwähnt werden, also die Koreferenzkette erst nach größerer Distanz wieder aufgegriffen wird, muss der Referent meistens durch einen lexikalischen referierenden Ausdruck reaktiviert werden (Kunz 2010:82). Dies wird unterstützt durch die Beobachtung, dass Demonstrativpronomen meist naheliegende Antezedenten haben, wohingegen lexikalische referierende Ausdrücke eher auf entfernte Antezedenten hinweisen (Kunz 2010:18). Steiners (2013) Studie zeigt, dass die referierenden Ausdrücke der untersuchten literaten Texte im Durchschnitt 8,8 Sätze auseinanderliegen, wogegen die Ausdrücke der oraten Texte nur einen Abstand von durchschnittlich 6,5 Sätzen aufweisen. Aus der Studie von Kunz (2010:388) geht hervor, dass lexikalische referierende Ausdrücke bevorzugt über größere Distanzen verwendet werden, ganz entgegen den pronominalen Formen, so dass die Elemente einer Koreferenzkette in nicht-öffentlichen Registern eine geringere Distanz und, in Einklang mit Hypothese 2, mehr pronominale Ausdrücke aufweisen sollten.

3. Korpusbasierte Studie

Korpora ermöglichen es, Phänomene neben der qualitativen Betrachtung auch quantitativ zu analysieren (Halliday & Matthiessen 2013:51). Systematische Unterschiede können so mit authentischen Belegen herausgearbeitet werden, denn „[w]hat people actually say is very different from what they think they say; and even more different from what they think they ought to say“ (Halliday & Matthiessen 2013:51). In Hinblick auf die These, dass sich das Verhalten deutscher Sprecher in der öffentlichen und nicht-öffentlichen Verwendung gesprochener Sprache bezüglich der Nutzung von Referenzstrategien unterscheidet, wurde daher eine korpusbasierte Studie durchgeführt, die die Hypothesen aus Abschnitt 2.4 untersucht. Der Fokus lag dabei auf nominalen Referenzstrategien im weiten Sinn, da sie im besonderen Maße von Diskursteilnehmern für die Struktur des Diskurses verwendet werden. Im Gegensatz zu non-nominalen Referenzausdrücken wird von ihnen eher erwartet, dass sie einen Bezug zur textuellen Diskurswelt herstellen; Verben denotieren dagegen primär neue Diskursentitäten (Bois 1987:817). Für die in dieser Studie durchgeführte Koreferenzresolution haben nominale Referenzstrategien außerdem den Vorteil, dass Referenten eindeutiger zugewiesen werden können als bei non-nominalen. Aus demselben Grund wurde darauf verzichtet, indirekte Anaphern mit einzubeziehen. Auf Basis eines Inter-Annotator-Agreements des MATE-Schemas stellte Poesio u. a. (1999:66) fest, dass es kaum Übereinstimmung bei der Identifikation von Referenten aus Bridgingrelationen gab, aber 95 % denselben Antezedenten von definiten Phrasen annotierten.

3.1 Daten

Für die Untersuchung der Register wurden zwei Teil-Korpora benötigt, von denen eines öffentliche Gespräche und das andere nicht-öffentliche Gespräche beinhalten sollte. Bei der Auswahl wurde auf bereits verfügbare Daten zurückgegriffen, da diese meist vorannotiert sind und die Erhebung komplett neuer Daten den Umfang der Vorstudie im Rahmen dieses Projektes übersteigen würde. Das hat den Nachteil, dass bestimmte Faktoren nicht kontrolliert werden können, besonders da verfügbare Daten für gesprochene deutsche Sprache im Vergleich zu geschriebenen Texten begrenzt sind. Die ausgewählten Daten stammen aus dem Freiburger Korpus, verfügbar über die Datenbank für Gesprochenes Deutsch (DGD) vom IDS Mannheim (2015a). Das Freiburger Korpus eignet sich aufgrund der detaillierten

Sprecher- und Ereignisannotation für die Registerdifferenzierung. Das Korpus ist transkribiert, tokenisiert, lemmatisiert und mit POS-tags versehen.

Bei der Auswahl wurde darauf geachtet, dass die Verteilung der einzelnen Sprechbeiträge in beiden Teil-Korpora möglichst gleich ist. Öffentliche Texte tendierten zu langen Sprechbeiträgen weniger Sprecher und nicht-öffentliche Texte dagegen zu kurzen, stark wechselnden Beiträgen. Um die Vergleichbarkeit zu erhöhen, wurden daher vor allem nicht-öffentliche Texte mit größeren Redeanteilen einzelner ausgewählt. Um die Unterschiede bei den Teil-Korpora möglichst nicht durch Faktoren der regionalen, sozialen oder historischen Variation zu beeinflussen, stammen alle Texte aus einem vergleichbaren Umfeld und Zeitraum; stark dialektale Sprecherereignisse wurden ausgeschlossen.

Das Teil-Korpus des öffentlichen Registers besteht aus dem Sprecherereignis FR--_E_00205_SE_01 (IDS 2015c). Hierbei handelt es sich um eine öffentliche Umweltschutzdebatte mit insgesamt 18 dokumentierten Sprechern aus dem Jahr 1971 in Freiburg. Die Einordnung liegt begründet in den Kriterien aus Abschnitt 2.3.2. Thematisch geht es um politische Entscheidungen in Umweltschutzbelangen und um die Kritik der Bürger. Daraus leitet sich auch die Zielorientierung ab, die als höchst argumentativ sowie überzeugend beschrieben werden kann. Eine genauere Analyse der bestimmenden Faktoren ist im Rahmen dieser Arbeit nicht möglich. Die agentive Rolle konzentriert sich auf einige hervorstechende Diskussionsteilnehmer, die sich in der sozialen Position durch das Verhältnis Bürger und Politiker unterscheiden, wozu auch ein unterschiedliches Machtverhältnis aufgrund der Autorität beiträgt. Allerdings ist bei der Ausbildung und Expertise kein Unterschied festzustellen. Die soziale Distanz entspricht einem formalen Verhältnis, mit geringer Zuneigung und eher einmaligem Kontakt für einen pragmatischen Zweck. Da elektronische Kommunikation für den Kanal auszuschließen ist, um die Ergebnisse dadurch nicht zu beeinflussen, handelt es sich im Sinne der These um phonische Kommunikation. Das Medium ist mündlich und die Rolle der Sprache dominant.

Für das nicht-öffentliche Teil-Korpus wurde das Sprecherereignis FR--_E_00025_SE_01 (IDS 2015b) gewählt, eine Unterhaltung mit fünf Gesprächspartnern über die Berufsaussichten der Mediziner aus dem Jahr 1966 in Göttingen. Thematisch geht es um die medizinische Berufsausbildung, wobei die Zielorientierung zwischen argumentativer und narrativer Struktur wechselt. Der hier nicht vermeidbare Unterschied bei der Zielorientierung zum öffentlichen Teil-Korpus muss bei

der Analyse einbezogen werden. Vor allem im ‚tenor of discourse‘ sollte sich das nicht-öffentliche Teil-Korpus unterscheiden, um davon ausgehen zu können, dass das verwendete Register im Gegensatz zu dem des öffentlichen Teil-Korpus steht. So ist die agentive Rolle hier zwischen den Sprechern gleich verteilt und sie befinden sich auf derselben sozialen Ebene als Studenten ohne erkennbar ungleiche Machtstrukturen. Da die These keine Aussage zu der sozialen Hierarchie macht, sollten die Sprecher beider Register sich darin möglichst wenig unterscheiden. Der Bildungsstand bei Politikern und Studenten kann als ähnlich angesehen werden, jedoch ist der Altersunterschied zwischen den Sprechern der Teil-Korpora ein möglicher Störfaktor. Bei der sozialen Distanz sind die Sprecher des nicht-öffentlichen Textes eher ungezwungen, was im Vergleich zu dem öffentlichen Text am Verzicht der Höflichkeitsform *Sie* erkennbar ist (Weinrich 2003:435). Das Zuneigungsverhältnis dürfte dem von Freunden und Kollegen entsprechen, nach Schikorsky (1990:35) ebenfalls ein Anzeichen der „nichtoffiziellen Sphäre des Gesellschaftsverkehrs“. Die Interaktionsgeschichte lässt sich nur schwer beurteilen, jedoch lassen die lockeren Umgangsformen mindestens auf einen diskontinuierlichen Kontakt schließen, der freiwillig zu sein scheint, aber als Ausgangspunkt den beruflichen Zweck hat. Die geringere Anzahl der Diskursteilnehmer verglichen mit den öffentlichen Gesprächen sorgt nach Schikorsky (1990:34) ebenfalls für eine persönlichere Bindung und ist damit Anzeichen für informelle Privatheit. Die Kriterien des ‚mode of discourse‘ sind bis auf die Rolle der Sprache identisch. Die Rolle der Sprache scheint in diesem Text weniger dominant - ein Merkmal, das als Anzeichen für private, nicht-öffentliche Kommunikation gesehen wird (1990:34). Die zu untersuchende Korpusgröße der Teil-Korpora richtet sich nach der Anzahl der ‚Markables‘ (siehe Abschnitt 3.2), da die Tokenanzahl bei den Hypothesen keine geeignete Normalisierungsgröße ist.

3.2 Annotationsrichtlinien

Um ein geeignetes Annotationsschema für die ausgewählten Texte zu finden, wurden die gängigsten Annotationsrichtlinien für die Koreferenzresolution wie unter anderem MUC-6 (Hirschmann & Chinchor 1995) und MUC-7 (Hirschmann & Chinchor 1997) verglichen. In den Richtlinien nach MUC wird Koreferenz allein auf der Basis der Identitätsrelation analysiert, was aber für viele Formen der nominalen Referenz unzureichend ist. Poesio u. a. (1999:66) schätzen sie außerdem als

ungeeignet für die Anwendung auf Dialogtexte ein.

Da das NoSta-D-Projekt sich auf die Annotation von Nichtstandardvarietäten spezialisiert hat und medial gesprochene Texte ebenfalls in diese Kategorie gehören, wurden die in dem Projekt entwickelten Richtlinien für die Annotation von Koreferenzen (Reznicek 2013a) hier verwendet. Die Richtlinien des NoSta-D-Projektes haben den Vorteil, dass sie ein Diskursmodell (vgl. Abschnitt 2.1) zugrundelegen. Außerdem basieren die NoSta-D-Richtlinien auf den Annotationsrichtlinien von PoCoS (Kaupat u. a. 2011) und Tüba-D/Z (Naumann 2007) und sind als Erweiterung zu verstehen.

Nach den Richtlinien werden im ersten Schritt sogenannte primäre Markables als ‚CM‘ (‚chain member‘) annotiert. Diese beinhalten nominale referierende Ausdrücke wie maximale definite Nominalphrasen, inklusive Attribute, Komplemente und Appositionen, sowie Personalpronomen, Possessivpronomen, Relativpronomen, Eigennamen, lexikalische Reflexive, gebundene Anaphern und pronominale Adverbien und einige quantifizierende Phrasen. Abgesehen von Relativpronomen, die bei PoCoS nicht annotiert wurden, ist eine ausführliche Beschreibung zu den primären Markables bei Kaupat u. a. (2011) zu finden. Diese Zuordnung entspricht auch der Vorstellung von Prince (1992:299) darüber, welche Ausdrücke Definitheit andeuten. Die primären Markables werden durch Referenzrelationen mit ihren Antezedenten verbunden. Falls keine Relation zu einem weiteren referierenden Ausdruck besteht, wird das CM-Label entfernt. Antezedenten, die nicht zu den primären Markables gehören, wie unter anderem indefinite Nominalphrasen, zählen als sekundäre Markables und werden zusätzlich als CM markiert.

Bei den Referenzrelationen wird zwischen ‚coreferential‘, ‚anaphoric‘, ‚cataphoric‘, ‚bound‘, ‚group-antecedent‘ und ‚group‘ unterschieden. Wie in 2.2.2 dargestellt, zeichnen sich koreferentielle Relationen dadurch aus, dass zwei referierende Ausdrücke denselben Referenten markieren, aber die Referenz unabhängig voneinander herstellbar ist. Anaphorische Relationen gelten, „wenn die Referenz des ersten nur unter Rückgriff auf die Referenz des letzteren erschließbar ist“ (Reznicek 2013a:5). Dies gilt auch für kataphorische Relationen, mit dem Unterschied, dass der Antezedent der kataphorischen Relation dem kataphorischen Ausdruck folgt (siehe 2.2.2). Als ‚bound‘ (gebunden) werden Relationen markiert, die zwar Referenzidentität markieren, aber deren Referenten nicht erschließbar sind (2013a:6), wie bei den quantifizierten Antezedenten aus 2.2.2. Das NoSta-D-

Schema unterscheidet außerdem Gruppen von Referenten („group“) und Verweise auf einzelne Referenten einer solchen Gruppe („group-antecedent“). Marx (2011:57) stellt Studien vor, die gezeigt haben, dass solche Gruppen direkt bei der Rezeption entstehen und daher auch als solche markiert werden müssen. Die andere Möglichkeit, dass erst plurale Pronomen in der mentalen Diskurswelt eine Repräsentation für Gruppen etablieren, konnte nicht bestätigt werden.

Für die Untersuchung der in 2.4 formulierten Hypothesen wurden folgende Annotationsebenen aus dem PoCoS-Schema ausgewählt: die Form der nominalen Phrase (NP-Form) mit den Kriterien keine, definite NP, indefinite NP, Eigennamen, Personalpronomen, Possessivpronomen, Pronominaladverbien, Demonstrativpronomen und andere, erweitert nach NoSta-D um Relativpronomen; sowie die grammatische Rolle unterschieden nach Subjekt, direktes Objekt, indirektes Objekt und andere (Kaupat u. a. 2011:35f.). Für die vierte Hypothese wird außerdem die Ebene Distanz von Kunz & Lapshinova-Koltunski (2015) und Kunz (2010) übernommen, bei der die Entfernung zum Antezedenten in gleicher Satz („same“), nächster Satz („next“), bis zu vier Sätze („several“), mehr als vier Sätze („text“) kategorisiert wird.

Die Vorschläge für den Umgang mit Zweifelsfällen des NoSta-D-Schemas werden hier übernommen (Reznicek 2013a:7ff.). Demnach denotieren Prädikationen und Vergleichskonstruktionen keine Referenzidentität mit dem Subjekt, denn wie Schwarz-Friesel & Chur (2014:91) angeben, charakterisieren sie eher. Dennoch können diese Entitäten wieder aufgegriffen werden und so als Antezedenten auftreten (van Deemter & Kibble 2000:5). Expletive und Vokative sind nicht referenziell, sie haben eher eine verstärkende Wirkung (Halliday & Matthiessen 2013:160). Mit Vokativen werden vielmehr Interaktionen wie die Forderung nach Aufmerksamkeit ausgeführt (2013:159).

3.3 Annotation

Auf der Grundlage dieser Richtlinien wurden die Daten mit MMAX2 annotiert, einem plattformunabhängigen Programm für die linguistische Annotation auf mehreren Ebenen (Müller & Strube 2006). Dafür mussten die Daten zunächst aus dem vorliegenden Folker-Format in das TCF-Format exportiert werden. Einige Informationen wie Sprecherzuordnung und Zeichensetzung gingen dabei verloren. Sprecher-IDs wurden daher als Umweg in Form eines segmentinitialen Tokens hinzu-

gefügt, denn ohne sie wären viele Referenzen nicht auflösbar. Mitsamt einem segmentfinalen Satzzeichen wurden daher Token nach den Informationen des Folker-Formats hinzugefügt und anschließend anhand dieser die Segmente neu segmentiert. In einigen Fällen wurde die Segmentierung an die Vorgaben des NoSta-D-Projektes (Reznicek 2013b:4) angepasst. Anschließend wurden die Daten mithilfe von SaltNPepper (Humboldt-Universität zu Berlin 2015) nach MMAX2 übertragen und jeweils 500 Markables annotiert; das entspricht im öffentlichen Teil-Korpus ca. 2150 Token und im nicht-öffentlichen Teil-Korpus ca. 2450 Token.

Da im NoSta-D-Projekt mit dem Web-Tool WebAnno gearbeitet wurde, musste mit den Richtlinien ein neues Annotationsschema für MMAX2 erstellt werden (siehe Abbildung 1 im Anhang). MMAX2 erlaubt drei Annotationsmechanismen: die Zuweisung von Attributen, die Einteilung in Sets und die Herstellung von Relationen. Nach den Angaben des PoCoS-Schemas, das ebenfalls auf MMAX2 basiert, wurden die Kategorien NP-Form und grammatische Rolle als Attributauswahl festgelegt, so dass für jedes Markable eine Auswahl getroffen werden kann. Ebenfalls als Attribut wurde das CM-Label vergeben. Referierende Ausdrücke, die keinen Antezedenten haben, mussten so nicht gelöscht werden, sondern erhielten das Label ‚none‘ für kein Mitglied einer Koreferenzkette. Dadurch beinhaltet die Anzahl der Markables alle nominalen referierenden Ausdrücke, so dass sie als Normalisierungsgröße genommen werden kann.

Abhängig von der Markierung als ‚cm‘ konnten folgende weitere Annotationen vorgenommen werden: Alle Markables mit demselben Referenten wurden in ein Koreferenzset eingeteilt, dies wurde in WebAnno über die Verbindung der Relationen gelöst, soll hier aber anhand einer Seteinteilung explizit gemacht werden. Außerdem wurde die Art der Relation, die ein Markable herstellt, als Attribut ausgewählt, gesondert nach anaphorisch, kataphorisch, koreferentiell, gebunden, ‚group-antecedent‘ und, falls es sich um das erste Element der Koreferenzkette handelt, ‚none‘. Je nach Referenztyp wurde eine entsprechende Relation zu genau einem Antezedenten hergestellt, so dass sowohl die komplette Koreferenzkette anhand des Sets als auch die einzelnen Relationspaare nach Relationstyp betrachtet werden können. Nach Zuordnung des Antezedenten wurde die Entfernung mit der Attributauswahl ‚same‘, ‚next‘, ‚several‘ oder ‚text‘ markiert. Der Referenztyp ‚group‘ entfällt bei MMAX2, da Token mehrfach zu unterschiedlichen Markables zugeord-

net werden können. Die Relation ‚group-antecedent‘ musste daher nur für Pluralausdrücke verwendet werden. Ein Markable, das eine ‚group-antecedent‘ Relation herstellt, eröffnet dabei eine neue Koreferenzkette, da ein Referent aus einer Menge von Referenten ausgewählt wird und die Relation allein die Beziehung anzeigt.

Trotz der ausführlichen Hinweise vor allem bei Naumann (2007) sind bei der Annotation diverse Problemfälle aufgetreten, zu denen die Tabelle 6 im Anhang Beispiele mit Lösungsvorschlägen auflistet. Mitunter konnten Referenzen vom Annotator nicht aufgelöst werden, weil Wissen über die Diskurswelt des Textes nicht vorhanden war; so fehlten unter anderem Kenntnisse über das ‚perceptive environment‘ sowie fachspezifisches Wissen und zwischenmenschliches Insiderwissen. Als Faustregel gilt, dass ambige Referenten nicht als ‚cm‘ markiert wurden.

- (9) Bei dem roten Auto war ein Reifen platt.
- (10) Hinter dem Haus liegt ein Reifen.
- (11) Hinter dem Haus liegen drei Reifen.

Wie Prince (1992:302) feststellt, können definite Phrasen auch diskursneue Informationen (‚hearer-new‘) präsentieren und indefinite Phrasen können in der Diskurswelt vorhandene Informationen darstellen (1992:303). Es muss also zwischen spezifischer und unspezifischer Lesart unterschieden werden. Bei der Nominalphrase *ein Reifen* kann *ein* sowohl indefiniter Artikel als auch Zahlwort sein (Averintseva-Klisch 2013:53), wodurch die Nominalphrase eine spezifische oder unspezifische Lesart bekommen kann. Zahlwörter beeinflussen Mengenverhältnisse (Vater 2005:111), sind aber selbst in spezifische (11) und unspezifische Lesarten (9) zu unterteilen. In dem Beispiel in (9) ist *ein* als Zahlwort mit einer unspezifischen Lesart zu sehen (Halliday & Matthiessen 2013:365), da es durch *irgendein* oder *einer der vier* ersetzt und dessen Referent nicht eindeutig erschlossen werden kann, so dass ein sich darauf beziehender referierender Ausdruck als ‚bound‘ annotiert wurde. In (10) und (11) kommt die spezifische Lesart vor, wobei die Nominalphrase in (10) ambig zwischen indefinit und Zahlwort ist, trotzdem aber als Antezedent einer nichtgebundenen Relation zählt. Zur Annotation der gebundenen Relation schreibt Naumann:

One question concerning this category was how to treat generic uses of NPs (e.g. ‚der Deutsche an sich, der...‘; ‚ein Mensch, der...‘). We decided to restrict the use of the category ‚bound‘ to quantified non-referential antecedents but to include the impersonal 3rd person singular pronoun ‚man‘. (Naumann 2007:23)

Nach der Definition von NoSta-D sind generische Nominalphrasen, die als Antezedenten fungieren, nicht auszuschließen, da sie Teil einer Relation sind. Hier wurde nach der Regel verfahren, kann der Artikel durch *irgendein* ersetzt werden, wird mit der ‚bound‘ Relation auf den Ausdruck verwiesen. Da Relationen immer zwischen zwei referierenden Ausdrücken zu betrachten sind, wurden nachfolgende Referenzausdrücke derselben Koreferenzkette entsprechend als anaphorisch bzw. kataphorisch markiert.

Halliday & Matthiessen (2013:372) schreiben, dass auch wiederholte Erwähnungen derselben Substanz oder Abstraktion eine Referenz herstellen, wenn auch eine generelle Referenz zum entsprechenden ‚Type‘ (2013:395), so dass konkrete Substanzen wie *Wasser* als ‚cm‘ und mit der koreferentiellen Relation markiert wurden. Ähnlich wurde auch mit Referenten verfahren, die im Laufe des Diskurses verändert werden, wie bei *ein altes Auto*, das nach der Schrottpresse nur noch als *das zusammengepresste Eisenpaket* bezeichnet wird (Averintseva-Klisch 2013:52).

Die Referenzidentifizierung des Pluralpronomens der ersten Person ist ebenfalls oft nicht eindeutig einem Antezedenten zuzuordnen, da es sowohl den Textproduzenten mit mehreren anderen Diskursteilnehmern einschließen kann, aber auch den Textproduzenten in einem anderen situativen Kontext sowie in kollektiver Funktion (Halliday & Matthiessen 2013:62). Bei dem kollektiven Gebrauch wird für eine ‚bound‘ Relation analog zu *alle Menschen* argumentiert, da die einzelnen Referenten nicht klar spezifiziert werden können. Auf der anderen Seite ist die Zugehörigkeit zu einer solchen Menge eindeutig, da jedes Element mit der mengenspezifischen Eigenschaft zugeordnet werden kann, ganz im Gegensatz zu quantifizierenden Phrasen wie *die Hälfte der Deutschen*, wo nicht klar ist, welche Elemente der Gesamtheit in diese Hälfte gehören. Die anderen zwei Verwendungen von *wir* denotieren dagegen eine konkrete Gruppe von Individuen und werden nicht als ‚bound‘ annotiert.

Das Paradigma des Indefinitpronomens *man* wird ebenfalls als gebunden annotiert. Als NP-Form wurde ‚other‘ ausgewählt, aber für weitere Studien dieser Art wäre die Aufnahme der Kategorie ‚Indefinitpronomen‘ sinnvoll. Da das Pronomen in oraten Diskursen häufig verwendet wird, entwickelt sich eine sehr lange Koreferenzkette mit gebundenen Relationen, die allerdings keinen aktiven Diskursreferenten, sondern vielmehr die Abwesenheit dessen oder eine Neutralisierung kennzeich-

nen (Weinrich 2003:102). Die Markierung von Indefinitpronomen in einer zusammenhängenden Kette ist daher problematisch und sollte für folgende Untersuchungen überdacht werden.

Potentielle Markables in Korrekturen, die einen Teil wiederholen, wurden nicht annotiert. Außerdem wurden für die Zuordnung der Distanzattribute solche Segmente ausgelassen, die lediglich Diskursmarker enthalten, da diese meist parallel geäußert werden und nicht zur selben Art von Äußerung gezählt werden sollten.

3.4 Ergebnisse und Auswertung

Zu den vorangegangenen Annahmen über die abhängige Verwendung von Referenzstrategien in öffentlichen und nicht-öffentlichen Diskursen wurden in Abschnitt 3.4 vier Hypothesen aufgestellt. Nach der ersten Hypothese sollen in nicht-öffentlichen Registern längere Referenzketten und weniger Referenzketten vorzufinden sein. Tabelle 1 zeigt die Ergebnisse der durchgeführten Studie in Bezug auf die Referenzkettenlänge und die Anzahl der Referenzketten. Mit 78 Referenzketten weist das ausgewählte nicht-öffentliche Teil-Korpus deutlich weniger Referenzketten auf als das öffentliche und auch die Referenzkettenlänge ist in diesem Korpus um einiges länger.

Tabelle 1: Anzahl und durchschnittliche Länge der Koreferenzketten der öffentlichen und nicht-öffentlichen Teil-Korpora.

	+öffentlich	-öffentlich
Anzahl	86	78
Länge	4,33	4,83

Diese Ergebnisse bestätigen also die erste Hypothese, allerdings hat die qualitative Analyse der Texte ergeben, dass die Referenzkettenlänge des nicht-öffentlichen Teil-Korpus wesentlich durch die nach den Richtlinien vorgegebene Kennzeichnung der wiederholten Verwendung des Indefinitpronomens beeinflusst wurde, so dass nach Ausschluss dieser Referenzkette (34 Elemente im nicht-öffentlichen und 6 im öffentlichen Teil-Korpus) das nicht-öffentliche Teil-Korpus nur gering größere Durchschnittswerte bei der Länge aufzeigte. Die Ergebnisse zeigen dennoch eine Tendenz des nicht-öffentlichen Textes, längere Koreferenzketten zu bilden, was darauf hindeutet, dass vertraute Sprecher leichter Bezüge zur Diskurswelt herstellen können, mitunter aufgrund der Relevanz des ‚perceptive environment‘ oder einvernehmlicher Kenntnisse über die gegenseitige Diskurswelt.

Tabelle 2: Verteilung der Markables nach der NP-Form mit gesonderter Auszeichnung für Mitglieder von Koreferenzketten (*cm*) und pronominalen und lexikalischen Markables für die beiden Teil-Korpora.

NP-Form	+öffentlich			-öffentlich		
	<i>cm</i>	<i>alle</i>	%	<i>cm</i>	<i>alle</i>	%
definite NP	122	200	40,00	65	133	26,60
indefinite NP	19	19	3,80	18	18	3,60
Eigennamen	7	14	2,80	6	11	2,20
Personalpronomen	145	153	30,60	141	145	29,00
Possessivpronomen	14	14	2,80	15	15	3,00
Demonstrativpronomen	23	40	8,00	54	91	18,20
Relativpronomen	14	14	2,80	15	15	3,00
Pronominaladverbien	24	39	7,80	24	33	6,60
andere	7	7	1,40	39	39	7,80
Pronomen insgesamt	220	260	52,00	249	299	59,80
Lexikal. insgesamt	148	233	46,60	89	162	32,40
Insgesamt	375	500	100	377	500	100

Für Hypothese 2 wurden die Frequenzen der Formen referierender Ausdrücke betrachtet, um festzustellen, ob nicht-öffentliche Register häufiger pronominale gegenüber lexikalischen Ausdrücken verwenden. In den Teil-Korpora sind mit 46,60 % mehr lexikalische Ausdrücke für Referenzstrategien in dem öffentlichen Text verwendet worden als in dem nicht-öffentlichen, jedoch fällt der Unterschied bei den Pronomen mit 52 % in dem öffentlichen und 59,80 % in dem nicht-öffentlichen geringer aus (vgl. Tabelle 2). Auffällig ist dabei allerdings, dass der nicht-öffentliche Text viele Ausdrücke der Kategorie ‚andere‘ beinhaltet, worunter vor allem Indefinitpronomen fallen. Das könnte die Normalisierung beeinflusst haben, da als Bezugsgröße die Menge aller Markables gewählt wurde. Weiterhin ist zu beobachten, dass die Sprecher in dem nicht-öffentlichen Text mit 18,20 % mehr als doppelt so viele Demonstrativpronomen einsetzen wie die in dem öffentlichen Text. Da viele dieser Demonstrativpronomen keiner nominalen Koreferenzkette angehören und somit andere Arten von Antezedenten wie Propositionen haben, scheint es tatsächlich eine Tendenz in nicht-öffentlichen Gesprächen zu einem Fokus auf satzartige Referenten zu geben (vgl. Abschnitt 2.4). Textproduzenten in nicht-öffentlichen Kommunikationssituationen haben dem Anschein nach also weniger Probleme damit, unkonkretere Ausdrücke zu benutzen, wie etwa durch reduzierte Referenzformen, i. e. Pronomen, die komplexe Einheiten ausdrücken. Hier

wäre es interessant, die Register in Hinblick auf andere als nur nominale Referenzstrukturen zu vergleichen. Des Weiteren zeigt die Betrachtung der NP-Form bei den anaphorischen und koreferentiellen Relationen (siehe Anhang: Tabelle 7), dass in dem öffentlichen Text ca. dreimal so viele definite Nominalphrasen für anaphorische und koreferentielle Referenzrelationen verwendet wurden. Insgesamt ist also zu schließen, dass eine Tendenz im Sinne der zweiten Hypothese durchaus erkennbar ist.

Tabelle 3: *Verteilung der Markables nach ihrer grammatischen Rolle, mit gesonderter Auszeichnung für Mitglieder von Koreferenzketten (cm) in den beiden Teil-Korpora*

grammatische Rolle	+öffentlich			-öffentlich		
	cm	alle	%	cm	alle	%
Subjekt	187	216	43,20	224	262	52,40
direktes Objekt	55	81	16,20	58	92	18,40
indirektes Objekt	17	17	3,40	8	9	1,80
andere	116	186	37,20	87	137	27,40
Insgesamt	375	500	100	377	500	100

Aus den Ergebnissen zur grammatischen Rolle, zusammengefasst in Tabelle 3, können auf den ersten Blick keine Anzeichen für die dritte Hypothese erkannt werden. Es wurde vermutet, dass die Subjektposition von Textproduzenten in öffentlichen Situationen mehr bevorzugt wird (vgl. Abschnitt 2.4). Die Mehrzahl aller referierenden Ausdrücke wurde in beiden Teil-Korpora in Subjektposition beobachtet, wobei sich bei dem nicht-öffentlichen Text mit 52,40 % die meisten referierenden Ausdrücke in Subjektposition befinden. Bei einer differenzierten Darstellung nach Pronomen und lexikalischen Ausdrücken zeigt sich jedoch (vgl. Tabelle 8 im Anhang), dass der Überschuss der referierenden Ausdrücke in Subjektposition bei dem nicht-öffentlichen Text eher den Pronomen geschuldet ist und hier die Zahl der lexikalischen Ausdrücke in Subjektposition (22,47 %) leicht unter der des öffentlichen Textes (25 %) liegt. Lexikalische Ausdrücke sind jedoch in beiden Teil-Korpora meist weder Subjekt noch Objekt, sondern Teil von Präpositionalphrasen und anderen Adverbialphrasen. Ein Blick auf die drei längsten Koreferenzketten

(ausgenommen das Paradigma von *man*) hat gezeigt, dass diese Referenten in beiden Teil-Korpora fast ausschließlich in Subjektposition markiert werden. Sie spezifizieren als Personaldeixis zum Großteil in der Kommunikationssituation anwesende Personen und es ist anzuraten, diese Referenzstrategie gesondert zu betrachten (vgl. Abschnitt 2.2.2). Bei den lexikalischen Ausdrücken, welche die grammatische Rolle des direkten Objekts einnehmen, zeigt sich ein größerer Unterschied zwischen den öffentlichen und nicht-öffentlichen Teil-Korpora, wobei der öffentliche Text ein Drittel weniger lexikalische Ausdrücke als der nicht-öffentliche Text in der Position des direkten Objekts aufweist. Es können also nur sehr wenige Indizien für Hypothese 3 in den Teil-Korpora beobachtet werden, allerdings sollte eine größere Untersuchung mehr Aufschluss über die Registervariation bringen.

Tabelle 4: *Verteilung der Markables, die Teil von Koreferenzketten sind und eine Referenzrelation zu einem Antezedenten herstellen, nach der Entfernung zum Antezedenten in beiden Teil-Korpora.*

Distanz	+öffentlich		-öffentlich	
	<i>n</i>	%	<i>n</i>	%
same	71	24,30	95	31,77
next	87	29,80	106	35,45
several	58	19,90	48	16,06
text	76	26,00	50	16,72
Insgesamt	292	100	299	100

Die letzte Hypothese besagt, dass die Distanz zwischen einem referierenden Ausdruck und seinem Antezedenten in nicht-öffentlichen Registern geringer ausfällt. Die Ergebnisse der Studie zeigen (vgl. Tabelle 4), dass mit den meisten Referenzrelationen in dem nicht-öffentlichen Teil-Korpus in der Tat Bezüge zum selben oder nächsten Satz hergestellt werden (zusammen 67,22 % der Relationen), wohingegen der öffentliche Text vermehrt ‚text‘-Distanzen, also solche mit einer Distanz von mehr als vier Sätzen, beinhaltet. Bei Betrachtung der Distanz zwischen Pronomen und lexikalischen Ausdrücken beider Teil-Korpora (vgl. Anhang: Tabelle 9) wird deutlich, dass Referenten öfter in dem öffentlichen Text mit Pronomen nach längeren Distanzen wiederaufgenommen werden. Ein deutlicher Unterschied zwischen den Teil-Korpora ist bei den lexikalischen Ausdrücken mit dem Attribut ‚next‘ zu erkennen, wonach die Textproduzenten des nicht-öffentlichen Textes mit

lexikalischen Ausdrücken überwiegend Referenzrelationen erzielen, die genau einen Satz überschreiten. Aus den Ergebnissen lässt sich eine Tendenz zu kürzeren Distanzen in dem nicht-öffentlichen Text nach der Hypothese ableiten.

Generell ist bei diesen Ergebnissen zu beachten, dass es sich um wenige Beobachtungen handelt und daher keine Aussagen über die angenommenen Register gemacht werden können. Es konnten zwar Tendenzen in den Texten nach drei der vier Hypothesen aufgezeigt werden, jedoch sind diese Tendenzen eher in Hinblick auf die ausgewählten Texte zu verstehen. Trotz der Klassifizierung nach bestimmten Kriterien beinhaltet jeder Text individuelle Merkmale, die die Ergebnisse beeinflussen können. Schikorsky (1990:98) findet so in seinen privaten Texten eine sehr hohe Schwankung und geht daher von einer großen Bandbreite aus. Die hier untersuchten Texte sind z. B. beide argumentativ, allerdings ist der nicht-öffentliche Text viel weniger auf Konfrontation aus, da es um keine wichtigen Entscheidungen geht wie in dem öffentlichen Text. Außerdem wurde ein Altersunterschied zwischen den Diskursteilnehmern der Texte festgestellt, der ebenfalls Einfluss auf die Ergebnisse nehmen kann. Für weitergehende Annahmen ist daher die Untersuchung weiterer Texte notwendig, nach Möglichkeit auch die Erhebung eigener Daten nach kontrollierten Bedingungen für die zu untersuchenden Faktoren, so dass eventuell eine noch größere Vertrautheit bei den Sprechern in den nicht-öffentlichen Texten erreicht wird, wodurch ein Kontrast besser zu erkennen sein könnte. Auch ein Inter-Annotator-Agreement sollte für Folgeuntersuchungen in Betracht gezogen werden. Aus den genannten Gründen ist die Studie hier als Voruntersuchung zu verstehen, die daher auch auf eine statistische Analyse verzichtet. Bei dem Vergleich weniger Texte sind generell Unterschiede zu erwarten und eine statistische Analyse würde nur wenig aussagen. Da die Unterschiede allerdings in Richtung der Hypothesen verlaufen, erscheint eine fortlaufende Betrachtung der Hypothesen sinnvoll.

4. Zusammenfassung

Um Sprache zielführend einzusetzen, haben Textproduzenten und Textrezipienten eine Vielzahl an Strategien entwickelt, mit denen der Aufwand verringert und der Informationstransfer erleichtert werden kann. Wie in der Einleitung ausgeführt wurde, benutzen Textproduzenten in Kommunikationssituationen unterschiedliche Strategien, bestimmt durch eine Reihe situativer Faktoren wie Beziehungen zwischen Diskursteilnehmern und Zweck des Diskurses. In dieser Arbeit wurde daher untersucht, auf welche Weise Diskursteilnehmer sich die Diskurswelt zu Nutze machen, um mit anderen Diskursteilnehmern erfolgreich zu kommunizieren. Es wurde festgestellt, dass der Umgang mit der Diskurswelt bedeutend von der Verwendung der Referenzstrategien gesteuert wird. Da der Austausch zwischen Diskursteilnehmern besonders bei der medial gesprochenen Kommunikation zum Tragen kommt, wurde eine Studie angestrebt, die zwei Register oratorischer Texte, die möglichst an den Polen eines Kontinuums der Vertrautheit anzusiedeln sind, kontrastiv in Hinblick auf die Verwendung der Referenzstrukturen beleuchtet. Dabei wurde die These zugrunde gelegt, dass Sprecher in öffentlichen Kontexten aufgrund spezifischer Bedürfnisse andere Referenzstrategien verwenden als Sprecher in nicht-öffentlichen Kontexten, bei denen das zwischenmenschliche Verhältnis von größerer Vertrautheit geprägt sein sollte.

Daraufhin wurde ein Überblick über gängige Diskurstheorien präsentiert (vgl. Abschnitt 2.1), wonach davon ausgegangen wird, dass referierende Ausdrücke eine overt Verbindung zur mentalen Repräsentation des Diskurses, i. e. der Diskurswelt, darstellen und damit der Diskurs entsprechend geleitet werden kann. Referenzielle Ausdrücke dienen dabei als Mittel, um Entitäten als Referenten im Diskurs zu etablieren und vorhandene bzw. herleitbare Referenten zu (re)aktivieren, nachdem sie wegen der Konkurrenz mit anderen Diskursreferenten in der Diskurswelt des Textrezipienten einen allmählichen Prozess der Deaktivierung durchlaufen haben. Außerdem fokussieren sie Referenten im Zentrum der Aufmerksamkeit.

Im weiteren Verlauf wurden die Formen der Referenzstrategien genauer betrachtet (vgl. Abschnitt 2.2). Zahlreiche referenzielle Ausdrücke können verwendet werden, um eine Beziehung zu denselben Referenten herzustellen, wodurch in der textuellen Diskurswelt eine Koreferenzkette entsteht, dessen Eigenschaften Aufschluss über die Annahmen der Textproduzenten über die Diskurswelt des Textre-

zipienten geben. Die verschiedenen Aufnahmestrategien innerhalb der Koreferenzkette wurden vorgestellt, darunter anaphorische und kataphorische Referenzrelationen sowie Identitätsrelationen mit unreduzierten Ausdrücken – in der Studie wurde hierfür die koreferentielle Relation verwendet –, aber auch deiktischen Ausdrücken. Bei der Auflösung der Referenten bestehen einige Schwierigkeiten, die in Voraussicht auf die durchzuführende Annotation im Rahmen der Studie besprochen wurden.

Für den Aufbau der Studie wurde schließlich das Konzept der Register vorgestellt und Formen der Registervariation aufgezeigt (vgl. Abschnitt 2.3). Als situativ bedingte Sprachverwendung unterliegt das Auftreten eines Registers zahlreichen Faktoren, die in Kriterien zusammengefasst wurden, um mit ihrer Hilfe Texte weitestgehend zu klassifizieren und bestimmte Unterschiede hervorzuheben. Damit konnten Texte für die Studie ausgewählt und Probleme bei der Registerzuordnung herausgestellt werden.

Als letzte Vorbereitung für die Studie wurden aus dem vorgestellten theoretischen Hintergrund und der allgemeinen These spezifische Hypothesen abgeleitet, die als Grundlage für die Untersuchung der Texte dienten (vgl. Abschnitt 2.4). Anschließend wurde das Vorgehen bei der Studie beschrieben, darunter die Zuordnung der ausgewählten Texte in Teil-Korpora, die Anwendung der Annotationsrichtlinien sowie das Vorgehen bei der Annotation selbst (vgl. Abschnitt 3).

Mit den Ergebnissen der Studie (vgl. Abschnitt 3.4) konnte gezeigt werden, dass die in nicht-öffentliche und öffentliche Teil-Korpora unterteilten Texte sich in wesentlichen Punkten bei der Verwendung von Referenzstrategien unterscheiden. Die als öffentlich klassifizierten Texte zeigten kürzere Koreferenzketten, eine häufigere Benutzung lexikalischer Ausdrücke für die Indikation von Referenzen und längere Distanzen zwischen den Elementen der Koreferenzketten. Im Gegenzug wurden in dem nicht-öffentlichen Teil-Korpus längere Koreferenzketten, mehr pronominal referenzielle Ausdrücke und kürzere Distanzen innerhalb der Koreferenzketten beobachtet.

Bei der Subjektposition standen die Ergebnisse zum Teil im Gegensatz zu der Hypothese, obwohl die Unterscheidung zwischen pronominalen und lexikalischen Ausdrücken einen minimalen Unterschied zugunsten der Hypothese zeigte, da lexikalische Ausdrücke in dem nicht-öffentlichen Teil-Korpus häufiger in Objektposition auftraten. Hier sind in weiteren Studien eventuell detailliertere Annotationen

anzustreben, die mehr Aufschluss über die syntaktischen Eigenschaften referierender Ausdrücke geben können. Unter anderem ist der Blick auf Präferenzen bei ergativen Strukturen erstrebenswert und die Überprüfung der Annahmen von Du Bois (1987:819ff.), der mit seiner ‚preferred argument structure‘ Theorie Einschränkungen bei der Verwendung von nominalen Ausdrücken in syntaktischen Positionen deklariert.

Insgesamt können die Ergebnisse allenfalls als Tendenzen für die entsprechenden Hypothesen gesehen werden, denn die Datenmenge ist nicht ausreichend, um allgemeine Folgerungen abzuleiten. Die Tendenzen motivieren allerdings, weiter den Faktor der Öffentlichkeit in gesprochener Sprache zu erforschen und die Ergebnisse mit weiteren Daten zu überprüfen.

Rückblickend lässt sich also eine Tendenz feststellen, dass Sprecher eventuell andere Referenzstrategien in öffentlichen und nicht-öffentlichen Diskursen verwenden. Unter der Annahme, dass die Daten tatsächlich Anzeichen dafür zeigen, wird vermutet, dass es einen nominalen Fokus in öffentlichen Diskursen gibt, der sich darin begründet, dass dem Inhalt eine größere Relevanz zugeschrieben wird und nominale Konstruktionen mit den konkreten Informationen der Diskurswelt in Verbindung stehen. Dagegen konzentrieren Textproduzenten in nicht-öffentlichen Kontexten sich möglicherweise häufiger auf abstrakte Entitäten, die bei der Interaktion mit vertrauten Diskursteilnehmern leichter in der Diskurswelt verwaltet werden können. Außerdem scheint Kohärenz in nicht-öffentlichen Texten eher durch reduzierte Formen hergestellt zu werden, so dass den Sprechern eine größere Einvernehmlichkeit zugesprochen werden kann. Sie können dadurch Ambiguitäten eher hinnehmen bzw. auflösen. Hinzu kommt die vermeintlich erhöhte Bezugnahme zu Textproduzenten und Textrezipienten in nicht-öffentlichen Kontexten, den zum Beispiel Schikorsky (1990:98) in seiner Studie feststellt. Dieser erhöht das Vorkommen deiktischer Begriffe, zeugt aber gleichzeitig auch von einer engeren Beziehung zwischen den Diskursteilnehmern. Dazu ist jedoch eine Unterscheidung zwischen deiktischen und textuellen Referenzen notwendig, die für die Weiterarbeit mit der untersuchten These empfohlen werden kann. Auch in Hinblick auf andere Faktoren der Registervariation wie syntaktische Komplexität und rhetorische Mittel sollten solche Diskurse untersucht werden, um einen besseren Einblick in die Faktoren der Registervariation zu erhalten.

Literaturverzeichnis

- Ariel, Mira 2001. Accessibility theory: An overview. In T. Sanders, J. Schilperoord, & W. Spooren *Human Cognitive Processing*. Amsterdam: John Benjamins Publishing Company, 29–87.
- Averintseva-Klisch, Maria 2013. *Textkohärenz*. Heidelberg: Winter.
- Azzam, Saliha, Humphreys, Kevin & Gaizauskas, Robert 1998. Coreference Resolution in a Multilingual Information Extraction System. In R. Mitkov & B. Boguraev *Proceedings of a Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*. Association for Computational Linguistics. Stroudsburg, PA, 75–81.
- Besnier, Niko 1988. The Linguistic Relationships of Spoken and Written Nukulaelae Registers. *Language* 64, 4, 707–736.
- Biber, Douglas 2008. Multidimensional Approaches. In A. Lüdeling & M. Kytö *Corpus Linguistics*. Mouton de Gruyter, 822–855.
- Biber, Douglas 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press. <http://e-books.cambridge.org/ref/id/CBO9780511621024> [Stand 2015-08-23].
- Bois, John W. Du 1987. The Discourse Basis of Ergativity. *Language* 63, 4, 805–855.
- Bosch, Peter, Katz, Graham & Umbach, Carla 2007. The non-subject bias of German demonstrative pronouns. In M. Schwarz-Friesel, M. Consten, & M. Knees *Anaphors in text: cognitive, formal and applied approaches to anaphoric reference*. Studies in language companion series. Amsterdam ; Philadelphia: J. Benjamins Pub. Co, 145–164.
- Chafe, Wallace L. 1987. Cognitive constraints on information flow. In R. S. Tomlin *Coherence and Grounding in Discourse: Outcome of a Symposium*, Eugene, Oregon, June 1984. Typological Studies in Language. Amsterdam: John Benjamins Publishing Company, 21-51.
- van Deemter, Kees & Kibble, Rodger 2000. On Coreferring: Coreference in MUC and Related Annotation Schemes. *Comput. Linguist.* 26, 4, 629–637.
- Dooley, Robert & Levinsohn, Stephen H. 2001. *Analyzing Discourse: A Manual of Basic Concepts*. Dallas, Tex: SIL International.
- Eggs, Suzanne & Slade, Diana 2004. Relevant approaches to analysing casual conversation. In *Analyzing Casual Conversation*. Equinox Publishing Ltd, 23-66.
- Gavins, Joanna 2007. *Text World Theory: An Introduction*. 1. Auflage Edinburgh: Edinburgh University Press.
- Grice, Herbert Paul 1975. Logic and conversation. In P. Cole & J. L. Morgan *Syntax and semantics*. New York u.a: Academic Press, 41–58.
- Grosz, Barbara J., Weinstein, Scott & Joshi, Aravind K. 1995. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics* 21, 2, 203–226.
- Gundel, Jeanette K., Hedberg, Nancy & Zacharski, Ron 1993. Cognitive Status and the Form of Referring Expressions in Discourse. *Language* 69, 2, 274-307.

- Halliday, M. A. K. & Hasan, Ruqaiya 1985. *Language, context, and text: aspects of language in a social-semiotic perspective*. Reprinted with corrections. Geelong: Deakin Univ.
- Halliday, M. A. K. & Matthiessen, Christian 2013. *Halliday's Introduction to Functional Grammar*. 4th edition. Milton Park, Abingdon, Oxon: Routledge.
- Hawkins, John A. 1984. A note on referent identifiability and co-presence. *Journal of Pragmatics* 8, 5, 649–659.
- von Heusinger, Klaus 2007. Accessibility and definite noun phrases. In M. Schwarz-Friesel, M. Consten, & M. Knees *Anaphors in text: cognitive, formal and applied approaches to anaphoric reference*. Studies in language companion series. Amsterdam ; Philadelphia: J. Benjamins Pub. Co, 123–144.
- Hirschmann, Lynette & Chinchor, Nancy A. 1995. MUC-6 Coreference task definition. In *Sixth Message Understanding Conference (MUC-6)*.
- Hirschmann, Lynette & Chinchor, Nancy A. 1997. MUC-7 Coreference task definition. In *Seventh Message Understanding Conference (MUC-7)*.
- House, Juliane 1997. *Translation quality assessment: a model revisited*. Tübingen: G. Narr.
- Humboldt-Universität zu Berlin, INRIA 2015. *SaltNPepper*. Berlin.
- Hymes, Dell 1972. Models of the interaction of language and social life. In J. Gumperz & D. Hymes *Directions in sociolinguistics: The ethnography of communication*. New York: Holt, Rhinehart & Winston, 35–71.
- IDS, Mannheim 2015a. *Datenbank für Gesprochenes Deutsch (DGD)*. Datenbank für Gesprochenes Deutsch (DGD). <http://dgd.ids-mannheim.de>.
- IDS, Mannheim 2015b. FR--_E_00025_SE_01_T_01_DF_01. *Datenbank für Gesprochenes Deutsch (DGD): Grundstrukturen: Freiburger Korpus 2*.
- IDS, Mannheim 2015c. FR--_E_00205_SE_01_T_01_DF_01. *Datenbank für Gesprochenes Deutsch (DGD): Grundstrukturen: Freiburger Korpus 2*.
- Kaiser, Elsi 2005. Different Forms Have Different Referential Properties: Implications for the Notion of 'Saliency'. In A. Branco, T. McEnery, & R. Mitkov *Current Issues in Linguistic Theory*. Amsterdam: John Benjamins Publishing Company, 261–282.
- Kaupat, David, Warzecha, Saskio & Stede, Manfred 2011. *Koreferenz, chapter 5: Erweiterung des PoCoS-Kernschemas*. Potsdam: Universität Potsdam.
- Kunz, Kerstin Anna 2010. *Variation in English and German nominal coreference: a study of political essays*. Frankfurt am Main ; New York: Peter Lang.
- Kunz, Kerstin & Lapshinova-Koltunski, Ekaterina 2015. Cross-linguistic analysis of discourse variation across registers. *Nordic Journal of English Studies* 14, 1, 258–288.
- Lambrecht, Knud 1994. *Information structure and sentence form: topic, focus, and the mental representations of discourse referents*. Cambridge ; New York, NY, USA: Cambridge University Press.
- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge [Cambridgeshire] ; New York: Cambridge University Press.

- Maas, Utz 2010. Literat und orat. Grundbegriffe der Analyse geschriebener und gesprochener Sprache. *Grazer Linguistische Studien* 73, 21–150.
- Marx, Konstanze 2011. *Die Verarbeitung von Komplex-Anaphern: neurolinguistische Untersuchungen zur kognitiven Textverstehenstheorie*. Berlin: Univ.-Verl. d. TU Berlin.
- Müller, Christoph & Strube, Michael 2006. Multi-level annotation of linguistic data with MMA2. In S. Braun, K. Kohn, & J. Mukherjee *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt a.M., Germany: Peter Lang, 197–214.
- Naumann, Karin 2007. *Manual of the annotation of in-document referential relations*. Seminar für Sprachwissenschaft, Abt. Computerlinguistik Universität Tübingen.
- Neumann, Stella 2014. *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. Berlin: Walter de Gruyter.
- Poesio, M., Bruneseaux, F. & Romary, L. 1999. The MATE meta-scheme for coreference in dialogues in multiple languages. In *Towards Standards and Tools for Discourse Tagging*. 65–74. <http://aclasb.dfki.de/nlp/bib/W99-0309> [Stand 2015-08-17].
- Prince, Ellen 1981. Toward a taxonomy of given-new information. In P. Cole *Radical pragmatics*. New York: Academic Press, 223–255.
- Prince, Ellen F. 1992. The ZPG Letter: Subjects, Definiteness, and Information-status. In W. C. Mann & S. A. Thompson *Pragmatics & Beyond New Series*. Amsterdam: John Benjamins Publishing Company, 295–325.
- Pu, Ming-Ming 2011. *Discourse anaphora: a cognitive-functional approach*. Muenchen: LINCOM Europa.
- Reznicek, Marc 2013a. *Linguistische Annotation von Nichtstandardvarietäten — Guidelines und „Best Practices“: Guidelines Koreferenz*. Berlin: Humboldt-Universität zu Berlin.
- Reznicek, Marc 2013b. *Linguistische Annotation von Nichtstandardvarietäten — Guidelines und „Best Practices“: Guidelines Vorverarbeitung*. Berlin: Humboldt-Universität zu Berlin.
- Sanders, Ted & Pander Maat, Henk 2006. Cohesion and Coherence: Linguistic Approaches. In *Encyclopedia of Language & Linguistics*. Elsevier, 591–595.
- Schikorsky, Isa 1990. *Private Schriftlichkeit im 19. Jahrhundert: Untersuchungen zur Geschichte des alltäglichen Sprachverhaltens kleiner Leute*. Tübingen: Niemeyer.
- Schwarz-Friesel, Monika 2007. Indirect anaphora in text. In M. Schwarz-Friesel, M. Consten, & M. Knees *Anaphors in text: cognitive, formal and applied approaches to anaphoric reference*. Studies in language companion series. Amsterdam ; Philadelphia: J. Benjamins Pub. Co, 3–20.
- Schwarz-Friesel, Monika & Chur, Jeannette 2014. *Semantik: ein Arbeitsbuch*. 6., grundlegend überarb. und erw. Aufl. Tübingen: Narr.
- Schwarz-Friesel, Monika & Consten, Manfred 2007. Anaphern. In L. Hoffmann *Handbuch der deutschen Wortarten*. De-Gruyter-Lexikon. Berlin: de Gruyter, 265–292.

- Schwarz, Monika 2000. *Indirekte Anaphern in Texten: Studien zur domänengebundenen Referenz und Kohärenz im Deutschen*. Tübingen: M. Niemeyer.
- Sperber, Dan & Wilson, Deirdre 1995. *Relevance: communication and cognition*. 2nd ed. Oxford ; Cambridge, MA: Blackwell Publishers.
- Steiner, Erich 2013. Contrasts in English-German cohesion – frequencies, functional motivations, registerial variation and some impacts on translations. [Präsentation] Saarbrücken: Saarland Universität.
- Steiner, Erich 2004. *Translated texts: properties, variants, evaluations*. Frankfurt am Main ; New York: P. Lang.
- Ting, Jen 2009. Situational Characteristics and Register Variation: A Case Study of the Particle *Suo* in Mandarin Chinese*. In S.-C. Tseng *Linguistic patterns in spontaneous speech*. Language and linguistics monograph series. Taipei: Institute of Linguistics, Academia Sinica, 51–75.
- Vater, Heinz 2005. *Referenz-Linguistik*. München: Fink.
- Ventola, Eija 1979. The structure of casual conversation in english. *Journal of Pragmatics* 3, 3–4, 267–298.
- Wagner, Laura, Greene-Havas, Maia & Gillespie, Rebecca 2010. Development in children's comprehension of linguistic register. *Child Development* 81, 6, 1678–1686.
- Weinrich, Harald 2003. *Textgrammatik der deutschen Sprache*. 4., revidierte Auflage. Hildesheim u.a.: Olms, Georg.

Anhang

Tabelle 5: *Referenzielle Ausdrücke und ihre Referenzstrategien*

Mentaler Status	Diskurswelt	Referent	Ausdruck
brand-new	textuell neu Konzept neu	Entität nicht aus Diskurswelt erschließbar	<ul style="list-style-type: none"> • indefinite Phrasen • komplexe Nominalphrasen
brand-new	textuell neu	Entität aus kontextueller Dis- kurswelt erschließbar	<ul style="list-style-type: none"> • indefinite Phrasen • komplexe Nominalphrasen
unused	textuell neu	unikale Entität aus kontextueller Dis- kurswelt erschließbar	<ul style="list-style-type: none"> • Eigennamen • definite Nominalphrasen
-	textuell neu	Klasse von Referenten aus kontextueller Dis- kurswelt erschließbar	<ul style="list-style-type: none"> • indefinite Nominalphrasen • definite Nominalphrasen
situationally evoked	textuell neu/alt	Entität aus kontextueller Dis- kurswelt erschließbar	<ul style="list-style-type: none"> • Personaldeixis • Demonstrativpronomen • Possessivpronomen • Lokaladverbien • Temporaladverbien
inferrable	textuell neu	Entität aus Diskurswelt er- schließbar	<ul style="list-style-type: none"> • Definite Phrasen • Komplexe Nominalphrasen
textually evoked	textuell alt	Entität aus textueller Diskurs- welt erschließbar	<ul style="list-style-type: none"> • Definite Phrasen • Pronomen • Adverbialpronomen

Abbildung 1: MMAX2 Annotationsschema

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<annotationschema>

  <attribute id="level_np_form" name="np_form" type="nominal_list" text="Choose noun phrase
  form!">
    <value id="value_np_form_none" name="none"/>
    <value id="value_np_form_NE" name="NE"/>
    <value id="value_np_form_defNP" name="defNP"/>
    <value id="value_np_form_indefNP" name="indefNP"/>
    <value id="value_np_form_PPER" name="PPER"/>
    <value id="value_np_form_PPOS" name="PPOS"/>
    <value id="value_np_form_PDS" name="PDS"/>
    <value id="value_np_form_PADV" name="PADV"/>
    <value id="value_np_form_PREL" name="PREL"/>
    <value id="value_np_form_other" name="other"/>
  </attribute>

  <attribute id="level_grammatical_role" name="grammatical_role" type="nominal_button"
  text="Choose grammatical form">
    <value id="value_grammatical_role_none" name="none"/>
    <value id="value_grammatical_role_SBJ" name="SBJ"/>
    <value id="value_grammatical_role_dirOBJ" name="dirOBJ"/>
    <value id="value_grammatical_role_indOBJ" name="indOBJ"/>
    <value id="value_grammatical_role_other" name="other"/>
  </attribute>

  <attribute id="level_markable_form" name="markable_form" type="nominal_button"
  text="Choose markable form!">
    <value id="value_markable_form_none" name="none"/>
    <value id="value_markable_form_cm" name="cm" next="level_type,level_corefclass"/>
  </attribute>

  <attribute id="level_type" name="Type" type="nominal_list" text="Choose type of expression!">
    <value id="value_type_none" name="none"/>
    <value id="value_type_anaphoric" name="anaphoric" next="level_distance,level_class_anapho-
    ric"/>
    <value id="value_type_cataphoric" name="cataphoric" next="level_distance,level_class_cata-
    phoric" />
    <value id="value_type_coreferential" name="coreferential" next="level_distance,le-
    vel_class_coreferential"/>
    <value id="value_type_bound" name="bound" next="level_distance,level_class_bound"/>
    <value id="value_type_group" name="group_antecedent" next="level_distance,le-
    vel_class_group"/>
  </attribute>

  <attribute id="level_corefclass" name="Ref_class_coreference" color="green" width="2"
  type="markable_set" add_to_markableset_text="Mark as coreferent" style="straight" re-
  move_from_markableset_text="Unmark as coreferent" adopt_into_markableset_text="Move this
  into current coreference set" merge_into_markableset_text="Merge both coreference sets into
  one">
    <value id="value_corefclass" name="Ref_class_coreference"/>
  </attribute>

  <attribute id="level_class_anaphoric" name="Ref_class_anaphoric" color="blue" width="2"
  type="markable_pointer" max_size="1" point_to_markable_text="Mark as anaphoric"
  style="lcurve" remove_pointer_to_text="Unmark as anaphoric" >
    <value id="value_class_anaphoric" name="not_set"/>
    <value id="value_class_anaphoric" name="set"/>
  </attribute>

```

```
<attribute id="level_class_cataphoric" name="Ref_class_cataphoric" color="yellow" width="2"
type="markable_pointer" max_size="1" point_to_markable_text="Mark as cataphoric"
style="xcurve" remove_pointer_to_text="Unmark as cataphoric" >
  <value id="value_class_cataphoric" name="not_set" />
  <value id="value_class_cataphoric" name="set" />
</attribute>
```

```
<attribute id="level_class_bound" name="Ref_class_bound" color="black" width="2"
type="markable_pointer" max_size="1" point_to_markable_text="Mark as bound" style="rcurve"
remove_pointer_to_text="Unmark as bound" >
  <value id="value_class_bound" name="not_set" />
  <value id="value_class_bound" name="set" />
</attribute>
```

```
<attribute id="level_class_coreferential" name="Ref_class_coreferential" color="orange"
width="2" type="markable_pointer" max_size="1" point_to_markable_text="Mark as coreferential"
style="lcurve" remove_pointer_to_text="Unmark as coreferential" >
  <value id="value_class_coreferential" name="not_set" />
  <value id="value_class_coreferential" name="set" />
</attribute>
```

```
<attribute id="level_class_group" name="Ref_class_group" color="grey" width="2"
type="markable_pointer" max_size="1" point_to_markable_text="Mark as group antecedent"
style="lcurve" remove_pointer_to_text="Unmark as group antecedent" >
  <value id="value_class_group" name="not_set" />
  <value id="value_class_group" name="set" />
</attribute>
```

```
<attribute id="level_distance" name="distance" type="nominal_button" text="Distance">
  <value id="value_text" name="text"/>
  <value id="value_next" name="next"/>
  <value id="value_several" name="several"/>
  <value id="value_same" name="same"/>
</attribute>
```

```
</annotationscheme>
```

Tabelle 6: Beschreibung der Problemfälle mit Lösungsvorschlägen, die in der vorliegenden Studie angewendet wurden.

Problemfall (Beispiel)	Ebene	Kommentar
<i>FR--_E_00205_SE_01_T_01_DF_01.</i>		
<u>ich</u> bin sehr gerne bereit <u>ich</u> werde mich mit Herrn Evers in V setzen	Markables	nicht innerhalb von Korrekturen an- notiert
<u>das</u> ist auch die Gefahr bei ...	Markables	Satzreferenten annotiert, aber nicht als CM ausgewiesen
<u>vorhin</u> , <u>dort</u> , <u>hier</u>	Markables	alle Pronominaladverbien annotiert, und evtl als CM
hat <u>insofern</u> großes Aufsehene er- regt, dass da die ...	Markables	Pronominaladverbien nicht an- notieren, wenn sie eine VP/CP kontrollieren (siehe NoSta-D)
Indefinitpronomen, Interroga- tivpronomen, etc. (<u>man</u> , <u>wer</u>)	NP-Form	“other” , da “cm” in bound-Relation
<u>_ ja ja</u>	Distanz	Segmente mit nur Interjektionen und Diskursmarkern ausgelassen
ich wollte nur sagen, vielleicht weil es nicht bekannt ist , dass	Distanz	Parenthesen werden zum eingebette- ten Satz gezählt
S2 <u>wir</u> / S2 <u>ich</u>	Korefer- enzset	group_antecedent macht ein neues Koreferenz-Set auf, da Sets transitiv sind. Eine Koreferenzkette mit nur der ursprünglichen Gruppe wäre sonst nicht möglich. In Analogie zur Erstel- lung eines neuen Markables bei Grup- pen, die dann ebenfalls Ausgang- spunkt für ein neues Set sind.
S2 <u>ich</u> / S2 <u>wir</u>	Relation	Wenn ein Individuum zuerst genannt wird und später noch einer Gruppe an- gehört, wird keine Relation hergestellt.
<u>der Reihe</u> nach	Markables	nicht annotiert, da idiomatisch
aus <u>ökonomischen und nicht so</u> <u>sehr regionalbezogenen Aspekten</u>	Markables	als definite NP annotiert, da definiten Artikel ohne Bedeutungsänderung ein- gesetzt werden kann
<u>den letzten Monaten</u>	Markables	annotiert, da definit, allerdings mit bound-Relation
eigentlich müsste <u>ein Vertreter</u> <u>des Innenministeriums</u> oder <u>ein</u> <u>Vertreter des Außenministeriums</u> hier sein. / Und was würde <u>der</u> sa- gen?	Markables Relation	<i>oder</i> verbindet eine Gruppe bound-Relation wird benutzt, da Refe- rent nicht eindeutig sind
<u>ein Vertreter des Innenministeri-</u> <u>ums, mindestens aber ein Vertre-</u> <u>ter des Regierungspräsidiums</u> / <u>der</u> würde	Markables Relation	s.o.

<u>ans Kultusministerium</u> , weil <u>die dort</u> zuständig sind	Relation	<i>die</i> erstellt eine semantische Relation, <i>dort</i> ist hier als Koreferenz zu bevorzugen
das Podium (ORT), das Podium (PERSONEN)	Koreferenzset	getrennte Koreferenzsets, da semantische Relation
<u>die Sache mit dem schweizer Dreck auf deutschem Boden</u> / <u>die Frage schweizer Dreck auf deutschem Boden</u>	Koreferenzset	Koreferenz, da <i>die Sache</i> und <i>die Frage</i> Paraphrasen sind, so dass der Referent identisch ist
<u>ins Uferlose</u>	Markables	idiomatisch
<u>hier</u> - bei <u>einer Podiumsdiskussion</u> - <u>hier</u>	Markables	da Teil der Koreferenzkette, wird die indefinite Phrase als sekundäres Markable hinzugefügt
jetzt aber <u>Sie</u> , bitte <u>Herr Decker</u>	Markables	Vokative und zugehörige Pronomen ergibt ein einziges Markable
das ist <u>ein ernstes Problem</u> , <u>das sich ergeben hat</u>	Relation	Prädikative können Antezedenten von Pronomen sein
<u>die Initiative des Freiburger Bundestagsabgeordneten</u> und auch <u>die gerade in den Zeitungen ...</u>	Koreferenzset	keine Koreferenzrelation, da unterschiedliche Initiativen ausgedrückt werden
<u>der Sache</u> / <u>das ist eine Sache</u> , <u>die ... den deutschen Zoll angeht</u>	Koreferenzset	<i>das</i> und <i>der Sache</i> sind koreferent sowie <i>eine Sache</i> und <i>die</i>
<u>die großen Grundstücke</u> / <u>derartig große Grundstücke</u>	Koreferenzset	keine Koreferenz, da <i>die großen Grundstücke</i> wie bei Prädikation eher einem Attribut gleicht
<u>man</u> hat s auf der Sonnenhalde oben ... wo <u>sie</u> hinschauen	Relation	bound, da <i>sie</i> indefinit verwendet wird; ersetzbar ohne Bedeutungsveränderung durch <i>wo man hinschaut</i>
<u>Hamburger Abladeangelegenheit</u> / <u>die Frage Schuttablade</u>	Koreferenzset	Paraphrase, daher selber Referent

FR--_E_00025_SE_01_T_01_DF_01

dreizehn Semester müssen sie machen. Und wenn <u>sie</u> fertig sind, machen <u>sie</u> ...	Relation	trotz Vorerwähnung des Referenten kataporische Relation, da Referenzresolution erst durch <i>sie</i> im Hauptsatz eindeutig ist
<u>dreizehn Semester</u>	Markables	als definite NP annotiert, da es um ein konkretes Konzept geht, was als Referent anzusehen ist; die Entität ist eine Zeitspanne
irgendwo in <u>einem Krankenhaus</u> , irgendwo in <u>einem Krankenhaus</u> , <u>im Krankenhaus</u> , <u>im Krankenhaus</u>	Markables	nicht annotiert, da generisch (NoSta-D S.9)
<u>die Mediziner</u> / daß man <u>sie</u> irgendwohin dirigiert, wo <u>er</u> dann drei Jahre bleiben muss	Relation	Wechsel des Numerus von <i>sie</i> zu <i>er</i> , aber daher group-antecedent

wo er <u>drei Jahre</u> bleiben soll / muß dort <u>drei Jahre</u> bleiben / nach <u>den drei Jahren</u>	Koreferenz- set	Koreferenz, da dieselbe Zeitspanne als Referent gelten kann
<u>ein Assistent oder so ein Arzt</u> , wenn er fertig ist, dass er	Relation	bound, s.o.
auch bei <u>einem Facharzt</u> in <u>dessen</u> Praxis, <u>der</u> dazu ...	Relation	bound zwischen <i>dessen</i> und <i>einem</i> <i>Facharzt</i> , da <i>einem</i> durch <i>irgendeinem</i> ersetzbar ist, danach anaphorisch

Tabelle 7: Verteilung der anaphorischen (*ana*) und koreferentiellen (*ko*) Markables nach ihrer NP-Form in beiden Teil-Korpora.

NP-Form	+öffentlich				-öffentlich			
	<i>ana</i>	%	<i>ko</i>	%	<i>ana</i>	%	<i>ko</i>	%
definite NP	14	7,40	63	94,00	5	2,33	27	87,10
indefinite NP	1	0,50	0	0,00	0	0,00	0	0,00
Eigennamen	0	0,00	4	6,00	0	0,00	4	12,90
Personal- pronomen	115	61,00	0	0,00	125	58,14	0	0,00
Possessiv- pronomen	10	5,30	0	0,00	12	5,58	0	0,00
Demonstrativ- pronomen	16	8,50	0	0,00	43	20,00	0	0,00
Relativ- pronomen	13	6,90	0	0,00	9	4,18	0	0,00
Pronominal- adverb	19	10,00	0	0,00	20	9,30	0	0,00
andere	0	0,00	0	0,00	1	0,47	0	0,00
total	188	100	67	100	215	100	31	100

Tabelle 8: Verteilung der pronominalen (*Pro*) und lexikalischen (*Lex*) Markables nach ihrer grammatischen Rolle in den beiden Teil-Korpora

grammatische Rolle	+öffentlich				–öffentlich			
	<i>Pro</i>	%	<i>Lex</i>	%	<i>Pro</i>	%	<i>Lex</i>	%
Subjekt	144	65,46	37	25,00	173	69,48	20	22,47
direktes Objekt	26	11,82	29	19,59	26	10,44	26	29,21
indirektes Objekt	12	5,45	5	3,38	6	2,41	2	2,25
andere	38	17,27	77	52,03	44	17,67	41	46,07
Insgesamt	220	100	148	100	249	100	89	100

Tabelle 9: Verteilung der pronominalen (*Pro*) und lexikalischen (*Lex*) Markables, die Teil von Koreferenzketten sind und eine Referenzrelation zu einem Antezedenten herstellen, nach der Entfernung zum Antezedenten in beiden Teil-Korpora.

Distanz	+öffentlich				–öffentlich			
	<i>Pro</i>	%	<i>Lex</i>	%	<i>Pro</i>	%	<i>Lex</i>	%
same	63	31,50	7	8,05	83	36,25	5	13,89
next	65	32,50	22	25,39	81	35,37	14	38,89
several	34	17,00	24	27,58	36	15,72	7	19,44
text	38	19,00	34	39,08	29	12,66	10	27,78
Insgesamt	200	100	87	100	229	100	36	100