

Entwicklung eines Nullsubjektresolutionssystems

im Koreanischen

Eine computer- und korpusbasierte Untersuchung
der Parameter zur Rekonstruierbarkeit von Nullsubjekten

Inaugural-Dissertation

zur

Erlangung des Grades eines Doktors der Philosophie

in der

Fakultät für Philologie

der

RUHR-UNIVERSITÄT BOCHUM

vorgelegt

von

Arum Park

Gedruckt mit der Genehmigung der Fakultät für Philologie der Ruhr-Universität Bochum

Referent: Prof. Dr. Tibor Kiss

Korreferent: Prof. Dr. Ralf Klabunde

Tag der mündlichen Prüfung: 19.07.2024

Inhaltsverzeichnis

Inhaltsverzeichnis	i
1. Einleitung.....	2
1.1. Gegenstandsbereich und Problemstellung	2
1.2. Zielsetzung und Methodik	6
1.3. Aufbau der Arbeit	12
2. Forschungsstand zu koreanischen Nullsubjekten.....	14
2.1. Deiktische und anaphorische Nullsubjekte im Koreanischen	14
2.1.1. Definitionen der Nullsubjekttypen	14
2.1.2. Häufigkeitsverteilungen der Nullsubjekttypen	18
2.1.3. Resolution der Verweisstrukturen koreanischer Nullsubjekte	21
2.1.3.1. Anaphorische Nullsubjekte	21
2.1.3.2. Deiktische Nullsubjekte.....	27
2.2. Fazit: Wahl einer ganzheitlichen Strategie für die Resolution von Nullsubjekten	33
2.2.1. Vorstellung der Studie von Hong Munpyo (2011)	33
2.2.2. Vorstellung der Studie von Park Arum & Hong Munpyo (2015)	37
3. Datengrundlage	43
3.1. Aufbereitung der Daten für die Annotation.....	43
3.1.1. Originalkorpus und Zusatzdaten von ETRI	43
3.1.2. Annotationstool und Rohdaten	45
3.2. Annotation	47
3.2.1. Etablierte und optimierte Annotationsschemata für koreanische Nullsubjekte	47
3.2.2. Annotationsprozess nach dem optimierten Annotationsschema.....	53
3.3. Charakteristiken der annotierten Daten	56
3.3.1. Eckdaten der annotierten Rohdaten.....	56
3.3.2. Ausgeschlossene annotierte Prädikate	59
4. Nullsubjektresolutionssystem mit heuristischen Regeln	64
4.1. Resolution von adressatenbezogenen Nullsubjekten.....	65
4.1.1. Imperativmarker in direkten Aufforderungen	65
4.1.2. Feste Wendung für indirekte Aufforderungen	68
4.1.3. Linguistische Mittel zum Ausdruck von Respekt.....	71
4.2. Resolution von sprecherbezogenen Nullsubjekten	74
4.2.1. Exhortativmarker	74
4.2.2. Sprecher- und subjektbezogene Modalität.....	75
4.2.3. Lexikalische Indikatoren für die Identifizierung von Sprecherreferenten	84
4.2.3.1. Performative Verben	84
4.2.3.2. Psychologische Verben	86
4.2.3.3. Honorifika	90

4.3. Ergänzende heuristische Regeln	94
4.3.1. Regeln zur Resolution von Nullsubjekten in Nebensätzen	94
4.3.2. Regeln zum Ausschluss der Subjekte in der 3. Person	103
4.3.3. Fazit: Reihenfolge der Regelanwendung.....	107
4.4. Output-Dataframe von Nullsubjektresolutionssystem	111
5. Bewertung der Leistung des Nullsubjektresolutionssystems.....	113
5.1. Überprüfung der Hypothese dieser Arbeit	114
5.1.1. Analyse der Leistung der Regeln.....	114
5.1.2. Zuverlässige Indikatoren für die korrekte Rekonstruktion sprecherbezogener Nullsubjekte	116
5.1.3. Zuverlässige Indikatoren für die korrekte Rekonstruktion adressatenbezogener Nullsubjekte ..	127
5.1.4. Fazit: Anzahl der Parameter zur Rekonstruierbarkeit von Nullsubjekten	133
5.2. Leistungsanalyse des Nullsubjektresolutionssystem	135
5.2.1. Bewertung der Leistung des Nullsubjektresolutionssystem	135
5.2.2. Leistungsvergleich der beiden Nullsubjektresolutionssysteme	140
5.2.2.1. Hintergründe und Leistungsanalyse des Basissystems.....	140
5.2.2.2. Evaluation der Leistung beider Systeme im Vergleich	144
6. Schlussfolgerung.....	148
6.1. Fazit	148
6.2. Ausblick	155
Literatur	157
Appendix	165
Appendix 1: Code für die gesamte Regel in Nullsubjektresolutionssystem.....	165
Appendix 2: Code für die gesamte Regel für den ersten Schritt des Basissystems	186
Lebenslauf	189

Danksagung

Diese Arbeit wäre ohne die Unterstützung und den Beistand vieler Menschen nicht möglich gewesen. An dieser Stelle möchte ich allen danken, die mich während meiner Promotionszeit begleitet und unterstützt haben.

Zunächst gilt mein aufrichtiger Dank meinem Erstbetreuer, Herrn Prof. Dr. Tibor Kiss, der mir die Möglichkeit eröffnete, mein Promotionsstudium in Deutschland zu beginnen. Durch seine Geduld und Unterstützung über die lange Zeit hinweg hat er maßgeblich dazu beigetragen, dass ich meine Arbeit erfolgreich abschließen konnte.

Mein Zweitbetreuer, Herr Prof. Dr. Ralf Klabunde, war immer eine freundliche und wohlwollende Präsenz, besonders in den frühen Tagen, als mir alles noch fremd war. Auch wenn meine Deutschkenntnisse in seinen Kursen noch begrenzt waren, schenkte er mir stets ein verständnisvolles Lächeln.

Ich möchte den Mitgliedern des Linguistic Data Science Lab danken, die mir in der finalen Phase meiner Dissertation und bei der Vorbereitung auf die Disputation geholfen haben, obwohl wir während meines Studiums wenig persönlichen Kontakt hatten. Mein Dank gilt insbesondere Dr. Claudia Roch, Dr. Alicia Katharina Börner, Simon Masloch und Johanna Poppek.

Ein herzliches Dankeschön geht an Annelie von Könemann für ihre Unterstützung bei den zahlreichen organisatorischen Angelegenheiten.

Im Rahmen der von der DFG geförderten Forschungsgruppe „Praktiken der Personenreferenz“ (Projektnr. 457855466) danke ich dem Hamburger Team, bestehend aus Prof. Dr. Wolfgang Imo, Dr. Irina Mostovaia und Marie-Louise Bartsch, für die Möglichkeit, neue und spannende Forschungsthemen im Bereich der Chatkommunikation zu erkunden.

Ein besonderer Dank gilt Dr. Soonha Hwang, der als ehemaliger Doktorand an der RUB seine Erfahrungen mit mir geteilt und mir von Anfang bis Ende mit wertvollen Tipps und Unterstützung zur Seite gestanden hat.

Ein weiterer besonderer Dank gilt Dr. Maike Edelhoff, deren harte Arbeit und unglaubliche Unterstützung mir geholfen haben, nicht aufzugeben und diese Arbeit bis zum Ende fertigzustellen. Ohne ihre unermüdliche Hilfe wäre dies nicht möglich gewesen.

Fabian Simonjetz möchte ich dafür danken, dass er mir, als ich an der Uni noch niemanden kannte, als Programmierlehrer zur Seite stand und mir half, mein Nullsubjektresolutionssystem zum Laufen zu bringen.

Ich danke meinem ewigen Betreuer in Korea, Prof. Dr. Munpyo Hong, der mich nie vergessen und stets ermutigt hat. Prof. Dr. Jeong-Jun Lee bin ich dankbar, dass er mich während meiner gesamten Universitätszeit als Mentor unterstützt hat. Prof. Dr. Myungwon Choi danke ich dafür, dass sie mich zur Linguistik geführt und mir die Freude an diesem faszinierenden Fachgebiet vermittelt hat. Mein Dank gilt auch Prof. Dr. Minhaeng Lee und Prof. Dr. Sung-Eun Lee, die mich trotz unterschiedlicher Universität stets unterstützt haben, sowie meinem Deutschlehrer Dr. Bruno Roßbach, der mir während meiner Universitätszeit die Freude am Deutschlernen zeigte und mich großzügig in seinem Haus in Deutschland willkommen hieß.

Mein Dank gilt auch dem koreanischen Forschungsinstitut ETRI, das mir Daten zur Verfügung gestellt und mir viele Forschungsprojekte in Korea ermöglicht hat.

Dem Machine Translation Core Team von Electronic Arts in Köln danke ich für die Gelegenheit, als Machine Translation Specialist Intern zu arbeiten. Besonders danken möchte ich meiner Managerin Cristina Anselmi und der Senior Machine Translation Specialist Lisa Honig, die mich mit Geduld und Unterstützung in das Team integrierten und mir halfen, meine Fähigkeiten weiterzuentwickeln und an mich selbst zu glauben.

Meiner Familie in Korea danke ich für ihre unerschütterliche Unterstützung, ohne die ich diese lange Zeit der Herausforderungen nicht überstehen könnte.

Ein besonderer Dank gilt der Familie Tran in Deutschland, meiner deutschen Familie, die mich immer herzlich aufgenommen hat, wenn ich die Wärme meiner eigenen Familie vermisste. Insbesondere möchte ich Jackie Tran danken, der mich von Anfang bis Ende meines Studiums unterstützt und ermutigt hat. Ohne seine unerschütterliche Unterstützung hätte ich diese Reise nicht zu Ende führen können.

Mein Freundeskreis war eine wertvolle Quelle der Unterstützung und Ermutigung, insbesondere in schwierigen Zeiten. Vielen Dank an all meine Freunde, die sich stets Zeit genommen haben, mich zu treffen und zu unterstützen.

Diese Arbeit ist das Ergebnis der Unterstützung all dieser Menschen und vieler weiterer, die hier nicht namentlich erwähnt sind. Ich bin zutiefst dankbar für die Hilfe und den Beistand, die ich auf diesem Weg erfahren habe. Allein hätte ich es niemals geschafft.

Arum Park

Essen, den 23.07.2024

1. Einleitung

1.1. Gegenstandsbereich und Problemstellung

Der Begriff *Nullsubjekt* bezeichnet ein implizites Subjekt in einem Satz. Sprachen, die Nullsubjekte erlauben, sind weitaus häufiger anzutreffen als solche, die dies nicht zulassen. Diese Tendenz wird durch Daten des *World Atlas of Language Structures* (WALS)¹ illustriert, wie die Karte in Abbildung 1 zur geografischen Verbreitung von Sprachen in Bezug auf die Zulässigkeit von Nullsubjekten veranschaulicht.

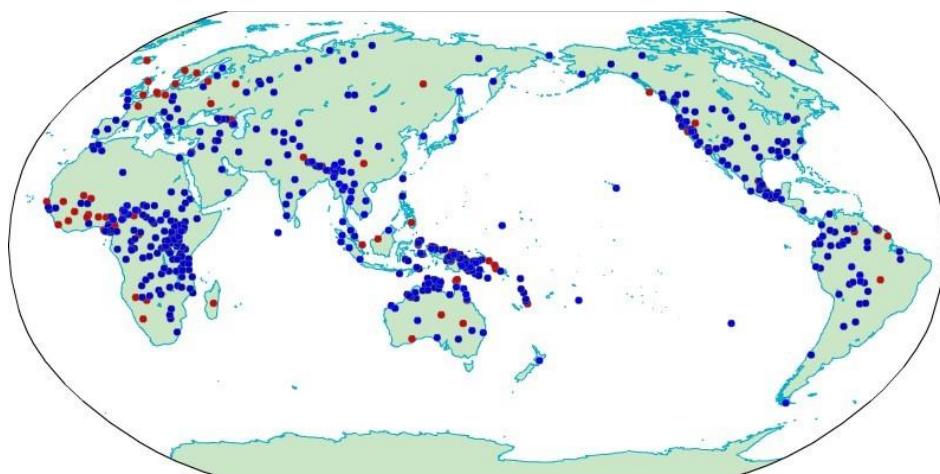


Abbildung 1: Geografische Verteilung der Sprachen in Bezug auf das Auftreten von Nullsubjekten (erstellt mit dem WALS Interactive Reference Tool²)

Die blauen Punkte, die die Nullsubjektsprachen repräsentieren, überwiegen deutlich gegenüber den roten Punkten, die die Nicht-Nullsubjektsprachen darstellen. Dies legt nahe, dass das Nullsubjekt ein weit verbreitetes Phänomen ist, das sprachübergreifend auftritt.

Im Rahmen der Prinzipien- und Parameter-Theorie (Chomsky 1981; Chomsky & Lasnik 1993) wird die Variation von syntaktischen Eigenschaften zwischen verschiedenen Sprachen als ein Ergebnis von Parametrisierung betrachtet. Diese Theorie geht davon aus, dass die Universalgrammatik durch eine begrenzte Anzahl von Prinzipien und Parametern dargestellt wird, die es ermöglichen, die Strukturen und Regularitäten aller Sprachen zu verstehen. Ein solcher Parameter könnte beispielsweise bestimmen, ob ein Subjekt obligatorisch explizit sein muss oder nicht.

Ausgehend von der Beobachtung, dass Nullsubjekte häufig in Sprachen mit einer starken Verbalflexion wie im Spanischen vorkommen, wurde eine reiche Verbalflexion als Schlüsselfaktor für die Lizenzierung und Identifizierung von Nullsubjekten angesehen (Taraldsen 1978; Hyams 1986; Rizzi

¹ Diese Datenbank, die sowohl als Buch (Haspelmath et al. 2005) als auch online verfügbar ist (Dryer & Haspelmath 2013, ist die aktuellste Auflage), bietet Einblicke in die sprachliche Vielfalt weltweit. Die aktuellen Daten über Nullsubjekte sind abzurufen unter: <https://wals.info/chapter/101>. Obwohl die Verteilung in WALS-Online wahrscheinlich einen realen Trend widerspiegelt, sollten die genauen Zahlen mit Vorsicht betrachtet werden. WALS stützt sich auf deskriptive Grammatiken von Sprachen, in denen möglicherweise nicht sämtliche Aspekte des tatsächlichen Sprachgebrauchs exakt wiedergegeben werden können. Kritische Kommentare zu diesem speziellen Merkmal der Nullsubjekte im WALS finden sich in Holmberg (2017: 364–368).

² Das *WALS Interactive Reference Tool* ist ein interaktives Nachschlagewerk, das es Nutzer:innen ermöglicht, Sprachkarten basierend auf der WALS-Datenbank in verschiedenen Formen anzuzeigen. Weitere Informationen zu diesem Tool unter: <https://www.eva.mpg.de/lingua/research/tool.php>

1986). Im Rahmen der Generativen Transformationsgrammatik unter Bezugnahme auf den Ansatz der Rektions- und Bindungstheorie (*Government and Binding Theory*) (Chomsky 1981; 1982) werden nicht realisierte Subjekte in finiten Sätzen in Sprachen mit reicher Flexion durch ein abstraktes Pronomen mit der Bezeichnung *pro* beschrieben. Die Auslassung der Pronomen wird als *pro-drop* bezeichnet. In dieser Arbeit wird der Begriff *Nullsubjekt* anstelle von *pro* verwendet, um zu verdeutlichen, dass der Fokus auf den Nullformen von Subjekten liegt. Darüber hinaus wird der Begriff *Nullpronomen (zero pronouns, im Folgenden ZP)* verwendet, um Nullformen von Pronomen zu beschreiben. Der Begriff *Nullelemente* wird als Oberbegriff eingeführt, um auf verschiedene Elemente mit Nullform zu verweisen, ohne dabei ihre spezifischen Wortarten zu betonen.

Die Unterscheidung, ob eine Sprache eine Pro-Drop-Sprache ist, erfolgt durch das Anlegen eines Pro-Drop-Parameters. Der Pro-Drop-Parameter ist definiert als ein Cluster von Eigenschaften, die zwei typologische Gruppen von Sprachen bestimmen. Wenn eine Sprache eine positive Besetzung des Parameters [+pro-drop] hat, wird die Sprache als Pro-Drop-Sprache klassifiziert, während eine Sprache mit einer negativen Besetzung des Parameters [–pro-drop] als Non-Pro-Drop-Sprache klassifiziert wird. Als prototypische Pro-Drop-Sprachen gelten beispielsweise das Spanische und das Italienische, da sie das Auslassen von Pronomen ermöglichen [+pro-drop] und über eine reichhaltige Verbalflexion verfügen. Im Gegensatz dazu werden Koreanisch, Japanisch und Chinesisch als radikale Pro-Drop-Sprachen klassifiziert (Huang 1984; 1989; Tomioka 2003; Neeleman & Szendrői 2007), da sie trotz ihrer wenigen Flexionsformen ebenfalls die Nullform *pro* zulassen.

Zur Ermittlung des Lizenzgebers von Nullsubjekten in beiden Arten von Pro-Drop-Sprachen schlagen Jaeggli & Safir (1989) einen Nullsubjekt-Parameter vor: „Nullsubjekte sind in allen und nur den Sprachen erlaubt, die morphologisch uniforme Paradigmen aufweisen“ (vgl. Jaeggli & Safir 1989: 29). Der Begriff der *morphologischen Uniformität* für den Nullsubjekt-Parameter ist wie folgt definiert: „Ein Flexionsparadigma P in einer Sprache L ist nur dann morphologisch uniform, wenn P entweder nur nicht derivierte oder nur derivierte Flexionsformen aufweist“ (vgl. Jaeggli & Safir 1989: 30). Der Nullsubjekt-Parameter und die morphologische Uniformität erfassen den Zusammenhang zwischen der Lizenzierung des Nullsubjekts und dem Flexionsparadigma. Tabelle 1 zeigt die Unterschiede in der Verbalflexion für das Verb *denken* im Präsensparadigma in vier Sprachen, um zu untersuchen, welche der vier Sprachen nach den Definitionen von Jaeggli & Safir (1989) als Nullsubjektsprache angesehen werden können.

Numerus	Person	Spanisch	Deutsch	Englisch	Koreanisch
<i>Singular</i>	1.	<i>cre-o</i>	<i>denk-e</i>	<i>think</i>	<i>saenggagha-nda</i>
	2.	<i>cre-es</i>	<i>denk-st</i>	<i>think</i>	<i>saenggagha-nda</i>
	3.	<i>cre-e</i>	<i>denk-t</i>	<i>think-s</i>	<i>saenggagha-nda</i>
<i>Plural</i>	1.	<i>cre-emos</i>	<i>denk-en</i>	<i>think</i>	<i>saenggagha-nda</i>
	2.	<i>cre-éis</i>	<i>denk-t</i>	<i>think</i>	<i>saenggagha-nda</i>
	3.	<i>cre-en</i>	<i>denk-en</i>	<i>think</i>	<i>saenggagha-nda</i>

Tabelle 1: Verbalflexion im Präsensparadigma für Spanisch, Deutsch, Englisch und Koreanisch

Nach der Definition von Jaeggli & Safir (1989) werden von den vier Sprachen in Tabelle 1 das Spanische und das Koreanische als morphologisch uniform betrachtet, da spanische Verben für jede grammatische Person und Numerus markiert sind und koreanische Verben in jeder Person und in jedem Numerus in den gleichen Formen auftreten. Im Vergleich zum Spanischen unterscheidet sich das Verbalparadigma im Deutschen nicht vollständig, da einige Konjugationsformen des Verbs identisch sind.³ Im Englischen wird nur die 3. Person Singular im Präsens mit -s markiert. Die Referenz

³ Es besteht Uneinigkeit darüber, ob das Flexionsparadigma im Deutschen morphologisch uniform ist. Dafür argumentieren Jaeggli & Hyams (1987), dagegen Hamann (1992). Die Klärung dieser Frage ist für die vorliegende Arbeit nicht von Interesse und wird daher nicht näher behandelt.

auf die Person und den Numerus wird im Deutschen und Englischen overt über Pronomen ausgedrückt, anstatt per Nullsubjekt wie in den beiden anderen Sprachen.⁴ Die Einführung der morphologischen Uniformität von Jaeggli & Safir (1989) ermöglicht es schließlich, das Phänomen des Nullsubjekts zu erklären, das in typologisch völlig unterschiedlichen Sprachen auftritt. Selbst wenn man davon ausgeht, dass die Nullsubjekte zweier Pro-Drop-Sprachen durch morphologische Uniformität lizenziert sind, ist die Identifizierung des Referenten der Nullsubjekte nötig.

Es wird argumentiert, dass die explizite Subjekt-Verb-Kongruenz von prototypischen Pro-Drop-Sprachen mit einer reichen Verbalflexion nicht nur Nullsubjekte lizenziert, sondern auch ihre Referenten identifiziert (Hyams 1986; Rizzi 1986). Die Frage ist dann, wie Nullsubjekte in radikalen Pro-Drop-Sprachen rekonstruiert werden können, in denen es kein grammatisches Mittel zur Rekonstruktion von Nullsubjekten gibt. Für diese Sprachen wird der Parameter Topik-Drop eingeführt. Aufgrund der Tatsache, dass die Nullform *pro* in Sprachen wie Koreanisch, Chinesisch und Japanisch auftreten kann, obwohl das flektierende Merkmal fehlt, ist eine Theorie entstanden, die die Nullform *pro* in diesen Sprachen als Diskurstopik betrachtet (Huang 1984; Hyams 1986). Durch die Einführung des Topik-Drop-Parameters können Nullsubjekte in radikalen Pro-Drop-Sprachen durch ein Diskurstopik identifiziert werden, während die Referenten der Nullsubjekte in Pro-Drop-Sprachen wie im Spanischen durch die Person/Numerus-Flexion des Verbs identifiziert werden können.⁵

Huang (1984) argumentiert, dass Koreanisch, Chinesisch und Japanisch typologisch als diskursorientierte Sprachen klassifiziert werden sollten, in denen ein Diskurstopik weggelassen werden darf. Die Topik-Weglassung wird als grammatische Operation erklärt, bei der eine Topik-Phrase in Bezug auf den gegebenen Kontext weggelassen wird. Dabei wird eine Topik-Kette gebildet, die aus dem ausgelassenen und dem explizit realisierten Topik im Diskurs besteht. Es wird erklärt, dass ein Null-Topik anhand der Topik-Kette identifiziert werden kann (Tsao 1977; Huang 1984). Die Auslassung eines Topiks in einer Topik-Kette wurde von Tsao (1977) als *Topik-NP-Deletion* bezeichnet.

Das Beispiel (1) (zitiert aus Oh Sun-Young⁶ 2007: 50) repräsentiert ein Gespräch zwischen den Personen A und B und veranschaulicht ein typisches Beispiel für koreanische Nullsubjekte, die als Topik-Drop interpretiert werden können. Die Topik-Drops im Diskurs (1) sind mit dem Symbol Ø gekennzeichnet.

⁴ In spezifischen Kontexten werden Nullsubjekte in Sprachen wie Deutsch und Englisch beobachtet, insbesondere in Verbindung mit der Auslassung von Subjektargumenten, die auf den Sprecher/Autor oder den Adressaten hinweisen. Im Englischen wird die Auslassung der Subjekte der 1. Person Singular als charakteristisches Merkmal des Tagebuchregisters angesehen, da Tagebücher per Definition von ihren Schreibenden handeln (Haegeman 2013: 89). Laut Androutsopoulos & Schmidt (2001: 20) ist einer der häufigsten Ellipsentypen in deutschen SMS-Nachrichten die Autor/Adressat-Ellipse. Frick (2017: 79) argumentiert, dass Vorfeld-Ellipsen der 1./2. Person Singular im Deutschen im Kontext der SMS-Kommunikation als kommunikativ und (morpho-)syntaktisch unproblematisch betrachtet werden.

⁵ In diesen Sprachen können nur Nullsubjekte der 1. und 2. Person Singular und Plural mithilfe der Verbalflexion rekonstruiert werden, da sie in der Regel nicht von einem Antezedens abhängig sind. Die Verbalflexion kann sogar zwischen Nullsubjekten der 3. Person Singular und Plural unterscheiden. Dennoch müssen sie letztendlich im Kontext identifiziert werden, um festzustellen, worauf sie sich tatsächlich beziehen.

⁶ Bei den Literaturangaben werden koreanische Namen auf besondere Weise behandelt, indem der vollständige Name im Text zitiert wird. Dies orientiert sich an den Richtlinien der *The Generic Style Rules for Linguistics*, da koreanische Nachnamen allein oft nicht sehr aussagekräftig sind.

- (1) A1. *Sinae_i-neun eotteo-ass-nayo-?*
 Sinae-TOP wie-PST-SFS.INT-SF
 'Wie ist es Sinae ergangen?'
- A2. *Ø_i bab gat-eun geos jal meog-eoss-nayo-?*
 (sie.SBJ) Reis so wie-NMZ gut essen-PST-SFS.INT-SF
 'Hat sie gut gegessen?'
- B1. *Ø_i jal meog-seubnida-.*
 (sie.SBJ) gut essen-SFS.DECL-SF
 'Sie isst gut.'

Um den Topik-Drop in (1) besser zu verstehen, ist es sinnvoll, die Diskurssituation zu kennen, denn Topik ist eine Größe in der Diskurspragmatik. Das Gespräch in (1) findet in der folgenden Situation statt (entnommen aus Oh Sun-Young 2007: 49):

Die Töchter von A und B wurden bei ihrer Geburt in einem Krankenhaus versehentlich vertauscht. Ohne dies zu wissen, zogen A und B die Töchter der jeweils anderen 14 Jahre lang auf. Nachdem sie die Wahrheit herausgefunden haben, treffen sie sich und tauschen Fragen über ihre Töchter aus. Eines der Gespräche beginnt damit, dass A B fragt, wie es ihrer leiblichen Tochter *Sinae* ergangen ist.

In (1) ist die Referenz auf *Sinae* in der ersten Äußerung (A1) explizit mit dem Topik-Marker *neun*⁷ realisiert und wird in den folgenden Äußerungen (A2 und B1) ausgelassen. Da sich das explizite Topik und die Null-Topiks auf dieselbe Entität *Sinae* beziehen, werden sie in (1) alle miteinander koindiziert. Bei der Anwendung der Topik-NP-Deletion von Tsao (1977) können die ausgelassenen Topiks in A2 und B1 als das Topik *Sinae* in A1 identifiziert werden, da sie alle in einer Topik-Kette stehen.

Es gibt jedoch auch koreanische Nullsubjekte, die nicht durch den Topik-Drop-Parameter erklärt werden können. Das Gespräch in (2) (zitiert aus Han Na-Rae 2006: 51) findet statt, als sich die zwei Freunde⁸ A und B zufällig treffen.

⁷ Es ist nicht so, dass der Topik-Marker *neun* im Koreanischen grundsätzlich eine Diskursentität als Topik in einem Diskurs qualifiziert. Eine Diskurs-Entität wird zu einem Topik in einem Diskurs, wenn die Gespräche über sie fortgesetzt werden. Es ist bekannt, dass die Verwendung des postpositionalen Suffixes *neun* eine zusätzliche pragmatische Bedeutung hat, wie z. B. Kontrast. Daher ist es sehr wahrscheinlich, dass die Diskurs-Entität mit dem Suffix *neun* diejenige ist, die der Sprecher betonen will.

Bei der Resolution koreanischer Nullsubjekte kann dementsprechend das Suffix *neun* verwendet werden, da eine NP mit *neun* im Koreanischen als die salienteste Entität gilt (vgl. Kapitel 2.1.3.1).

⁸ In der vorliegenden Arbeit werden generische Termini für den Verweis auf theoretische Referenten (wie Sprecher, Adressat) zur besseren Lesbarkeit im generischen Maskulinum belassen. Nur die Referenz auf lebensweltliche Personen geschieht mit einem Gendermarker.

- (2) A1. \emptyset *eodi* *ga-ni-?*
 (du⁹.SBJ) irgendwo gehen-SFS.INT-SF
 'Gehst du irgendwo hin?'
- B1. *eung-*, \emptyset *doseogwan-e* *ga-*.
 ja-NF (ich.SBJ) Bibliothek-in gehen.SFS¹⁰-SF
 'Ja, ich gehe in die Bibliothek.'
- A2. *geuleom* \emptyset *gachi* *ga-ja-*.
 dann (wir.SBJ) zusammen gehen-SFS.EXH-SF
 'Dann gehen wir zusammen.'

In den drei Äußerungen beziehen sich alle Subjekte auf die Diskursteilnehmer, werden aber nicht realisiert: Dies betrifft sowohl den Adressaten in A1 als auch den Sprecher in B1 und den Sprecher und Adressaten in A2. Koreanische Muttersprachler:innen würden diese Konversation als natürlich betrachten, selbst wenn keins der Subjekte in (2) ausgedrückt würde.

Probleme ergeben sich, wenn man versucht, die Nullsubjekte im Diskurs (2) als Topik-Drops zu erklären. Erstens findet sich keine Topik-NP, die bereits im Diskurs erwähnt wurde, und keiner der Referenten der Nullsubjekte kann als Topik des Diskurses angesehen werden. Bei der Identifizierung der Referenten dieser Nullsubjekte in (2) geht es darum, die Rolle der Referenten in den Äußerungen zu identifizieren, z. B. Sprecher oder Adressat. Es geht nicht darum, eine bestimmte Entität zu finden, auf die sich die Nullsubjekte beziehen, wie im Fall des Topik-Drops in (2).

Zusammenfassend lässt sich sagen, dass das Koreanische als eine Pro-Drop-Sprache angesehen wird, in der Nullsubjekte aufgrund der Uniformität der Verbalmorphologie zugelassen sind. Zudem spricht dafür, dass koreanische Nullsubjekte in einem Diskurs mithilfe der Einführung des Topik-Drop-Parameters als Null-Topik identifiziert werden können. Es scheint jedoch zusätzlich eine Art von koreanischen Nullsubjekten zu geben, die nicht allein durch den Topik-Drop erklärt werden kann. Dies liegt daran, dass die Klassifizierung koreanischer Nullsubjekte als Topik-Drop bislang auf die Diskussion von Subjektausslassungen im Diskurskontext beschränkt war. Es kann davon ausgegangen werden, dass die Auslassung von Subjekten der 1. und 2. Person, die sich auf die Diskursteilnehmer:innen beziehen, nicht dasselbe Phänomen ist, wie wenn die Referenten im Diskurs identifiziert werden können (Hong Min-pyo 2000; Kim Mi-Young 2003; Han Na-Rae 2006).

1.2. Zielsetzung und Methodik

Das Ziel dieser Arbeit ist es, Parameter zu identifizieren, die die Rekonstruktion von diskursteilnehmerbezogenen Nullsubjekten im Koreanischen ermöglichen sollen. Die Hypothese dieser Arbeit besagt, dass das Phänomen der koreanischen Nullsubjekte, insbesondere solcher, die

⁹ Das Nullsubjekt in A1 in (2) wird mit *du* glossiert, da das Verbalsuffix *eola* zur informellen Sprechweise gehört. In Kapitel 4.2.3.3 werden die Unterschiede in der Sprechweise im Koreanischen je nach Verbalsuffix genauer erläutert.

¹⁰ Die Verbalphrase *ga* in B1 ist eine Kombination aus dem Stamm *ga* des Verbs *ga-da* 'gehen' und dem Verbalsuffix *a*. Für die Verbalsuffixe in A1 und A2 wird in der Glossierung die Information über den Satzmodus hinzugefügt, da sie zur Markierung des Interrogativs oder Exhortativs verwendet werden. Im Gegensatz dazu wird das Verbalsuffix in B1 bei allen Satzmodi universell verwendet. Daher wurde die Satzmodusinformation für dieses Verbalsuffix nicht zusätzlich glossiert. Die Tatsache, dass Satzmodi im Koreanischen durch Verbalsuffixe markiert werden können, wird genutzt, um koreanische Nullsubjekte wiederherzustellen (vgl. Kapitel 2.1.3.2).

sich auf Diskursteilnehmer beziehen, aufgrund ihrer Rekonstruierbarkeit durch explizite Hinweise in verbalen Komplexen zurückzuführen ist. Daher sollten diese Hinweise eine wichtige Rolle bei der Identifizierung der Referenten in der Sprecher- oder Adressatenrolle spielen. Diese Annahme basiert auf der Erkenntnis, dass im Koreanischen der Satzmodus und die sprecher-adressaten-orientierte Modalität durch linguistische Mittel in den Verbalkomplexen herausgelesen werden können. Das folgende Beispiel (entnommen aus Han Na-Rae 2006: 52) verdeutlicht, welche linguistischen Mittel dazu beitragen können, die Referenten von diskurstteilnehmerbezogenen Nullsubjekten zu bestimmen.

- (3) a. Ø jeomsim meog-eola-.
(du.SBJ) Mittagessen essen-SFS.IMP-SF
'Iss Mittagessen.'
- b. Ø gachi ga-ja-.
(wir.SBJ) zusammen gehen-SFS.EXH-SF
'Lass uns zusammen gehen.'

Die Verbalsuffixe in (3) spielen eine entscheidende Rolle bei der Identifizierung der Referenten von Nullsubjekten als Diskurstteilnehmer:innen im Koreanischen. Das Verbalsuffix (*sentence-final suffix*, im Folgenden SFS) *eola* in (3)a ist ein Imperativmarker (SFS.IMP), und *ja* in (3)b ist ein Exhortativmarker (SFS.EXH). Dieses Beispiel verdeutlicht, dass die Identifizierung der Referenten von diskurstteilnehmerbezogenen Nullsubjekten anhand morphologischer Hinweise erfolgen kann. Frühere Studien haben betont, wie wichtig diese Hinweise für die Interpretation von diskurstteilnehmerbezogenen Nullsubjekten sind. Han Na-Rae (2006) postuliert beispielsweise, dass bestimmte Satzmodi wie der Imperativ und der Exhortativ im Koreanischen eng mit deiktischen Referenzen in Bezug auf Sprecher und Adressat verbunden sind und weist darauf hin, dass diese Satzmodi durch Verbalsuffixe markiert werden, was die mögliche Verwendung von Satzmodusmarkern zur Resolution von Nullsubjekten mit Referenten in der Sprecher- und Adressatenrollen nahelegt. Kim Mi-Young (2003) argumentiert, dass Nullsubjekte in Sätzen mit bestimmten Satzmodi von Koreanisch-Muttersprachler:innen als natürlich betrachtet werden, da die Referenten der Nullsubjekte aufgrund der Informationen in den Verbalsuffixen leicht identifiziert werden können. Sie argumentiert, dass dies einem Phänomen ähnelt, das in Sprachen wie Spanisch und Italienisch beobachtet wird, wo die Rekonstruierbarkeit von Nullsubjekten auf die reiche Verbalmorphologie zurückzuführen ist.

Es bleibt jedoch fraglich, ob die Verbalsuffixe allein der entscheidende Faktor sind, der im Koreanischen die Rekonstruktion der diskurstteilnehmerbezogenen Nullsubjekte ermöglicht. In dieser Studie wird angenommen, dass einzelne Parameter wie das Verbalsuffix allein nicht ausreichen, um den Mechanismus der koreanischen Nullsubjekte mit Referenz auf Diskurstteilnehmer:innen zu erklären, und dass eine Kombination von Parametern berücksichtigt werden muss. Das folgende Beispiel verdeutlicht, welche Faktoren dazu führen können, dass Nullsubjekte als Sprecher oder Adressat interpretiert werden.

- (4) a. Ø jeongmal haengbogha-bnida-.
(ich.SBJ) wirklich glücklich sein-SFS.DECL-SF
'Ich bin wirklich glücklich.'
- b. Ø jeongmal haengbogha-bnikka-?
(Sie.SBJ) wirklich glücklich sein-SFS.INT-SF
'Sind Sie wirklich glücklich?'
- c. Ø jeongmal haengbogha-si-gess-seubnida-.
(Sie.SBJ) wirklich glücklich sein-HON-IFX(Vermutung)-SFS.DECL-SF
'Sie müssen wirklich glücklich sein.'

Die drei Äußerungen unterscheiden sich in ihren Verbalkomplexen, abhängig von den linguistischen Mitteln, die dem Verbalstamm des Verbs *haengboghada* 'glücklich sein' angehängt werden. Ein wesentlicher Faktor, der es ermöglicht, das Nullsubjekt in (4)a als Sprecher und das Nullsubjekt in (4)b als Adressat zu interpretieren, ist eine Kombination aus dem psychologischen Verb¹¹ *haengboghada* und der Satzmodusinformation, die als Verbalsuffix markiert ist.

Frühere Studien haben postuliert, dass psychologische Verben in Deklarativsätze auf das Subjekt der 1. Person (Sprecher) und in Interrogativsätze auf das Subjekt der 2. Person (Adressat) beschränkt sind (Yang In-Seok 1972; Lee Chung-min 1976; Kim Heungsoo 1989; Yang Jungseok 1995; Yoo Hyun-kyung 1998). Diese Subjektbeschränkungen bei psychologischen Verben resultieren hauptsächlich daraus, dass nur der Sprecher angeben kann, was er gerade wahrnimmt, denkt und fühlt. Im Gegensatz dazu kann nur der Adressat die Frage nach seinem Empfinden, Denken und Fühlen beantworten. Diese Beobachtung legt nahe, dass lexikalische Indikatoren in Kombination mit morphologischen Hinweisen wie den Verbalsuffixen eine bedeutende Rolle bei der Resolution der Verweisstruktur von Nullsubjekten im Koreanischen spielen.

Es ist jedoch wichtig zu beachten, dass solche Subjektbeschränkungen, die durch die Kombination von Verbtyp und der Satzmodusinformation entstehen, nicht unveränderlich sind. Das Nullsubjekt in (4)c wird nicht mehr als Sprecher interpretiert, obwohl das psychologische Verb *haengboghada* 'glücklich sein' in einem Deklarativsatz verwendet wird. Der Grund dafür liegt am Honorativinfix *si* (*honorific*, im Folgenden HON) im Prädikat. Das Honorativinfix dient als Subjekt-Honorifikum und markiert die Kategorie Respekt am Subjekt eines Satzes. Das Subjekt des Prädikats mit dem Honorativinfix *si* muss sich nicht auf einen Adressaten beziehen. Es wird jedoch in der Literatur betont, dass das Honorativinfix *si* auch eine Funktion als Honorifikum in Bezug auf den Adressaten hat. Es kann also auch direkt gegenüber einem Adressaten die Kategorie Respekt anzeigen (vgl. Wang Hahn-Sok 1986; Lim Dong Hoon 2000; Eom Gyeong-ok 2002; Han Kil 2002). Aus diesem Grund kann das Vorhandensein des Honorativinfixes in einem Prädikat dazu beitragen, ein Nullsubjekt in Adressatenrolle zu rekonstruieren. Darüber hinaus kann das Honorativinfix dazu beitragen, dass das Nullsubjekt nicht als Sprecher interpretiert wird, da es in der Regel nicht verwendet wird, um höflich auf den Sprecher selbst zu verweisen. Zudem kann die Nutzung des Verbalprefixes *gess* (*infix*, IFX), das in (4)c als Markierung von Vermutungen verwendet wird, dazu führen, dass das Nullsubjekt nicht als Sprecher interpretiert wird. Die Äußerung einer Vermutung stellt eine Verletzung der Grice'schen Maximen dar, wonach man nicht weniger ausdrücken soll, als man weiß. Das Beispiel (4)c verdeutlicht, dass bei der Interpretation eines Nullsubjekts im Koreanischen sogar eine Kombination der Verbalaffixe berücksichtigt werden sollte. Bisher wurde jedoch wenig über ihre spezifische Bedeutung im Zusammenhang mit dem koreanischen Nullsubjekt diskutiert.

Beispiel (4) verdeutlicht, dass es entscheidend ist, herauszufinden, welche Kombination linguistischer Mittel die Interpretation der diskursteilnehmerbezogenen Nullsubjekte im Koreanischen beeinflusst, um ein tieferes Verständnis für das Phänomen zu erlangen. Dies impliziert, dass es eine Vielzahl verschiedener Kombinationen von linguistischen Mitteln in Verbalkomplexen zur Erklärung des Phänomens geben könnte.

Um die Hypothese dieser Studie zur Rekonstruierbarkeit von diskursteilnehmerbezogenen Nullsubjekten anhand expliziter Hinweise in Verbalkomplexen zu überprüfen, werden in dieser Arbeit drei Analyseschritte durchgeführt:

¹¹ Psychologische Verben sind eine Klasse von Verben, bei denen das Subjekt ein Experiencer ist, der eine Situation oder einen mentalen Zustand erlebt, die bzw. der durch den Stimulus verursacht wird (vgl. Belletti & Rizzi 1988). Zu den psychologischen Verben gehören zum Beispiel Verben der Kognition (z. B. im Deutschen *wissen*, *sich erinnern*), der Wahrnehmung (z. B. *sehen*, *hören*) und der Emotion (z. B. *lieben*, *hassen*).

1. **Untersuchung der linguistischen Mittel und ihrer Kombinationen:** Diese Untersuchung erfolgt durch Literaturrecherche und empirische Befunde. Das Ziel ist, die linguistischen Merkmale und ihre Kombinationen zu identifizieren, die zur Rekonstruktion von Nullsubjekten mit Referenz auf Diskursteilnehmer:innen beitragen können.
2. **Zuordnung expliziter Hinweise zu spezifischen Referenten:** Basierend auf den Ergebnissen der ersten Untersuchung werden explizite Hinweise in den Verbalkomplexen bestimmten Referenten in Sprecher- oder Adressatenrolle zugeordnet. Dies geschieht durch die Formulierung heuristischer Regeln, die einem „wenn-dann“-Prinzip folgen. Wenn ein Prädikat mit einem Nullsubjekt einen identifizierten Hinweis aufweist, wird diesem Prädikat ein vordefinierter Referent zugewiesen.
3. **Bewertung der Effektivität der Hinweise zur Resolution von Nullsubjekten:** Die in der zweiten Untersuchung entwickelten Regeln werden in ein Nullsubjektresolutionssystem (im Folgenden NRS) integriert und die Leistung des NRS wird bewertet, um die Effektivität der Regeln bei der Resolution von Nullsubjekten zu untersuchen. Es wird analysiert, inwieweit das NRS in der Lage ist, Nullsubjekte korrekt zu rekonstruieren. Diese Analyse soll klären, ob das Auftreten von Nullsubjekten mit Referenz auf Diskursteilnehmer:innen aufgrund ihrer Rekonstruierbarkeit mittels expliziter Hinweise in Verbalkomplexen zurückzuführen ist. Die Regeln und das NRS werden mithilfe der Programmiersprache Python implementiert.

Für die Untersuchung der Leistung des NRS ist eine annotierte Datengrundlage notwendig, die die Information über die Nullsubjekte und ihre Verweisstruktur enthält. Bislang existieren allerdings keine geeigneten annotierten Daten, die interaktive Alltagsgespräche abdecken, in denen voraussichtlich viele Nullsubjekte in Sprecher- oder Adressatenrolle auftreten.

Die Penn Korean Treebank (PKT)¹² ist das einzige Korpus, das annotierte Daten zu Dialogsequenzen enthält und koreanische Nullsubjekte beinhaltet. Allerdings ist sie für die Analyse in der vorliegenden Arbeit nicht geeignet. Der Grund dafür ist, dass die PKT auf Daten in militärischen Kontexten basiert, die andere linguistische Aspekte als alltägliche Konversationen aufweisen. Im Militär wird beispielsweise für Befehle oder Verbote oft eine Form des Futurs verwendet, während für Anweisungen elliptische Sätze genutzt werden. Dieser Unterschied ist wahrscheinlich auf die Charakteristika der militärischen Gemeinschaft zurückzuführen, in der strenge Anforderungen an Gehorsam und Befehlsgewalt bestehen (vgl. Lee Phil Young 2003: 349; Lee Ji Soo 2016: 38, 42, 44). Auch der Gebrauch von Verbalsuffixen unterscheidet sich zwischen alltäglichen und militärischen Konversationen, z. B. wird das Verbalsuffix *eula* selten in alltäglichen Konversationen verwendet, aber häufig in hierarchischen Kontexten wie dem Militär, was dem Gesagten einen antiquierten Klang verleiht (vgl. Kang Hyoun Hwa et al. 2016: 346).

Angesichts der deutlichen Unterschiede in den Registern zwischen militärischen Interaktionen und Alltagsgesprächen ist es notwendig, auf andere annotierte Daten zurückzugreifen. Aus diesem Grund wird für diese Studie ein annotiertes Korpus zusammengestellt, das die Charakteristika von Alltagsgesprächen adäquat widerspiegeln soll. Die Datengrundlage für das hier verwendete Korpus wurde vom Electronics and Telecommunications Research Institute (ETRI)¹³ bereitgestellt. Dieses

¹² Die PKT setzt sich aus Texten militärischer Ausbildungshandbücher zusammen und wurde speziell für ein maschinelles Übersetzungsprojekt an der University of Pennsylvania erstellt. Dieses Korpus beinhaltet Dialoge zwischen dem Militärpersonal und den Kommandanten:innen sowie detaillierte Anweisungen. Für weitere Informationen über den Aufbau und die Größe des Korpus vgl. Han Chung-hye et al. (2011).

¹³ ETRI ist ein führendes Forschungsinstitut in Daejeon, Südkorea, das sich auf die Entwicklung und Forschung in verschiedenen Bereichen wie Informations- und Kommunikationstechnologie sowie Künstliche Intelligenz

Korpus stellt ein Parallelkorpus für das Sprachpaar Koreanisch-Spanisch dar, wobei die koreanischen Gespräche die Grundlage bilden, die in einem weiteren Schritt ins Spanische übersetzt wurden. Bei den koreanischen Daten handelt es sich nicht um transkribierte Daten aus Alltagsgesprächen, sondern um authentische Konversationen, die gezielt in simulierten Gesprächssituationen während Reiseszenarien erzeugt wurden. Diese Szenarien umfassen verschiedene Situationen wie die Buchung von Unterkünften oder Flügen, Notfälle während der Reise, den Souvenirkauf oder das Erfragen von Wegbeschreibungen. Diese Daten wurden ursprünglich für die Entwicklung von Softwareanwendungen wie automatischen Übersetzern oder Suchmaschinen erstellt und umfassen insgesamt 100.000 Sätze. Ein Subkorpus, das etwa 30 % des Gesamtkorpus ausmacht, wurde exklusiv für diese Arbeit zur Verfügung gestellt. Dabei wurde ausschließlich die koreanischen Datensatz für die vorliegende Arbeit verwendet.

Aufgrund der Nutzung authentischer Gesprächssituationen ist zu erwarten, dass das Korpus als Datenmaterial für die vorliegende Studie geeignet ist. Zum einen ist die Nutzung eines gesprochensprachlichen Korpus nötig, da sich das gesprochene Koreanisch durch die häufige Auslassung von Satzbestandteilen wie Subjekten auszeichnet, was im geschriebensprachlichen Kontext weniger verbreitet ist (vgl. Noh Dae Kyu 1996; Kim Mi-Kyung 2002; Mok Jungsoo 2011). Zum anderen ist das Setup der Gespräche geradezu prädestiniert für die Frage nach der Resolution von Nullsubjekten, da in interaktiven Gesprächen oft Sprecher und Adressat nicht explizit genannt werden, weil sie als bereits bekannte Informationen vorauszusetzen sind (siehe Kapitel 2.1.1). Gerade in Bezug auf die diskursteilnehmerbezogenen Subjekte ist das Korpus also durchaus für die Analyse geeignet. Ein zusätzlicher Vorteil dieses Korpus besteht darin, dass keine Textvorverarbeitungsschritte wie eine Bereinigung und Normalisierung erforderlich sind, da es sich nicht um ein transkribiertes Korpus handelt.

Bei der Generierung der annotierten Daten wird in dieser Arbeit ein speziell entwickeltes Annotationsschema verwendet, da die bestehenden Annotationsschemata die Herausforderungen bei der Identifizierung und Klassifikation koreanischer Nullsubjekte aufweisen (siehe Kapitel 3.2.1). Dies unterstreicht die Notwendigkeit einer Neuauflage solcher Annotationsprogramme in dreierlei Hinsicht, wie die folgenden Fragen verdeutlichen sollen. Es gilt für eine erfolgreiche Annotation zu klären:

1. Welche Elemente sollen annotiert werden?
2. Wie sollten die Nullsubjekte annotiert werden, um die Zuverlässigkeit und Genauigkeit der Annotationen sicherzustellen?
3. Nach welchen Kriterien sollten die Nullsubjekte klassifiziert werden, um ein klares Verständnis der verschiedenen Arten von Nullsubjekten im Koreanischen zu ermöglichen?

Das für die vorliegende Studie ausgewählte Annotationstool ist die Software MMAX2¹⁴. MMAX2 ist eine Software für multimodale Annotation in XML (MMAX), die von der EML Research gGmbH speziell für Koreferenzannotationen entwickelt wurde. Die Entscheidung für dieses Tool basiert auf seiner hohen Anpassungsfähigkeit und Flexibilität hinsichtlich der Annotationsschemata.

Die in dieser Studie erzeugten annotierten Daten dienen zwei wesentlichen Zwecken. Erstens werden die Prädikate, bei denen in den annotierten Daten Nullsubjekte identifiziert wurden, als Input für das entwickelte NRS verwendet. Hierbei kommen heuristische Regeln zum Einsatz, die das Grundgerüst des NRS bilden. Zweitens werden die annotierten Daten genutzt, um die Leistung des NRS zu evaluieren. Dies erfolgt durch den Vergleich der annotierten Referenten oder Antezedenzen der Nullsubjekt-Prädikate in den annotierten Daten mit den Referenten, die durch die heuristischen Regeln

spezialisiert hat. Für weitere Informationen über das Institut vgl. die offizielle Website unter <https://www.etri.re.kr/eng/main/main.etri>.

¹⁴ <https://mmax2.net/>

des NRS zugewiesen werden. Die folgende Abbildung veranschaulicht die Funktionsweise des NRS in dieser Studie.

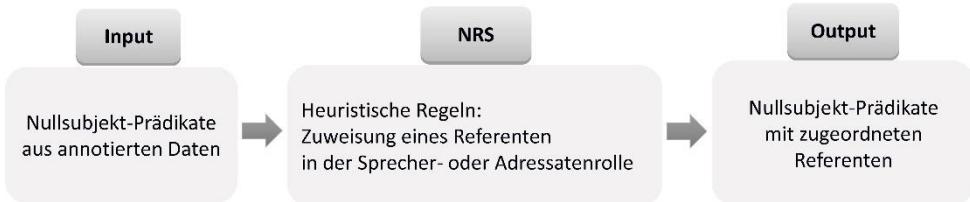


Abbildung 2: Flussdiagramm des regelbasierten Nullsubjektresolutionssystem in dieser Studie

Neben der Bewertung der Leistung des NRS zur Überprüfung der Hypothese dieser Studie ist auch die Effektivität des NRS von Interesse. Um die Wirksamkeit der in dieser Arbeit vorgeschlagenen Resolutionsstrategie für die Nullsubjekte objektiv beurteilen zu können, ist ein Vergleich mit der Leistung von Resolutionsstrategien anderer Studien von entscheidender Bedeutung. Die Resolutionsstrategie von Hong Munpyo (2011) (siehe Kapitel 2.2.1) erscheint dabei als die am besten geeignete Vergleichsgrundlage. Ein direkter Leistungsvergleich dieser Resolutionsstrategien gestaltet sich jedoch als herausfordernd, da die Art und Menge der verwendeten Daten zur Bewertung der Resolutionsstrategien unterschiedlich sind. Während Hong Munpyo (2011) seine Strategie anhand von 207 Nullsubjekten aus Dramenskripten evaluiert, basiert das NRS der vorliegenden Arbeit auf 15.883 Nullsubjekten aus simulierten Reisegesprächen. Um einen möglichst objektiven Vergleich der Leistung beider Resolutionsstrategien zu ermöglichen, wird die von Hong Munpyo (2011) verwendete Strategie in dieser Studie als Basissystem (im Folgenden BS) implementiert. Durch die Bewertung, welche der beiden Resolutionsstrategien – NRS oder BS – besser geeignet ist, um 15.883 Nullsubjekte im gleichen Analysekörper wiederherzustellen, kann eine solide Grundlage für die Bewertung und Auswahl der am besten geeigneten Strategien zur Resolution der Nullsubjekte für die koreanische Sprache gebildet werden.

Die Glossierung sämtlicher Beispiele in dieser Studie erfolgt gemäß der *Leipzig Glossing Rule*. Bei der Transkription des Koreanischen wird die *Revised Romanization of Korean* (RR) angewendet. Die Entscheidung für die RR anstelle der weit verbreiteten *McCune-Reischauer Romanization* (MR) basiert darauf, dass die RR im Gegensatz zur MR keine Sonderzeichen wie Apostrophe verwendet, was die Anwendung vereinfacht. Die *Yale-Romanization* (Yale) wird ebenfalls nicht verwendet, da die RR die Standardaussprache der koreanischen Sprache darstellen kann, während die Yale-Romanisierung lediglich Buchstaben und nicht den Klang wiedergibt, da es sich um eine Transliteration handelt. Eom Iksang (2013: 4–12) führte eine Umfrage unter 180 im Ausland lebenden koreanischen Expert:innen durch, um festzustellen, welches Romanisierungssystem sie für die koreanische Sprache verwendeten und aus welchen Gründen. Die Umfrageergebnisse zeigen, dass die MR das von den Expert:innen am häufigsten verwendete Romanisierungssystem war, gefolgt von der RR, während nur sehr wenige Yale verwendeten. Diejenigen, die sich für die RR entschieden haben, gaben als Vorteil an, dass es leicht zu erlernen sei, die koreanische Orthografie besser widerspiegele als andere Systeme und dass es sich um ein von der Regierung zertifiziertes Romanisierungssystem handelt.

1.3. Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sechs Hauptkapitel. Nach der Einleitung wird in Kapitel 2 die Forschungsliteratur referiert, um anhand der dort gewonnenen Erkenntnisse eine Arbeitshypothese zu den Funktionen der Nullsubjekte im Koreanischen zu erarbeiten. Das Kapitel beginnt mit der Einführung von zwei Arten von Nullsubjekten und der Darstellung von Häufigkeitsverteilungen der Nullsubjekte entsprechend ihres Typs. Bei dieser Betrachtung wird darauf hingewiesen, dass verschiedene Arten von Nullsubjekten unterschiedliche Resolutionsstrategien erfordern (Kapitel 2.1). Der Forschungsstand zu den Resolutionsstrategien wird eingehend betrachtet, insbesondere im Kontext der diskursteilnehmerbezogenen Nullsubjekte. Dabei stehen die Vorteile einer regelbasierten Resolutionsstrategie im Vordergrund (Kapitel 2.2).

In Kapitel 3 wird die Datengrundlage für die Generierung der Annotationsdaten dieser Studie beschrieben. Zunächst werden die Verarbeitungsschritte der Rohdaten für die Annotationen vorgestellt (Kapitel 3.1). Anschließend werden zwei Ansätze bestehender Annotationsschemata für koreanische Nullelemente präsentiert. Danach wird das entwickelte Annotationsschema dieser Studie mit seinen Vorteilen vorgestellt (Kapitel 3.2.1). Daraufhin wird der Annotationsprozess mit dem entwickelten Annotationsschema detailliert beschrieben (Kapitel 3.2.2), woraufhin abschließend die Charakteristika der annotierten Daten dargestellt und ihre Merkmale mit anderen Korpora in früheren Arbeiten verglichen werden, um die gewonnenen Erkenntnisse über koreanische Nullsubjekte zu kontextualisieren (Kapitel 3.3).

In Kapitel 4.1 und Kapitel 4.2 werden verschiedene linguistische Mittel und ihre Kombinationen vorgestellt, die bei der Resolution von diskursteilnehmerbezogenen Nullsubjekten herangezogen werden können sowie darauf basierende Regeln eingeführt. Es wird erklärt, wie sie dazu beitragen, Referenten in der Sprecher- oder Adressatenrolle zu bestimmen. In Kapitel 4.3.1 und 4.3.2 werden ergänzende heuristische Regeln eingeführt, die nicht nur dazu dienen, Nullsubjekte in Nebensätzen wiederherzustellen, sondern auch Prädikate auszuschließen, wenn ihre Nullsubjekte eher in der 3. Person interpretiert werden. Anschließend wird die Anpassung der entwickelten Regeln im NRS an ihre Anwendungsreihenfolge beschrieben, um Konflikte zwischen den Regeln zu vermeiden und sicherzustellen, dass die Zielreferenten korrekt zugeordnet werden (Kapitel 4.3.3). Zudem wird der Output des NRS vorgestellt, der als wichtige Ressource zur Bewertung der Systemeffektivität dient (Kapitel 4.4).

Kapitel 5 liefert detaillierte Angaben zur Evaluierung des Resolutionserfolges, wobei der Schwerpunkt auf einer umfassenden Leistungsbewertung des NRS liegt. Zunächst steht dabei die Leistung der Regeln im NRS im Fokus, um die Hypothesen dieser Arbeit zu überprüfen (Kapitel 5.1). Die Bewertung der Vorhersagengenauigkeit der Regeln wird Aufschluss darüber geben, wie präzise die expliziten Hinweise zu einer erfolgreichen Resolution der Nullsubjekte mit Bezug auf die Diskursteilnehmer:innen beitragen. Des Weiteren wird der Anteil der identifizierten Nullsubjekte durch die Regeln Aufschluss darüber geben, inwiefern diese Hinweise das Phänomen des diskursteilnehmerbezogenen Nullsubjekts erklären können (Kapitel 5.1.1). Anschließend folgt eine detaillierte Untersuchung der einzelnen Regeln, um ihren Beitrag zur Resolutionsgenauigkeit zu bewerten und Schlüsselparameter für die erfolgreiche Resolution von Nullsubjekten mit Bezug auf Diskursteilnehmer:innen zu untersuchen. Zudem wird analysiert, ob die Kombination der Parameter eine wichtige Rolle bei der Resolution spielt (Kapitel 5.1.2 and Kapitel 5.1.3). In Kapitel 5.1.4 wird die Anzahl der Parameter abgeschätzt, die im Koreanischen relevant sind, um Nullsubjekte mit Referenz auf Diskursteilnehmer:innen zu identifizieren. Diese Abschätzung erfolgt durch die Analyse der genutzten und ungenutzten Regeln im NRS. In Kapitel 5.2 wird die Leistungsanalyse des NRS durch einen Vergleich mit dem BS erweitert, um die Effektivität der Resolutionsstrategie des NRS objektiv zu bewerten. Dies

umfasst eine detaillierte Untersuchung der Resolutionsstrategien beider Systeme und ihrer jeweiligen Leistung bei der Identifizierung und Resolution von Nullsubjekten in verschiedenen Kontexten.

Im abschließenden Kapitel 6 werden die wichtigsten Erkenntnisse dieser Arbeit zusammengefasst und ein Ausblick auf verwandte Bereiche gegeben sowie offene Fragen angesprochen.

2. Forschungsstand zu koreanischen Nullsubjekten

Im folgenden Kapitel werden zunächst die beiden unterschiedlichen Typen von Nullsubjekten vorgestellt, die in dieser Arbeit als *anaphorischer* und *deiktischer Gebrauch* bezeichnet werden (Kapitel 2.1). Dabei wird erläutert, welche Funktionen anaphorische und deiktische Nullsubjekte im Koreanischen in Bezug auf ihre Resolution haben und welche Verwendungsmotivationen infragekommen. Diese Unterscheidung legt offen, dass je nach Nullsubjekttyp unterschiedliche Resolutionsstrategien erforderlich sind (Kapitel 2.1.1). Bereits in früheren Studien wurde die Verteilung der koreanischen Nullsubjekte nach ihrem Typ beschrieben. Durch das Hinzuziehen dieser Arbeiten kann der Anteil der anaphorischen und deiktischen Nullsubjekttypen an den gesamten koreanischen Nullsubjekten abgeschätzt werden (Kapitel 2.1.2). Anschließend wird näher darauf eingegangen werden, welche Resolutionsstrategie für anaphorische Nullsubjekte im Koreanischen vorgeschlagen wird. Dadurch ist eine Aussage darüber möglich, welche Methoden wesentlich zur Resolution von koreanischen anaphorischen Nullsubjekten beitragen (Kapitel 2.1.3.1). Der Fokus liegt weiterhin auf den diskursteilnehmerbezogenen Nullsubjekten, deren Rekonstruktion im Koreanischen auf Satzebene geschehen muss (Kapitel 2.1.3.2). Abschließend wird der Forschungsstand zu den Resolutionsstrategien von Nullsubjekten unter Einbeziehung der vorgenannten diskursteilnehmerbezogenen Nullsubjekte festgehalten. Dabei ist zu bemerken, dass es nur sehr wenige diesbezügliche Studien gibt, obwohl die diskursteilnehmerbezogenen Nullsubjekte einen großen Anteil an den gesamten koreanischen Nullsubjekten ausmachen dürften und eindeutige Hinweise zur Identifizierung ihrer Referenten liefern. Zuletzt wird mit der vorliegenden Arbeit für eine regelbasierte Resolutionsstrategie koreanischer Nullsubjekte plädiert, deren Herleitung und Vorteile Thema sein werden (Kapitel 2.2).

2.1. Deiktische und anaphorische Nullsubjekte im Koreanischen

Im Folgenden stehen die Gründe für das Auftreten der beiden Nullsubjekttypen sowie die damit zusammenhängenden Strategien zur Resolution im Mittelpunkt.¹⁵

2.1.1. Definitionen der Nullsubjekttypen

Nullsubjekte in anaphorischen Beziehungen beziehen sich auf ein linguistisches Element, das vorgenannt ist und als *Antezedens* bezeichnet wird.¹⁶ Um eine korrekte Zuordnung zwischen Anapher und Antezedens vornehmen zu können, ist eine Resolution der Verweisstruktur, d. h. eine sogenannte Anaphernresolution notwendig (vgl. Mitkov 1999: 2). Das Ziel einer Anaphernresolution ist es, das korrekte Antezedens einer Pro-Form zu finden. Nullsubjekte, die als Topik-Drop beschrieben werden, werden auch als anaphorische Verweise klassifiziert.

Die Funktion von Anaphern besteht bekanntermaßen darin, die Kohäsion eines Diskurses zu bewahren. Zur Erklärung der Funktion anaphorischer Nullpronomen wird weithin die Centering-Theorie (CT), die die lokale Kohärenz von Äußerungen in einem Diskurs modelliert, verwendet (Grosz et al. 1983; Grosz & Sidner 1986 u. a.). Die CT geht davon aus, dass es ein salientes Diskuselement gibt, das tendenziell

¹⁵ Für die ausführliche Beschreibung und Beispiele für andere Arten koreanischer Nullsubjekte vgl. Han Na-Rae (2006), insbesondere Kapitel 2.4 „Korean Zero Pronouns by Reference Types“.

¹⁶ Für detailliertere Begriffsdefinitionen in Bezug auf die Anapher, vgl. Mitkov (1999), insbesondere Kapitel 1.1 „Basic notions and terminology“.

in einer folgenden Äußerung durch overte Pronomen (z. B. im Englischen) oder Nullpronomen (z. B. im Koreanischen) realisiert wird. Aus diesem Grund wurde in mehreren Studien versucht, koreanische Nullpronomen im Rahmen der CT zu erklären (Kim Mi-Young 1994; 2003; Ryu Byungryul 2001; Han Na-Rae 2006).¹⁷

Das folgende Beispiel (5) ist eine Variante des Diskurses (1) in Kapitel 1.1 (zitiert aus Oh Sun-Young 2007: 55), bei der alle Nullsubjekte explizit realisiert werden.

- (5) A1. *Sinae_i-neun eotteo-ass-nayo-?*
 Sinae-CONT¹⁸ wie-PST-SFS.INT-SF
 'Wie ist es Sinae ergangen?'
- A2. *Sinae_i bab gat-eun geos jal meog-eoss-nayo-?*
 Sinae.SBJ Reis so wie-NMZ gut essen-PST-SFS.INT-SF
 'Hat sie gut gegessen?'
- B1. *geu ae_i jal meog-seubnida-.*
 das Kind.SBJ gut essen-SFS.DECL-SF
 'Sie isst gut.'

In A2 und B1 werden zwei Realisierungsformen anstelle von Nullsubjekten verwendet, um die Referentin *Sinae* in A1 zu bezeichnen. In A2 wird der Eigenname *Sinae* wiederholt, während in B1 das Pronomen *geu ae*¹⁹ verwendet wird. Die Verwendung von NPs wie in A2 und B1 ist pragmatisch möglich, aber Nullsubjekte wie in Beispiel (1) werden im Koreanischen eher bevorzugt (vgl. Oh Sun-Young 2007: 56).

Kim Mi-Young (2003) argumentiert, dass das Topik eines Diskurses im Koreanischen höchstwahrscheinlich das Zentrum des Diskurses ist, da es im Allgemeinen als die salienteste Entität in einem Diskurs angesehen wird. Da im Koreanischen die Nullform gegenüber der overten Form bevorzugt wird, um das Zentrum des Diskurses zu bewahren, ist zu erwarten, dass die Nullsubjekte gegenüber den Overtsujekten bevorzugt werden, um in den folgenden Äußerungen auf das eingeführte Diskurstopik zu referieren.

Die Referenz deiktischer Ausdrücke ergibt sich ausschließlich aus der jeweils pragmatisch situierten Sprechsituation heraus. In diesem Zusammenhang ist die Kontextabhängigkeit eine der charakteristischen Eigenschaften deiktischer Nullsubjekte. Ihr Referent kann ein Objekt oder eine dritte Person sein, die in der Szene des Diskurses anwesend ist, oder es handelt sich, wie im Koreanischen eher üblich, um Diskursteilnehmer:innen einschließlich des Sprechers und des Adressaten (vgl. Han Na-Rae 2006: 51).

¹⁷ Wie die CT zur Resolution der anaphorischen Nullsubjekte im Koreanischen verwendet wurde, wird in Kapitel 2.1.3.1 ausführlich erläutert.

¹⁸ Wie in Fußnote 7 in Kapitel 1.1 kurz erwähnt, wird das postpositionale Suffix *neun* nicht nur zur Markierung von Topik, sondern für eine pragmatische Bedeutung wie Kontrast verwendet. Ihre Funktion muss also je nach dem Kontext des Diskurses interpretiert werden. Daher wird das Suffix *neun* in (5) entsprechend ihrer Funktion in der Diskurssituation als Kontrasttopik (CONT) glossiert, im Gegensatz zur Glossierung von *neun* als Topik (TOP) in (1) in Kapitel 1.1 nach der traditionellen Ansicht.

¹⁹ Auch wenn *gue ae* als *das Kind* glossiert ist, wird diese Phrase im Koreanischen pronominal für die 3. Person gebraucht und deshalb im Deutschen mit *sie* übersetzt. Die Phrase besteht aus dem Demonstrativpräfix *gue* 'das' und dem Nomen *ai* 'Kind'. Auch eine abgekürzte Form *gyae* für *gue ai* ist laut Studien häufig in gesprochenen koreanischen Daten zu finden ist, um auf die 3. Person zu referieren (s. Kim Ahrim (2018) für eine detaillierte Diskussion).

Während anaphorische Nullsubjekte in Hinblick auf die lokale Kohärenz in einem Diskurs erklärt werden, sind deiktische Nullsubjekte unter Bezugnahme auf die Informationsstruktur zu verstehen. Chafe (1976: 31–32) argumentiert, dass Sprecher und Adressat, die bereits im gegenseitigen Bewusstsein vorhanden sind, als alte Informationen behandelt werden. Diese aktive Information gilt als gegebene Information und wird im Englischen als unbetontes Pronomen und im Koreanischen als Nullpronomen realisiert (vgl. Kim Mi-Young 2003: 107). Da Sprecher und Adressat immer die wesentlichen Entitäten sind, die sich im Bewusstsein der Diskurteilnehmer:innen befinden, ist es verständlich, dass diskurteilnehmerbezogene Nullsubjekte im Koreanischen häufig als deiktische Nullsubjekte zu finden sind, wie auch Han Na-Rae (2006) feststellt.

Lee Narah (2014) analysiert die Verteilungen von Nullsubjekten und overten Subjekten in Bezug auf die Diskurteilnehmer in 1. und 2. Person Singular und Plural in natürlichen Gesprächen, die aus zwei koreanischen Korpora²⁰ extrahiert wurden.

Person	Nullsubjekt	Overtes Subjekt	Total
1. <i>(Sprecher)</i>	70,74 % ²¹ (n=701)	29,26 % (n=290)	100,00 % (n=991)
	79,35 % (n=369)	20,65 % (n=96)	100,00 % (n=465)

Tabelle 2: Verteilung der Nullsubjekte und overten Subjekten in der 1. und 2. Person (nach Lee Narah 2014: 150)

Tabelle 2 zeigt, dass die Referenz auf die 1. Person in den Korpora doppelt so häufig geschieht wie die Referenz auf die 2. Person. Bei den Subjekten der 1. Person wird das Subjekt häufiger ausgelassen (in rund 71 % der Belege) als explizit ausgedrückt. Im Fall der Subjekte in der 2. Person macht das Nullsubjekt etwa 79 % der Gesamtbelege aus, womit die overteten Subjektreferenzen noch stärker in der Minderheit sind. Basierend auf der Verteilung der Subjekttypen, die sich auf die Diskurteilnehmer beziehen, argumentiert Lee Narah (2014), dass Nullsubjekte im Koreanischen im Vergleich zu overteten Subjekten die unmarkierte Option sind. Darüber hinaus erklärt sie, dass die Verwendung von overteten Subjekten zur Referenz auf die Diskurteilnehmer eine zusätzliche pragmatische Funktion übernimmt, die der Sprecher hinzufügen möchte.

Zur Illustration dieses pragmatischen Zusatzes sind im folgenden Diskurs (zitiert aus Lee Narah 2014: 146) sowohl Nullsubjekte als auch overtete Subjekte enthalten, so dass ihre Unterschiede deutlich erkennbar sind. Dieser Diskurs besteht aus einem lockeren Gespräch zwischen dem Senior A und einem Junior B an einer Universität. Die deiktischen Nullsubjekte in (6) wurden mit Δ markiert, um sie von anaphorischen Nullsubjekten zu unterscheiden.

²⁰ Das erste Korpus, das von Lee Narah (2014) verwendet wurde, ist das "Sejong spoken corpus". Dabei handelt es sich um ein umgangssprachliches Korpus mit etwa 4,2 Millionen Wörtern, das vom *Nationalen Institut für koreanische Sprache* zwischen 1997 und 2006 gesammelt und zugänglich gemacht wurde. Das zweite Korpus ist das "CallFriend Korean corpus", das von der amerikanischen Forschergruppe *Linguistic Data Consortium* zugänglich gemacht wurde. Dieses Korpus enthält 60 Transkriptionsdateien, die auf Aufzeichnungen von Telefongesprächen zwischen koreanischen Sprecher:innen beruhen.

²¹ In der vorliegenden Arbeit werden Primzahlen in den Tabellen auf drei Dezimalstellen gerundet.

- (6) A1. Δ *bab meog-eoss-eo-?*
 (du.SBJ) Reis essen-PST-SFS-SF
 'Hast du gegessen?'
- B1. *ani-, Δ an meog-eoss-eoyo-*
 Nein-NF (ich.SBJ) nicht essen-PST-SFS-SF
 'Nein, ich habe nicht gegessen.'
- A2. *na-n²² bab meoggo o-ass-neunde-*
 ich.SBJ-CONT Reis essen-PST-SFS-SF
 'ICH habe gegessen.'

In den ersten beiden Äußerungen werden die Subjekte nicht realisiert. Das Nullsubjekt in A1 bezieht sich auf den Adressaten und das in B1 auf den Sprecher. Nur das Subjekt in der letzten Äußerung wird realisiert und bezieht sich auf den Sprecher. Der Sprecher verwendet in A2 das overte Subjekt, um auf sich selbst zu referieren, und den Gegensatz zum Adressaten zu betonen (hier: Er hat gegessen, der Adressat nicht). Diese kontrastierende Bedeutung wurde in der deutschen Übersetzung mit 'ICH' wiedergegeben. Die Subjektposition in A2 wird jedoch nicht mit einem reinen Pronomen, sondern einer Verschmelzung des Pronomens mit dem präpositionalen Suffix *neun* gefüllt. Entgegen den traditionellen Ansichten handelt es sich bei der Funktion des Suffixes in diesem Kontext allerdings nicht um einen Topik-Marker, sondern um die Markierung von Kontrastfokus. In den früheren Studien wurde erklärt, dass das Suffix *neun* die abstrakte Bedeutung *habe*, den Bereich der aufgerufenen Objekte für die Diskursteilnehmer:innen zu begrenzen, und daher eine kontrastive Implikation aufweise (Sung Kwang Soo 1978; Choi Suk-Jae 2013; Park Chul-woo 2015; Han Song Hwa 2017). Durch die Verwendung des overten Subjekts in Kombination mit dem Suffix *neun* in A2 kann der Sprecher also betonen, dass er derjenige ist, der bereits gegessen hat, im Gegensatz zum Adressaten.

Bisher wurde erklärt, dass Nullsubjekte im Koreanischen häufiger als overte Subjekte verwendet werden, um auf Diskursteilnehmer:innen zu verweisen, weil sie als aktive Referenten oder alte Informationen gelten. Eine weitere Erklärung ist laut Studienergebnissen, dass explizite Hinweise zur kommunikativen Funktion oder zu semantischen Kategorien wie Satzmodus oder Modalität über Verbalsuffixe auf der Satzebene die verbreiteten diskursteilnehmerbezogenen Nullsubjekte im Koreanischen ermöglichen (Hong Min-pyo 2000; Kim Mi-Young 2003; Han Na-Rae 2006; Hong Munpyo 2011; Park Arum & Hong Munpyo 2015).

Das folgende Gespräch (7) stammt aus Yun Sun Young (2016: 63). Die Autorin erklärt, dass es sich um ein sehr typisches Gespräch handelt, das in einem koreanischen Restaurant stattfinden kann und dabei die natürliche Satzstruktur im Koreanischen zeigt. Im ersten Satz sind Subjekt und indirektes Objekt nicht ausgedrückt, während im zweiten Satz sowohl das Subjekt als auch das indirekte Objekt und sogar das direkte Objekt weggelassen wurden. Als eines der Merkmale des Satzbau im Koreanischen erläutert Herrmann (1994), dass insbesondere in der gesprochenen Sprache alle Satzglieder außer dem Prädikat weggelassen werden können, wenn sie sich aus dem Kontext erschließen lassen (vgl. Herrmann 1994: 70).

²² *nan* ist eine abgekürzte Form von *na-neun*, die sich aus dem Pronomens *na* 'ich' und dem postpositionalen Suffix *neun* zusammensetzt.

- (7) A. $\Delta \quad \Delta \quad mul_i \quad jom \quad ju\text{-}seyo\text{.}$
 (Sie.SBJ) (mir.IO) Wasser.DO bitte geben-SFS.IMP.HON-SF
 'Geben Sie mir bitte Wasser.'
- B. $ye\text{-}, \quad \Delta \quad \Delta \quad \emptyset_i \quad deuli\text{-}lgeyo\text{.}$
 ja-NF (Ich.SBJ) (Ihnen.IO) (Wasser.DO) geben.HON-SFS(Wille)-SF
 'Ja, ich werde Ihnen Wasser geben.'

In (7) bittet der Kunde A um Wasser, woraufhin der Kellner B sagt, er werde dem Kunden Wasser bringen. In der Äußerung von A wurde das Subjekt des Imperativsatzes, das sich auf den Adressaten bezieht, weggelassen. In koreanischen Imperativsätzen ist die Auslassung des Subjekts nicht zwingend, aber im Allgemeinen wird das Subjekt nicht syntaktisch realisiert (vgl. Park Young Me 2005: 184; Lee Ji Soo 2016: 89). Eine Besonderheit von Imperativsätzen im Koreanischen ist, dass der Imperativ am Verbalsuffix *seyo* im Prädikat zu erkennen ist. Dadurch können die Referenten der Nullsubjekte in Imperativsätzen anhand von Imperativmarkern im Prädikat als Adressaten wiederhergestellt werden. In der Äußerung von B drückt der Sprecher in seiner Antwort auf die Bitte von A seinen Willen aus, Wasser zu bringen. Das Nullsubjekt der Äußerung bezieht sich auf den Sprecher und kann durch das Verbalsuffix *lgeyo* im Prädikat identifiziert werden. Dies liegt daran, dass die Absicht des Sprechers im Koreanischen durch ein Verbalsuffix im Prädikat einer Äußerung erkannt werden kann. Das Wort *lgeyo* in der Äußerung von B ist eines der typischen Verbalsuffixe, mit denen ein Sprecher seinen Willen auszudrücken kann.

Hong Min-pyo (1991) weist darauf hin, dass morphologische Informationen, die den Satzmodus oder Sprechakte anzeigen, direkte Hinweise sein können, um diskursteilnehmerbezogene Nullsubjekte rekonstruieren zu können, die mit der CT nicht rekonstruierbar sind.

2.1.2. Häufigkeitsverteilungen der Nullsubjekttypen

Im Folgenden stehen die Häufigkeitsverteilungen der anaphorischen und deiktischen Nullsubjekte im Vergleich mit anderen Nullsubjekttypen im Mittelpunkt der Betrachtung. Da die herangezogenen Studien unterschiedliche Klassifikationsparameter für die Nullsubjekte benutzen, muss analysiert und abgeleitet werden, welche Typen den anaphorischen und deiktischen Nullsubjekten entsprechen könnten.

Zunächst teilt Hong Min-pyo (2000) die in den Daten²³ gefundenen koreanischen Nullsubjekte in fünf Kategorien ein, je nach ihren Referenztypen und ihrer Verfügbarkeit im Kontext. Er unterscheidet

1. solche, die sich auf Diskursteilnehmer beziehen (DP: „Discourse Participants“)
2. solche, die sich auf zuvor erwähnte Entitäten beziehen (PM: „Previously Mentioned Entities“),
3. solche, die sich auf die allgemeine Öffentlichkeit beziehen (GP: „General Public“),
4. solche, die sich auf Entitäten beziehen, die außersprachlich zu rekonstruieren sind (EL: „Extra-Linguistically Recoverable“), und
5. andere.

Die folgende Tabelle zeigt die Verteilung der Nullsubjekte nach ihrer Kategorie, wobei sie in der Rangfolge des größten Anteils aufgeführt sind.

²³ Die von Hong Min-pyo (2000) verwendeten Daten bestehen aus Transkripten eines einstündigen Dialogs, der von öffentlichen Radiosendern aufgezeichnet wurde.

Rang	Kategorie	Anteil
1	DP	33,90 % (n=238)
2	PM	25,78 % (n=181)
3	GP	23,22 % (n=163)
4	EL	16,52 % (n=116)
5	Andere ²⁴	0,57 % (n=4)
Total		100,00 % (N=702)

Tabelle 3: Verteilung der Nullsubjekte nach ihren Kategorien gemäß Hong Min-pyo (2000: 14)

Die am häufigsten vorkommenden Nullsubjekte gehören zur Kategorie DP, d. h. die Diskursteilnehmer:innen sind ihre Referenten. Sie machen etwa 34 % der insgesamt 702 Nullsubjekte aus. Die am zweithäufigsten gefundene Kategorie von Nullsubjekten sind PM, sie machen etwa 26 % der Gesamtzahl aus. Nullsubjekte in dieser Kategorie sind die Fälle von Anaphorik, in denen die Antezedenzen in der vorangehenden Äußerung oder im selben Diskurssegment zu finden sind. Die Nullsubjekte, die den drittgrößten Anteil ausmachen, gehören zur Kategorie GP, in der sich Nullsubjekte auf Referenten der 3. Person beziehen, insbesondere auf verallgemeinernd- oder unbestimmt-persönliche Subjekte. Ihr Anteil an der Gesamtzahl betrug etwa 23 %. Die Nullsubjekte an vierter Stelle gehören zur Kategorie EL, bei der die Referenten der Nullsubjekte nur durch den gesamten Diskurs erschlossen werden können. Sie machen etwa 17 % der Gesamtzahl aus.

Aufgrund der unterschiedlichen Referenztypen von Nullsubjekten betont Hong Min-pyo (2000), dass bei der Resolution der verschiedenen Nullsubjekttypen unterschiedliche Ansätze eingeführt werden müssen. Er argumentiert, dass die CT nur auf diejenigen Nullsubjekte angewandt werden kann, die als Mittel zur Fokussierung auf das zentrale Thema des Diskurses verwendet werden, wie z. B. auf Nullsubjekte in der Kategorie PM. Darüber hinaus nimmt er an, dass Nullsubjekte, die sich auf Diskursteilnehmer:innen beziehen, anders behandelt werden sollten als anaphorische Nullsubjekte, da sie nicht dazu dienen, die lokale Kohärenz eines Diskurses zu erhöhen. Basierend auf der Datenauswertung schlägt er vor, morphologische und lexikalische Informationen zu nutzen, um mit diskursteilnehmerbezogenen Nullsubjekten umzugehen.²⁵

Basierend auf Hong Min-pyos (2000) Erklärungen zu den Nullsubjekten in jeder Kategorie können diejenigen, die zur Kategorie DP gehören, als deiktische Nullsubjekte und die in der Kategorie PM als anaphorische Nullsubjekte klassifiziert werden. In Anbetracht der Verteilungen in Tabelle 3 ist dementsprechend zu erwarten, dass mehr als 60 % der Nullsubjekte zu deiktischen und anaphorischen Nullsubjekten gehören. Darunter haben die deiktischen Nullsubjekte, die Diskursteilnehmer:innen als Referenten haben, den größten Anteil. Auf Grundlage von Hong Min-pyos (2000) Datenauswertung ist anzunehmen, dass im besten Fall bei etwa 34 % der Nullsubjekte die Hinzuziehung von morphologischen und lexikalischen Informationen zur Resolution der Verweisstruktur ausreicht und bei etwa 26 % auf die CT zurückgegriffen werden kann.

Zu einer anderen Verteilung der Nullsubjekttypen kommt Hong Munpyo (2011) bei seiner Analyse diskursbezogener schriftlicher Textsorten. Bei der Analyse von Messenger-Chats und Dramenskripten

²⁴ In Hong Min-pyo (2000) wurden keine Erläuterungen oder Beispiele für den Fall geliefert, der zur letzten Kategorie „Andere“ gehört.

²⁵ Weitere Einzelheiten zu diesen morphologischen und lexikalischen Informationen, die als Anhaltspunkte für die Resolution von diskursteilnehmerbezogenen Nullsubjekten dienen können, werden in Kapitel 2.1.3.2 behandelt.

setzt er drei Typen von Nullsubjekten an, wie sie auch für das Japanische (vgl. u. a. Nakaiwa et al. 1995; Nariyama 2002) unterschieden werden.

Diese Klassifikation setzt die Art der Antezedenzen in den Mittelpunkt:

1. „Intrasentential null subject“ (kurz „Intra-NS“),
2. „Intersentential null subject“ (kurz „Inter-NS“),
3. „Extrasentential null subject“ (kurz „Extra-NS“).

Als Nullsubjekte des Typs Intra-NS sind diejenige Nullsubjekte zu verstehen, deren Antezedens im selben Satz zu finden ist, in dem das Nullsubjekt selbst vorkommt. Als Nullsubjekte des Typs Inter-NS werden diejenige klassifiziert, deren Antezedens an anderer Stelle im Text zu finden ist. Da es bei der Resolution sowohl der Nullsubjekte des Typs Intra-NS als auch des Typs Inter-NS darum geht, die jeweiligen Antezedenzen im selben Satz oder im vorangehenden Texten zu finden, können sie für die vorliegende Arbeit als anaphorische Nullsubjekte eingestuft werden. Als Nullsubjekte des Typs Extra-NS werden wiederum diejenigen Nullsubjekte bezeichnet, deren Referenten durch außersprachliche Faktoren wie den Kontext wiederhergestellt werden können. Laut Nakaiwa et al. (1995: 99) verweisen die Nullsubjekte des Typs Extra-NS auf deiktische Referenten. Aus diesem Grund können sie nach Hong Munpyos (2011) Klassifikationssystem als deiktische Nullsubjekte klassifiziert werden. Hong Munpyo (2011: 423–424) weist darauf hin, dass die koreanischen Nullsubjekte des Typs Extra-NS nicht nur durch extralinguistische Faktoren,²⁶ sondern auch anhand morphologischer und lexikalischer Hinweise rekonstruiert werden können. Da Hong Munpyo (2011) vor allem Vorschläge für die Resolution diskursteilnehmerbezogener Nullsubjekte macht, ist anzunehmen, dass es sich bei den meisten der Nullsubjekte des Typs Extra-NS um diskursteilnehmerbezogene Nullsubjekte handelt.

Tabelle 4 präsentiert die Verteilung der Nullsubjekte in den beiden von Hong Munpyo (2011) verwendeten Datensätzen, basierend auf den zuvor beschriebenen Nullsubjekttypen.

Nullsubjekttyp	Messenger-Chats	Dramaskripte	Total
<i>Extra-NS</i>	84,41 % (n=162)	78,28 % (n=191)	81,14 % (n=353)
<i>Inter-NS</i>	9,94 % (n=19)	19,67 % (n=48)	15,4 % (n=67)
<i>Intra-NS</i>	5,23 % (n=10)	2,05 % (n=5)	3,44 % (n=15)
Total	100,00 % (N=191)	100,00 % (N=244)	100,00 % (N=435)

Tabelle 4: Verteilung der koreanischen Nullsubjekte in den Daten von Hong Munpyo (2011: 422)

In beiden Datensätzen ist der Anteil der Nullsubjekte, die zur Kategorie Extra-NS gezählt werden, am größten, gefolgt von den Nullsubjekten, die zu Inter-NS gehören und zuletzt denjenigen, die als Intra-NS gelten. Zur Kategorie Extra-NS zählen rund 84 % aller Nullsubjekte in den Messenger-Chats und 78 % in den Dramenskripten sowie insgesamt 81 % über beide Textsorten. Dies bedeutet, dass bei einer bedeutsamen Anzahl der koreanischen Nullsubjekte der Referent möglicherweise mithilfe morphologischer und lexikalischer Anhaltspunkte wiederhergestellt werden kann, wenn man mit Hong Min-pyo (2000) und Hong Munpyo (2011) annimmt, dass die Referenten diskursteilnehmerbezogener Nullsubjekte im Koreanischen durch derartige Hinweise bestimmt werden können.²⁷

Die Nullsubjekte der Kategorien Inter-NS und Intra-NS, bei deren Resolution ein Antezedens im Diskurs zu finden sein soll, machen dagegen nur einen Anteil von etwa 19 % der insgesamt 435 Nullsubjekte aus. Daher ist davon auszugehen, dass sowohl Topik-Drop als auch die CT, die die Antezedenzen von

²⁶ Vielmehr ähneln Nullsubjekte, die von Hong Min-pyo (2000) als EL klassifiziert werden, eher den japanischen Extra-NS, da ihre Referenten nur durch die Art oder den Inhalt des gesamten Diskurses abgeleitet werden können und es keine vorangehende NP gibt, die ein Antezedens sein kann.

²⁷ In Kapitel 2.2 werden die Ergebnisse zu Strategien der Nullsubjektresolution in früheren Studien anhand dieser Hinweise im Detail dargestellt.

anaphorischen Nullsubjekten identifizieren konnten, letztendlich nur einen kleinen Teil der koreanischen Nullsubjekte abdecken.

Zusammenfassend lässt sich sagen, dass anaphorische und deiktische Nullsubjekte im Koreanischen laut Hong Min-pyos (2000) Datenanalyse mehr als die Hälfte aller Nullsubjekte ausmachen dürften. Darüber hinaus ist zu erwarten, dass diskursteilnehmerbezogene Nullsubjekte nach der Datenanalyse von Hong Munpyo (2011) zu etwa 81 % auftreten können. Aufgrund dieser Ergebnisse ist anzunehmen, dass eine signifikante Anzahl von Referenten koreanischer Nullsubjekte einfach durch morphologische und lexikalische Hinweise auf der Satzebene wiederhergestellt werden kann.

2.1.3. Resolution der Verweisstrukturen koreanischer Nullsubjekte

Im Folgenden wird anhand einschlägiger Beispiele ausführlicher erläutert, wie die in Kapitel 2.1.1 kurz vorgestellten Strategien zur Wiederherstellung der Referenz bei den zwei besprochenen Arten koreanischer Nullsubjekte beitragen.

2.1.3.1. Anaphorische Nullsubjekte

Die CT ist ein häufig verwendetes Ansatz zur Resolution koreanischer anaphorischer Nullsubjekte. Das Centering-Modell, entwickelt von Grosz, Joshi & Weinstein (1983; 1995), bildet die Grundlage der CT und zielt darauf ab, die Beziehung zwischen lokaler Kohärenz und der Verwendung von verweisenden Ausdrücken zu erfassen. Die Verwendung von Pronomen ist eines der Mittel zur Wahrung der Kohärenz in einem Diskurs, und damit auch die weit verbreitete Nullform im Koreanischen (vgl. Kim Mi-Young 2003: 69). Gemäß dem Centering-Modell ist jeder Äußerung in einem Diskurs einem Set von Diskursentitäten zugeordnet, die Forward-Looking Centers (Cfs) genannt werden. Das Backward-Looking Center (Cb) entspricht wiederum der Diskursentität, auf die sich die Äußerung zentral bezieht und das die aktuelle Äußerung mit dem vorherigen Diskurs verbindet. Das Set der Cfs wird nach der Diskurssalienz geordnet, was als *CF-Ranking* bezeichnet wird. Walker, Iida & Cote (1994) nehmen an, dass die Kriterien für die Bestimmung des CF-Rankings in den einzelnen Sprachen unterschiedlich sind. Es wird angenommen, dass die salienteste Diskursentität das Zentrum der aktuellen Äußerung ist und mit hoher Wahrscheinlichkeit in der folgenden Äußerung im Koreanischen als Nullform realisiert wird (Jang Seokjin 1994; Kim Mi-Kyung 2002; Kim Mi-Young 2003). Unter der Voraussetzung, dass die Diskursentität auf dem höchsten Rang im CF-Ranking mit großer Wahrscheinlichkeit dem Antezedens des Nullsubjekts entspricht, kann das CF-Ranking verwendet werden, um die Verweisstruktur anaphorischer Nullsubjekte wiederherzustellen. Die folgende Tabelle zeigt die CF-Rankings, die in früheren Studien für das Koreanische vorgeschlagen wurden.

CF-Ranking	Literatur
Topik mit Topik-Marker > Subjekt ohne Subjekt-Marker > Objekt ohne Objekt-Marker > andere	No Youngkyoon (1991)
Topik mit Topik-Marker > Subjekt mit Subjekt-Marker > Objekt2 mit Objekt-Marker > Objekt1 mit Objekt-Marker> andere	Ryu Byungryul (2001)
a. Diskurseinleitende Äußerung: Topik mit Topik-Marker > andere gegebene Information > neue Information b. Ab der zweiten Äußerung: gegebene Information (null > Pronomen > definiter Ausdruck > wiederholtes Nomen) > neue Information (Subjekt > Objekt > Lokativ > Obliquus)	Kim Mi-Young (2003)

Tabelle 5: CF-Rankings für Koreanisch in der Literatur

Wie in Tabelle 5 zu sehen ist, ist im Koreanischen ein als Topik markiertes Element die salienteste Diskursentität vor allen anderen grammatischen Argumenten in der Äußerung einschließlich des Subjekts. Dies hängt damit zusammen, dass das Topik für eine topik-prominente Sprache wie das Koreanische als salientestes Element in einem Diskurs angesehen wird (Li & Thompson 1976). Es gibt einige Studien, die Sprecher- und Adressatenkonzepte in die koreanischen CF-Rankings einführen und davon ausgehen, dass sie eine höhere Position als das Topik einnehmen (Choe Jae-Woong & Lee Min-Haeng 1999; Lee Ik-Hwan & Lee Min-Haeng 2000). Die vorliegende Arbeit vertritt dagegen die Ansicht, dass diskursteilnehmerbezogene Nullsubjekte ein anderes Phänomen als anaphorische sind und daher in Bezug auf ihre Resolution anders behandelt werden sollten, eben genau nicht mithilfe der CT (vgl. Jang Seokjin 1994; Hong Min-pyo 2000; Kim Mi-Young 2003). In der vorliegenden Arbeit wird daher davon abgesehen, das Konzept der Diskursteilnehmer in die koreanischen CF-Rankings aufzunehmen.²⁸

Hinsichtlich der folgenden Elemente des CF-Rankings in Tabelle 5 bringen No Youngkyoon (1991) und Ryu Byungryul (2001) leicht unterschiedliche Meinungen zum Ausdruck. So legt No Youngkyoon (1991) fest, dass die hochrangigen Subjekt- und Objekt-NPs „bare NPs“ ohne Kasusmarker sein müssen, während Ryu Byungryul (2001) postuliert, dass sie alle durch postpositionale Suffixe markiert sein müssen.²⁹ Trotz dieser unterschiedlichen Meinungen stimmen die Autoren darin überein, dass das Subjekt salienter ist als das Objekt und das Objekt salienter als alle anderen Elemente.

Im Gegensatz zu den ersten beiden CF-Rankings von No Youngkyoon (1991) und Ryu Byungryul (2001) in Tabelle 5 wendet Kim Mi-Young (2003) informationstheoretische Konzepte auf das koreanische CF-Ranking an. Grundsätzlich geht sie davon aus, dass gegebene Informationen höher in der Hierarchie anzuordnen sind als neue Informationen. Der Grund dafür ist, dass mehrere Äußerungen zu einem kohärenten Diskurs werden, wenn die in der vorangegangenen Äußerung gegebene Information in der aktuellen Äußerung und in den nächsten Äußerungen fortgeführt wird. Entsprechend schlägt Kim Mi-Young (2003) zwei CF-Rankings vor, die auf der Position der Äußerung in einem Diskurs basieren. Am Diskursanfang wird die als Topik markierte NP unter den gegebenen Informationen als das salienteste Element betrachtet. Im Koreanischen kann das Suffix *neun* nicht für neue Informationen verwendet werden, sondern nur für Informationen, die der Adressat bereits kennt (vgl. Kim Mi-Young 2003: 109). Kim Mi-Young (2003) erklärt, dass das Nullelement ab der zweiten Äußerung in einem Diskurs den höchsten Platz im koreanischen CF-Ranking einnimmt. Die Autorin erläutert weiter, dass die Hierarchie der linguistischen Ausdrucksformen bei den gegebenen Informationen, die in der nicht initialen Äußerung im Diskurs stehen, auf der Grundlage der Givenness-Hierarchie von Gundel (1998) und der EXP ORDER von Kameyama (1998) entstanden ist.³⁰ Es ist zu beobachten, dass die beiden von Kim Mi-Young (2003) vorgeschlagenen CF-Rankings die Tendenz widerspiegeln, dass das Nullelement im Koreanischen als Topik in einem Diskurs interpretiert wird und auch als Zentrum zur Verknüpfung von Äußerungen in einem Diskurs fungiert.

²⁸ Da die Erstellung eines CF-Rankings für die koreanische Sprache nicht Gegenstand der vorliegenden Arbeit ist, werden die Vorschläge bisheriger Studien zum koreanischen CF-Ranking hier nur vorgestellt.

²⁹ Die Frage, welches Element als salienter angesehen wird, je nachdem, ob ein postpositionales Suffix angehängt ist, liegt nicht im Interesse der vorliegenden Studie. Generell wird jedoch davon ausgegangen, dass das Suffix *i/ga* zur Anzeige von Subjektargumenten verwendet werden kann, aber auch Bedeutungen wie Exklusivität oder kontrastiver Fokus ausdrückt (Kim Mi-Young 2003; Park Jae-yon 2006; Choi Yoonji 2016; Kim Min-gook 2016 u. a.). Parallel dazu wird erklärt, dass das Suffix *eul/leul* nicht nur die grammatische Funktion von Objekten kennzeichnen kann, sondern auch Bedeutungen wie Aufmerksamkeit oder kontrastierender Fokus trägt. So können Sprecher:innen sich dafür entscheiden, den Marker in Situationen zu verwenden, wenn diese zusätzlichen Bedeutungen benötigt werden (Shin Hyon Sook 1982; Lee Namsoon 1998; Ko Eon-Suk 2000; Lee Hanjung 2006 u. a.). Wenn man bedenkt, dass eine markierte Topik-NP mit dem Suffix *eun/neun* als salienteste Diskursentität im Koreanischen angesehen wird, kann man auch davon ausgehen, dass ein Subjekt oder Objekt mit einem postpositionalen Suffix salienter sein könnte als ohne Suffix.

³⁰ Siehe Kim Mi-Young (2003: 92) für die Details.

Das folgende Beispiel (zitiert aus Yang Jaehyung 1997: 116) zeigt, wie die CF-Rankings verwendet werden, um das Antezedens eines anaphorischen Nullsubjekts zu bestimmen.

- (8) U1. *cheolsui-neun achime bogdo-eseo yeonghui-leul manna-ass-da-*
Chulsoo.SBJ-TOP³¹ am Morgen Flur-auf Younghee-OBJ treffen-PST-SFS-SF
'Chulsoo traf Younghee am Morgen auf dem Flur.'
- U2. *halu jongil gaseum-i dugeungeoli-eo Ø_{i/*j} amu il-do ha-l su eobs-eoss-da-*
den ganzen Tag Herz-SBJ klopften-CNJ nichts-sogar machen-PPC(Möglichkeit)-
PST-SFS-SF
'Sein Herz klopfte den ganzen Tag und er konnte nicht einmal etwas tun.'

Die Sätze U1 und U2 in (8) entsprechen der ersten und zweiten Äußerung, die nacheinander in einem Diskurssegment auftreten. Nach der CT ist das Zentrum der aktuellen Äußerung (U2) mit hoher Wahrscheinlichkeit das hochrangigste Element in der vorherigen Äußerung (U1). Zwei Diskuselemente in U1 können dementsprechend Kandidaten für das CF-Ranking sein: (Cfs: *Chulsoo*, *Younghee*)³². Unter der Annahme, dass als Topik markierte NPs höher in der Hierarchie rangieren als Objekte, ist *Chulsoo* ein salienteres Element als *Younghee*. Daher wird das Nullsubjekt in U2 bevorzugt als *Chulsoo* interpretiert, da er als Zentrum des Diskurses vorausgesetzt wird.

Theoretisch ist es sehr wahrscheinlich, dass das saliente Element in der vorherigen Äußerung das Antezedens eines anaphorischen Nullsubjekts in der aktuellen Äußerung ist. Dies ist jedoch nicht die einzige Interpretationsmöglichkeit für ein Nullsubjekt, sondern nur die präferierte, weshalb die Genauigkeitsrate für die Bestimmung des Antezedens nicht sehr hoch ist, wenn nur die Interpretation mittels des CF-Rankings vorgenommen wird.

Hong Munpyo (2011) führt zur Genauigkeitsbestimmung des CF-Rankings ein kleines Experiment durch, um herauszufinden, wie viele anaphorische Nullsubjekte mit dessen Hilfe rekonstruiert werden können. Insgesamt extrahiert er dafür 46 Nullsubjekte mit Antezedens aus Messenger-Chats und Dramen skripten und suchte nach deren Antezedenz in den vier vorangehenden Sätzen vor dem Nullsubjektsatz. Die Größe des Suchfensters legte er auf Grundlage der Erkenntnis von Okumura & Tamura (1996) fest, dass die Leistung der Anaphernresolution für japanische Nullsubjekte dadurch gesteigert werden kann, dass das Antezedens in den vier vorherigen Sätzen gesucht wird. Nach der Extraktion der NPs in den vorangegangenen vier Sätzen bestimmte er das saliente Element unter den Antezedens-Kandidaten als das Antezedens eines Nullsubjekts. Dabei wurde das CF-Ranking wie folgt angenommen: [Topik-NP > Subjekt-NP > Objekt-NP > Adverb-NP].³³ Abschließend berichtet Hong Munpyo (2011), dass nur bei 30 von insgesamt 46 Nullsubjekten (entspricht 65,22 %) das Antezedens mithilfe des CF-Rankings korrekt gefunden werden konnte. Von den 16 fehlerhaft zugeordneten Nullsubjekten gehörten 10 zu den Fällen, in denen die Antezedenzen der Nullsubjekte in den Sätzen gefunden wurden, die weiter als die vorangegangenen vier Sätze lagen. Bei den restlichen 6 fehlerhaften Zuordnungen handelte es sich um Fälle, in denen das saliente Element nicht das Antezedens des entsprechenden Nullsubjekts war, auch wenn das passende Antezedens in den vorangegangenen vier Sätzen vorkam. Aus den Ergebnissen von Hong Munpyo (2011) lässt sich

³¹ Der Grund, warum das Suffix *neun* als TOP glossiert wurde, ist, dass *Chulsoo* als das Topik des Diskurses beurteilt wurde, weil in den Äußerungen U1 und U2 weiterhin über *Chulsoo* gesprochen wird.

³² Wenn das Element *bogdo* 'Flur' als ein potenzieller Kandidat für das CF-Ranking betrachtet wird, wird es in jedem Fall in der Hierarchie niedriger eingestuft als die beiden Kandidaten *Chulsoo* und *Younghee*. Dies liegt daran, dass es entweder in die Kategorie "andere" gemäß dem CF-Ranking von No Youngkyoon (1991) und Ryu Byungryul (2001) oder in die Kategorie "neue Information" gemäß dem CF-Ranking von Kim Mi-Young (2003) in Tabelle 5 fällt.

³³ Hong Munpyo (2011) erwähnt die Realisierungsform der einzelnen Elemente in Bezug auf die postpositionalen Suffixe nicht.

ableiten, dass die CT allein möglicherweise nicht ausreicht, um anaphorische Beziehungen für koreanische Nullsubjekte zu erklären.

Kim Mi-Young (2003: 126–130) weist darauf hin, dass das CF-Ranking in der CT für die Resolution koreanischer Nullelemente nicht ausreicht, weil das Diskurselement mit der höchsten Hierarchie in den vorangehenden Äußerungen nicht immer dem Antezedens des Nullelements entspricht. Das folgende Beispiel (zitiert aus Kim Mi-Young 2003: 128) zeigt, dass das CF-Ranking nicht erklären kann, warum das Nullelement in U2 als das Element mit dem niedrigeren Rang in U1 interpretiert wird.

- (9) U1. *abeoji-neun yeongsu-ege koteuk-leul sa ju-eoss-da-*.
Vater.SBJ-CONT Youngsu-IO Mantel-DO kaufen-PST-SFS-SF
'Der Vater kaufte Youngsu einen Mantel.'
- U2. *geuligo eomeoni-neun Ø*i/*j/*k moja-leul sa ju-eoss-da-*.
und Mutter.SBJ-CONT ihm Mütze-DO kaufen-PST-SFS-SF
'Und die Mutter kaufte ihm eine Mütze.'

Das Beispiel (9) umfasst zwei Äußerungen in einem Diskurssegment. In Äußerung U1 finden sich drei Diskuselemente, wovon das erste Element *abeogi* 'Vater' höher eingestuft ist als die anderen beiden Elemente, da eine als Topik markierte NP in den CF-Rankings eine höhere Hierarchiestufe einnimmt als ein Objekt. Das Antezedens des Nullelements in U2 ist jedoch nicht *abeogi* 'Vater', sondern der Eigename *Youngsu*, der in U1 ein indirektes Objekt war. Dies zeigt, dass das salienteste Element in der vorherigen Äußerung nicht unbedingt das Antezedens des Nullelements in der nächsten Äußerung sein muss.

Kim Mi-Young (2003) stellt fest, dass die strukturelle Parallelität zwischen U1 und U2 erklären kann, warum das Antezedens des Nullelements in U2 als *Youngsu* interpretiert werden kann. Gemäß der Parallelitätspräferenz wird das Nullelement in U2 bevorzugt als das indirekte Objekt *Youngsu* in U1 identifiziert, da beide Elemente in U1 und U2 das indirekte Objekt zum Verb *sa juda* 'kaufen' sind.

Die Parallelitätspräferenz ist eine Präferenzregel, die verwendet wird, um das am besten geeignete Antezedens unter den Antezedens-Kandidaten für eine Anapher zu finden. Kameyama (1998) führt diese Präferenzregel ein, die auf der Beobachtung beruht, dass zwei benachbarte Äußerungen in einem Diskurs tendenziell grammatisch parallel aufgebaut sind. Im Detail besagt die Regel, dass die Subjekte oder Objekte zweier aufeinanderfolgender Äußerungen dazu neigen, sich auf die gleiche Entität zu beziehen. Die Regel erhöht das Gewicht eines Antezedens-Kandidaten, wenn er die gleiche syntaktische Funktion wie die Anapher hat.

Unter anderem diese Regel hat Kim Mi-Kyung (2003) auf ein Korpus aus Volksmärchen (431 Sätze) und Zeitungstexten (249 Sätze) angewandt, um die beste Resolutionsstrategie herauszuarbeiten. Sie wendet insgesamt zwei wesentliche Resolutionsstrategien an: die Center Continuation (CC) und Parallelism Preference (PP). Die CC folgt der Annahme, dass das Thema einer Äußerung auch das Thema der nächsten sein wird. Die PP folgt der These, dass sich ein Nullsubjekt oder -objekt einer Äußerung auf dieselbe Entität bezieht wie das Subjekt oder Objekt der vorherigen Äußerung. Tabelle 6 zeigt die Ergebnisse von Kim Mi-Kyungs (2003) Analyse im Detail.

Textart	ZP-Typ	Resolutionsstrategie			Total
		CC	PP	andere	
Volksmärchen	Nullsubjekt	77,29 % (n=177)	14,85 % (n=34)	7,86 % (n=18)	100,00 % (N=229)
	Nullobject	13,89 % (n=5)	58,33 % (n=21)	27,78 % (n=10)	100,00 % (N=36)
Zeitungstexte	Nullsubjekt	51,00 % (n=51)	29,00 % (n=29)	20,00 % (n=20)	100,00 % (N=100)
	Nullobject	18,18 % (n=2)	72,73 % (n=8)	9,09 % (n=1)	100,00 % (N=11)

Tabelle 6: Anteil der rekonstruierten Nullsubjekte und -objekte aufgeschlüsselt nach Kim Mi-Kyungs (2003: 135)
Resolutionsstrategie

Was in Tabelle 6 zunächst festgestellt werden kann, ist, dass in beiden Textsorten viel mehr Nullsubjekte als Nullobjecte vorkommen: 229 Nullsubjekte vs. 36 Nullobjecte in den Volksmärchen, 100 Nullsubjekte vs. 11 Nullobjecte in den Zeitungstexten.

In einem ersten Analyseschritt stehen die Nullsubjekte im Vordergrund. In beiden Textarten konnten die Antezedenzen von mehr als der Hälfte der Nullsubjekte mithilfe der CC identifiziert werden, was etwa 77 % bei den Volksmärchen bzw. 51 % bei den Zeitungstexten ausmacht. Das bedeutet, dass bei solchen Nullsubjekten das Zentrum der vorherigen Äußerung das Antezedens des Nullsubjekts der aktuellen Äußerung ist. Mithilfe der PP konnten etwa 15 % (Volksmärchen) bzw. 29 % (Zeitungstexte) der Referenzen der Nullsubjekte wiederhergestellt werden. Die betrifft solche Nullsubjekte, bei denen das Subjekt der vorherigen Äußerung das Antezedens in der aktuellen Äußerung war.

Aus diesen Daten ist herauszulesen, dass selbst mit nur zwei Präferenzregeln (CC und PP) etwa 92 % der Nullsubjekte im Volksmärchen und etwa 80 % der Nullsubjekte in Zeitungstexten korrekt rekonstruiert werden konnten, auch wenn es weitere anaphorische Nullsubjekte gibt, die nicht anhand dieser Regeln interpretiert werden können. Dieses Ergebnis kann dahingehend interpretiert werden, dass ein signifikanter Anteil der koreanischen anaphorischen Nullsubjekte tatsächlich das Diskurszentrum oder die Subjekt-NP der vorherigen Äußerung ist.

Die Anwendung der PP zeigt in Bezug auf die Identifizierung der Antezedenzen von Nullobjecten in beiden Textarten eine höhere Erfolgsrate als die Verwendung von CC. Dies bedeutet, dass in diesem kleinen Datensample die Fälle, in denen das Objekt der vorherigen Äußerung das Antezedens des Nullobjects der aktuellen Äußerung war, häufiger sind als die Fälle, in denen das Zentrum der vorherigen Äußerung das Antezedens des Nullobjects der aktuellen Äußerung war. Auch wenn die Auswertung der Resolutionsstrategien für die Identifizierung der Referenten von Nullobjecten nahelegt, dass die PP bessere Erfolgsaussichten liefert als die CC, ist das Datensample für eine belastbare Aussage zu klein. Im Fall der Nullsubjekte ist dagegen eindeutig zu sehen, dass die CC in höherem Maße zur Resolution der Anaphern beiträgt als die PP, was bedeutet, dass das salienteste Element in der vorherigen Äußerung mit hoher Wahrscheinlichkeit das Antezedens des folgenden Nullsubjekts ist.

Während Kim Mi-Kyung (2003) versucht hat, den Nutzen von CC und PP für die Resolution koreanischer anaphorischer Nullsubjekte mit einer regelbasierten Methode nachzuweisen, wählt Han Na-Rae (2006) ein statistisches Modell. Dafür erstellt sie ein statistisch trainiertes Modell unter Verwendung von Merkmalen, die häufig als entscheidend für die Anaphernresolution diskutiert werden. Zu diesen Merkmalen gehören u. a. Informationen wie die Topikalität, die grammatische Funktion und die strukturelle Parallelität. Im Detail unterscheidet Han Na-Rae (2006: 189) die Fälle:

1. Das Nullsubjekt und die NP haben die gleiche grammatische Funktion.
2. Die NP ist die am nächsten links vom Nullsubjekt stehende, als Topik markierte Einheit („Nearest Topic“)
3. Die NP ist die am nächsten links vom Nullsubjekt stehende Subjekt-NP („Nearest Subject“)

Die oben genannten Merkmale erfassen die Beziehung zwischen einem Nullsubjekt und einer Antezedenz-NP. Das erste Merkmal ist mit der PP von Kim Mi-Kyung (2003) zusammenzubringen, da sowohl ein Nullelement als auch sein Antezedenz in der Regel dieselbe grammatische Funktion wie Subjekt oder Objekt erfüllen. Das zweite Merkmal „Nearest Topic“ ist in Zusammenhang zu bringen mit dem Topik-Drop und der CC von Kim Mi-Kyung (2003). Dieses Merkmal bedeutet zunächst, dass das Topik in einem Diskurs das salienteste Element ist und eine hohe Wahrscheinlichkeit besteht, dass es das Antezedenz für das folgende Nullelement ist. Laut Han Na-Rae (2006) wird eine NP mit einem Topik-Marker auch als Topik in einem Diskurs angesehen. Darüber hinaus spiegelt dieses Merkmal auch empirische Befunde wider, wonach ein Antezedenz-Kandidat in der Nähe der Anapher mit höherer Wahrscheinlichkeit das zugehörige Antezedenz sein wird als ein Kandidat, der weiter entfernt ist (Lappin & Leass, 1994; Fisher et al., 1995; McCarthy, 1996; Mitkov, 1998; 2002; Strube et al., 2002; Holen, 2006). Das dritte Merkmal bringt zunächst zum Ausdruck, dass das Subjekt nach dem als Topik markierten Element das zweit-salienteste Element in einem Diskurs ist und es bestätigt zusätzlich die obengenannte Korrelation zwischen der üblichen Nähe von Anapher und Antezedenz.

Zum Trainieren ihres Modells verwendete Han Na-Rae (2006) zwei annotierte Korpora, die *Penn Korean Treebank 1* und *2*. Die *Penn Korean Treebank 1* besteht aus Dialogsequenzen zwischen Militärangehörigen und entspricht demselben Korpus wie das im Kapitel 1.2 vorgestellte PKT. Die Dialoge enthalten viele Frage-Antwort-Paare im umgangssprachlichen Stil. Die *Penn Korean Treebank 2* enthält kurze Nachrichtenberichte über Weltereignisse und Politik. Die Sätze sind formeller und enthalten im Allgemeinen mehr Wörter pro Satz als die Texte des ersten Korpus.³⁴ Als Trainingsmodell wählte sie die Maximum-Entropie-Methode und formte sie zu einer Ranking-Methode um. Auf diese Weise können alle möglichen Paare aus einem Nullelement und seinen potenziellen Antezedenzen nach dem Grad ihrer Wahrscheinlichkeit geordnet und als „coreferential“ eingestuft werden. Han Na-Rae (2006) berichtet, dass das trainierte Modell im Vergleich mit dem regelbasierten Ansatz, der auf der Grundlage des Hobbs-Algorithmus entwickelt wurde,³⁵ eine höhere Genauigkeitsrate erzielt. Das Ergebnis zeigt, dass das trainierte Modell um 8 % (in der *Penn Korean Treebank 1*) bzw. 4 % (in der *Penn Korean Treebank 2*) besser abschneidet als das Basismodell, das eine Genauigkeitsrate von 46,1 % (*Penn Korean Treebank 1*) bzw. 62,6 % (*Penn Korean Treebank 2*) erreichte. Eine der Entdeckungen von Han Na-Rae (2006) ist, dass die linguistischen Kontexte, die oft als entscheidend für das Phänomen der koreanischen Nullelemente diskutiert werden, in der Tat signifikante Prädiktoren für die NP-Antezedenzen in einem statistisch trainierten Modell sind. Insbesondere das Merkmal „Nearest Subject“ erweist sich als das Merkmal mit der größten Vorhersagekraft für Nullelemente in eingebetteten Sätzen. Allein durch die Auswahl des nächstgelegenen Subjekts für jedes Nullelement wird ein Resolutionssystem voraussichtlich in fast 65 % der Fälle erfolgreich sein. Das Merkmal „Nearest Topic“ ist ebenfalls ein starker Prädiktor, insbesondere für Nullelemente auf Matrixebene. Allein durch die Auswahl des nächstgelegenen Topiks konnte ein Resolutionssystem eine Erfolgsquote von 55,3 % erzielen.

Zusammenfassend lässt sich sagen, dass die Hauptstrategie zur Resolution des anaphorischen Nullsubjekte im Koreanischen in den bisherigen Studien unabhängig von ihren Ansätzen darin besteht, den am besten geeigneten Antezedenzen unter den Antezedenz-Kandidaten auszuwählen. Basierend

³⁴ Für eine ausführlichere Beschreibung der Korpora siehe Han Na-Rae (2006: 145).

³⁵ Für weitere Informationen über die Anwendung des Hobbs-Algorithmus auf die koreanischen Daten siehe Han Na-Rae (2006: 161).

auf den Resolutionsergebnissen der bisherigen Studien ist das Antezedens eines anaphorischen Nullsubjekts mit hoher Wahrscheinlichkeit in Äußerungen zu finden, die nah am Nullsubjekt selbst liegen. Darüber hinaus ist die NP mit dem postpositionalen Suffix *neun* oder in Subjektposition mit hoher Wahrscheinlichkeit das Antezedens eines anaphorischen Nullsubjekts.

Es bleibt jedoch zu bedenken, dass anhand der vorgestellten Strategien zur Resolution anaphorischer Nullsubjekte nur ein kleiner Teil der Koreanischen Nullelemente abgedeckt werden kann. Basierend auf den Ergebnissen von Hong Min-pyo (2000) und Hong Munpyo (2011) ist zu erwarten, dass letztendlich nur etwa 15–26 % der koreanischen Nullsubjekte so rekonstruiert werden können. Dementsprechend müssen weitere Resolutionsstrategien für Nullsubjekte eingeführt werden, die sich z. B. auf die Diskursteilnehmer:innen beziehen, die laut Hong Munpyo (2011) bis zu 81 % der koreanischen Nullsubjekte ausmachen. In Kapitel 2.1.3.2 wird näher beschrieben werden, welche Strategien zur Wiederherstellung der Referenten koreanischer deiktischer Nullsubjekte gegeben werden, insbesondere für diskursteilnehmerbezogene Nullsubjekte.

2.1.3.2. Deiktische Nullsubjekte

Frühere Studien haben darauf hingewiesen, dass es möglich ist, diskursteilnehmerbezogene Nullsubjekte anhand von expliziten morphologischen Hinweisen zu rekonstruieren. Auf Grundlage einer Datenanalyse folgert Hong Min-pyo (2000), dass der Satzmodus einen signifikanten Einfluss auf die Auslassung des Adressatenarguments hat. Auffallend häufig beobachtet er die Auslassung von Adressaten in Interrogativ-, Imperativ- und Exhortativsätzen. Er argumentiert, dass dieses Phänomen eng mit direktiven Sprechakten wie Aufforderungen, Vorschlägen oder Fragen verbunden ist.³⁶ Da der Satzmodus und die Sprechakte im Koreanischen bis zu einem gewissen Grad mithilfe morphologischer Informationen identifiziert werden können, ist zu erwarten, dass sie direkte Hinweise zur Wiederherstellung der Referenten von diskursteilnehmerbezogenen Nullsubjekten bieten können, was mit der CT nicht zu leisten ist (vgl. Hong Min-pyo 2000: 20–21).

Kim Mi-Young (2003) wiederum erklärt, dass deiktische Nullsubjekte im Koreanischen aufgrund von Verbalsuffixen in Prädikaten rekonstruiert werden können, insbesondere wenn sie in Sätzen vorkommen, die Sprechakte wie Befehle, Bitten und Vorschläge ausdrücken. Darüber hinaus können ihr zufolge die Nullsubjekte in Sätzen, in denen entweder das Morphem *sip* oder *gess* enthalten ist, als Subjekt der 1. Person Singular in Deklarativsätzen und als Subjekt der 2. Person Singular in Interrogativsätzen identifiziert werden. Die Autorin weist auch darauf hin, dass zusätzlich zu den morphologischen Informationen auch der Verbtyp dazu beitragen kann, deiktische Nullsubjekte im Koreanischen zu identifizieren, da das auf den Sprecher bezogene Subjekt in Deklarativsätzen und das auf den Adressaten bezogene Subjekt in Interrogativsätzen weggelassen werden, wenn psychologische Verben verwendet werden (vgl. Kim Mi-Young 2003: 36–37).

Auch Jang Seokjin (1986) stellt fest, dass Nullsubjekte bei psychologischen Prädikaten wie *gippeuda* 'glücklich sein', *jeulgeobda* 'fröhlich sein' und *bangabda* 'erfreut sein' in Deklarativsätzen als die 1. Person Singular interpretiert werden, während sie im Interrogativsatz als 2. Person Singular interpretiert werden. Han Na-Rae (2006) nimmt an, dass dies darauf zurückzuführen ist, dass im Koreanischen vermieden wird, einer Person einen emotionalen Zustand zuzuschreiben, der nicht selbst erlebt wird. Der Sprecher kann die Emotion einer anderen Person als ein beobachtetes Ereignis nur dann ausdrücken, wenn er ein Prädikat verwendet, bei dem das Hilfsverb *a/eo hada* angehängt wird, z. B. *gippeohada* 'glücklich sein'. Die folgenden Beispiele (entnommen aus Han Na-Rae 2006: 55–56) zeigen, dass Nullsubjekte nur bei psychologischen Prädikaten je nach Satztyp als Sprecher oder als Adressat interpretiert werden können. Darüber hinaus gilt diese Einschränkung bereits nicht mehr für

³⁶ Wie diese Sprechakte in der morphologischen Ebene identifiziert und zur Resolution adressatenbezogener Nullsubjekte im Koreanischen verwendet werden können, wird in Kapitel 3 ausführlicher behandelt.

die Prädikate *gippeohada*, welches durch Kombination des Verbstamms *gippeu* von *gippeuda* und des Hilfsverbs *eo hada* gebildet wird.

- (10) a. *hangug tim-i igi-eoseo* Δ *gippeu-da-*.
 koreanisch Team-SBJ gewinnen-CNJ (ich.SBJ) glücklich sein-SFS.DECL-SF
 'Ich bin glücklich, weil das koreanische Team gewonnen hat.'
- b. *hangug tim-i igi-eoseo* Δ *gippeu-ji-?*
 koreanisch Team-SBJ gewinnen-CNJ (du.SBJ) glücklich sein-SFS.INT-SF
 'Bist du glücklich, weil das koreanische Team gewonnen hat?'
- c. *hangug tim-i igi-eoseo yeonghui-neun gippeoha-nda-*.
 koreanisch Team-SBJ gewinnen-CNJ Younghee.SBJ.DP³⁷ glücklich sein-SFS.DECL-SF
 'Younghee ist glücklich, weil das koreanische Team gewonnen hat.'

In (10)a und (10)b wird das psychologische Verb *gippeuda* verwendet, während in (10)c das Verb *gippeohada* verwendet wird. In (10)a kann nur der Experiencer, der sich in diesem Moment glücklich fühlt, seinen aktuellen Gefühlszustand angeben, daher muss der Sprecher das Subjekt zum Prädikat *gippeuda* sein. In (10)b sollte das Subjekt im Fragesatz als Adressat interpretiert werden, da nur der Adressat die Frage beantworten kann, ob er sich derzeit glücklich fühlt. In (10)c wird das Prädikat *gippeohada* verwendet, was dem Sprecher ermöglicht, die Gefühle einer anderen Person wie *Younghee* zu beschreiben. Selbst wenn das Subjekt in (10)c ausgelassen würde, wird es als Drittperson und nicht als Sprecher oder Adressat interpretiert, da das Verb *gippeohada* verwendet wird.

Anhand eines gesprochensprachlichen Korpus des Koreanischen, das interaktive Gespräche enthält, versuchen Mok Jungsoo & You Hyunjo (2003) zu beweisen, dass die grammatische Person im Koreanischen auf morphologischer Ebene unterscheidbar ist, so dass die Referenten der Nullsubjekte durch einen grammatischen Mechanismus wiederhergestellt werden können. Die Autoren klassifizieren für ihre Studie drei Kategorien von Verben:

1. subjektive Verben,
2. Handlungsverben,
3. deskriptive Verben.

Aus jeder Kategorie wählen sie ein repräsentatives Verb aus, um die Verteilung der Nullsubjekte und der overten Subjekten ebenso wie die Hinweise auf die grammatische Person in jeder Verbkategorie zu analysieren.

Zu den subjektiven Verben können Mok Jungsoo & You Hyunjo (2003) herausarbeiten, dass sie grundsätzlich die 1. und 2. Person³⁸ implizieren und fest als Frage-Antwort-Paare realisiert werden, in dem die 2. Person in den Fragen und die 1. Person in den Antworten prominent ist. Sie erklären weiter, dass subjektive Verben die 3. Person implizieren, wenn das Hilfsverb *a/eo hada* mit den Verben kombiniert wird. Diese Beobachtung hängt auch mit der Feststellung zusammen, dass die deiktischen Nullsubjekte im Koreanischen durch die Verwendung der Verbtypen identifiziert werden können, da der Sprecher in Deklarativsätzen und der Adressat in Interrogativsätzen als Subjekt der

³⁷ Da es keinen Diskurs zur Funktionsbestimmung des Suffixes *neun* gibt, wird sie hier als Diskurs-Partikel (DP) glossiert. Die koreanischen Partikeln unterscheiden sich fundamental von den deutschen Partikeln, die als freie Lexeme mit abstrakter Semantik fungieren. Im Koreanischen handelt es sich bei den Partikeln um gebundene Morpheme, die als Suffixe auftreten können und sowohl nach Substantiven als auch nach Adverbien oder Phrasen positioniert werden (vgl. Lee Hansol 1989: 41).

³⁸ Mok Jungsoo & You Hyunjo (2003) lieferten Informationen zur grammatischen Person der Subjekte, jedoch fehlten Angaben zum Numerus. In den von ihnen präsentierten Beispielen für Subjekte der 1. und 2. Person überwog der Singular, während bei Subjekten der 3. Person sowohl Singular als auch Plural gemischt vorkamen.

psychologischen Verben weggelassen wird (Kim Mi-Young 2003; Han Na-Rae 2006). Mok Jungsoo & You Hyunjo (2003) wählen hierfür das subjektive Verb *silhda* 'hassen' und das Verb *silheohada* 'hassen', wobei Letzteres eine Kombination von *silhda* und *eo hada* ist, und analysieren die Verteilungen von Nullsubjekt und overtem Subjekt sowie die grammatische Person für beide Prädikate.

Person	Total	<i>silhda</i>		Total	<i>silheohada</i>		Total
		Null	Overt		Null	Overt	
1. (Sprecher)	29	21	2	23	1	5	6
2. (Adressat)	7	7	0	7	0	0	0
3.	21	1	6	7	5	9	14
Total	57	29	8	37	6	14	20

Tabelle 7: Verteilung des Nullsubjekts und des overt Subjekts bei den Verben *silhda* und *silheohada* 'hassen' nach der Person (vgl. Mok Jungsoo & You Hyunjo 2003: 541)

Zunächst steht die Frage im Zentrum der vorliegenden Analyse, ob das subjektive Verb *silhda* tatsächlich die 1. und 2. Person implizieren kann, während das Verb *silheohada* die 3. Person impliziert, wie Mok Jungsoo & You Hyunjo (2003) argumentieren. Als nächstes ist zu prüfen, ob die Subjekte der 1. und 2. Person (sprecher- und adressatenbezogene Subjekte) beim subjektiven Verb *silhda* häufig weggelassen werden, wie in früheren Studien argumentiert wurde. Von den insgesamt 29 Subjekten in der 1. Person liegen für *silhda* insgesamt 23 Subjekte vor, für *silheohada* 6. Darüber hinaus sind für die 2. Person 7 Subjekte nur beim Verb *silhda* belegt. Zudem stehen von den insgesamt 21 Subjekten in der 3. Person 7 Subjekte mit *silhda*, 14 mit *silheohada*.

Im Fall des subjektiven Verbs *silhda* sind von den insgesamt 37 Subjekten 29 (78,38 %) null realisiert, während 8 (21,62 %) overt realisiert werden. Insgesamt teilen sich die 29 Nullsubjekte asymmetrisch auf die grammatischen Personen auf: 21 Nullsubjekte (72,41 %) in der 1. Person, 7 Nullsubjekte (24,14 %) in der 2. Person und ein Nullsubjekt (3,45 %) in der 3. Person. Das bedeutet, dass etwa 97 % der Nullsubjekte beim Verb *silhda* in der 1. und 2. Person stehen. Es ist also davon auszugehen, dass bei subjektiven Verben wie *silhda* häufiger Subjekte ausgelassen als realisiert werden, und wenn sie weggelassen werden, sind die meisten Nullsubjekte auf die 1. und 2. Person bezogen. Wenn die Referenten der Nullsubjekte von *silhda* über ein Subjekt der 1. oder 2. Person wiederhergestellt werden, geschieht nur in einem Fall eine Fehlinterpretation, da das Subjekt auf die 3. Person referiert. Bei *silheohada* entsprechen 6 Subjekte der 1. Person und 14 Subjekte der 3. Person. Von den insgesamt 20 Subjekten werden 6 Subjekte (~30 %) nicht und 14 Subjekte (~70 %) overt realisiert. Daher kann für ein Verb wie *silheohada* festgehalten werden, dass ein Nullsubjekt bis zu einem gewissen Grad ein Subjekt der 3. Person implizieren kann, da 70 % der Subjekte in der 3. Person gefunden wurden. Außerdem ist zu beobachten, dass das Subjekt bei *silheohada* häufiger realisiert (~70 %) als ausgelassen (~30 %) wird, im Gegensatz zum subjektiven Verb *silhda*, bei dem das Subjekt häufiger ausgelassen (78,38 %) als realisiert (21,62 %) wird.

Natürlich ist es nicht möglich, diese Beobachtung für alle subjektiven Verben zu verallgemeinern, jedoch ist davon auszugehen, dass ein Nullsubjekt bei subjektiven Verben mit hoher Wahrscheinlichkeit als Referent der 1. oder 2. Person interpretiert werden kann. Bei Kombinationen aus subjektivem Verb und *a/eo hada* steht das Nullsubjekt eher in der 3. Person (in 5 von 6 Fällen). Daher kann es sinnvoll sein, den Verbtyp zur Identifizierung der Verweisstruktur koreanischer Nullsubjekte zu verwenden. Obwohl Mok Jungsoo & You Hyunjo (2003) die Verteilung der Nullsubjekte bei subjektiven Verben nicht nach dem Satzmodus analysieren, kann mit Rückgriff auf frühere Diskussionen möglicherweise sogar zwischen der 1. und der 2. Person für Nullsubjekte bei subjektiven Verben unterschieden werden, indem weitere Informationen über die Satztypen wie Deklarativsätze oder Interrogativsätze hinzugezogen werden.

Im Gegensatz zu den oben diskutierten subjektiven Verben können Mok Jungsoo & You Hyunjo bei den Handlungsverben keine implizierte spezifische grammatische Person herausarbeiten, da sie anscheinend mit verschiedenen grammatischen Personen kombiniert werden können. Die Autoren wählten in ihrer Studie das Handlungsverb *ilgda* 'lesen' aus. Das Ergebnis ihrer Untersuchung ist in Tabelle 8 dargestellt.

Person	<i>ilgda</i>		Total
	Null	Overt	
1. (Sprecher)	36	23	59
2. (Adressat)	8	2	10
3.	2	4	6
Total	46	29	75

Tabelle 8: Verteilung des Nullsubjekts und des overt Subjekts bei *ilgda* 'lesen' nach grammatischer Person (nach Mok Jungsoo & You Hyunjo 2003: 547)

In Tabelle 8 fällt zunächst auf, dass die Mehrheit der Subjekte des Handlungsverbs *ilgda* in der 1. Person vorliegt (59 von 75, entspricht ca. 79 %). 10 Subjekte (entspricht ca. 13 %) sind der 2. Person und 6 Subjekte (8 %) der 3. Person zuzuordnen. Wie von Mok Jungsoo & You Hyunjo (2003: 547) festgestellt, bieten die Subjekte der 2. und 3. Person im Vergleich zur 1. Person möglicherweise nicht ausreichend Anhaltspunkte, um bestimmte Schlussfolgerungen zu ziehen. Es ist jedoch wichtig zu prüfen, ob es in den Verbalkomplexen Hinweise darauf gibt, dass die Referenten von diskursteilnehmerbezogenen Nullsubjekten wiederhergestellt werden können.

Im Hinblick auf die Realisierung der Subjekte ist festzustellen, dass von den insgesamt 75 Subjekten 46 nicht (61 %) und 29 (39 %) overt realisiert werden. Bei den Nullsubjekten entfallen 36 Fälle (78,26 %) auf die 1. Person, 8 Fälle auf die 2. Person (17,39 %) und nur 2 (4,35 %) auf die 3. Person. Mok Jungsoo & You Hyunjo (2003: 547) erläutern, dass die 8 Fälle, in denen die Nullsubjekte des Handlungsverbs *ilgda* in der 2. Person stehen, anhand von Satzmodus und Modalität leicht identifiziert werden können. Sie zeigen in ihrem Beitrag jedoch nur die acht entsprechenden Sätze, ohne im Detail zu erläutern, wie dieser Identifizierungsprozess tatsächlich funktioniert. Bei der eigenen Untersuchung dieser Sätze stellt sich heraus, dass zwei der acht Sätze als Imperativsätze zu kategorisieren sind, die sich leicht durch die Imperativmarker unterscheiden lassen. Wie bereits festgestellt, kann dies ein Hinweis sein, der zur Wiederherstellung der Referenz adressatenbezogener Nullsubjekte beitragen kann. Das bemerkenswerte Phänomen in den verbleibenden sechs Sätzen ist, dass 4 das Honorativ infix (*eusi*) im Verbalkomplex enthielten. Die folgenden Beispiele zeigen zwei der vier Sätze aus Mok Jungsoo & You Hyunjo (2003: 547).

- (11) a. Δ *jamkkān* Ø *ilg-eo bo-si-nayo?*
(Sie) kurz (es) lesen-AUX-HON-SFS.INT-SF
'Lesen Sie es kurz?'
- b. Δ Ø *ilg-eusi-myeonseo-do* Δ *mol-ayo?*
(Sie) (es) lesen-HON-wenn-selbst (Sie) nicht wissen-SFS-SF
'Sie wissen nicht, selbst wenn Sie es lesen?'

Im Beispielsatz (11)a fragt der Sprecher den Adressaten, ob er kurz etwas lesen würde. Der Verbalkomplex in (11)a setzt sich aus vier Bestandteilen zusammen:

1. dem Stamm des Verbs *ilgda*,
2. dem Stamm des Hilfsverbs *eo boda*, das die Bedeutung des Versuchens hinzufügt,
3. dem Honorativ infix *si*, und
4. dem Verbalsuffix von *nayo*, das einen Interrogativsatz markieren kann.

Unter den Bestandteilen, die den Verbalkomplex bilden, ist das Honorativinfix *si* dasjenige, dass das Subjekt in (11)a als Honorifikum in der 2. Person (*Sie* statt *du*) identifiziert. Bei (11)b ermöglicht das Honorativinfix *eusi* wiederum die Interpretation des Subjekts als Honorifikum der 2. Person. Auf der Grundlage der Datenanalyse berichtet Hong Min-pyo (2000), dass in den meisten Fällen, in denen die adressatenbezogenen Subjekte weggelassen wurden, das Honorativinfix *(eu)si* zu finden war. Er erklärt, dass der Grund dafür darin liegt, dass außer in extrem informellen Gesprächen zwischen Freunden eine respektvolle Haltung gegenüber der anderen Person in Form eines Honorativinfixes ausgedrückt wird (vgl. Hong Min-pyo 2000: 21–22). Daraus lässt sich ableiten, dass die Wahrscheinlichkeit, dass sich ein Nullsubjekt sich auf ein Honorifikum in der 2. Person, also den Adressaten, bezieht, recht hoch ist, wenn morphologische Informationen wie das Honorativinfix *(eu)si* im Prädikat gefunden werden. Es ist daher insgesamt zu erwarten, dass morphologische Hinweise wie Imperativmarker oder Honorativfixe dabei helfen können, Nullsubjekte mit Referenten in der Adressatenrolle wiederherzustellen.

Für die Nullsubjekte in der 1. Person argumentieren Mok Jungsoo & You Hyunjo (2003: 548–549), dass das Agens des Handlungsverbs *ilgda* durch Hilfsverben und Verbalaffixe als Subjekt identifiziert werden kann. Sie präsentierten sechs Verbalkomplexe, mittels derer 28 von insgesamt 36 Nullsubjekten in der 1. Person abgedeckt werden können, allerdings ohne entsprechende Beispiele. In der folgenden Tabelle sind die von Mok Jungsoo & You Hyunjo (2003) vorgestellten Verbalkomplexe aufgeführt, die Nullsubjekte nach sich ziehen, die als Subjekte in der 1. Person rekonstruiert werden können.

Nr.	Verbalkomplex	Anzahl
1	<i>ilg-eo deuli-</i> lesen-AUX-	11
2	<i>ilg-eo deuli-gess-</i> lesen-AUX-IFX(Wille)-	7
3	<i>ilg-gess-</i> lesen-IFX(Wille)-	6
4	<i>ilg-go sip-</i> lesen-PPC(Wille)-	2
5	<i>ilg-eo bo-ass-</i> lesen-AUX-PST-	1
6	<i>ilg-eo bo-gess-</i> lesen-AUX-IFX(Wille)-	1
Total		28

Tabelle 9: Verbalkomplexe, die auf Nullsubjekte in der 1. Person hinweisen können (nach Mok Jungsoo & You Hyunjo 2003: 547–548)

In allen sechs Verbalkomplexen in Tabelle 9 ist der Stamm des Verbs *ilgda* der erste gemeinsame Bestandteil. Der auffälligste Bestandteil ist wiederum das Verbalinfix *gess*, das im 2., 3. und 6. Verbalkomplex enthalten ist. Wie auch das Morphem *sip* im 4. Verbalkomplex ist das Verbalinfix *gess* ein linguistisches Mittel, das den Willen des Sprechers ausdrücken kann, was bereits im Zusammenhang mit Kim Mi-Young (2003) erwähnt wurde. Kim Mi-Young (2003) erklärt, dass das Subjekt des Prädikats, das das Morphem *gess* oder *sip* enthält, in Deklarativsätzen als Sprecher und in Interrogativsätzen als Adressat interpretiert werden kann. Im 1. und 2. Verbalkomplexen ist der Stamm des Hilfsverbs *eo deulida* enthalten, das ausdrücken kann, dass jemand etwas für eine andere Person tut. Da bekannt ist, dass die Verbalkomplexe mit dem Subjekt in der 1. Person verwendet werden, ist davon auszugehen, dass die Äußerungen dazu dienen, dass der Sprecher dem Adressaten vorschlägt, etwas für ihn zu lesen. Die 5. und 6. Verbalkomplexe enthalten den Stamm des Hilfsverbs *eo boda*, das anzeigt, dass eine Handlung als Versuch ausgeführt wird. Wird das Hilfsverb wie im 5. Verbalkomplex in der Vergangenheitsform verwendet, kann es anzeigen, dass die Handlung schon einmal erlebt wurde. Da bekannt ist, dass die Verbalkomplexe mit Subjekten in der 1. Person stehen,

ist es wahrscheinlich, dass der Sprecher im Fall des 5. Verbalkomplexes seine Leseerfahrung in der Vergangenheit ausdrückt oder wie im Fall des 6. Verbalkomplexes seine Bereitschaft äußert, etwas zu lesen zu versuchen.

Die oben genannten Strategien scheinen mit der sprecherorientierten Modalität in Verbindung zu stehen, was darauf hindeutet kann, dass die Äußerung, insbesondere bei den Aussagesätzen, an den Sprecher gerichtet ist. Obwohl Handlungsverben selbst keine spezifische grammatische Person implizieren können, besteht dennoch die Möglichkeit, dass eine morphologische Information in Verbindung mit der sprecherorientierten Modalität einen Hinweis darauf liefert, wie das Nullsubjekt im Koreanischen auf den Sprecherreferenten bezogen und die Verweisstruktur wiederhergestellt werden kann.³⁹

Der letzte Verbtyp von Mok Jungsoo & You Hyunjo (2003) sind deskriptive Verben, die einen Zustand oder eine Form beschreiben. Sie werden auch als *objektive Adjektive* bezeichnet. Wörter wie *jeogda* 'wenig sein', *manhda* 'viel sein', *jagda* 'klein sein' und *keuda* 'groß sein' gehören zu dieser Art von Prädikat. Im Allgemeinen werden sie als einstelliges Prädikat in der Form [Themenrolle Argument – Prädikat] verwendet (vgl. Jeong Yeon ju 2011: 220). Mok Jungsoo & You Hyunjo (2003) haben für ihre Studie das deskriptive Verb *mahnda* 'viel sein' ausgewählt, um die Verteilung der Subjekte in den Daten zu analysieren. Sie berichten, dass im Vergleich zu den ersten beiden Verbtypen die Mehrheit der Subjekte von *mahnda* in der 3. Person steht (etwa 55 %). Wie aus Tabelle 7 und Tabelle 8 ersichtlich ist, stand die Mehrheit der Subjekte des subjektiven Verbs und des Handlungsverbs in der 1. Person (rund 62 %–79 %). Im Fall von *manhda* werden etwa 77 % der Subjekte overt realisiert, während die Mehrheit der Subjekte für das subjektive Verb und das Handlungsverb nicht realisiert wurde (etwa 78 % bzw. 61 %). Für deskriptive Verben ist also festzuhalten, dass Subjekte häufiger realisiert als ausgelassen werden und dass sie tendenziell eher in der 3. Person auftreten. Angesichts dessen, dass objektive Adjektive im Allgemeinen dazu dienen, einen Zustand oder eine Form zu beschreiben, ist es verständlich, dass ihre Subjekte häufiger in der 3. Person stehen. Daher kann in Erwägung gezogen werden, diese Art von Prädikat von der Analyse von diskursteilnehmerbezogenen Nullsubjekten auszuschließen.⁴⁰

Zusammenfassend lässt sich sagen, dass im Koreanischen der Satzmodus und die Modalität eine entscheidende Rolle bei der Wiederherstellung der Verweisstrukturen von sprecher- und adressatenbezogenen Nullsubjekten spielen, da sie mithilfe von morphologischen Informationen rekonstruierbar sind. Des Weiteren wurde erörtert, dass Verbtypen wie das subjektive Verb ebenfalls nützliche Anhaltspunkte für die Wiederherstellung diskursteilnehmerbezogener Referenten bei Nullsubjekten darstellen können. Dies resultiert daraus, dass nur derjenige, der die Emotion gerade erlebt, seinen psychologischen Zustand als Experiencer ausdrücken kann. Obwohl diskursteilnehmerbezogene Nullsubjekte im Koreanischen häufig vorkommen und anhand von Informationen im Verbalkomplex bestimmbar sind, gab es bisher nur wenige Versuche, ihre Referenten anhand der in früheren Studien vorgeschlagenen Strategien wiederherzustellen.

³⁹ Der Begriff der *sprecherorientierten Modalität* und sein Zusammenhang mit der Resolution von diskursteilnehmerbezogenen Nullsubjekten wird in Kapitel 3 ausführlicher behandelt.

⁴⁰ In Kapitel 3 wird erläutert, welche Anhaltspunkte im Koreanischen darauf hindeuten, dass Nullsubjekte mit geringerer Wahrscheinlichkeit als Diskursteilnehmer:innen interpretiert werden und wie sie in dieser Arbeit als Regeln im NRS formuliert werden.

2.2. Fazit: Wahl einer ganzheitlichen Strategie für die Resolution von Nullsubjekten

Bisherige Studien, die sich mit der Nullsubjektresolution beschäftigt und dabei auch diskursteilnehmerbezogene Nullsubjekte berücksichtigt haben, sind äußerst selten. Dies ist überraschend, da diskursteilnehmerbezogene Nullsubjekte einen erheblichen Anteil an den koreanischen Nullsubjekten ausmachen und klare Strategien für die Wiederherstellung von sprecher- oder adressatenbezogenen Referenten bieten.

Im Folgenden werden zwei Studien mit unterschiedlichen Ansätzen vorgestellt. In Kapitel 2.2.1 wird die Studie von Hong Munpyo (2011) präsentiert, die sich für einen regelbasierten Ansatz zur Resolution der Nullsubjekte entscheidet, während in Kapitel 2.2.2 die Studie von Park Arum & Hong Munpyo (2015) vorgestellt wird, die einen statistischen Ansatz verwendet. Durch die detaillierte Untersuchung dieser beiden Ansätze kann ermittelt werden, welcher Ansatz besser geeignet ist, um die Referenten koreanischer diskursteilnehmerbezogener Nullsubjekte zu identifizieren. Des Weiteren ermöglicht sie die Bewertung der Effektivität der angewandten Strategien für die Resolution dieser Nullsubjekte.

2.2.1. Vorstellung der Studie von Hong Munpyo (2011)

Hong Munpyo (2011) argumentiert in seinem regelbasierten Ansatz, dass für verschiedene Typen von Nullsubjekten unterschiedliche Resolutionsstrategien erforderlich sind. Die von ihm verwendeten Nullsubjekttypen sind, wie oben bereits erwähnt, die Kategorien „Intra-NS“, „Inter-NS“ und „Extra-NS“ (vgl. Kapitel 2.1.2). Er geht davon aus, dass die Resolution des koreanischen Nullsubjekts die Leistung maschineller Übersetzungssysteme verbessert, insbesondere wenn Koreanisch als Ausgangssprache und Englisch oder Deutsch als Zielsprachen mit subjektorientierter Struktur verwendet werden. In Hong Munpyos Diskussion wird die Hypothese aufgestellt, dass ein Nullsubjekt bereits im Eingabesatz des maschinellen Übersetzungssystems identifiziert wird.

Hong Munpyo präsentiert drei Resolutionsstrategien für die drei Typen von Nullsubjekten sowie eine zusätzliche Default-Strategie zur Resolution koreanischer Nullsubjekte.⁴¹ Für Nullsubjekte des Typs Intra-NS schlägt er vor, auszunutzen, dass im Koreanischen zwei durch bestimmte Konjunktionen verbundene Sätze ihre Subjekte teilen. Das bedeutet, dass die Subjekte der beiden verbundenen Sätze gleich sind, wenn sie durch eine bestimmte Konjunktion verbunden sind. Hong Munpyo hat insgesamt sieben Konjunktionen für diesen Zweck verwendet.⁴² Diese Resolutionsstrategie für die Kategorie Intra-NS ist inspiriert von den Constraints bei der Bestimmung der Antezedenzen der japanischen Nullsubjekte der Kategorie Intra-NS, bei der die Eigenschaft genutzt wird, dass zwei durch bestimmte Konjunktionen verbundene Sätze den *ga*-Kasus teilen, der für den Nominativ verwendet wird (vgl. Nakaiwa & Ikehara 1995: 100). Diese Eigenschaft koreanischer und japanischer Konjunktionen, ihre Subjekte zu teilen, wird in einschlägigen Arbeiten als *Subject-Sharing*⁴³ bezeichnet. Diese Strategie

⁴¹ Die von Hong Munpyo (2011) vorgeschlagenen Resolutionsstrategien werden in der vorliegenden Studie als das BS implementiert, um die Leistung des in dieser Arbeit vorgeschlagenen Ansatzes zu vergleichen und zu bewerten (Siehe Kapitel 1.2). Weitere Details dazu finden sich in Kapitel 5.

⁴² Die sieben Konjunktion aus Hong Munpyo (2011) für die Resolution der Nullsubjekte des Typs Intra-NS sind: 1. -*go*, 2. -*eoseo*, 3. -*myonseo*, 4. -*myeo*, 5. -*lyeogo*, 6. -*lyeomyeon* und 7. -*neula(go)*. Auf die Eigenschaft dieser sieben Konjunktionen, ihre Subjekte zu teilen, wird in Kapitel 4 näher eingegangen, wenn vorgestellt wird, wie Nullsubjekte in Nebensätzen in der vorliegenden Studie behandelt werden. Siehe Hong Munpyo (2011: 430–432) für weitere Informationen darüber, wie er diese Eigenschaft von Konjunktionen für die Resolution der koreanischen Nullsubjekte des Typs Intra-NS nutzt.

⁴³ Die Eigenschaft des „Subject Sharing“ ist ein Merkmal einiger Konjunktionen im Koreanischen und Japanischen, die im Deutschen allerdings nicht existiert. Da hierfür in einschlägigen Studien der englische Begriff verwendet wird, wird das Phänomen auch in der vorliegenden Studie diese Eigenschaft als „Subject-Sharing“ bezeichnet.

führt zur Resolution der Nullsubjekte in Hauptsätzen (S2), die durch eine der sieben Konjunktionen mit direkt vorausgehendem Satz (S1) verbunden sind, unter der Annahme, dass die Nullsubjekte in S2 dieselben wie die expliziten Subjekte in S1 sind.

Für die Nullsubjekte der Typen Inter-NS und Extra-NS nutzt Hong Munpyo (2011) die bereits betrachteten Strategien aus früheren Studien. In seiner Untersuchung nutzt er das CF-Ranking der CT in der Abfolge [Topik-NP > Subjekt-NP > Objekt-NP > Adverb-NP], um das optimale Antezedens unter den Antezedens-Kandidaten für die Resolution von Nullsubjekten des Typs Inter-NS zu identifizieren. Mit dieser Strategie wird die saliente NP unter den NPs in den vorangegangenen vier Sätzen als das Antezedens eines Nullsubjekts ausgewählt. Morphologische und lexikalische Informationen werden verwendet, um Nullsubjekte des Typs Extra-NS zu rekonstruieren, wobei der Fokus auf der Resolution von diskursteilnehmerbezogenen Nullsubjekten liegt. Dabei werden lediglich linguistische Mittel einbezogen, die in früheren Studien oft als starke Indikatoren für die Resolution von diskursteilnehmerbezogenen Nullsubjekten angesehen wurden, wie z. B. Imperativ- oder Exhortativmarker und subjektive Adjektive.

Die Default-Strategie dient der Zuweisung von Referenten in der Sprecher- oder Adressatenrolle. Nach dieser Strategie werden Satztypen vorausgesetzt, da das Nullsubjekt in Deklarativsätzen als Sprecher wiederhergestellt werden kann, während es in Interrogativ- und Imperativsätzen auf den Adressaten referiert. Dabei werden die Satztypen anhand von Satzzeichen klassifiziert: Sätze mit einem Schlusspunkt wurden als Deklarativsätze, solche mit einem Fragezeichen als Interrogativsätze und solche mit einem Ausrufezeichen als Imperativsätze klassifiziert.

Hong Munpyo präsentiert einen Resolutionsprozess, bei dem je nach Kontext, in dem ein Nullsubjekt auftritt, eine der vier Resolutionsstrategien angewandt wird. Die Auswahl der Resolutionsstrategie erfolgt unabhängig vom Typ des Nullsubjekts, da der Typ des Nullsubjekts möglicherweise in einem maschinellen Übersetzungsprozess nicht bekannt ist. Wenn ein Nullsubjekt im Eingabesatz identifiziert wird, wird zunächst die erste Resolutionsstrategie angewandt. Diese Strategie ordnet dem Nullsubjekt anhand von linguistischen Merkmalen im Verbalkomplex einen diskursteilnehmerbezogenen Referenten zu. Wenn die erste Strategie auf das Nullsubjekt nicht angewandt werden kann, variiert der Resolutionsprozess je nachdem, ob das Nullsubjekt in einem einfachen oder einem zusammengesetzten Satz erscheint. Bei einem einfachen Satz, der nur einen eigenständigen Hauptsatz enthält, wird die Resolutionsstrategie für die Resolution des Typs Inter-NS angewandt, um die saliente NP auszuwählen. Falls das Nullsubjekt in einem zusammengesetzten Satz gefunden wird, berücksichtigt der Prozess die Konjunktion, die den zusammengesetzten Satz verbindet, und die Resolutionsstrategie für den Typ Intra-NS kommt zum Einsatz. Bei Verwendung einer der sieben Konjunktionen mit Subject-Sharing wird die Resolutionsstrategie für den Typ Intra-NS angewandt, um das explizite Subjekt in S1 als Referent des Nullsubjekts in S2 zu identifizieren. Falls andere Konjunktionen verwendet werden, kommt wieder die Resolutionsstrategie für den Typ Inter-ZP zum Einsatz und wird mithilfe des CF-Rankings die saliente NP ausgewählt. Sollte das Nullsubjekt durch diesen Prozess nicht rekonstruiert werden, wird abschließend die Default-Strategie angewandt, um diskursteilnehmerbezogene Referenten mithilfe von Satzzeichen zuzuweisen. Nach dieser Default-Strategie wird das Nullsubjekt in Sätzen mit einem Schlusspunkt als Sprecher wiederhergestellt, während es in Sätzen mit einem Frage- oder Ausrufezeichen als Adressat wiederhergestellt wird.

Hong Munpyo führt die Bewertung verschiedener Resolutionsstrategien anhand eines Korpus aus Dramentexten durch, aus denen 207 Nullsubjekte extrahiert wurden. Der Nutzen der einzelnen Resolutionsstrategien wird durch den Vergleich von fünf Kombinationen dieser Strategien bewertet, wobei jede Strategie nacheinander hinzugefügt wird. Anhand der Bewertungsparameter *Precision* 'Präzision', *Recall* 'Recall' und *F1-Score* 'F-Maß' wird ermittelt, wie viele Referenten der Nullsubjekte wiederhergestellt werden konnten und wie genau die Resolution vonstatten ging.

Precision gibt an, wie wahrscheinlich es ist, dass die positive Identifizierung durch ein Modell (eine Strategie bei Hong Munpyo) tatsächlich korrekt ist. *Recall* repräsentiert das Verhältnis der richtig

erfassten Fälle zu allen tatsächlichen Fällen. Der *F1-Score*, der den harmonischen Mittelwert aus *Precision* und *Recall* darstellt, gibt Aufschluss darüber, wie effektiv das Modell positive Fälle identifiziert, ohne dabei falsche Positive zu generieren.

Nr.	Strategie	Precision (%)	Recall (%)	F1-score (%)
1	<i>Default</i>	46,41	74,40	57,15
2	<i>Verbalsuffix + Default</i>	62,50	77,29	69,08
3	<i>Verbalsuffix + subjektives Adjektiv + Default</i>	63,98	77,78	70,19
4	<i>Verbalsuffix + subjektives Adjektiv + Konjunktion + Default</i>	64,94	84,06	73,26
5	<i>Verbalsuffix + subjektives Adjektiv + Konjunktion + CT + Default</i>	73,29	85,02	78,72

Tabelle 10: Leistungsergebnis der Strategien zur Nullsubjektreolution von Hong Munpyo (2011: 435)

Tabelle 10 zeigt die Bewertungsparameter *Precision*, *Recall* und *F1-Score* für die fünf Kombinationen der Resolutionsstrategien. Die erste Strategie „Default“ wird als Basis für die Bewertung der Resolutionsstrategie verwendet. Der *Recall*-Wert in dieser Baseline liegt bei 74,40 %, was recht hoch ist, wenn man bedenkt, dass er sich nur um etwa 11 % vom *Recall*-Wert der letzten Kombination aus Resolutionsstrategien unterscheidet, der bei 85,02 % liegt. Daraus lässt sich schließen, dass bereits eine ganze Reihe von Herleitungen der Referenten von Nullsubjekten allein durch die Verwendung der Satztypinformation abgedeckt werden kann. Der *Precision*-Wert dieser Baseline beträgt jedoch 46,41 %, was einen Unterschied von etwa 27 % im Vergleich mit der letzten Strategiekombination bedeutet. Dies kann dahingehend interpretiert werden, dass die mit dieser Default-Strategie wiederhergestellten Referenten der Nullsubjekte nicht ganz korrekt waren. Der erste Grund für den niedrigen *Precision*-Wert der Default-Strategie liegt darin, dass die Satztypen nur anhand der Satzzeichen beurteilt wurden. Ein Imperativsatz endet beispielsweise nicht unbedingt mit einem Ausrufezeichen, sondern kann auch mit einem Schlusspunkt enden. Daher kann es Fälle geben, in denen Nullsubjekte als Adressaten wiederhergestellt werden müssen, weil sie tatsächlich in Imperativsätzen vorkommen, auch wenn diese mit einem Punkt enden. Der zweite Grund ist, dass der Satzmodus und die Sprechakte eine wichtigere Rolle bei der Resolution der Nullsubjekte spielen als die Satztypen. Wenn zum Beispiel ein Satz mit einem Ausrufezeichen den Satzmodus Exhortativ aufweist, sollte das Nullsubjekt als 1. Person Plural rekonstruiert werden, nicht der Adressat. Wenn ein Interrogativsatz für einen Vorschlag, etwas gemeinsam zu tun, verwendet wird, sollte das Nullsubjekt ebenfalls als 1. Person Plural rekonstruiert werden.⁴⁴ Daher ist zu erwarten, dass falsche Vorhersagen auftreten, wenn Nullsubjekte in Sätzen mit Ausrufe- und Fragezeichen einfach als Adressaten wiederhergestellt werden.

Die zweite Resolutionsstrategie in Tabelle 10 berücksichtigt zusätzlich den Satzmodus oder den Sprechakt, was in der Default-Strategie fehlt. Die hinzugefügte Strategie wird als *Verbalsuffix* bezeichnet, weil der Satzmodus, einige Sprechakte und die diskursteilnehmerbezogene Modalität im Koreanischen durch Verbalsuffixe erkannt werden können. Die auf die Verbalsuffixe bezogene

⁴⁴ Der Zusammenhang zwischen dem Satzmodus, den Sprechakten und der Modalität sowie bestimmten Referenten von Nullsubjekten wird in Kapitel 4 ausführlicher behandelt.

Strategie wird in diesem Fall zuerst angewandt, danach die Default-Strategie für den Rest der Fälle. Bei der Leistungsbewertung fällt auf, dass der *Recall*-Wert der zweiten Resolutionsstrategie im Vergleich zur Default-Strategie nur um ca. 3 %, der *Precision*-Wert aber um ca. 16 % gestiegen ist. Dies ist die größte Steigerung des *Precision*-Wertes unter den fünf in Tabelle 10 dargestellten Kombinationen. Aus diesem Ergebnis lässt sich ableiten, dass bei koreanischen Nullsubjekten, insbesondere bei diskursteilnehmerbezogenen Nullsubjekten, morphologische Informationen, die auf den Satzmodus, den Sprechakt und die Modalität hinweisen können, in hohem Maße zur genauen Wiederherstellung der Referenten beitragen.

Bei der dritten Resolutionsstrategie in Tabelle 10 wird der Parameter der subjektiven Adjektive hinzugefügt. Zuerst wird in dieser Strategienkombination die auf Verbalsuffixe bezogene Strategie angewandt, danach der Parameter der subjektiven Adjektive. Für die verbleibenden Fälle wird dann die Default-Strategie angewandt. Im Vergleich zur zweiten Resolutionsstrategienkombination sind die *Precision*- und *Recall*-Werte leicht um 1,5 % bzw. 0,5 % angestiegen. Auf der Grundlage dieses Ergebnisses könnte man interpretieren, dass der Parameter des subjektiven Adjektivs nicht viel zur Nullsubjektresolution beiträgt. Allerdings gibt Hong Munpyo (2011) nicht an, wie viele Nullsubjekte in den subjektiven Prädikaten gefunden wurden. Daher ist nicht abschließend zu bewerten, ob der Hinweis auf die subjektiven Adjektive nicht doch wesentlich zur Wiederherstellung der Referenz der koreanischen Nullsubjekte beiträgt. Zudem könnte es sein, dass subjektive Prädikate nur einen sehr geringen Anteil am gesamten Evaluationsdatensatz ausmachen, der ohnehin bereits sehr begrenzt ist.

In der vierten Stufe in Tabelle 10 werden die Konjunktionen als Resolutionsstrategie zu Rate gezogen und zur dritten Resolutionsstrategie vor der Default-Strategie hinzugefügt. Die hinzugefügte Strategie verwendet die Eigenschaft des *Subject-Sharing* von Konjunktionen. Bemerkenswert ist, dass der *Recall*-Wert im Vergleich zur dritten Resolutionsstrategie um etwa 6 % gestiegen ist, was den höchsten Anstieg in Bezug auf den *Recall*-Wert unter allen Strategienkombinationen in Tabelle 10 darstellt. Der mögliche Grund für den signifikanten Anstieg des *Recall*-Wertes bei der vierten Resolutionsstrategie könnte darin liegen, dass durch die Hinzufügung der Konjunktionsstrategie mehr Nullsubjekte in zusammenhängenden Sätzen erfasst werden, die von der dritten Resolutionsstrategie möglicherweise nicht abgedeckt wurden. Anders als der *Recall*-Wert ist der *Precision*-Wert nur um etwa 1 % angewachsen. Daraus lässt sich ableiten, dass die Anzahl der korrekt wiederhergestellten Referenten der Nullsubjekte durch die Hinzunahme von konjunktionsbezogenen Strategien nicht signifikant erhöht wurde. Einer der Gründe für den geringen Anstieg des *Precision*-Wertes könnte sein, dass auch bei Verwendung einer Konjunktion, von der bekannt ist, dass sie die Eigenschaft des *Subject-Sharing* besitzt, nicht in jedem konkreten Fall *Subject-Sharing* auftritt (vgl. Kapitel 4).

Die CT ist die letzte Resolutionsstrategie, die in Tabelle 10 hinzugefügt wird. Betrachtet man den *F1-Score*, der den harmonischen Mittelwert aus *Precision*- und *Recall*-Wert repräsentiert, so ist er bei dieser fünften Resolutionsstrategie der höchste aller Strategienkombinationen. Im Vergleich zur vierten Resolutionsstrategie steigt der *Recall*-Wert jedoch nur um etwa 0,9 %. Es ist davon auszugehen, dass es insgesamt nicht viele anaphorische Nullsubjekte gibt. Beim *Precision*-Wert ist im Vergleich zur vierten Resolutionsstrategie ein Anstieg von etwa 8,4 % zu beobachten, was die zweitgrößte Steigerung dieses Wertes darstellt. Dies kann dahingehend interpretiert werden, dass die Wahrscheinlichkeit hoch ist, dass die nach dem CF-Ranking höchstrangige NP unter den Antezedens-Kandidaten tatsächlich das Antezedens des Nullsubjekts ist.

Alle Strategien in Bezug auf die Satztypen, den Satzmodus, die Sprechakte, die Modalität und psychologische Verben, die in früheren Studien als gute Anhaltspunkte für die Interpretation von Nullsubjekten in Bezug auf Diskursteilnehmer:innen vorgeschlagen wurden, sind in der dritten Resolutionsstrategie von Hong Munpyo (2011) enthalten. Der *Recall*-Wert dieser Methode liegt bei 77,78 %, was darauf schließen lässt, dass fast alle Referenten der Nullsubjekte in der Kategorie Extra-

NS auf diese Weise wiederhergestellt werden konnten, wenn man bedenkt, dass die Kategorie Extra-NS etwa 81 % aller Nullsubjekte gemäß der Verteilung von Hong Munpyo (2011) ausmacht (vgl. Tabelle 4, Kapitel 2.1.2).

Allerdings ist der *Precision*-Wert dieser Resolutionsstrategie mit ca. 63,98 % nicht sehr hoch, im Gegensatz zur Annahme der vorherigen Studien, dass die Hinweise recht eindeutig sind, um Nullsubjekte den Diskursteilnehmern zuordnen zu können. Einer der möglichen Gründe für den niedrigen *Precision*-Wert dieser Resolutionsstrategie ist, dass diese Methode eine potentielle Richtungsänderung einer Äußerung nicht berücksichtigt, wenn ein Morphem zum Verbalkomplex hinzugefügt wird. Wenn z. B. in einem Deklarativsatz ein Morphem hinzugefügt wird, das auf eine Vermutung oder ein Honorifikum schließen lässt, kann das Subjekt der Äußerung als Adressat und nicht als Sprecher interpretiert werden (vgl. Beispiel (4), Kapitel 1.2). Betrachtet man die Kombination der verschiedenen Bestandteile von Verbalkomplexen im Hinblick auf die Nullsubjektreolution, so wird erwartet, dass nicht nur die richtig rekonstruierten Nullsubjekte, sondern auch die Anzahl der abgedeckten Nullsubjekte erhöht wird, selbst wenn nur die morphologischen und lexikalischen Informationen in Prädikaten verwendet werden.

Bisher war zu sehen, dass Hong Munpyo (2011) versucht hat, das Nullsubjekt mit einem regelbasierten Ansatz zu rekonstruieren. Er verwendet dafür verschiedene Strategien, von denen bekannt ist, dass sie zur koreanischen Nullsubjektreolution beitragen. Dabei bewertet er den Nutzen dieser Strategien für die Resolution, indem er verschiedene individuelle Strategien in Resolutionsstrategien gruppier. Aufgrund dieser Herangehensweise ist es jedoch schwierig, eine detaillierte Analyse der einzelnen Strategien durchzuführen, um festzustellen, welche zur Wiederherstellung bestimmter Referenten verwendet werden kann und wie genau eine Strategie eine belastbare Vorhersage für bestimmte Referenten zusichert. Daher kann argumentiert werden, dass Hong Munpyo (2011) den Vorteil des regelbasierten Ansatzes, der eine direkte Beziehung zwischen Eingabe und Ausgabe aufweist, hier zwischen der Regel und der Vorhersage, nicht vollständig ausnutzt.

2.2.2. Vorstellung der Studie von Park Arum & Hong Munpyo (2015)

Park Arum & Hong Munpyo (2015) versuchen ihrerseits, die Referenten koreanischer Nullsubjekte mit fast denselben Strategien wie Hong Munpyo (2011) wiederherzustellen, allerdings unter Verwendung von maschinellem Lernen (kurz „ML“). Die Autor:innen betonen die Problematik des Nullsubjekt-Phänomens im Koreanischen, insbesondere im Kontext der maschinellen Übersetzung von Koreanisch nach Deutsch. Dies ergibt sich aus der Notwendigkeit, implizite Subjekte im Koreanischen explizit ins Deutsche zu übersetzen. Daher unterstreichen sie die Wichtigkeit der Resolution koreanischer Nullsubjekte, um die Qualität der maschinellen Übersetzung vom Koreanischen ins Deutsche zu verbessern. Die Autor:innen erklären, dass die Verwendung von ML notwendig sei, um die Einschränkungen deterministischer Ansätze zu überwinden. Bei einem deterministischen Ansatz wird bei einer bestimmten Eingabe immer die gleiche Ausgabe erzeugt, wenn die Eingabe eine bestimmte Bedingung erfüllt. Es können jedoch Fälle auftreten, in denen keine Bedingung erfüllt ist, oder die Ausgabe selbst dann nicht korrekt ist, wenn eine Bedingung erfüllt ist (vgl. Park Arum & Hong Munpyo 2015: 203).

Insgesamt präsentierten sie zwölf ML-Merkmale für die Aufgabe. In Tabelle 11 werden die ML-Merkmale von Park Arum & Hong Munpyo (2015) mit den von Hong Munpyo (2011) verwendeten Resolutionsstrategien verglichen. Dies ermöglicht einen schnellen Überblick darüber, wie die Resolutionsstrategien von Hong Munpyo (2011) als ML-Merkmale genutzt wurden und welche Strategien von Park Arum & Hong Munpyo (2015) hinzugefügt wurden. Die ML-Merkmale von Park Arum & Hong Munpyo (2015) können entweder den Wert 1 oder 0 haben, abhängig davon, ob die in den ML-Merkmalen beschriebene Bedingung erfüllt ist oder nicht.

ML-Merkmal in Park Arum & Hong Munpyo (2015)		Strategie in Hong Munpyo (2011)
F1	Imperativmarker	
F2	Interrogativmarker	Verbalsuffix
F3	Exhortativmarker	
F4	subjektives Adjektiv im Deklarativsatz	Subjektives Adjektiv
F5	objektives Adjektiv im Deklarativsatz	N/A
F6	Verbalsuffix mit sprecherorientierte Modalität	Verbalsuffix
F7	Honorativ infix (<i>eu</i>) <i>si</i>	N/A
F8	subjektives Adjektiv + den Interrogativmarker <i>sibnikka</i> / <i>seyo</i>	Subjektives Adjektiv
F9	Topik-NP innerhalb der vorherigen vier Sätze	
F10	Subjekt-NP innerhalb der vorherigen vier Sätze	
F11	Objekt-NP innerhalb der vorherigen vier Sätze	
F12	Adverb-NP innerhalb der vorherigen vier Sätze	CT

Tabelle 11: Gegenüberstellung der ML-Merkmale von Park Arum & Hong Munpyo (2015: 204) und der Resolutionsstrategie von Hong Munpyo (2011)

Von den insgesamt zwölf von Park Arum & Hong Munpyo (2015) vorgeschlagenen ML-Merkmalen kodieren zehn dieselben Eigenschaften, die von Hong Munpyo (2011) verwendet wurden. Zunächst rücken die zehn ML-Merkale in den Fokus, die die gleichen Eigenschaften wie die von Hong Munpyo (2011) verwenden. Die Merkmale F1 bis F3 beziehen sich auf den Satzmodus, wobei der Wert 1 zugewiesen wird, wenn Prädikate mit Nullsubjekten Verbalsuffixe haben, die den Imperativ, Interrogativsätze und den Exhortativ markieren. Diese entsprechenden Hinweise werden in Hong Munpyos (2011) Resolutionsstrategien als *Verbalsuffix* bezeichnet. Das Merkmal F6 ist ebenfalls mit Hong Munpyos (2011) Resolutionsstrategie *Verbalsuffix* verbunden und weist den Wert 1 zu, wenn in Prädikaten mit Nullsubjekten Verbalsuffixe auftreten, die eine sprecherorientierte Modalität ausdrücken können, wie eine Absicht, einen Plan oder ein Versprechen vonseiten des Sprechers. Die Merkmale F4 und F8 beziehen sich wiederum auf die Eigenschaften der subjektiven Adjektive. Beim Merkmal F4 wird der Wert 1 zugewiesen, wenn ein subjektives Adjektiv in einem Deklarativsatz vorkommt, und beim Merkmal F8, wenn ein subjektives Adjektiv in einem Interrogativsatz vorkommt.⁴⁵ Bei einer genaueren Untersuchung des Merkmals F8 wird der Fokus auf Interrogativmarkern *sibnikka* und *seyo* gelegt, in denen zusätzlich das Honorativprefix *si* enthalten ist. Dies spiegelt den empirischen Befund wider, dass das Honorativprefix *si* ein nützlicher Hinweis für Nullsubjekte sein kann, um sie als Adressaten zu interpretieren. Allerdings scheint das Merkmal F8 die Eigenschaften subjektiver Adjektive nicht vollständig zu kodieren, da Nullsubjekte von subjektiven Adjektiven in Interrogativsätzen auch ohne das Honorativprefix *si* als Adressaten interpretiert werden können. Die ML-Merkmale F9 bis F12 beziehen sich auf die CT, die auch von Hong Munpyo (2011) verwendet wird. In Park Arum & Hong Munpyo (2015) werden diese Merkmale so kodiert, dass ihnen der Wert 1 zugewiesen wird, wenn in den vorherigen vier Sätzen vor dem Nullsubjekt eine Topik-NP, Subjekt-NP, Objekt-NP oder Adverb-NP vorkommt.

Die ML-Merkmale, die neue Strategien im Vergleich zu den Resolutionsstrategien von Hong Munpyo (2011) kodieren, sind F5 und F7. Das Merkmal F5 erhält den Wert 1, wenn Nullsubjekte in Prädikaten mit objektiven Adjektiven gefunden werden. Park Arum & Hong Munpyo (2015) betrachten die Kombination von subjektiven Adjektiven und dem Hilfsverb *eo hada* als objektive Adjektive. Dies steht im Zusammenhang mit dem Argument von Mok Jungsoo & You Hyunjo (2003), dass die 3. Person impliziert wird, wenn *eo hada* mit subjektiven Verben kombiniert wird.⁴⁶ Schließlich wird beim

⁴⁵ Hierfür wurde die Liste der subjektiven Adjektive von Yoo Hyun-kyung (1998) genutzt, die im Anhang von Park Arum & Hong Munpyo (2015) zu finden ist.

⁴⁶ Die entsprechenden Ergebnisse der Korpusanalyse von Mok Jungsoo & You Hyunjo (2003) sind in Tabelle 7 in Kapitel 2.1.3.2 zu finden.

Merkmal F7 der Wert 1 zugewiesen, wenn das Honorativinfix *si* in den Prädikaten von Nullsubjekten enthalten ist. Dies basiert auf der Argumentation früherer Studien, dass die Respekthaltung gegenüber dem Adressaten im Koreanischen mit linguistischen Mitteln ausgedrückt wird, die in hohem Maße zur Wiederherstellung eines Adressatensubjekts beitragen können. Da das Merkmal F7 bereits das Auftreten des Honorativinfix *si* in Prädikaten behandelt, kann argumentiert werden, dass die Informationen zum Merkmal F8 das wiederholen, was bereits von Merkmal F7 erfasst wird.

Park Arum & Hong Munpyo (2011) extrahieren insgesamt 1.000 Nullsubjekte aus einem Korpus von Dialogsätzen im Bereich des Tourismus⁴⁷, um die Effektivität der vorgeschlagenen ML-Merkmale für die Resolution koreanischer Nullsubjekte zu bewerten. Für diese 1.000 koreanischen Nullsubjekte wurden manuell vier deutsche Subjektklassen zugewiesen. Die vier Subjektklassen sind

1. *ich*,
2. *Sie*⁴⁸
3. *der/die/das*,
4. *wir*.

Diese Subjektklassen werden während der Trainingsphase eines ML-Modells verwendet, um die Merkmale jeder Subjektklasse zu erlernen. In der Leistungsbewertungsphase des ML-Modells dienen sie als Referenz, um die vom Modell ausgewählte Subjektklasse zu vergleichen. Die Verteilung der einzelnen Subjektklassen für die 1.000 koreanischen Nullsubjekte wird im Folgenden dargestellt. Die Anzahl jeder Klasse entspricht der Anzahl der Trainingsbeispiele, die dem ML-Modell zur Verfügung stehen, um die Merkmale der Beispiele für jede Subjektklasse zu erlernen.

Subjektklasse	Anteil
<i>ich</i>	47,20 % (n=472)
<i>Sie</i>	47,20 % (n=472)
<i>der/die/das</i>	5,30 % (n=53)
<i>wir</i>	0,30 % (n=3)

Tabelle 12: Verteilung der richtigen Subjektklassen in den Daten von Park Arum & Hong Munpyo (2015: 208)

Betrachtet man die Verteilungen der Subjektklassen in Tabelle 12, so zeigt sich ein großer Unterschied zwischen denjenigen, die für das ML verwendet werden. Die zwei Subjektklassen *ich* und *Sie*, die dem Sprecher und dem Adressaten entsprechen, machen jeweils etwa 47 % der Gesamtzahl aus, und kommen zusammen etwa auf 94 %. Die geringe Repräsentation der verbleibenden beiden Subjektklassen *der/die/das* und *wir* mit nur etwa 6 % in den Trainingsdaten bedeutet, dass es für ein ML-Modell nur sehr wenige Trainingsbeispiele gibt, um die Merkmale dieser beiden Klassen zu erlernen. Ein erhebliches Ungleichgewicht im Verhältnis der Trainingsbeispiele, die jede Klasse in den gesamten Trainingsdaten repräsentieren, kann sich nachteilig auf die Klassifizierungsleistung des ML-Modells auswirken. Es ist bekannt, dass zwischen Mehrheits- und Minderheitsklassen eine Leistungsverzerrung besteht, wobei die Ergebnisse in der Mehrheitsklasse tendenziell genauer und die

⁴⁷ Das von Park Arum & Hong Munpyo (2011) verwendete Korpus ist identisch mit dem in Kapitel 1.2 beschriebenen Analysekörper dieser Studie. Trotz dieser Gemeinsamkeit unterscheidet sich die Gesamtzahl der Nullsubjekte, die zur Bewertung der Resolutionsstrategie herangezogen wurden, erheblich zwischen Park Arum & Hong Munpyo (2011) und dieser Studie. Hier wurden insgesamt 15.883 Nullsubjekte verwendet, um die Leistung des NRS zu bewerten, was mehr als das 15-fache der Anzahl der Nullsubjekte ist, die in Park Arum & Hong Munpyo (2011) für die Bewertung verwendet wurden.

⁴⁸ Laut Park Arum & Hong Munpyo (2015: 208) wurde entschieden, die auf den Adressaten bezogene Subjektklasse als *Sie* festzulegen, um das Problem der Datenknappheit zu vermeiden, obwohl sich im Deutschen sowohl "Sie" als auch "du" auf den Adressaten beziehen können.

Ergebnisse in der Minderheitsklasse tendenziell weniger genau sind (vgl. Kaur et al. 2019: 1). Die Auswirkungen der unterschiedlichen Klassengewichtungen können im Folgenden anhand der Leistungsbewertung nach Subjektklassen beobachtet werden.

Park Arum & Hong Munpyo (2015) verwenden die Open-Source-Software Weka⁴⁹, die eine Vielzahl von ML-Modellen bereitstellt, für die Nullsubjektresolution. Die Autor:innen entscheiden sich dabei für die Support Vector Machine (SVM) aus den von Weka angebotenen ML-Modellen. Zur Ergänzung einer vergleichsweise geringen Menge an Trainingsdaten setzen sie die 10-fache Kreuzvalidierungsmethode ein. Hierbei werden beispielsweise 1.000 Sätze in zehn Gruppen unterteilt, wobei auf 900 Sätzen das Training durchgeführt und die Trainingsleistung auf den verbleibenden 100 Sätzen bewertet wird. Nachdem dieser Prozess für alle Gruppen wiederholt wurde, wird das durchschnittliche Leistungsergebnis durch Summierung aller Bewertungsergebnisse ermittelt. Die Ergebnisse der ML-Leistungsbewertung von Park Arum & Hong Munpyo (2015) durch diese Methode lauten wie folgt.

Subjektklasse	Precision (%)	Recall (%)	F1-score (%)
<i>ich</i>	85,70	93,90	89,60
<i>Sie</i>	96,10	89,00	92,40
<i>der/die/das</i>	65,20	56,60	60,60
<i>wir</i>	0,00	0,00	0,00

Tabelle 13: Precision- und Recall-Werte sowie F1-Score nach Subjektklassen von Park Arum & Hong Munpyo (2015: 208)

Wie erwartet ist der *F1-Score* für die Subjektklassen *ich* und *Sie*, die als Mehrheitsklassen angesehen werden können, signifikant höher als für die Minderheitsklassen *der/die/das* und *wir*. Insbesondere in der Subjektklasse *wir* liegt er bei 0, was darauf zurückzuführen ist, dass nur drei Beispiele für diese Klasse in den Daten enthalten sind, was nicht genug ist, um Schlüsse zu dieser Subjektklasse ziehen zu können, wie auch Park Arum & Hong Munpyo (2015: 210) feststellen. Vergleicht man die *Precision*- und *Recall*-Werte der Subjektklassen *ich* und *Sie*, so ist der *Recall*-Wert für die Klasse *ich* höher als der für die Klasse *Sie* (93,90 % vs. 89,00 %). Auf der anderen Seite ist der *Precision*-Wert für die Klasse *ich* geringer als der für die Klasse *Sie* (85,70 % vs. 96,10 %). Das bedeutet, dass die Anzahl der vorhergesagten Fälle der Nullsubjekte in der Klasse *ich* höher ist als die Anzahl der vorhergesagten Fälle der Nullsubjekte in der Klasse *Sie*. Es gibt jedoch mehr korrekte Vorhersagen für Nullsubjekte in der Klasse *Sie* verglichen mit denen der Klasse *ich*. Angesichts dieser Ergebnisse kann davon ausgegangen werden, dass die ML-Merkmale zur Klassifizierung adressatenbezogener Nullsubjekte eine wichtigere Rolle bei der korrekten Vorhersage der Nullsubjekte mit Adressatenreferenten spielen könnten als die ML-Merkmale zur Klassifizierung sprecherbezogener Nullsubjekte.

Park Arum & Hong Munpyo (2015) verwenden zudem den vom Weka bereitgestellten *InfoGain Attribute Evaluator*⁵⁰, um eine Rangliste der zwölf ML-Merkmale auf der Grundlage ihrer Auswirkungen auf die Vorhersagen der Subjektklassen zu erstellen.

⁴⁹ <https://www.cs.waikato.ac.nz/ml/weka/>

⁵⁰ Diese Methode misst die Signifikanz eines Merkmals anhand des Informationsgewinns, der in Bezug auf die Zielklasse berechnet wird. Für weitere Informationen siehe Gnanambal et al. (2018).

Rang	ML-Merkmal in Park Arum & Hong Munpyo (2015)	
1	F7	Honorativinfix (<i>eu</i>) <i>si</i>
2	F6	Verbalsuffix mit sprecherorientierter Modalität
3	F1	den Imperativmarker
4	F9	Topik-NP innerhalb der vorherigen vier Sätze
5	F10	Subjekt-NP innerhalb der vorherigen vier Sätze
6	F2	den Interrogativmarker
7	F11	Objekt-NP innerhalb der vorherigen vier Sätze
8	F4	subjektives Adjektiv im Deklarativsatz
9	F3	den Exhortativmarker
10	F8	subjektives Adjektiv + den Interrogativmarker <i>sibnikka</i> / <i>seyo</i>
11	F5	objektives Adjektiv im Deklarativsatz
12	F12	Adverb-NP innerhalb der vorherigen vier Sätze

Tabelle 14: ML-Merkmale in der Reihenfolge des größten Einflusses auf die Klassifizierung von Subjektklassen (nach Park Arum & Hong Munpyo 2015: 210–211)

Die ML-Merkmale auf den Plätzen 1 bis 3 in Tabelle 14, die als besonders einflussreich für die korrekte Klassifizierung von Subjektklassen bewertet werden, enthalten Anhaltspunkte, von denen erwartet wird, dass sie zur Wiederherstellung von Adressaten- oder Sprecherreferenten beitragen können. An erster Stelle steht dabei das Honorativinfix (*eu*)*si*. Dies kann als empirischer Nachweis für die Beobachtungen aus früheren Studien gesehen werden, dass das Honorativinfix (*eu*)*si* in hohem Maße zur Resolution von adressatenbezogenen Nullsubjekten im Koreanischen beiträgt. Park Arum & Hong Munpyo (2015: 211) berichten, dass das Honorativinfix (*eu*)*si* in den insgesamt 1.000 Sätzen 432 Mal gefunden wurde, während die Subjektklasse *Sie* 413 Mal vorkam. Daher vermuten sie, dass das ML-Modell wahrscheinlich die Subjektklasse *Sie* als die am wahrscheinlichsten zutreffende Klasse lernen würde, wenn das ML-Merkmal F7 den Wert 1 hat. Darüber hinaus berichten sie, dass einige der Sätze mit Imperativmarkern das Honorativinfix *si* enthalten, weshalb sie überlegen, ob die Kombination der beiden ML-Merkmale F7 und F1 zur Vorhersage der Subjektklasse *Sie* beigetragen haben könnte. An zweiter Stelle steht das ML-Merkmal in Bezug auf verschiedene Verbalsuffixe, die bekanntermaßen hilfreich sind, um Nullsubjekte mit Referenz auf den Sprecher wiederherzustellen. Park Arum & Hong Munpyo (2015: 211) geben an, dass die im ML-Merkmal F6 kodierten Verbalsuffixe in den 1.000 Sätzen insgesamt 195-mal gefunden wurden, wovon 188 Fälle sprecherbezogene Subjekte aufweisen. Daraus schließen sie, dass das ML-Modell wahrscheinlich die Subjektklasse *ich* als die am wahrscheinlichsten zutreffende Klasse gelernt hätte, wenn die Hinweise, die die sprecherorientierte Modalität ausdrücken können, auftraten.

Zusammenfassend kann festgestellt werden, dass die ML-Merkmale, die einen erheblichen Einfluss auf die korrekte Klassifizierung von Subjektklassen ausüben, in Beziehung zu den Strategien stehen, die in früheren Studien als hilfreich bei der Wiederherstellung der Referenten von diskursteilnehmerbezogenen Nullsubjekten genannt wurden. Allerdings ist bei der Verwendung von ML nicht immer klar, ob bestimmte in den ML-Merkmalen kodierte Hinweise tatsächlich einen Beitrag zur Klassifizierung bestimmter Klassen leisten. Dies liegt daran, dass ML-Modelle als Blackboxes bekannt sind und ihre Vorhersagen nicht in einer für Menschen verständlichen Weise erklärt werden können (vgl. Rudin 2019: 1).

In der vorliegenden Arbeit wird darum argumentiert, dass regelbasierte Ansätze besser geeignet sind, um nachvollziehbar zu zeigen, ob bestimmte morphologische und lexikalische Hinweise in Prädikaten tatsächlich auf bestimmte Referenten von Nullsubjekten verweisen können. Daher wird in der vorliegenden Arbeit ein NRS mit heuristischen Regeln entwickelt, da es als vorteilhafter angesehen wird als nicht deterministische Ansätze, um den Nutzen einzelner Strategien bei der Zuordnung bestimmter Referenten zu analysieren. Das in dieser Arbeit entwickelte NRS hat den Vorteil, dass es einfach zu implementieren ist, da es auf leicht auffindbaren Hinweisen in Prädikaten basiert. Es

erfordert keine zusätzliche syntaktische oder semantische Begutachtung und auch keine Diskursanalyse. Darüber hinaus ist es nicht notwendig, die Art der Nullsubjekte zu klassifizieren und dementsprechend unterschiedliche Resolutionsstrategien anzuwenden. Es wird erwartet, dass das NRS einen Großteil der Referenten der koreanischen Nullsubjekte in Gesprächen rekonstruieren kann, insbesondere derjenigen, die sich auf Diskursteilnehmer beziehen. Ein weiterer Vorteil dieser Studie besteht darin, dass die Leistung der Regeln anhand einer großen Menge von Nullsubjekten bewertet wird, um die Wirksamkeit der unterschiedlichen Strategien zu bestimmen. Im Vergleich zu Hong Munpyo (2011), der 207 Nullsubjekte für die Evaluierung eines regelbasierten Ansatzes verwendet hat, greift die vorliegende Arbeit auf mehr als 15.000 Nullsubjekte als Input für das NRS zurück. Dies ermöglicht zuverlässigere Ergebnisse bei der Überprüfung der Leistung der heuristischen Regeln, um die Hypothese dieser Arbeit zu überprüfen, dass explizite Hinweise im Verbalkomplex eine wichtige Rolle bei der Rekonstruierbarkeit diskurstteilnehmerbezogener Nullsubjekte im Koreanischen spielen.

3. Datengrundlage

Die Notwendigkeit neuer Annotationsdaten für die Untersuchung dieser Studie wurde bereits betont, und sowohl die Datenquelle des Korpus zur Generierung der Annotationsdaten als auch der Zweck der annotierten Daten wurden in Kapitel 1.2 vorgestellt. Im folgenden Kapitel werden Einzelheiten darüber gegeben, wie die Rohdaten für die Annotationen über die Nullsubjekte und ihre Verweisstruktur aufbereitet wurden (Kapitel 3.1), sowie wie die Annotationen gemäß einem optimierten Annotationsschema verarbeitet wurden (Kapitel 3.2). Abschließend werden die wichtigsten Informationen und Charakteristika der in dieser Studie generierten annotierten Daten erklärt (Kapitel 3.3).

3.1. Aufbereitung der Daten für die Annotation

Dieses Kapitel bietet einen Überblick über die Struktur der verwendeten Korpora und beschreibt die Schritte zur Verarbeitung der Rohdaten, um ein besseres Verständnis für die Aufbereitung eines annotierbaren Datensatzes für koreanische Nullsubjekte zu vermitteln.

3.1.1. Originalkorpus und Zusatzdaten von ETRI

Der koreanische Datensatz, der als Grundlage für die Generierung der Annotationsdaten in dieser Arbeit dient, wird als *Originalkorpus* bezeichnet. Diese Bezeichnung dient zur Unterscheidung von den zusätzlichen Satzdaten, die speziell für die vorliegende Studie bei ETRI, dem Forschungsinstitut, das auch den Originalkorpus bereitstellt, angefordert wurden und hier als *Zusatzdaten* bezeichnet werden. Diese Zusatzdaten enthalten wichtige Informationen über die Wortart (im Folgenden POS) und syntaktische Strukturen, die zur Vorbereitung der Rohdaten für die Annotation verwendet werden. Sie sind besonders wichtig für die Markierung von Verb- und Nomen-Markables, die wiederum die Identifizierung von Prädikat-Kandidaten ermöglichen, die potenzielle Nullsubjekte haben können, sowie von Antezedens-Kandidaten für Nullsubjekte. Diese Informationen sind auch von entscheidender Bedeutung für die Entwicklung der heuristischen Regeln für das NRS, wie sie im detaillierten Kapitel 4 beschrieben werden.

Im Folgenden liegt der Schwerpunkt auf der Strukturierung des Originalkorpus und der Zusatzdaten. Abbildung 3 zeigt den koreanischen Datensatz aus dem Originalkorpus.

	한국어 문장
S000001	길이 왜 이렇게 막히는 겁니까?
S000002	지금은 러시아워라서 차가 많습니다.
S000003	막히지 않는 다른 길로 돌아갈 수 없나요?
S000004	이 시간에는 어느 도로든지 막혀서 어쩔 수 없습니다.
S000005	이 속도로 가다가는 비행기를 놓치겠는데요.
S000006	공항까지 빨리 가는 길이 있지만 너무 돌아가야 해서 탑승비가 더 듭니다.
S000007	그렇게 해서 빨리 갈 수 있다면 더 저렴할게요.
S000008	알겠습니다. 그럼 우회도로를 이용해서 가겠습니다.
S000009	3시 안에 공항에 도착할 수 있도록 해주세요.
S000010	네, 최대한 빨리 가보겠습니다.
S000011	여보세요? 거기 택시회사인가요?
S000012	네, 무엇을 도와드릴까요?
S000013	호텔 앞으로 10시까지 택시를 예약했는데 오지 않았어요.
S000014	예약자 성함이 어떻게 되시나요?
S000015	홍길동입니다. 저희가 식구와 짐이 많아서 두 대를 예약했는데요.
S000016	아, 홍길동님. 택시가 갔다가 손님이 없다고 도로 있습니다.
S000017	저희는 호텔 앞에서 계속 기다리고 있었는데요?
S000018	뉴욕호텔이 맞으신가요?
S000019	제가 뉴욕호텔이라고 했나요? 제가 실수했네요. 죄송합니다. 워싱턴호텔이에요.
S000020	그럼 그곳으로 지금 곧 보내드리도록 하겠습니다.

Abbildung 3: Beispiel für die koreanischen Sätze aus dem Originalkorporus

Abbildung 3 zeigt zwei Dialoge aus dem Originalkorporus, die durch eine Leerzeile getrennt sind. Um die Dialogeinheiten voneinander zu unterscheiden, wurde in dieser Studie jedem Dialog eindeutige Nummer zugewiesen. Zum Beispiel wurde der erste Dialog mit der Nummer "1" und der zweite Dialog mit der Nummer "2" gekennzeichnet. Der erste Dialog besteht aus insgesamt zehn Zeilen, die im Originalkorporus mit der ID S000001 bis S0000010 nummeriert sind. Jede Zeile kann jedoch mehrere Sätze enthalten, da sie einer Rederunde eines Sprechers entspricht. Ein Beispiel hierfür ist die Zeile mit der ID S000008, die zwei Sätze enthält, die durch den Trennstrich | getrennt sind, weil sie von einem Sprecher stammen.

Zur Anwendung in dieser Studie wurde das Originalkorporus bearbeitet, indem die Zeilen aufgeteilt wurden, wenn sie mehrere Sätze enthielten. Jeder Satz erhielt eine eindeutige Satz-ID. Durch diese Aufteilung der Zeilen mit mehreren Sätzen wurde ermittelt, dass das Originalkorporus insgesamt 2.407 Dialoge und 30.241 Sätze umfasst. Weiterhin wurden die Sprecherinformationen aus dem Originalkorporus extrahiert, die den Gesprächsteilnehmern zugeordnet werden konnten. Da das Originalkorporus ausschließlich Gespräche zwischen zwei Teilnehmern enthält, wurden den Beteiligten abwechselnd die Bezeichnungen "A" und "B" zugeordnet. Abbildung 4 veranschaulicht diese Arbeitsschritte anhand des ersten Dialogs aus dem Originalkorporus.

1	1	A
2	1	B
3	1	A
4	1	B
5	1	A
6	1	B
7	1	A
8	1	B
9	1	B
10	1	A
11	1	B

Abbildung 4: Beispiel für die verarbeiteten Informationen über die Satz-ID, Dialog-ID und die beteiligten Sprecher

In der ersten Spalte von Abbildung 4 sind die Satz-IDs von "1" bis "11" aufgeführt, da sie zu den elf Sätzen des ersten Dialogs gehören. Die zweite Spalte enthält die Dialog-ID, die als "1" angegeben ist. Die dritte Spalte zeigt den jeweiligen Sprecher jedes Satzes, entweder "A" oder "B". Zum Beispiel wurden den Sätzen mit den Satz-IDs 8 und 9 der Buchstabe "B" zugeordnet, da sie beide von Sprecher B stammen.

Die folgende Abbildung stellt die Datenstruktur eines Satzes in den Zusatzdaten dar.

```

9 그럼 우회도로를 이용해서 가겠습니다.
GIZAInput: 그림 우회도로를 이용하 아서 가 겠 습니다 .
1 4 modi 그러면 그림/C[pos=C, subpos=coor, morphpos=C]
2 3 obj 우회도로/N[pos=N, sem=길, subpos=base, morphpos>NN]+를/P[pos=P, subpos=obje, morphpos=FO]
3 4 clau 이용하+어서 이용하/V[form=verb, pos=V, subpos=base, morphpos=NN]+아서/E[pos=E, subpos=oend, morphpos=EC]
4 0 root 가+겠+으니다 가/v[mood=decl, pos=V, subpos=base, tense=futu, morphpos=VB]+겠/E[pos=E, subpos=pend, tense=futu, morphpos=EP]+습니
다/E[mood=decl, pos=E, subpos=send, morphpos=EE]
5 4 punc . ./S[pos=S, subpos=punc, morphpos=SF]

```

Abbildung 5: Datenstruktur eines Satzes in den Zusatzdaten des Originalkorpus

In der ersten Zeile befinden sich die Satz-ID und der Satz selbst. In der zweiten Zeile wird der tokenisierte Satz-String angezeigt. Das Unterkapitel der zweiten Zeile enthält Informationen zur Dependenz und Morphologie⁵¹ der im Satz enthaltenen Tokens. Der in Abbildung 5 gezeigte Satz besteht aus insgesamt fünf Tokens mit den entsprechenden Token-IDs von 1 bis 5. Die Dependenzbeziehung jedes Tokens wird auf der rechten Seite der Token-ID durch Zahlen angegeben. Darüber hinaus werden Informationen zur syntaktischen Funktion jedes Tokens bereitgestellt. Die morphologischen Segmente jedes Tokens sind durch das Symbol + getrennt, und für jedes Morphem werden zusätzliche Informationen in eckigen Klammern "[]" angegeben, einschließlich der POS ("form", "pos", "subpos", "morphpos"), semantischer Information ("sem"), des Satzmodus ("mood") und des Tempus ("tense").

Im nächsten Kapitel wird detaillierter erläutert, wie die Tokens und die zugehörigen POS-Informationen zur Markierung von Markables bei der Generierung der Rohdaten für die Annotation verwendet werden.

3.1.2. Annotationstool und Rohdaten

MMAX2 ist das ausgewählte Annotationstool für die Annotationen in dieser Arbeit (siehe Kapitel 1.2). Um die Rohdaten für die Annotation in MMAX2 aufzubereiten, wurden die Zusatzdaten in das CoNLL-Format konvertiert. Das CoNLL-Format ist ein standardisiertes Dateiformat zur Darstellung linguistischer Annotationen. Dabei wird jedem Token im Text eine Zeile in einer tabellarischen Struktur zugeordnet, wobei die verschiedenen Spalten der Tabelle unterschiedliche Informationen für jedes Token repräsentieren. In der CoNLL-Datei, die in dieser Arbeit erzeugt wurde, werden die Informationen sowohl aus dem Originalkorpus als auch aus den Zusatzdaten genutzt. Abbildung 6 zeigt beispielhaft, wie der Satz aus Abbildung 5 in das CoNLL-Format umgewandelt wurde.

1	그림	9	그림	C	1	B
2	우회도로를	9	우회도로	N	1	B
3	이용해서	9	이용하	V	1	B
4	가겠습니다	9	가	V	1	B
5	.	9	.	S	1	B

Abbildung 6: Die Äußerung aus Abbildung 5 in der Datei im CoNLL-Format

In der ersten Spalte sind aufsteigende Zahlen ab 1 aufgeführt, die den Token-IDs in Abbildung 5 entsprechen. Die zweite Spalte zeigt die Token-Strings, während die dritte Spalte die Satz-ID angibt. In den Spalten vier und fünf finden sich der Kopf jedes Tokens und die zugehörigen POS-Informationen, die aus der fünften Spalte der Zusatzdaten stammen und durch den Schrägstrich "/" getrennt sind. Die sechste und siebte Spalte beinhalten die Dialog-ID des Satzes sowie die Sprecherinformationen, die aus dem Originalkorpus extrahiert wurden.

Die erstellte CoNLL-Datei diente als Grundlage für die Rohdaten der Annotation. In den Rohdaten wurden die Sätze jedes Dialogs jeweils in einer Zeile dargestellt und zwischen den Dialogen zwei

⁵¹ Für weitere Informationen über den vom ETRI verwendeten POS-Tagger und Dependenzparser, die zur Bereitstellung der Zusatzdaten genutzt wurden, vgl. die Arbeiten von Na Seung-Hoon (2014; 2015).

Leerzeilen eingefügt, um die Dialoggrenzen zu kennzeichnen. Darüber hinaus wurden Verb- und Nomen-Markables definiert. Die Verb-Markables repräsentieren Prädikat-Kandidaten, die potenziell Nullsubjekte aufweisen können, während die Nomen-Markables als Kandidaten für das Antezedens von Nullsubjekten dienen. Basierend auf den POS-Informationen aus der CoNLL-Datei wurden Tokens, die die POS-Informationen V (Verb) oder J (Adjektiv) aufwiesen, als Verb-Markables klassifiziert, während Tokens mit der POS-Information N (Nomen) als Nomen-Markables eingestuft wurden. Aufgrund von Beschränkungen der Dateigröße, die vom MMAX2-Programm auf einmal geladen werden kann, wurde die CoNLL-Datei in Projekte aufgeteilt, wobei jeder Dialog bis zu 5.000 Tokens pro Projekt enthielt (vgl. Abbildung 7).



Abbildung 7: Erster Teil des ersten MMAX2-Projekts zur Annotation von Nullsubjekten

In Abbildung 7 sind alle Sätze des ersten Dialogs sowie einige Sätze des zweiten aus dem Originalkorpus zu sehen. Die beiden Dialoge sind durch Leerzeilen getrennt, wie sie bei der Generierung der Rohdaten festgelegt wurden. Die jeweiligen Sätze jedes Dialogs sind in aufeinanderfolgenden Zeilen angeordnet. Da der erste Dialog insgesamt elf Sätze umfasst, werden in Abbildung 7 insgesamt elf Zeilen für den ersten Dialog angezeigt. Es ist zu beachten, dass Tokens, die zu den Nomen-Markables gehören, in Fettdruck in eckigen Klammern "[]" hervorgehoben sind, während Tokens, die zu den Verb-Markables gehören, mit den geschweiften Klammern "{}" gekennzeichnet sind. Dies erleichtert die Identifizierung von Prädikaten und potenziellen Antezedens-Kandidaten.

Im Annotationsschema dieser Studie erfolgt eine direkte Markierung von Prädikaten, wenn das zugehörige Subjekt nicht explizit ausgedrückt ist. Diese Herangehensweise unterscheidet sich vom bislang üblichen Annotationsschema zur Kennzeichnung koreanischer Nullsubjekte, bei dem nicht realisierte Subjekte identifiziert und direkt in die entsprechenden Sätze eingefügt werden (vgl. die Erläuterungen zur Nutzung in Kapitel 3.2). Wenn demnach ein Prädikat annotiert wird, weil es ein Nullsubjekt hat, wird das Verb-Markable in MMAX2 blau hervorgehoben, wie in Abbildung 8 veranschaulicht wird.



Abbildung 8: Beispiel für das annotierte MMAX2-Projekt mit hervorgehobenen Prädikaten mit Nullsubjekt

Details zum Annotationsprozess stehen im folgenden Kapitel im Fokus. Zusätzlich wird erörtert, welche Kriterien zur Identifizierung von Nullsubjekten verwendet werden und warum in der vorliegenden Studie die Prädikate nur annotiert werden, anstatt die identifizierten Nullsubjekte direkt einzufügen.

3.2. Annotation

Die Annotation der Daten erfolgt durch Identifizierung von Prädikaten, die entweder anaphorische oder deiktische Nullsubjekte aufweisen, sowie durch deren Zuordnung zu Referenten oder Antezedenzen. Hierbei kommt das bereits erwähnte Tool MMAX2 zum Einsatz. In diesem Kapitel werden zunächst zwei Ansätze bestehender Annotationsschemata für koreanische Nullelemente vorgestellt, bevor das in dieser Arbeit entwickelte Schema im Detail erläutert wird. Dabei werden die Vorteile dieses Schemas im Vergleich zu den existierenden Schemata diskutiert (Kapitel 3.2.1). Anschließend wird ein Überblick über den Annotationsprozess gegeben, in dem erläutert wird, wie die beiden Typen von Nullsubjekten nach dem in dieser Arbeit entwickelten Annotationsschema annotiert werden (Kapitel 3.2.2).

3.2.1. Etablierte und optimierte Annotationsschemata für koreanische Nullsubjekte

In Bezug auf die Annotation von Nullelementen im Koreanischen existieren zwei Vorgehensweisen, die sich danach richten, ob Nullelemente markiert werden oder nicht. Das erste Schema, das als "zeroless"-Annotationsschema bezeichnet wird, zeichnet sich dadurch aus, dass Nullelemente nicht speziell markiert werden. Ein Beispiel für die Anwendung dieses Schemas findet sich in der Sejong Treebank (ST)⁵². Im zweiten Schema wird im Gegensatz dazu die Bedeutung einer Nullmarkierung betont. Forscher der Penn Korean Treebank (PKT) haben das sogenannte "zero mark-up"-

⁵² Die ST wurde im Rahmen des *21st Century Sejong Project* erstellt, das im Jahr 1998 mit dem Ziel gestartet wurde, umfangreiche nationale koreanische Korpora zu entwickeln. Diese Korpora umfassen eine breite Palette von Texten, darunter Romane, Zeitungsartikel, Märchen und mehr. Eine umfassende Beschreibung dieses Projekts und der enthaltenen Korpustypen findet sich bei Kim Hung-gyu et al. (2007).

Annotationsschema eingeführt, um die Nutzbarkeit eines annotierten Korpus zu steigern.⁵³ Das folgende Beispiel (12)a, in dem weder das Subjekt noch das Objekt realisiert ist, verdeutlicht die Unterschiede zwischen den beiden Analyseansätzen von ST und PKT (zitiert aus Lee Sun-Hee et al. 2005: 589).

- (12) a. *eojobam 12si-e bad-ass-seubnida-*
 gestern Nacht 12 Uhr-ADV erhalten-PST-SFS.DECL-SF
 'Gestern Nacht um 12 Uhr erhielt (ich/er/sie) (es).'

- b. STs Analyse
 (VP (AP *eojobam*/MAG)
 (VP (NP_AJT 12/SN + *si*/NNB + *e*/JKB)
 (VP *bad*/VV + *ass*/EP + *seubnida*/EF + ./SF)))
 c. PKTs Analyse
 (S (NP-SUBJ **pro**))
 (VP (NP-ADV *eojobam*/NNC)
 (NP-ADV 12/NNU *si*/NNX + *e*/PAD)
 (VP (NP-OBJ **pro**))
bad/VV + *ass*/EPF + *seubnida*/EFN)). ./SFN).

In (12)b erfolgt die Analyse des Satzes ohne die Berücksichtigung des Nullsubjekts und des Nullobjekts, da im ST-Schema keine Nullelemente annotiert werden. In (12)c wird das Annotationsergebnis für (12)a in der PKT präsentiert, wobei das Nullsubjekt und das Nullobjekt als "pro" eingefügt wurden. Das "zeroless"-Annotationsschema wie in (12)b kann die Belastung der Annotationsaufgabe reduzieren, weil Nullelemente in koreanischen Treebanks nicht berücksichtigt werden. Dieses "zeroless"-Annotationsschema stellt jedoch keine Informationen über Nullsubjekte bereit. Hinsichtlich der Resolution von Nullsubjekten erweist sich das "zero mark-up"-Annotationsschema wie in (12)c als die vorteilhaftere Option, weil es anzeigt, welcher Satz ein Nullsubjekt hat, was der erste Schritt zur Resolution ist. In Lee Sun-Hee et al. (2005) wird ebenfalls betont, dass die Kennzeichnung fehlender Subjekte in Treebanks für verschiedene NLP-Aufgaben wichtig ist, wie beispielsweise die Anaphernresolution. Die PKT-artige Annotation enthält jedoch keine Informationen darüber, worauf sich ein Nullsubjekt bezieht.

Han Na-Rae (2006) schlug ihrerseits ein "zero mark-up"-Annotationsschema vor und wandte dies in der PKT an. Ihr Annotationsschema beinhaltet nicht nur die Identifizierung der Nullelemente, sondern auch die Analyse ihrer Verweisstrukturen. Zur Identifizierung der Nullelemente stützt sie sich auf bereits in der PKT eingefügte Nullelemente, die als "pro" markiert sind. Im nächsten Schritt wurden diese Nullelemente danach kategorisiert, ob sie eine anaphorische oder deiktische Beziehung darstellen. Wenn ein Nullelement in einer anaphorischen Beziehung steht, wird es zusammen mit seinem Antezedens mit demselben numerischen Index annotiert. Ein deiktisches Nullelement hingegen erhält einen alphabetischen Index. Zum Beispiel wird der Sprecher mit dem Index "i" und der Adressat mit dem Index "y" annotiert.⁵⁴ Falls sich ein deiktisches Nullelement auf eine bereits im Diskurs realisierte Entität bezieht, wird sowohl dem Nullelement als auch der Entität zusätzlich

⁵³ Die Unterschiede zwischen den ST- und den PKT-Annotationen in Bezug auf Kasusmarker, subjektlose Konstruktionen, Verbalsuffixe und das Wh-Pronomen-Tagging werden in Lee Sun-Hee et al. (2005: 594) ausführlich beschrieben. Diese Quelle bietet tiefgehende Einblicke in die spezifischen Unterschiede und Annotationstechniken in den beiden Korpora.

⁵⁴ Für das vollständige Kodierungsschema, das bei der Annotation verwendet wurde, siehe Han Na-Rae (2006: 157).

derselbe numerische Index zugeordnet. Das folgende Beispiel (zitiert aus Han Na-Rae 2006: 158) zeigt das Ergebnis der Anwendung dieses erweiterten Annotationsschemas anhand einiger Äußerungen in einem Diskurs in der PKT.

- (13) A1. *ja-*, *Kim teugmujang₂* \emptyset_{y2} *anj-eula-*.
 Also-NF Kim Officer (SBJ) sich setzen-IMP-SF
 'Also, Offizier Kim, setz dich.'
- A2. \emptyset_{y2} *mwo₃* *jom* *meog-eoss-na?*
 (SBJ) etwas bisschen essen-PST-SFS.INT-SF
 'Hast du etwas gegessen?'
- B1. *ye-*, \emptyset_{i2} \emptyset_3 *meog-eoss-seubnida-*.
 ja-NF (SBJ) (OBJ) essen-PST-SFS.DECL-SF
 'Ja, ich habe etwas gegessen.'

Das Beispiel (13) ist ein Auszug aus einem Gespräch zwischen einem Kommandanten A und einem Unteroffizier B. Dieser Dialog umfasst drei Äußerungen, in denen A zunächst B auffordert, sich zu setzen (A1), und anschließend A B fragt, ob er etwas gegessen hat (A2). In seiner Antwort bestätigt B, dass er gegessen hat (B1). In allen drei Äußerungen sind die Subjekte nicht explizit ausgedrückt und könnten sich gleichermaßen auf den Sprecher oder den Adressaten beziehen. Zusätzlich wird in der letzten Äußerung das Objekt des Verbs *meogda* 'essen' nicht ausdrücklich genannt.

Han Na-Rae (2006) verwendet das Symbol Ø, um diese Nullelemente zu markieren, und fügt sie in die entsprechenden Sätze ein.⁵⁵ Darüber hinaus hat sie den Nullelementen gemäß ihres Kodierungsschemas numerische und alphabetische Indizes zugewiesen. Im Fall des Nullsubjekts in A1 wird der Buchstaben-Index "y" verwendet, da es sich auf den Adressaten bezieht. Gleichzeitig erhält das Nullsubjekt den Zahlen-Index "2", weil es auf die vorherige NP *Kim teugmujang* 'Offizier Kim' verweist, die ebenfalls mit "2" gekennzeichnet ist. Beim Nullsubjekt in A2 wird genauso wie in A1 der Index "y2" verwendet, da es auf den Adressaten verweist und ebenso auf die NP *Kim teugmujang* in A1 Bezug nimmt. Im Falle des Nullsubjekts in B1 wird dagegen der Buchstaben-Index "i" verwendet, da es sich auf den Sprecher bezieht. Das Nullobject in B1 erhält wiederum es den Zahlen-Index "3", weil es sich auf das Nomen *mwo* 'etwas' in A2 bezieht, das ebenfalls mit dem Zahlen-Index "3" gekennzeichnet ist.

Die annotierten Daten, die unter Verwendung des Annotationsschemas von Han Na-Rae (2006) für die PKT erstellt wurden, scheinen sich gut für die Untersuchung der Anaphernresolution zu eignen, da sie Informationen über Nullelemente und deren Referenten oder Antezedenzen enthalten. Das Annotationsschema von Han Na-Rae (2006) bzw. das "zero mark-up"-Annotationsschema weist jedoch drei wesentliche Probleme auf.

Erstens existiert keine klare Richtlinie dafür, was als Nullsubjekt annotiert werden sollte, was die Zuverlässigkeit der Annotationsergebnisse von PKT bei der Identifizierung von Nullsubjekten verringert und somit auch die Genauigkeit der Annotationsergebnisse von Han Na-Rae beeinträchtigt, da diese auf den Ergebnissen von PKT basieren. Der Hauptgrund dafür liegt in der Uneinheitlichkeit der Grammatikschreibung bezüglich dessen, was überhaupt als Nullsubjekt gilt. Dieses Problem ist keineswegs neu: Auch Lee Sun-Hee et al. (2005) berichten, dass die unklare Zuordnung des Phänomens zu Inkonsistenzen in der PKT-Annotation geführt hat. Die gleichen Sätze werden unterschiedlich klassifiziert, entweder als Sätze mit Subjektellipse, subjektlose Konstruktionen oder sogar als

⁵⁵ In der einschlägigen Literatur, in der das "zero mark-up"-Annotationsschema verwendet wird, ist nicht eindeutig angegeben, an welcher Stelle und nach welchen Kriterien Nullelemente eingefügt werden.

Routineformeln. Die Definition dieser drei Interpretationsarten variiert ebenfalls unter Wissenschaftlern, wobei das entscheidende Unterscheidungskriterium darin besteht, ob eine NP als Subjekt eingesetzt werden kann oder nicht. In Sätzen mit Subjektellipse kann ein overttes Subjekt die ausgelassenen Subjekte ersetzen, während in subjektlosen Konstruktionen oder Routineausdrücken ein overttes Subjekt undenkbar ist.⁵⁶

Die Diskussionen zu diesem Thema konzentrieren sich oft auf das Prädikat *ida* 'sein', dessen Nutzung unterschiedliche Interpretationen zulässt. Beispiel (14) (zitiert aus Ko Seok-Ju 2011: 4) illustriert diese Problematik unterschiedlicher Klassifikationsansätze.

- (14) a. *bul-i-ya!*

Feuer-sein-SFS-SF

'Da ist ein Feuer!'

- b.⁵⁷ *ja-, seonmul-i-ya-*.

hier-NF Geschenk-sein-SFS-SF

'Hier, (das) ist (dein) Geschenk.'

Beim Beispiel (14)a handelt es sich um eine Redewendung, die typischerweise dazu verwendet wird, auf einen Brand aufmerksam zu machen, während (14)b eine Äußerung darstellt, die ein Sprecher benutzen kann, wenn er dem Adressaten ein Geschenk überreicht. Abhängig von der Perspektive des Forschers können die Äußerungen in (14) entweder als subjektlose Konstruktion (vgl. Lee Seung-Wook 1969; Kim Mi-Young 2003) oder als Sätze mit weggelassenem Subjekt (vgl. Park Yang Kyu 1980) analysiert werden. Alternativ können sie als Routineformeln betrachtet werden, wodurch eine Diskussion über ihr Subjekt überflüssig scheint (vgl. Hong Jaesung 1997; Nam Kisim 2001).

Das Hauptkriterium für die Annotation von Nullsubjekten in dieser Studie ist, ob ein overttes Subjekt in einen Satz eingefügt werden könnte, der ein Nullsubjekt haben sollte. Nach diesem Kriterium wird (14)a in dieser Studie nicht annotiert, da es im Koreanischen ungewöhnlich klingt, ein Subjekt hinzuzufügen (z.B. *?igeos-i* 'das-SBJ' *bul-i-ya!* 'ist Feuer'). Äußerungen wie in (14)b, in denen explizite Pronomen wie *igeo* 'das' und *neo* 'dein' hinzugefügt werden, klingen dagegen ebenso natürlich wie die Varianten ohne overttes Subjekt. Daher werden die letzteren Fälle in der vorliegenden Studie als Sätze mit einem Nullsubjekt analysiert und dementsprechend annotiert. In dieser Studie werden solche Äußerungen als Annotationsgegenstand betrachtet, die natürlicher klingen, wenn ihnen eine explizite Subjektversion hinzugefügt wird, wie auch im Beispiel (14)b. Der Annotationsgegenstand dieser Studie ist daher der Fall, in dem eine NP nicht in der Subjektposition eines Satzes realisiert wird und ein overttes Subjekt denkbar ist. Da die Beurteilung, ob ein overttes Subjekt denkbar ist, sehr subjektiv sein kann, schlägt diese Studie vor, die Äußerungsabsicht in der Sprechsituation, in der der Satz tatsächlich verwendet wird, zu berücksichtigen, anstatt sich nur auf die Struktur des Satzes selbst und die Bedeutung des Prädikats zu konzentrieren. Auch wenn dieses Kriterium verwendet wird, gibt es immer noch Fälle, in denen eine Routineformel in das Annotationsergebnis aufgenommen wird, weil ein overttes Subjekt denkbar ist. Solche Fälle werden später durch manuelle Überprüfung aus dem Annotationsergebnis ausgeschlossen und dienen nicht als Input für NRS. Entsprechende Beispiele sind in Kapitel 3.3.2 dokumentiert.

Das folgende Beispiel veranschaulicht die in dieser Studie entwickelten Kriterien zur Identifizierung von Sätzen mit Nullsubjekt. In (15) werden in zwei Äußerungen Wünsche für eine gute Nacht an den

⁵⁶ Die Unstimmigkeiten, die in vorherigen Untersuchungen zum koreanischen Nullsubjekt aufgetreten sind, werden ausführlich in Kapitel 2 von Ko Seok-Ju (2011) behandelt.

⁵⁷ Ab dieser Stelle werden ausgelassene Nullelemente, einschließlich Nullsubjekte, nicht direkt in die koreanischen Beispiele eingefügt, sondern in der deutschen Übersetzung mit Klammern markiert. Dies geschieht, da in dieser Studie nicht die Position der ausgelassenen Elemente im Satz von Interesse ist, sondern vielmehr die Informationen im Prädikat, die zur Resolution von Nullsubjekten beitragen. Dies entspricht auch dem Annotationsschema für Nullsubjekte, das in der vorliegenden Arbeit vorgeschlagen wird.

Adressaten gerichtet. In der ersten Äußerung wird das Prädikat *doeda* 'werden' verwendet, welches neben dem Prädikat *ida* 'sein' häufig in Diskussionen über koreanische subjektlose Konstruktionen erwähnt wird (vgl. Park Chang-Young 2014: 84). Hingegen wird in der zweiten Äußerung das Prädikat *bonaeda* 'verbringen' verwendet.

- (15) a. *joheun bam doe-seyo-*.
 gut Nacht werden-SFS.IMP-SF
 'Schönen Abend noch.'
- b. *joheun bam bonaе-seyo-*.
 gut Nacht verbringen-SFS.IMP-SF
 'Schönen Abend noch.'

Aus der Perspektive der Betrachtung von (15)a als subjektlose Konstruktion erfordert das Prädikat *doeda* 'werden' zwei Substantivargumente, X und Y. Dabei nimmt es die syntaktische Struktur [X-i/ga Y-i/ga *doeda*] 'X wird Y' an, wobei X die Subjektfunktion und Y die Komplementfunktion des Subjekts übernimmt. Die semantische Beziehung zwischen X und Y wird als im Wesentlichen äquivalent analysiert (vgl. Park Chang-Young 2014: 83–84). Ein Beispiel für diese typische syntaktische Struktur von *doeda*, das aus Kim Minsun (2005: 51) entnommen ist, lautet: "*yeonghui-ga* (X, 'Yeonghui') *seonsaengnim-i* (Y, 'Lehrerin') *doe-eoss-da*" 'Younghee ist Lehrerin geworden'.

Die Analyse von (15)a als subjektlose Konstruktion basiert darauf, dass es nicht möglich ist, eine Subjekt-NP anzunehmen, die semantisch äquivalent zu der NP *joheun bam* 'gute Nacht' ist. Beispielsweise könnte eine NP, die sich auf den Adressaten bezieht, als Subjekt-NP für (15)a betrachtet werden. Allerdings können der Adressat und *joheun bam* 'gute Nacht' aus der Perspektive der Analyse von (15)a als subjektlose Konstruktion nicht als semantisch äquivalent betrachtet werden.

Es sei jedoch darauf hingewiesen, dass das Prädikat *doeda* in einer Vielzahl von Bedeutungen verwendet wird. Laut Kim Minsun (2005: 39) findet sich das Prädikat *doeda* in drei koreanischen Wörterbüchern mit insgesamt 14 verschiedenen Bedeutungen, wobei die erforderlichen Argumente je nach Bedeutung erheblich variieren. Kim Minsun (2005: 146–148) klassifiziert Sätze wie (15)a als eine spezielle Verwendung von *doeda*, wobei *doeda* in (15)a 'Zeit verbringen' bedeutet. Sie erklärt, dass das Prädikat *doeda* in dieser Bedeutung besonders häufig in Befehlen vorkommt. Da *doeda* in Satz (15)a im Sinne von Zeit verbringen verwendet wird, lässt sich dieser Satz als äquivalent zu Satz (15)b analysieren, in dem das Prädikat *bonaeda* 'verbringen' verwendet wird, um den Wunsch des Sprechers auszudrücken, dass der Adressat eine gute Nacht haben möge. Im Fall von (15)b dürfte die Auffassung, dass sich das nicht realisierte Subjekt auf den Adressaten bezieht, wenig umstritten sein. Im Fall von (15)a, der in derselben Äußerungssituation austauschbar verwendet werden kann, ist die Analyse, dass sich das nicht realisierte Subjekt auf den Adressaten bezieht, nicht unvernünftig.

Es kann also davon ausgegangen werden, dass beide Äußerungen in Beispiel (15) die beiden in dieser Studie vorgeschlagenen Kriterien zur Identifizierung von Sätzen mit Nullsubjekt erfüllen. Erstens sind ihre Subjekte nicht realisiert, und zweitens können die nicht realisierten Subjekte als NPs betrachtet werden, die sich auf den Adressaten beziehen. Daher werden beide als Nullsubjektsätze analysiert und in dieser Studie als Annotationsgegenstand berücksichtigt.

In dieser Studie wird postuliert, dass konsistente Annotationsergebnisse erreicht werden können, wenn Sätze mit Nullsubjekten anhand der beiden in dieser Studie vorgeschlagenen Kriterien bestimmt werden. Es ist jedoch schwierig zu sagen, ob der Vorschlag dieser Studie den gewünschten Effekt haben wird, da nur eine Annotatorin die Annotationsaufgabe für diese Studie auf der Grundlage der oben genannten Kriterien durchgeführt hat. Dieser Aspekt sollte in zukünftigen Studien überprüft werden, indem untersucht wird, wie konsistent die Identifizierung von Nullsubjekten durch verschiedene Annotatoren ist, wenn sie die vorgeschlagenen Kriterien bei der Bestimmung von Sätzen mit koreanischen Nullsubjekten verwenden.

Zweitens besteht eine Herausforderung bei der Markierung von Nullelementen mit Platzhaltern wie dem Symbol Ø oder "pro" an einer bestimmten Stelle in den entsprechenden Sätzen. Wie bereits von Han Na-Rae (2004) betont, ist die genaue Position unsichtbarer Elemente schwer zu bestimmen, insbesondere im Koreanischen, wo das sogenannte *Scrambling* eine Herausforderung darstellt. Satzbestandteile einschließlich der Subjekte können je nach Sprecherabsicht frei angeordnet werden, und dementsprechend stehen Nullsubjekte nicht zwangsläufig vor der Verbalphrase (kurz: VP).⁵⁸ Die präzise Bestimmung der Position frei angeordneter Nullelemente ist bereits eine anspruchsvolle Aufgabe, und diese Schwierigkeiten können die Annotationsaufgaben komplexer machen, was wiederum zu Inkonsistenzen in den annotierten Daten führen kann.

In dieser Studie wird ein Annotationsschema vorgeschlagen, das die Annotation eines Verbalkomplexes mit Nullsubjekt ermöglicht, ohne dass das Nullsubjekt direkt an einer bestimmten Stelle im zugehörigen Satz eingefügt wird. Die reine Markierung sichtbarer Verbalkomplexe im vorgeschlagenen Annotationsschema dient dazu, die Komplexität der Annotation zu verringern. Annotatoren können Verbalkomplexe einfach mit Nullsubjekten kennzeichnen, ohne die exakte Position der Nullsubjekte im Auge behalten zu müssen. Dieser Ansatz wird voraussichtlich den Annotationsprozess beschleunigen, indem die Last der Annotation, die durch unsichtbare und flexible Elemente verursacht wird, reduziert wird.

Darüber hinaus erweist sich das vorgeschlagene Annotationsschema als vorteilhaft für beide Arten von koreanischen Nullsubjekten: in deiktischer und anaphorischer Verwendung. Für deiktische Nullsubjekte ist bekannt, dass Prädikate verschiedene morphosyntaktische Informationen enthalten, um ihre Referenten zu identifizieren (vgl. Kapitel 2.1.3.2). Die Markierung von Verbalkomplexen, die solche Informationen enthalten, ermöglicht einen direkten Zugriff auf diese identifikationsrelevanten Daten. Die Entscheidung, Prädikate zu annotieren anstatt Nullsubjekte in den Satz einzufügen, erweist sich auch im Fall anaphorischer Nullsubjekte als vorteilhaft. Durch die Verwendung definierter Positionen von Verbalkomplexen anstelle undefinierter Positionen von Nullsubjekten wird die koreferentielle Verknüpfung zwischen Prädikaten mit Nullsubjekten und ihren Antezedenzen zuverlässiger hergestellt. Dies gewinnt insbesondere dann an Bedeutung, wenn der Abstand zwischen einem Nullsubjekt und seinem potenziellen Antezedens berechnet wird, da die Verwendung unklarer Positionen von Nullsubjekten zu Inkonsistenzen bei der Referenzpunktzuordnung führen kann. Aus diesem Grund bietet das vorgeschlagene Annotationsschema zuverlässige und konsistente Annotations, indem es klare Referenzpunkte für die Verknüpfung koreferentieller Beziehungen bereitstellt. Der genaue Ablauf der Annotation gemäß diesem Schema wird im Überblick über den Annotationsprozess in MMAX2 beschrieben (siehe Kapitel 3.2.2).

Drittens ist die Kategorisierung der koreanischen Nullelemente in deiktische und anaphorische Typen keine optimale Option, da diese Unterscheidung unscharf ist. Ähnlich äußert sich auch Han Na-Rae (2006), die in ihrem Annotationsschema die unscharfe Unterscheidung zwischen anaphorischen und deiktischen Nullsubjekten hervorhebt. Zur besseren Handhabung schlägt sie in ihrem Annotationsschema vor, dass beide Verwendungen in solchen Fällen gleichzeitig mit Buchstaben- und Zahlenindizes markiert werden sollten.

Anstatt also eine Zweiteilung der Nullsubjekttypen anzunehmen, folgt die vorliegende Studie der Einteilung von Hong Munpyo (2011) in drei Subtypen Extra-NS, Inter-NS und Intra-NS (vgl. Kapitel 2.1.2). Im Vergleich zum binären Kategoriensystem für koreanische Nullsubjekte, wie es von Han Na-Rae (2006) verwendet wurde, bietet das in dieser Studie eingeführte Kategoriensystem entscheidende

⁵⁸ Mok Jungsoo (2004) erläutert, dass im gesprochenen Koreanisch nicht nur die Satzkonstituenten weggelassen werden können, sondern auch die Position der Konstituenten nicht festgelegt ist. Als Beispiel nennt er verschiedene Äußerungen, bei denen die Position der Konstituenten im gleichen Satz variiert, und zeigt auf, dass solche Sätze in gesprochenen koreanischen Daten häufig auftreten und als natürliche Äußerungen betrachtet werden müssen.

Vorteile. Es zeichnet sich durch die Fähigkeit aus, die Motivationen hinter der Verwendung der Nullform im Koreanischen sowie das sprachliche Umfeld der koreanischen Nullsubjekte präziser widerzuspiegeln. Ein bedeutender Fortschritt besteht darin, dass dieses Kategoriensystem den Nachteil der Unschärfe des bisherigen Systems von Han Na-Rae überwindet, indem es Nullsubjekten erlaubt, lediglich einer Unterkategorie zugeordnet zu werden. Zudem ermöglicht die Übertragung der Ergebnisse aus dem neuen Kategoriensystem auf das binäre System eine verbesserte Vergleichbarkeit und Interpretation, insbesondere durch die klare Zuordnung von Nullsubjekten des Typs Extra-NS zu deiktischen Nullsubjekten und solchen des Typs Inter-NS oder Intra-NS zu anaphorischen Nullsubjekten.

3.2.2. Annotationsprozess nach dem optimierten Annotationsschema

Die in Kapitel 3.2.1 genannten Herausforderungen bei der Identifizierung von koreanischen Nullsubjekten und bei der Bestimmung ihrer genauen Position in Sätzen sowie ihre klassifikatorischen Probleme zeigen die Notwendigkeit auf, für die Analyse der Nullsubjekte eine praxisorientierte Schablone zu entwerfen, die die Nullsubjekte einwandfrei identifiziert und voneinander abgrenzt.

In dieser Studie schlage ich ein "zero mark-up"-Annotationsschema vor, bei dem explizite Prädikate annotiert werden, anstatt die Nullsubjekte direkt an bestimmte Positionen in den Sätzen einzufügen. Der Grund für diesen Vorschlag ist, dass es im Sinne der Nullsubjektresolution wichtig ist, in den annotierten Daten Informationen darüber zu liefern, welches Prädikat ein Nullsubjekt hat, und nicht, wo es sich im Satz befindet. Zum besseren Verständnis des optimierten Annotationsschemas wird im Folgenden ein Überblick über den Annotationsprozess gegeben, wie anaphorische und deiktische Nullsubjekte, die in dieser Arbeit im Mittelpunkt der Annotation stehen, mit dem Annotationstool MMAX2 annotiert werden.

Um den Annotationsprozess für die beiden Typen von Nullsubjekten zu illustrieren, werden zwei Äußerungen aus einem Dialog des Originalkorpus extrahiert. Die Äußerung in Beispiel (16) dient als Beispiel für ein anaphorisches Nullsubjekt, da ein explizites Antezedens im Diskurs vorhanden ist. Zwei wichtige Aufgaben bei der Annotation von anaphorischen Nullsubjekten sind zunächst die Bestimmung eines Prädikats, das ein Nullsubjekt hat, und dann die Suche nach seinem Antezedens.

(16)	[<i>taegsi-ga</i>	<i>ga-ass-daga</i>]S1	[<i>sonnim-i</i>	<i>eobs-dago</i>]S2	[<i>dolo</i>	<i>o-ass-seubnida-</i> .]S3
	Taxi-SBJ	hinfahren-PST-	Fahrgäste-	nicht sein-CNJ	wieder	zurückkommen-PST-
	CNJ	SBJ				SFS.DECL-SF

'Das Taxi ist hingefahren und (es) ist wieder zurückgekommen, weil keine Fahrgäste (da) waren.'

Die Äußerung in (16) besteht aus insgesamt drei Sätzen, wobei das Prädikat in S3 ein Nullsubjekt aufweist. Wenn ein Prädikat mit einem Nullsubjekt identifiziert wird, ist es erforderlich, auch das Antezedens des Nullsubjekts zu bestimmen. Gemäß der Gesprächssituation bezieht sich das Nullsubjekt in S3 auf *taegsi* 'Taxi', das wiederum das Subjekt von S1 ist. Daher sollte das Nomen *taegsi* in S1 als Antezedens des Nullsubjekts in S3 annotiert werden.⁵⁹

Abbildung 9 zeigt, wie das Nullsubjekt und sein Antezedens in (16) mit dem in dieser Studie optimierten Annotationsschema annotiert werden.

⁵⁹ Beim Annotieren des Antezedens eines Nullsubjekts in einem anaphorischen Kontext wurde das nächstgelegene Antezedens des Nullsubjekts ausgewählt, falls es im Diskurs mehrere potenzielle Antezedenzen gab. Diese Vorgehensweise entspricht der in früheren Studien beobachteten Tendenz, dass Antezedenzen tendenziell in der Nähe von Nullpronomen liegen, wie in Kapitel 2 erläutert.

[택시가] (갔다가) [손님이] (없나고) 도로 {왔습니다} .

Abbildung 9: Annotation eines Nullsubjekts und seines Antezedens in MMAX2

In Abbildung 9 befinden sich drei Verb-Markables, die durch geschweifte Klammern gekennzeichnet sind. Das letzte Verb-Markable entspricht dem dritten Prädikat in (16) und es verfügt über ein Nullsubjekt. Nach dem Annotationsschema in dieser Studie wird das Prädikat als ein neues Markable auf der "Zero"-Ebene erstellt, die als eine vordefinierte Ebene fungiert, auf der alle Prädikate mit Nullsubjekten gespeichert werden. Das Ergebnis der Annotation zeigt sich dann in der Anzahl der neuen Markables auf der "Zero"-Ebene, was Auskunft darüber gibt, wie viele Prädikate in den analysierten Daten Nullsubjekte aufweisen. Nachdem das Verb-Markable auf der "Zero"-Ebene erstellt wurde, was auf ein Nullsubjekt hinweist, ist es notwendig, sein Antezedens zu annotieren. Insgesamt sind zwei Nomen-Markables mit eckigen Klammern markiert, wobei das erste Nomen-Markable als Antezedens des Nullsubjekts des dritten Verb-Markable identifiziert wurde. Dieses Nomen-Markable wird dann als das Antezedens des Nullsubjekt durch MMAX2 festgelegt, und die anaphorische Verwendung des Nullsubjekts wird in MMAX2 durch eine blaue Linie zwischen dem Nomen-Markable und dem Verb-Markable mit dem Nullsubjekt dargestellt, wie in Abbildung 9 gezeigt.

Der Annotationsprozess für Nullsubjekte in deiktischer Verwendung erfolgt auf fast die gleiche Weise wie bei anaphorischen Nullsubjekten hinsichtlich der Annotation der Prädikate, die Nullsubjekte aufweisen. Die Äußerung in (17) stammt aus einem Dialog des Originalkorpus und enthält ein deiktisches Nullsubjekt. Diese Äußerung erfolgte in einer Situation, in der der Sprecher dem Adressaten erklärt, dass er bald ein Taxi zu dem Ort schicken würde, an dem sich der Adressat befindet.

- (17) *geuleom geugos-eulo jigeum god bonae-deuli-dolog ha-gess-seubnida-*.
 dann dort-LOC jetzt gleich chicken-AUX-PPC-IFX-SFS.DECL-SF
 'Dann schicke (ich) (Ihnen) jetzt gleich (ein Taxi).'

Das fettgedruckte koreanische Prädikat in (17) wird in MMAX2 auf der "Zero"-Ebene als ein neues Markable erstellt, da es ein Nullsubjekt aufweist. In diesem Fall ist es erforderlich, seinen Referenten als den Sprecher zu identifizieren. Für die Annotation des Referenten als deiktisches Nullsubjekt wird ein zusätzliches Fenster aktiviert:

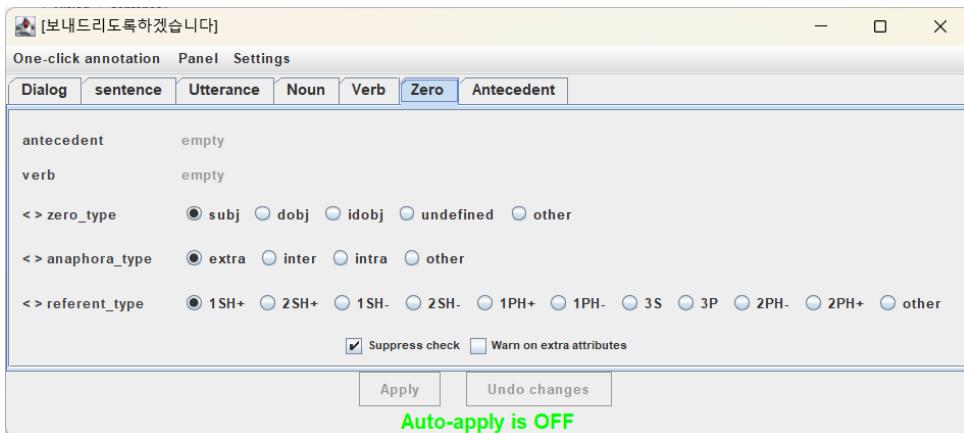


Abbildung 10: Annotationsfenster für zusätzliche Informationen über Nullsubjekte mit deiktischem Referenten

Drei Annotationsoptionen sind verfügbar, die in dem Annotationsfenster zusätzlich annotiert werden:

1. "zero_type",
2. "anaphora_type",
3. "referent_type".

Die ersten beiden Annotationsoptionen gelten nicht nur für deiktische, sondern auch für anaphorische Nullsubjekte. Die erste Annotationsoption "zero_type" dient der Annotation der grammatischen

Funktion eines Nullelements. Im Rahmen des Annotationsprojekts dieser Arbeit wird ausschließlich der Wert "subj" gewählt, da sich die Studie auf die Nullsubjekte unter den Nullelementen konzentriert. Wenn der Analysegegenstand auf andere Nullelemente wie Nullobjekte erweitert wird, können für die erste Annotationsoption auch andere Werte gewählt werden.

Die zweite Annotationsoption "anaphora_type" ist für die Annotation der Nullsubjekttypen und bei dieser Annotationsoption stehen die vier Werte "extra", "inter", "intra" und "other" zur Verfügung. Die Bestimmung der Nullsubjekttypen basiert auf den Kriterien von Hong Munpyo (2011), die in Kapitel 2.1.2 erläutert wurden. Demnach sind Nullsubjekte für den Wert "extra" solche Fälle, in denen die Referenten durch linguistische Mittel, die in den Prädikaten enthalten sind, abgeleitet werden können. Falls die Referenten der Nullsubjekte nicht durch linguistische Mittel identifiziert werden, sondern sie explizite Antezedenzen in Diskurs oder abgeleitete kontextuelle Referenten haben, werden die weiteren Werten "inter", "intra" oder "other" gewählt. So wird ein Nullsubjekt als "extra" annotiert, wenn im Prädikat ein Hinweis auf seinen Referenten vorhanden ist, selbst wenn eine Entität, die als mögliches Antezedens des Nullsubjekts fungieren könnte, bereits im Diskurs realisiert wurde. Dies erfolgt aufgrund der Annahme, dass in diesem Fall das explizit realisierte linguistische Mittel im Prädikat die primäre Motivation dafür darstellt, dass das Subjekt nicht ausgedrückt wird.

Wenn in einem Prädikat keine besonderen Hinweise vorliegen und ein Nullsubjekt dazu verwendet wird, sich auf etwas bereits im Diskurs Realisiertes zu beziehen, aber nicht im direkt vorangehenden Satz, erfolgt die Klassifizierung des Nullsubjekts als "inter". Wenn zum Beispiel zu Beginn des Diskurses ein bestimmtes Gesprächstopik eingeführt wird und das Gespräch später auf dieses Topik zurückkommt, wird ein Nullsubjekt, das sich auf das Topik bezieht, als "inter" annotiert.

Wenn im Prädikat keine speziellen Hinweise auf den Referenten des Nullsubjekts vorhanden sind und im unmittelbar vorangehenden Prädikat bereits ein explizites Antezedens existiert, wird der Typ des Nullsubjekts als "intra" annotiert. Diese Zuordnung basiert auf der Hypothese, dass eine enge Verknüpfung zwischen dem vorangehenden und dem nachfolgenden Satz möglicherweise die Ursache dafür ist, dass das Subjekt des nachfolgenden Satzes nicht ausgedrückt wird. In Hong Munpyo (2011) wurde die Motivation für diese Art von Nullsubjekt als Subject-Sharing von Konjunktionen beschrieben. Sonstige Fälle werden mit dem Wert "other" annotiert. Dies betrifft Nullsubjekte, die weder durch linguistische Mittel eindeutig bestimmbar sind noch ein explizites Antezedens im Diskurs aufweisen, sondern deren Referenten aus dem Kontext erschlossen werden müssen. Ein Beispiel hierfür könnte ein Nullsubjekt sein, das sich auf eine allgemeine Öffentlichkeit bezieht.

Die dritte Annotationsoption "referent_type" wird aktiviert, wenn der Wert "extra" für die zweite Annotationsoption ausgewählt wird. Die Werte dieser Annotationsoption beziehen sich auf die grammatische Person bzw. das Numerus der Referenten. An dieser Stelle muss eines von zehn Pronomen oder ein Freitext hinterlegt werden.

Tabelle 15 gibt eine Übersicht über die deutschen und koreanischen Personalpronomen im Nominativ, auf die sich die Werte der Annotationsoption "referent_type" beziehen:

Person/ Numerus	Deutsch	Koreanisch	"referent_type"	Koreanisch	"referent_type"
		[-Honorativ]		[+Honorativ]	
1. Person Singular	'ich'	<i>na</i>	1sh-	<i>jeo</i>	1sh+
1. Person Plural	'wir'	<i>uli</i>	1ph-	<i>jeohui</i>	1ph+
2. Person Singular	'Sie/ du'	<i>neo</i>	2sh-	<i>dangsin</i>	2sh+
2. Person Plural	'Sie'	<i>neohuideul</i>	2ph-	<i>dangsindeul</i>	2ph+
3. Person Singular/Plural	'er/ sie/ es/ sie'	<i>geu/</i> <i>geunyeo/</i> <i>geugeos</i>	3s	<i>geudeul/</i> <i>geunyeodeul/</i> <i>geugeosdeul</i>	3p

Tabelle 15: Werte von "referent_type" und die entsprechenden deutschen und koreanischen Personalpronomen

In Tabelle 15 sind fünf verschiedene Arten von Personalpronomen im Nominativ im Deutschen aufgeführt, die auf der Grundlage der grammatischen Person und des Numerus unterschieden werden. Im Koreanischen gibt es aufgrund der Kategorie Respekt zehn verschiedene Arten von Personalpronomen, die als Werte in der Annotationsoption "referent_type" zur Verfügung stehen. Zum Beispiel bezieht sich der Referent des Nullsubjekts in (17) auf den Sprecher und sollte, wenn er explizit ausgedrückt wird, im Koreanischen in der Höflichkeitsform realisiert werden, da das Verbalsuffix *seubnida* die Verwendung eines Pronomens in der Höflichkeitsform erfordert. In diesem Fall wird für die Annotationsoption "referent_type" der Wert "1sh+" gewählt, der das Pronomen *jeo* (1. Person, Singular, +Honorativ) kodiert.

3.3. Charakteristiken der annotierten Daten

Im vorliegenden Kapitel werden grundlegende statistische Informationen zu den annotierten Daten präsentiert, um die Charakteristika des Analysekorpus in Bezug auf die Nullsubjekte zu erfassen. Dies umfasst einen Überblick über die gewonnenen Erkenntnisse (Kapitel 3.3.1). Darauf folgt eine detaillierte Darstellung der verschiedenen Arten von aus den annotierten Daten ausgeschlossenen Prädikaten. Diese sorgfältige Bereinigung ist von entscheidender Bedeutung, um sicherzustellen, dass ausschließlich relevante Prädikate mit Nullsubjekten in die Analyse einfließen und als Input für das NRS dienen (Kapitel 3.3.2).

3.3.1. Eckdaten der annotierten Rohdaten

Das Originalkorpus, das als Grundlage für die Generierung der Annotationsdaten diente, umfasste insgesamt 2.407 Dialoge, wovon 58 Duplikate identifiziert wurden. Diese Dialoge bestanden aus insgesamt 30.241 Sätzen (siehe Kapitel 3.1.1). Nach Entfernung der Duplikate blieben 2.349 Dialoge mit 29.506 Sätzen übrig, die für die Generierung der Rohdaten zur Verfügung standen. Aufgrund der begrenzten Datenkapazität, die in ein MMAX2-Projekt geladen werden kann, wurden die Dialoge in Einheiten von bis zu 5.000 Tokens unterteilt, weshalb insgesamt 26 separate MMAX2-Projekte als Rohdaten für Annotation erstellt wurden.

In diesen Rohdaten waren insgesamt 38.157 Markables auf der "Verb"-Ebene vorhanden, was mit der Anzahl der Prädikate in den Zusatzdaten übereinstimmt, die entweder als Adjektive (J) oder Verben (V) getaggt waren. Diese Prädikatskandidaten wurden als potenzielle Prädikate betrachtet, die Nullsubjekte aufweisen können, wobei bei insgesamt 17.266 neuen Markables auf der "Zero"-Ebene erstellt wurden. In dieser Kohorte wurden 1.383 Prädikate von der Analyse manuell ausgeschlossen.

Dies geschah entweder, weil sie nicht dem Fokus dieser Studie entsprachen und somit keine Nullsubjekte in finiten Sätzen darstellten, oder aufgrund von Fehlern. Zu diesen Fehlern zählten Segmentierungs- und Parsingfehler sowie andere Ungenauigkeiten, die bereits in den Zusatzdaten vorhanden waren. Diese Zusatzdaten enthalten zusätzliche Informationen zum POS und zur syntaktischen Struktur aus dem Originalkorporus (vgl. Kapitel 3.1.1). Diese Fehler machten die Einbeziehung der betroffenen Prädikate in die Analyse unzulässig. Eine umfassende Erklärung zu den Kategorien der ausgeschlossenen Prädikate ist im Kapitel 3.3.2 dokumentiert.

Folglich verbleiben insgesamt 15.883 Prädikate mit der gleichen Menge Nullsubjekte. Diese Zahlen verdeutlichen, dass im so generierten Teilkörper durchschnittlich 0,54 Nullsubjekte pro Satz vorhanden sind, da 15.883 Nullsubjekte in 29.506 Sätzen identifiziert wurden.⁶⁰ In Tabelle 16 finden sich die Ergebnisse des Vergleichs der durchschnittlichen Anzahl von Nullsubjekten pro Satz im Analysekörper dieser Studie im Vergleich zu den Ergebnissen aus der Literatur.

Korpus	Genre	Quelle	Anzahl Nullsubjekte	Anzahl Sätze	Nullsubjekte pro Satz
PKT	Dialogsequenzen im militärischen Bereich	Han Na-Rae (2006)	4.074	5.083	0,80
KTB2	Nachrichtenartikel		7.820	5.010	1,56
Korpus 1	Chats im Messenger	Hong Munpyo (2011)	191	358	0,53
Korpus 2	Dramaskripte		244	357	0,68
Originalkorporus	Dialogsequenzen im Reisebereich	Die vorliegende Arbeit	15.883	29.506	0,54

Tabelle 16: Anzahl der koreanischen Nullsubjekte in verschiedenen Korpora

In der Studie von Han Na-Rae (2006) wird die Frequenz von Nullsubjekten in zwei Korpora analysiert: dem PKT⁶¹ (Penn Korean Treebank1) und dem KTB2 (Penn Korean Treebank2). Das Korpus PKT besteht hauptsächlich aus Dialogsequenzen aus dem militärischen Kontext und enthält insgesamt 5.083 Sätze. In PKT wurden insgesamt 4.074 Nullsubjekte ermittelt, was einer durchschnittlichen Anzahl von 0,80 Nullsubjekten pro Satz entspricht. Das Korpus KTB2 setzt sich aus 647 Nachrichtenartikeln zusammen und umfasst insgesamt 5.010 Sätze. In diesem Korpus wurden insgesamt 7.820 Nullsubjekte identifiziert. Die Analyse ergab, dass im KTB2 durchschnittlich 1,56 Nullsubjekte pro Satz vorkommen. Han Na-Rae (2006: 146) erklärt, dass die größere Häufigkeit von Nullelementen pro Satz im KTB2 darauf zurückzuführen ist, dass die Sätze in diesem Korpus etwa 2,5-mal länger sind als im Korpus PKT. Gemäß Han Na-Rae (2006) beläuft sich die Wortanzahl für das Korpus PKT auf 54.503, während das KTB2 132.040 Wörter umfasst.⁶² Wenn man die Anzahl der Nullsubjekte durch die Anzahl der Wörter in beiden Korpora teilt, ergibt sich eine durchschnittliche Anzahl von Nullsubjekten pro Wort von 0,07 in PKT und 0,06 in KTB2. Dies bedeutet, dass die durchschnittliche Anzahl von Nullsubjekten pro Wort in PKT höher ist.

⁶⁰ Wenn man die Anzahl der Nullsubjekte pro Satz berechnet, indem man die 1.383 ausgeschlossenen Prädikate von den insgesamt 29.506 Sätzen ausschließt (wobei jedes ausgeschlossene Prädikat als ein eigener Satz betrachtet wird), ergibt sich ein Durchschnitt von 0,56 Nullsubjekten pro Satz. Ich werde das Ergebnis der Berechnung von 0,54 Nullsubjekten pro Satz verwenden, um die Anzahl der Nullsubjekte pro Satz mit anderen Korpora zu vergleichen. Der geringfügige Unterschied zwischen 0,54 und 0,56 Nullsubjekten pro Satz hat keine signifikante Auswirkung auf den Vergleich mit anderen Korpora.

⁶¹ Das Korpus PKT entspricht dem Korpus, das zuvor in Kapitel 1.2 als PKT (Penn Korean Treebank) bezeichnet wurde. In dieser Arbeit wird es weiterhin als PKT bezeichnet, obwohl Han Na-Rae (2006) es als KTB1 bezeichnet, um es von KTB2 zu unterscheiden.

⁶² Die Wortanzahl für die Korpora PKT und KTB2 beinhaltet laut Han Na-Rae (2006) nicht nur die Anzahl der lexikalischen Wörter, sondern es werden auch die Tokens einschließlich der Symbole und Satzzeichen berücksichtigt, die neben den Wörtern während der Tokenisierung erfasst wurden.

Hong Munpyo (2011) wiederum wählt zwei Korpora aus, um die Häufigkeit des Auftretens von Nullsubjekten zu untersuchen. Sein Korpus 1 besteht aus Messenger-Chats und umfasst insgesamt 358 Sätze. Sein Korpus 2 setzt sich aus Dramaskripten zusammen und enthält insgesamt 357 Sätze. Im Korpus 1 wurden insgesamt 191 Nullsubjekte gefunden, was einem Durchschnitt von 0,53 Nullsubjekten pro Satz entspricht. Im Korpus 2 wurden insgesamt 244 Nullsubjekte identifiziert, was einem Durchschnitt von 0,68 Nullsubjekten pro Satz entspricht. Die vermehrte Auslassung von Subjekten in Dramaskripten im Vergleich zu Chats wird von Hong Munpyo auf visuelle Elemente wie Szenenbeschreibungen zurückgeführt, die in Dramaskripten vorhanden sind und mehr Kontext zur Handlung bieten. Diese visuelle Information könnte das vermehrte Auslassen von Subjekten gegenüber textbasierten Chats erklären (vgl. Hong Munpyo 2011: 421).

Das in der vorliegenden Untersuchung verwendete Originalkorpus beinhaltet Dialogsequenzen, die in Reisesituationen auftreten können. Insgesamt wurden 15.883 Nullsubjekte in 29.506 Sätzen identifiziert, was bedeutet, dass die durchschnittliche Anzahl von Nullsubjekten pro Satz bei 0,54 liegt. Im Vergleich zu den anderen Korpora in Tabelle 16 zeigt das Originalkorpus eine ähnlich hohe durchschnittliche Anzahl von Nullsubjekten pro Satz wie in Hong Munpyos (2011) Korpus aus Messenger-Chats (0,54 gegenüber 0,53).

Wenn man die Merkmale der fünf in Tabelle 16 aufgeführten Korpora vergleicht, fällt auf, dass lediglich KTB2 ein Korpus ist, das keine Dialoge beinhaltet. Stattdessen enthält es kurze Nachrichten über aktuelle Themen wie Weltnachrichten oder Politik. Die anderen vier Korpora ähneln sich hingegen insofern, als sie aus Gesprächen bestehen, bei denen Sprecher und Adressaten bzw. Sender und Empfänger miteinander interagieren.

Han Na-Rae (2006) analysiert das Vorkommen von Nullelementen im PKT, insbesondere in Bezug auf Diskurse, da einige Teile von PKT isolierte Sätze darstellen, die keine zusammenhängenden Diskurssegmente bilden. Im diskursbezogenen Abschnitt des PKT, der 2.689 Sätze umfasst, wurden insgesamt 2.468 Nullelemente identifiziert, was zu einer durchschnittlichen Häufigkeit von etwa 0,9 Nullelementen pro Satz führt. Besonders bemerkenswert ist, dass Nullsubjekte etwa 93,4 % der identifizierten Nullelemente im PKT ausmachen, was zu einer durchschnittlichen Häufigkeit von etwa 0,86 Nullsubjekten pro Satz führt.

Zusammenfassend lässt sich feststellen, dass in den vier Korpora, die interaktive Gespräche enthalten, die Häufigkeit von Nullsubjekten pro Satz zwischen mindestens 0,53 (Korpus 1, Hong Munpyo 2011) und höchstens 0,86 (PKT, Han Na-Rae 2006) liegt. Dies legt nahe, dass in koreanischen Gesprächen, bei denen Menschen miteinander interagieren, entsprechend den Werten dieser vier Korpora durchschnittlich 0,65 Nullsubjekte auftreten können. Ob die Auftretenshäufigkeit von Nullsubjekten im Vergleich zu anderen Sprachen hoch ist, müssen weitere Untersuchungen klären.⁶³

Neben der durchschnittlichen Häufigkeit des Auftretens von Nullsubjekten pro Satz ist auch die Auftretenshäufigkeit der Nullsubjekte nach Typen von Interesse. In Tabelle 17 wird die Verteilung der Werte für den "anaphora_type" in den annotierten Daten dieser Studie mit der Verteilung der Nullsubjekte nach Typen in den beiden Korpora von Hong Munpyo (2011) verglichen. Da die Kriterien für die Klassifizierung der Nullsubjekttypen in dieser Arbeit auf den Kriterien von Hong Munpyo (2011) basieren, ist ein direkter Vergleich der Auftretenshäufigkeiten in den drei Korpora möglich. Es ist

⁶³ Han Na-Rae (2006: 146) berichtet, dass in der Penn Chinese Treebank durchschnittlich 0,036 chinesische Nullelemente einschließlich *pro* und *PRO* pro Wort gefunden wurden. Dies legt nahe, dass koreanische Nullelemente tatsächlich ein weit verbreitetes Phänomen sein könnten. Han Na-Rae (2006) verzeichnet 0,080 Nullelemente pro Wort in PKT und 0,062 Nullelemente pro Wort in KTB2, selbst wenn nur *pro* in den koreanischen Korpora betrachtet werden. Um jedoch einen detaillierten Vergleich mit dem koreanischen Phänomen durchzuführen, bedarf es weiterführender Analysen, da Han Na-Rae (2006) weder den Kontext der chinesischen Treebank ausführlich erörtert noch die syntaktischen Funktionen der chinesischen Nullelemente im Detail beschreibt.

jedoch zu beachten, dass ein Vergleich mit Vorsicht zu genießen ist, da die Korpora unterschiedliche Größen haben und somit auch unterschiedliche Ergebnisse liefern können.

Korpus	Quelle	Nullsubjekttyp			Total
		Extra-NS	Inter-NS	Intra-NS	
Korpus 1	Hong Munpyo (2011)	84,82 % (n=162)	9,95 % (n=19)	5,24 % (n=10)	100,00 % (N=191)
Korpus 2		78,28 % (n=191)	19,67 % (n=48)	2,05 % (n=5)	100,00 % (N=244)
Originalkorpus	Die vorliegende Arbeit	90,26 % (n=14.336)	7,51 % (n=1.193)	2,23 % (n=354)	100,00 % (N=15.883)

Tabelle 17: Anteil der Nullsubjekte nach Typen in drei Korpora

Es gibt insgesamt weniger Nullsubjekte in den beiden Korpora von Hong Munpyo (2011), weil sie weniger Sätze umfassen, und zwar lediglich etwa 1,2–1,5 % der Nullsubjekte, die im Originalkorpus in der vorliegenden Studie erfasst wurden. Darüber hinaus unterscheiden sich die drei Korpora, wie in Tabelle 17 dargestellt, in Bezug auf ihre jeweiligen Domänen: Korpus 1 umfasst Messenger-Chats, Korpus 2 Dramaskripte und das Originalkorpus Dialoge in Reisesituationen. Obwohl die Anzahl der gefundenen Nullsubjekte in den einzelnen Korpora variiert und die Domänen unterschiedlich sind, zeigt das Verhältnis der Nullsubjekte jedes Typs zur Gesamtzahl der Nullsubjekte in den drei Korpora eine gewisse Ähnlichkeit. Dies könnte darauf zurückzuführen sein, dass alle drei Korpora aus interaktiven Gesprächen zwischen den Teilnehmern bestehen.

In allen drei Korpora traten die Nullsubjekte des Typs Extra-NS am häufigsten auf, gefolgt von Inter-NS, während die Nullsubjekte des Typs Intra-NS am seltensten anzutreffen waren. Diese Beobachtung lässt sich dahingehend interpretieren, dass bis zu 90 % der Subjekte ohne explizites Antezedenz im Diskurs nicht ausdrücklich genannt werden. Angesichts der Tatsache, dass im Analysekörper dieser Studie bis zu 90 % der Nullsubjekte des Typs Extra-NS gefunden wurden, ist es wahrscheinlich, dass bis zu 90 % dieser Nullsubjekte mithilfe der expliziten Hinweise in den Verbalkomplexen rekonstruiert werden können. Dies ergibt sich daraus, dass eines der Kriterien zur Identifizierung von Nullsubjekten in der Kategorie Extra-NS darin besteht, ob in den Prädikaten Hinweise zur Wiederherstellung ihrer Referenten vorhanden sind. Dies deutet auch darauf hin, dass bis zu 90 % der Nullsubjekte mit den in Kapitel 4 vorgestellten heuristischen Regeln erkannt werden können, da sie auf linguistischen Merkmalen basieren, die explizit in den Prädikaten vorhanden sind. Eine detaillierte Leistungsbewertung in Kapitel 5 wird zeigen, ob und inwieweit die expliziten Hinweise im Verbalkomplex zur Rekonstruierbarkeit der diskursteilnehmerbezogenen Nullsubjekte beitragen. Dabei wird analysiert, wie viele Nullsubjekte durch die heuristischen Regeln im NRS erkannt wurden und wie viele davon korrekt ihrem Referenten zugeordnet werden konnten.

3.3.2. Ausgeschlossene annotierte Prädikate

Der zentrale Zweck der erstellten annotierten Daten in dieser Studie besteht darin, die Prädikate, bei denen Nullsubjekte identifiziert wurden, als Input für NRS zu verwenden (siehe Kapitel 1.2). Von den insgesamt 17.266 auf der "Zero"-Ebene annotierten Prädikaten wurden nach manueller Überprüfung 1.383 Prädikate von der Eingabe in das NRS ausgeschlossen, da sie als ungeeignet für die Nutzung im NRS befunden wurden. Tabelle 18 stellt die Art und Anzahl der Prädikate dar, die nicht als Input in NRS eingegeben wurden.

Art der ausgeschlossenen Prädikate		Anteil
(1)	Prädikate mit weiteren syntaktischen Funktionen	67,39 % (n=932)
(2)	Segmentierungsfehler	25,31 % (n=350)
(3)	Routineformel	5,64 % (n=78)
(4)	Satzfehler	0,94 % (n=13)
(5)	Parsing-Fehler	0,72 % (n=10)
Total		100,00 % (N=1.383)

Tabelle 18: Art der ausgeschlossenen Prädikate aus dem Resolutionsprozess

Die nicht ausgedrückten Subjekte der Prädikate, die zu den Arten in Zeile (1) gehören, sind keine Nullsubjekte in finiten Sätzen. Diese Prädikate wurden somit von der Eingabe in das NRS ausgeschlossen, da sie nicht Gegenstand dieser Studie sind. Die Prädikate, die zu den Typen in den Zeilen (2), (4) und (5) gehören, weisen Fehler auf, die bereits im Originalkorpus oder in den Zusatzdaten vorhanden sind. Diese Fehler könnten in den entsprechenden Daten korrigiert werden, aber aus Gründen der Praktikabilität wurden diese Fälle vom Input in NRS ausgeschlossen. Die Prädikate, die den Typen in Zeile (3) zugeordnet sind, haben keine Subjekte. Trotzdem wurden sie nicht als Eingabe für das NRS verwendet, da es unangemessen wäre, das Subjekt explizit anzugeben.

Die ausgeschlossenen Prädikate in Zeile (1) sind solche, die zusätzliche syntaktische Funktionen innerhalb der Sätze ausüben. In dieser Arbeit werden Prädikate dann nicht für die Resolution ausgewählt, wenn sie als satzwertige Erweiterung von Satzgliedern dienen oder Satzglieder modifizieren. Beispiel (18) veranschaulicht die zusätzlichen syntaktischen Funktionen der ausgeschlossenen Prädikate aus Zeile (1) der Tabelle 18.

- (18) a. *jigeum dangjang gochyi-eo deuli-gi-ga* *himdeu-l geos gat-eunde-*,
 jetzt sofort reparieren-AUX.HON-NMZ-SBJ schwierig-PPC(Vermutung)-NF
 'Es jetzt zu reparieren, wäre vermutlich schwierig.'

b. *ne-, eolmakeum hwanjeonha-gi-leul* *wonha-bnikka-?*
 ja-NF wie viel umtauschen-NMZ-OBJ wollen-SFS.INT-SF
 'Ja, wie viel Umtausch möchten Sie?'
 'Wie viel Geld möchten Sie umtauschen?'

c. *chaj-eun jeongbo-neun ttalo jeog-eoya ha-nayo-?*
 finden-ADN Information-TOP extra schreiben-PPC(Pflicht)-SFS.INT-SF
 'Soll ich die gefundenen Informationen separat aufschreiben?'

Die Prädikate *gochyida* 'reparieren', *hwanjeonhada* 'umtauschen' und *chajda* 'finden' in (18) wurden alle in den annotierten Daten auf der "Zero"-Ebene als neue Markables erfasst, da ihre Subjekte nicht ausgedrückt wurden. Dennoch werden sie aus dem Resolutionsprozess ausgeschlossen, da die nicht ausgedrückten Subjekte zu den Prädikaten gehören, die zusätzliche syntaktische Funktionen der finiten Verben in den Sätzen übernehmen. In (18)a fungiert die Phrase mit dem Prädikat *gochyida* 'reparieren' als Subjekt des Prädikats *himdeulda* 'schwierig sein', in (18)b fungiert die Phrase mit dem Prädikat *hwanjeonhada* 'umtauschen' als Objekt des Prädikats *wonhada* 'wollen', und in (18)c fungiert die Phrase mit dem Prädikat *chajda* 'finden' als attributives Adjektiv für das folgende Nomen *jeongbo* 'Information'. In (18)a und (18)b kann die syntaktische Funktion der Prädikate jeweils am Subjektmarker *ga* und Objektmarker *leul* erkannt werden. Der Nominalisator *gi* vor dem Subjekt- und Objektmarker weist zudem auf die Nominalisierung der Prädikate hin. In (18)c zeigt der

Adnominalisator *eun* nach dem Verbstamm *chaj* 'finden' den Gebrauch des Prädikats als Adjektivattribut an.

Die zweite Kategorie der ausgeschlossenen Fälle in Zeile (2) der Tabelle 18 bezieht sich auf Segmentierungsfehler. Hierbei handelt es sich um Fehler, bei denen die Grenzen eines Satzgliedes in den Zusatzdaten, die bei der Aufbereitung der Rohdaten für die Annotation verwendet wurden, falsch identifiziert wurden. Insbesondere wird der Fall ausgeschlossen, dass der Parser gelegentlich einen Verbalkomplex, der als ein Markable auf der "Zero"-Ebene annotiert wurde, nicht als VP analysiert. Es scheint, dass der Parser besonders viele Segmentierungsfehler verursacht hat, wenn die Prädikate periphrastische Konstruktionen enthalten. Diese Art der Fehlsegmentierung wird in Beispiel (19) exemplarisch dargestellt.

- (19) a. *ileum-gwa jeonhwabeonho-leul al su iss-eulkkayo-?*
 Name-und Telefonnummer-OBJ wissen PPC(Möglichkeit)-SFS.INT-SF
 'Kann (ich) (ihren) Namen und (ihre) Telefonnummer wissen?'
 b. *nalja-wa jangso-leul al su iss-eulkkayo-?*
 Datum-und Ort-OBJ wissen PPC(Möglichkeit)-SFS.INT-SF
 'Kann (ich) das Datum und den Ort wissen?'

Die beiden Beispiele in (19) haben eine ähnliche Satzstruktur, da sie jeweils aus zwei NP, die durch die Konjunktion (*g*)*wa* 'und' verbunden sind, und dem Verbalkomplex *al su isseulkkayo* 'wissen' bestehen. Dieser Verbalkomplex umfasst den Verbstamm *al-*, die periphrastische Konstruktion *su iss-* und das Verbalsuffix *eulkkayo*. In Beispiel (20) werden die syntaktischen Analysen dieser beiden Sätze vom Parser in den Zusatzdaten dargestellt. Wie in (20)a zu sehen ist, wurde die VP in (19)a vom Parser nicht als ein Token erfasst, sondern als drei separate Tokens analysiert. (20)b zeigt die korrekte Analyse, bei der der Parser den Verbalkomplex als ein einheitliches Token erkannt hat.

- (20) a. 1 2 modi *ileum+wa*⁶⁴
 2 3 dobj *jeonhwabeonho-leul*
 3 4 modi ***al+l***
 4 5 subj ***su***
 5 0 root ***iss+Ikkayo***⁶⁵
 6 5 punc ?
 b. 1 2 modi *nalja+wa*
 2 3 dobj *jangso-leul*
 3 0 root ***al+Iss+Ikkayo***
 4 3 punc ?

Der Ausschluss von Prädikaten mit Segmentierungsfehlern ist von entscheidender Bedeutung, um die Leistung des in dieser Studie vorgeschlagenen NRS präzise bewerten zu können, da die heuristischen Regeln, die die Grundlage des NRS bilden, Referenten zuordnen, indem sie eine Kombination linguistischer Merkmale in einem Verbalkomplex identifizieren, wie es in den Regeln beschrieben ist. So wird zum Beispiel nach einer Regel einem Prädikat der Referent in der Sprecherrolle zugewiesen, wenn es die periphrastische Konstruktion (*eu*)/*su-iss* in einem Interrogativsatz aufweist. Selbst wenn eine Regel vorhanden ist, die einem Prädikat den richtigen Referenten zuordnen kann, kann sie nicht angewandt werden, wenn ein Verbalkomplex wie in (20)a falsch erkannt wird. Daher wurden

⁶⁴ *wa* ist die Grundform von *gwa* in den Zusatzdaten.

⁶⁵ *Ikkayo* ist die Grundform von *eulkkayo* in den Zusatzdaten.

Prädikate mit Segmentierungsfehlern aus dem Resolutionsprozess ausgeschlossen, um die Regeln optimal nutzen zu können. Auf diese Weise kann auch die Leistung der Regeln angemessen bewertet werden.

Der dritte Typ von Prädikaten (Zeile (3) in Tabelle 18), die von der Analyse ausgeschlossen wurden, betrifft Routineformeln.

- (21) a. ***hwanyeongha-bnida-*** *liseu leseutolang-i-bnida-*.
 willkommen-SFS-SF Lees Restaurant-sein-SFS.DECL-SF
 'Willkommen! Wir sind Lees Restaurant.'
- b. ***sillyeha-jiman-***, *geujjog-eulo ga ju-seyo-*.
 Entschuldigung-aber-NF dort-zu gehen AUX-SFS.IMP-SF
 'Entschuldigung, aber gehen (Sie) bitte dorthin.'

Die fett hervorgehobenen Prädikate in (21) stellen typische Beispiele für Routineformeln dar, die bewusst aus dem NRS ausgeschlossen wurden. Die Prädikate *hwanyeonghada* 'willkommen' und *sillyehada* 'entschuldigen' werden von den Sprecher:innen als feste Ausdrücke oder formelhafte Wendungen wahrgenommen, in denen ein Subjekt in der Sprecherrolle vorstellbar ist, eine explizite Nennung des Subjekts aber unangemessen wäre. Daher erfolgte die bewusste Nichtintegration dieser Routineformeln in das NRS.

Der vierte Typ der ausgeschlossenen Prädikate (Zeile (4) in Tabelle 18) bezieht sich auf Satzfehler, die auf die Verfasser:innen des Originalkorpus zurückzuführen sind. Prädikate werden nicht in das NRS aufgenommen, wenn Teile fehlen, wie in (22)a, oder wenn unnötige Teile hinzugefügt wurden, wie in (22)b:

- (22) a. *eumsig-i manhi nam-aseo geuleo-eonde ssaga-I (su issnayo)-?*
 Essen-SBJ viel übrig sein-weil so sein-und mitnehmen-PPC(Möglichkeit)-SFS-SF
 'Kann ich das Essen mitnehmen, da viel übriggeblieben ist?'
- b. ***guibha-gi-geona*** *hangug-ui jepum-eul loming seobiseu-lo jeonhwanha-si-yeoya ha-bnida-*.
 kaufen-NMZ-oder Korea- Produkt- Roaming Service-zu wechseln-HON-PPC(Pflicht)-
 POSS OBJ SFS.DECL-SF
 'Sie sollten entweder (eine Sim-Karte) kaufen oder zu einem Roaming-Service für das koreanische Produkt wechseln.'

In (22)a fehlt der letzte Teil der periphrastischen Konstruktion, was den Satz unvollständig macht. In (22)b wurde der inkompatible Nominalisator *gi* an den Verbstamm *guibha* angehängt, was den Satz ungrammatisch macht.

Der letzte Typ ausgeschlossener Prädikate (Zeile (5) in Tabelle 18) bezieht sich auf Parsing-Fehler, bei denen die Wortart einer VP falsch zugewiesen wurde. Dies geschieht, weil die Realisierungsform ohne Kontext mehrdeutig interpretiert werden kann. Diese Art von Fehler betrifft jedoch weniger als 1 % der ausgeschlossenen Prädikate. Beispiel (23) veranschaulicht einen Fall mit einem Parsing-Fehler:

Die Phrase *jaego* in (23) kann entweder als die VP 'messen und' oder als die NP 'Vorrat' interpretiert werden. Wie in (23)b zu sehen ist, hat der Parser die Phrase fälschlicherweise als NP identifiziert, obwohl sie tatsächlich eine VP ist.

Auf diese Weise konnten schlussendlich insgesamt 15.883 Prädikate in NRS eingegeben werden.

⁶⁶ In den Zusatzdaten sind die Bestandteile des Prädikats, die dem Schrägstrich / vorausgehen, in Hangul, dem koreanischen Alphabet, geschrieben, aber für die Beispiele in Fließtext der vorliegenden wurden sie romanisiert. Dies gilt auch für die weiteren Beispiele.

4. Nullsubjektresolutionssystem mit heuristischen Regeln

Im folgenden Kapitel stehen die Resolutionsstrategie des NRS in dieser Studie im Fokus. Dabei werden Regeln für verschiedene Kontexte, wie beispielsweise Sätze vom Sprechaktyp Aufforderung oder Verben mit Exhortativmarkern, aufgeführt. Diese Regeln basieren auf linguistischen Merkmalen des Koreanischen, die im Zusammenhang mit Nullsubjekten stehen, und verfolgen die Hypothese, dass in koreanischen Verbalkomplexen Hinweise zur Identifizierung ihrer Referenten zu finden sind (vgl. Kapitel 1.2). Insgesamt werden 212 heuristische Regeln aufgestellt, um Prädikaten, die bestimmte Bedingungen erfüllen, einen Referenten zuzuweisen, wobei das primäre Ziel die Resolution der diskursteilnehmerbezogenen Nullsubjekte ist. Nach der Aufstellung dieser Regeln ist es wichtig zu bestimmen, in welcher Reihenfolge die Regeln im NRS angewandt werden sollen, damit es nicht zu Konflikten zwischen den Regeln kommt.

In Kapitel 4.1 werden die Regeln vorgestellt, die für die Resolution von adressatenbezogenen Nullsubjekten angewandt werden. Dabei stehen Nullsubjekte bei Sprechakten wie direkten (Kapitel 4.1.1) und indirekten Aufforderungen (Kapitel 4.1.2) im Fokus sowie Nullsubjekte bei Verben mit Honorifikum (Kapitel 4.1.3).

In Kapitel 4.2 werden die Regeln zur Wiederherstellung der Referenz von sprecherbezogenen Nullsubjekten vorgestellt. Zu Beginn steht die Regel im Fokus, die sich auf Exhortativmarker konzentriert (Kapitel 4.2.1). Des Weiteren wird besondere Aufmerksamkeit auf Verbalsuffixe mit sprecherorientierter Modalität gelegt (Kapitel 4.2.2). Dabei werden erstmals periphrastische Konstruktionen als Merkmal in den Regeln einbezogen, da diese Mehrwortausdrücke insbesondere in natürlichen Gesprächen mit der Modalität im Koreanischen in Verbindung gebracht werden können. Obwohl diese Konstruktionen vielversprechend für die Wiederherstellung der Referenz sind, wurden sie in der Forschung zu koreanischen Nullsubjekten bisher vernachlässigt. Schließlich werden auch lexikalische Hinweise für die Resolution von Nullsubjekten betrachtet, die darauf hinweisen, dass ein Subjekt als Sprecher zu interpretieren ist (Kapitel 4.2.3).

In Kapitel 4.3 werden die ergänzenden heuristischen Regeln ausführlich erläutert, die auf der Annahme beruhen, dass die in den bisherigen Arbeiten genannten linguistischen Merkmale nicht ausreichen, um die Referenten der im Koreanischen weit verbreiteten diskursteilnehmerbezogenen Nullsubjekte zu identifizieren. Zunächst werden die im Rahmen der vorliegenden Arbeit neuformulierten Regeln für die Erfassung von Nullsubjekten in Nebensätzen und die Vorteile der entsprechenden Regeln im Vergleich zu den Resolutionsstrategien von Hong Munpyo (2011) erläutert (Kapitel 4.3.1). Anschließend werden die Regeln zum Ausschluss von Prädikaten, die aller Wahrscheinlichkeit nach Subjekte in der 3. Person fordern, vorgestellt. Damit soll der Fokus auf den Resolutionsgegenstand der vorliegenden Arbeit gerichtet werden, und zwar auf diskursteilnehmerbezogene Nullsubjekte (Kapitel 4.3.2). Abschließend wird erläutert, wie die vorgeschlagenen Regeln im NRS in ihrer Anwendungsreihenfolge angepasst werden müssen, um Konflikte zwischen den Regeln zu vermeiden, damit sie die Zielreferenten korrekt zuordnen können (Kapitel 4.3.3).

Abschließend präsentiert Kapitel 4.4 den Output des NRS, der Informationen zu den 15.883 Input-Prädikaten des NRS sowie deren zugewiesenen Referenten enthält. Dieser Output ist von entscheidender Bedeutung für eine umfassende Bewertung der Leistung der heuristischen Regeln im NRS.

4.1. Resolution von adressatenbezogenen Nullsubjekten

4.1.1. Imperativmarker in direkten Aufforderungen

Hong Min-pyo (2000) zufolge sind morphosyntaktische Hinweise zur Wiederherstellung der Referenz von Nullsubjekten, die auf den Adressaten referieren, häufig mit direktiven Sprechakten verbunden (siehe Kapitel 2.1.3.2). Ein direktiver Sprechakt gehört zu den fünf Klassen illokutionärer Sprechhandlungen, die von Searle (1976) klassifiziert werden.⁶⁷ Er charakterisiert den direktiven Sprechakt wie folgt:

The illocutionary point of these consists in the fact that they are attempts (of varying degrees, and hence more precisely, they are determinates of the determinable which includes attempting) by the speaker to get the hearer to do something. (Searle 1976: 355)

Laut obiger Beschreibung beabsichtigt der Sprecher mit einem direkten Sprechakt, den Adressaten zu einer bestimmten Handlung zu veranlassen. Der Satztyp, der mit dem direktiven Sprechakt in Verbindung gebracht wird, ist der Imperativsatz, der vor allem zum Vollzug von Aufforderungen dient. Im Allgemeinen ist die Subjektelellipse in einem Imperativsatz obligatorisch, wenn es sich nicht um eine Situation handelt, in der das Subjekt hervorgehoben werden soll. Im Aufforderungssatz wäre es der Adressat, der die vom Sprecher geforderte Handlung ausführt, auch wenn er im Satz selbst nicht explizit realisiert wird. Grammatikalische Kennzeichen des entsprechenden Satztyps im Koreanischen sind hauptsächlich durch Verbalsuffixe gekennzeichnet, die den Imperativ ausdrücken. Wenn also in einem koreanischen Satz ein Imperativmarker vorkommt, ist es möglich zu erkennen, dass das Nullsubjekt des Satzes der Adressat ist. Bei der Formulierung der entsprechenden Regel zur Nullsubjektreolution wird die Satzmodusinformation aus den Zusatzdaten des Originalkorpus (vgl. Kapitel 3.1.1) genutzt, die als Wert des Merkmals "mood" zu finden ist und einem Prädikat anhand des im Prädikat enthaltenen Verbalsuffixes durch den für die Zusatzdaten verwendeten syntaktischen Parser zugeordnet wird. Zum Beispiel weist der syntaktische Parser einem Prädikat den Wert "impe" für das Merkmal "mood" zu, wenn das Prädikat einen Imperativmarker wie ein Verbalsuffix enthält. Abbildung 11 zeigt den Aufbau der Regel, um Prädikaten im Imperativ einen Adressatenreferenten "2sh+" (2. Person, Singular, +Honorativ⁶⁸) zuzuweisen:

⁶⁷ Zur Beschreibung der anderen vier Klassen illokutionärer Sprechhandlungen siehe Searle (1976: 354–361).

⁶⁸ Der Grund, warum der durch diese Regel zugewiesene Referent die Eigenschaft [+Honorativ] hat, liegt darin, dass die Wahrscheinlichkeit höher ist, dass das Analysekorpus in dieser Studie Imperativsätze enthält, die durch Verbalsuffixe markiert sind und Respekt gegenüber dem Adressaten ausdrücken. Dies liegt daran, dass das Korpus Äußerungen in öffentlichen Situationen unter Fremden enthält, die gegenseitig Respekt bekunden. Die typischen Imperativmarker im Koreanischen und ihre Klassifizierung aufgrund der unterschiedlichen Höflichkeitsgrade werden im Zusammenhang mit der zweiten heuristischen Regel eingeführt (siehe Tabelle 19). Das entsprechende Pronomen der 2. Person für die Imperativsätze mit der Eigenschaft [+Honorativ] ist das Pronomen *dangsin*, nicht aber *neo*. Die Realisierung des Pronomens *dangsin* wird in unmarkierten Gesprächssituationen als unhöflich angesehen (vgl. Hwang Juck-Ryoon 1990; Park Mae-Ran 1990), auch wenn es als die Honorativform von *neo* eingestuft wird.

```

if datapoint["zp_mood"] == "impe":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "mood derivable"
    datapoint["rule"] = "impe"
    datapoint["rule_order"] = "1"

```

Abbildung 11: Die erste heuristische Regel⁶⁹ für Prädikate im Imperativ

Die Regel in Abbildung 11 beinhaltet fünf Datenpunkte:

1. "zp_mood",
2. "subject_by_rule",
3. "group",
4. "rule",
5. "rule_order".

Der erste Datenpunkt "zp_mood" enthält die Satzmodusinformation eines Input-Prädikats, die dem Merkmal "mood" aus den Zusatzdaten entspricht, sofern eine Regel nach der Suche nach einem Prädikat die Satzmodusinformation verwendet. Der zweite Datenpunkt "subject_by_rule" zeigt den Referenten, der dem Prädikat gemäß der entsprechenden Regel zugeordnet ist. Die Datenpunkte "group" und "rule" zeigen den Gruppennamen der Regel und den Regelnamen, die in dieser Studie nach dem Regelmerkmal gruppiert und benannt werden. Der letzte Datenpunkt "rule_order" gibt die Anwendungsreihenfolge der Regel im NRS an.

Die Regel in Abbildung 11 identifiziert zunächst ein Prädikat mit dem Wert "impe" im Datenpunkt "zp_mood", was darauf hinweist, dass das Input-Prädikat im Imperativ steht, da es Imperativmarker enthält. Wenn ein Prädikat mit dem Wert "impe" im Datenpunkt "zp_mood" gefunden wird, weist die Regel dem Datenpunkt "subject_by_rule" den Wert "2sh+" zu. Dieser Wert gibt an, dass sich das Nullsubjekt des Prädikats auf ein Subjekt der 2. Person Singular in Honorativform bezieht. Wie im Datenpunkt "rule_order" zu sehen ist, wird die Regel in Abbildung 11 als erste aller heuristischen Regeln im NRS angewandt, da Verbalsuffixe als Imperativmarker einer der offensichtlichsten Hinweise auf adressatenorientierte Referenten von Nullsubjekten solcher Prädikate sind.

In den Zusatzdaten kann es zu Fehlern bei der Information zum Satzmodus kommen, bei denen ein Prädikat im Imperativ fälschlicherweise nicht als Imperativ getaggt wird, obwohl es einen typischen Imperativmarker enthält. Die Beispiele (24)a und (24)b repräsentieren Äußerungen aus dem Originalkorpus, deren Prädikate im Imperativ stehen. Die Varianten (24)a' und (24)b' geben die unterschiedlichen Informationen aus den Zusatzdaten über die Prädikate einschließlich der Satzmodusinformation wieder.

- (24) a. *ssaijeu-leul malseumha-e ju-seyo-*.
 Größe-OBJ sagen.HON-AUX-SFS.IMP-SF
 'Sagen (Sie) (mir) die Größe.'
- a'. *seyo/E[hono=plus,mood=impe, pos=E, subpos=send, morphpos=EP]*
- b. *geogjeongha-ji ma-seyo-*.
 sich Sorgen machen-AUX.NEG-SFS.IMP-SF
 'Machen (Sie) sich keine Sorgen.'
- b'. *seyo/E[hono=plus,mood=decl, pos=E, subpos=send, morphpos=EP]*

⁶⁹ Der Code für die gesamte Regel befindet sich in Appendix 1.

Obwohl beide Prädikate in (24)a und (24)b im Imperativ stehen, wie am Imperativmarker *seyo* zu erkennen ist, wird dem Verbalsuffix in (24)b fälschlicherweise die Satzmodusinformation "decl", zugewiesen, die für Deklarativsätze steht (vgl. (24)b'). Aus diesem Grund wurde eine ergänzende Regel erstellt, um Prädikate im Imperativ zu erfassen, die übersehen wurden. Dem Prädikat in (24)b, das fälschlich als Deklarativsatz analysiert wurde, wird durch die Regel in Abbildung 12 voraussichtlich der richtige Referent "2sh+" zugewiesen.

```
elif re_morphmatch(morphs,morph="^(으?세요|으?세요여|으?십시오|으?십시오) $", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "mood derivable"
    datapoint["rule"] = "impe"
    datapoint["rule_order"] = "3"
```

Abbildung 12: Die zweite Regel für Prädikate im Imperativ

Die Regel in Abbildung 12 weist einem Prädikat den Referenten der 2. Person Singular in Honorativform ("2sh+") zu, wenn es eine der Varianten der Imperativmarker -(eu)seyo oder -(eu)sibsio enthält und sein Datenpunkt "zp_mood" den Wert "decl", also den Deklarativsatz, aufweist.

Um die Auswahl der beiden Imperativmarker für die Regel in Abbildung 12 zu erklären, ist es zunächst erforderlich, die Verbalsuffixe zu analysieren, die im Koreanischen gängige Marker für den Imperativ sind. Lee Ji Sun (2014) stellt die folgenden sieben Verbalsuffixe als Imperativmarker vor, die in früheren Studien am häufigsten genannt werden.⁷⁰ Die sieben Verbalsuffixe in Tabelle 19 lassen sich weitgehend in zwei Gruppen unterteilen, je nachdem, ob sie die Kategorie Respekt gegenüber dem Adressaten ausdrücken.

Nr.	Imperativmarker	Respekt
1	-ala/eola/yeola	[-Honorativ]
2	-ge/gena	
3	-(eu)ji	
4	-a/eo/yeo	
5	-a-yo/eoyo/yeoyo	[+Honorativ]
6	-(eu)seyo	
7	-(eu)sibsio	

Tabelle 19: Verbalsuffixe zur Markierung des Imperativs im Koreanischen nach Respektausprägung (aus Lee Ji Sun 2014: 20)

Im Gegensatz zu den ersten vier Verbalsuffixen (mit [-Honorativ]) in Tabelle 19 werden die letzten drei Verbalsuffixe verwendet, um den Adressaten um etwas zu bitten, womit eine Respektsbekundung gegenüber dem Adressaten gezeigt wird. Aufgrund der Eigenschaften des Korpus dieser Studie, das hauptsächlich aus Gesprächen zwischen Fremden während einer Reisesituationen besteht, ist die Wahrscheinlichkeit hoch, dass der Imperativmarker mit der Eigenschaft [+Honorativ] genutzt wird. Von den drei Verbalsuffixen werden das sechste und das siebte Verbalsuffix für die Regel in Abbildung 12 benutzt, weil sie das Honorativsuffix *si* enthalten, das Respekt für den Adressaten ausdrücken kann und daher im Korpus dieser Studie wahrscheinlich häufig zu finden ist. Das fünfte Verbalsuffix -a-yo/eoyo/yeoyo wird nicht für die Regel eingesetzt, obwohl es die Eigenschaft [+Honorativ] aufweist, da es nicht nur für Imperativsätze, sondern auch für andere Satztypen genutzt werden kann (vgl. Lee Ji Soo 2016: 30). Dementsprechend ist das Verbalsuffix nicht mehr eindeutig auf die Wiederherstellung des Adressaten anzuwenden, weshalb sie in der vorliegenden Regel ausgelassen wird.

⁷⁰ Die von Lee zur Extraktion der relevanten Verbalsuffixe verwendete Literatur wird in Lee Ji Sun (2014: 20) ausführlich dargestellt.

Die Beispiele (25)a und (25)b bestehen aus Imperativsätzen unter Verwendung des sechsten bzw. siebten Verbalsuffixes zur Imperativmarkierung, wie in Tabelle 19 präsentiert. Dadurch lässt sich bestimmen, welchen Prädikaten durch die in Abbildung 12 dargestellte Regel ein Adressat zugeordnet werden kann. Beide Äußerungen in (25) (zitiert aus Cho Yongkil 2005: 54) können in einer Situation benutzt werden, in der ein Polizist einen Fahrer auffordert, seinen Führerschein vorzuzeigen.

- (25) a. *myeonheojeung jom boyeoju-sibsio-*.
Führerschein bitte zeigen-SFS.IMP-SF
'Bitte zeigen (Sie) (mir) (Ihren) Führerschein.'
- b. *myeon-heo-jeung jom boyeoju-seyo-*.
Führerschein bitte zeigen-SFS.IMP-SF
'Bitte zeigen (Sie) (mir) (Ihren) Führerschein.'

Der Unterschied zwischen den beiden Äußerungen besteht in der Auswahl des Imperativmarkers, also zwischen der formalen Form *sibsio* und der informellen Form *seyo*. Die formelle Form wird in öffentlichen Gesprächssituationen z. B. unter Fremden verwendet, während die informelle Form üblicherweise in einer privaten oder relativ engen Beziehung benutzt wird (vgl. Cho Yongkil 2005: 54). Da aber der informelle Imperativmarker den Grad der Handlungsintensität des formellen abschwächen kann (Cho Choon-Hak 1982), kann er auch in einer öffentlichen Gesprächssituation verwendet werden. Cho Yongkil (2005) erklärt, dass der Satz in (25)a unter Umständen als unhöflich angesehen werden könnte, weil die Handlungsfunktion des Imperativs durch die Verwendung des Verbalsuffixes *sibsio* deutlich angezeigt wird. Andererseits kann (25)b im Vergleich zu (25)a als höflich betrachtet werden, da die Äußerung durch die Verwendung des Verbalsuffixes *seyo* eher wie eine Empfehlung oder ein Vorschlag als ein direkter Befehl klingt (vgl. Cho Yongkil 2005: 55). Obwohl sich die beiden Imperativmarker *-(eu)seyo* und *-(eu)sibsio* in ihrer Handlungsintensität unterscheiden, wurden sie zur Erstellung der Regel in Abbildung 12 herangezogen, weil sie definitive Anhaltspunkte dafür sind, dass Nullsubjekte als Adressaten zu interpretieren sind.

Die in Abbildung 11 und Abbildung 12 vorgestellten Regeln beziehen sich auf Fälle, in denen die Satztypen mit den illokutionären Akten übereinstimmen. Dies steht im Zusammenhang mit der sog. „Hypothese der wörtlichen Kraft“, die besagt, dass die illokutionäre Kraft einer Äußerung aus der wörtlichen Bedeutung des Satztyps abgeleitet werden kann (vgl. Levinson 1983: 263). Im Koreanischen sind illokutionäre Indikatoren morphologisch realisiert, um Satztypen unterscheiden zu können, so dass es möglich ist, auf der morphologischen Ebene zu erkennen, welche kommunikative Kraft mit einer Äußerung des Sprechers auf den Adressaten ausgedrückt wird. Dies ermöglicht es, die Regeln in Abbildung 11 und Abbildung 12 zu erstellen, da die in den Prädikaten enthaltenen Imperativmarker für die Zuweisung der Adressatenrolle genutzt werden.

4.1.2. Feste Wendung für indirekte Aufforderungen

Es gibt keine eindeutige 1:1-Beziehung zwischen Satztypen und illokutionärem Akt. So können z. B. Interrogativsätze, die im Rahmen der Sprechakttheorie dem Sprechakt Frage zugeordnet werden, auch für andere Sprechakte, wie Aufforderungen, verwendet werden. Wenn der mit linguistischen Mitteln ausgedrückte illokutionäre Akt nicht tatsächlich der beabsichtigten Kommunikationsfunktion entspricht, wird der Sprechakt als indirekt kategorisiert (vgl. Ehrich & Saile 1972; Franck 1975). Unter einem indirekten Sprechakt ist somit zu verstehen, dass der primäre illokutionäre Akt indirekt durch die Ausführung des sekundären illokutionären Aktes vollzogen wird (vgl. Searle 1979: 143). Das indirekte Ausdrücken der intendierten Illokution durch alternative Satztypen stellt eine Kommunikationsstrategie dar, die primär dazu dient, Höflichkeit zu bekunden. Dabei steht im Fokus,

das negative Gesicht des Adressaten zu wahren, um dessen Handlungsfreiheit und individuelle Persönlichkeitsrechte zu respektieren (vgl. Brown & Levinson 2007).

Für den Sprechakt Aufforderung sind verschiedene indirekte Ausdrucksvarianten vorhanden, da der Illokutionsakt als potenziell gesichtsbedrohlich betrachtet wird. Ein Sprecher versucht, gesichtsbedrohliche Akte möglichst zu vermeiden, um potenziellen Konflikten mit seinen Gesprächspartnern in einer Kommunikationssituation aus dem Weg zu gehen. Die wichtigste Strategie hierfür ist Indirektheit (vgl. Brown & Levinson 1987: 132–144). So kann ein Sprecher zwischen einer direkten und einer indirekten Form wählen, wenn er beispielsweise um Salz bittet und sagen: *Kannst du mir das Salz reichen?* oder *Gib mir das Salz!* Die Handlungsaufforderung, die im ersten Fall mit dem Interrogativsatz ausgedrückt wird, ist indirekter als die mit dem Imperativsatz geäußerte. Es gilt als eine universale Kommunikationsstrategie, dass ein Sprecher einen Interrogativsatz anstelle eines Imperativsatzes äußert, um die illokutionäre Kraft abzuschwächen. Mit dem so geäußerten Interrogativsatz fordert der Sprecher den Adressaten dazu auf, eine bestimmte Handlung auszuführen, indem er ihn fragt, ob er die Fähigkeit zur Ausführung dieser Handlung besitzt. Der Adressat versteht dennoch die Implikatur, weil er aus der sekundären Illokution die primäre erschließen kann. Nach dem Sprechaktszenario-Ansatz von Panther & Thornburg (1998)⁷¹ erfolgen die Schlussfolgerungen des Adressaten metonymisch, weil eine Teilkomponente eines Sprechaktszenarios für Bitten die primäre Illokution aktivieren kann.

Im Koreanischen werden Interrogativsätze konventionell mit spezifischen Ausdrücken gebraucht, um eine Aufforderung oder Bitte höflich auszudrücken. Da der illokutionäre Akt an den Adressaten gerichtet ist, können die verwendeten festen Wendungen als heuristische Regeln gefasst werden, die einem Prädikat mit Nullsubjekt einen Adressaten zuweisen.

Bei der Formulierung der entsprechenden Regel wurden die zehn Muster von Chung Jin Woo & Park Jong Chul (2009: 83) herangezogen, die dazu verwendet werden können, Aufforderungen an einen Adressaten auszudrücken.

Nr.	Muster
1	-[a/eo][ju-gess-ni? / ju-(si)-gess-eoyo?]
2	-[a/eo][ju-ji ahn-gess-ni? / ju-(si)-ji ahn-(eus)i)gess-eoyo?]
3	-[a/eo][ju-llae(yo)? / ju-si-llae-yo?]
4	-[a/eo][ju-ji anh-eulla(eyo)? / ju-si-llaeyo?]
5	-[a/eo][ju-(si)-gess-seubnikka?]
6	-[a/eo][ju-(si)-ji anh-(eus)i)-gess-seubnikka?]
7	-[a/eo][ju-si/ju]-lsu [iss-(gess)-ni? / iss-(eus)i)(gess)eo(yo)?]
8	-[a/eo][ju-si/ju]-lsu [eob-(gess)-ni? / eob-(eus)i)(gess)eo(yo)?]
9	-[a/eo][ju-si/ju]-lsu [iss/eob]- (eus)i)-gess-seubnikka?]
10	-[a/eo] ju-(si)-myeon an[doe-llka? / doe-gess-ni? / doe-gess-eoyo? / doe-gess-seubnikka?]

Tabelle 20: Muster zum Ausdruck von Aufforderungen in Chung Jin Woo & Park Jong Chul (2009: 83)

Die Muster in Tabelle 20 werden konventionell dafür genutzt, den Adressaten um etwas zu bitten, indem man ihn nach seiner Fähigkeit fragt. Da es zwischen den zehn Mustern einen leichten Unterschied im Höflichkeitsgrad gibt, wählen Sprecher je nach Äußerungssituation die am besten geeignete Option. Die vier wichtigsten linguistischen Mittel aus den oben dargestellten Mustern sind:

1. das Hilfsverb *-a/eo juda*,
2. die periphrastische Konstruktion *-i su-iss/eobsda* oder *-myeon andeoda*,
3. das Verbalinfix *-gess-*, und
4. das Honorativinfix *-si-*.

⁷¹ Das konkrete Szenario für eine Bitte besteht aus drei Teilkomponenten und einem pragmatischen Ergebnis. Wenn der Sprecher eine der Elemente im Szenario für Bitten zum Ausdruck bringt, wird erwartet, dass der Adressant die ganze illokutive Kategorie ohne Schwierigkeiten verstehen kann. Für weitere detaillierte Sprechaktszenarien, siehe Panther & Thornburg (1998: 759).

Das Honorativinfix *si* kann optional an alle Muster angehängt werden, was den Höflichkeitsgrad bei Bitten erhöht. Das Hilfsverb *a/eo juda* ist das linguistische Mittel, das in allen Mustern in Tabelle 20 zu finden ist. Laut Ho Jung-hwan (2015) kann das Hilfsverb *a/eo juda* den Druck des Adressaten, der durch den direkten Befehl entsteht, verringern, weil es impliziert, dass der Adressat jemandem einen Gefallen tut (vgl. Ho Jung-hwan 2015: 28–29). Son Se-mo-dol (1989) argumentiert, dass die Aufforderung mit diesem Hilfsverb sich vom Befehl unterscheidet, da sie einer Bitte um einen Gefallen nahekommt (vgl. Son Se-mo-dol 1989: 183). Die periphrastische Konstruktion */ su-iss/eobsda* oder *-myeon andeoda* und das Verbalinfix *gess* werden optional zu den Mustern in Tabelle 20 hinzugefügt. Mit der periphrastischen Konstruktion */ su-iss/eobsda* oder *-myeon andeoda* wird nach der Fähigkeit des Adressaten gefragt, etwas zu tun. Mit dem Muster, das diese periphrastischen Konstruktionen enthält, hat der Sprecher die Möglichkeit, eine Äußerung in euphemistischer Form zu formulieren. Gleichzeitig erhält der Adressat die Freiheit, die Bitte des Sprechers ohne spürbaren Druck abzulehnen, indem er klarstellt, ob er in der Lage ist, sie zu erfüllen (vgl. Ho Jung-hwan 2015: 36). Die oben erläuterte periphrastische Konstruktion wird als höflicher angesehen, wenn das Verbalinfix *gess* hinzugefügt ist: z. B. *-jul su iss-gess-eyo(eob-gess-eyo)*. Im Koreanischen kann das Verbalinfix *gess* im Allgemeinen verwendet werden, um die Bedeutung VERMUTUNG oder WILLE anzuzeigen. Ho Jung-hwan (2015) erklärt, dass die Verwendung von indirekten Ausdrücken, die *gess* enthalten, anstelle von direkten Befehlen den Höflichkeitsgrad erhöht, da *gess* als Mittel zur Abstufung fungiert, um die Wirkung der Äußerung auf den Adressaten zu reduzieren (vgl. Ho Jung-hwan 2015: 38). Jeon Hye Young (1995: 133) erläutert dazu, dass die Bedeutung der Ungewissheit durch *gess* zur Bedeutung der Höflichkeit führt und so die Belastung für den Adressaten verringert.

In der vorliegenden Studie wird die Regel für die Zuweisung der Adressatenrolle unter Verwendung der in Tabelle 20 vorgestellten Muster wie folgt erstellt.

```
indirect_str_list = ["주시겠","주실수","주시면안","시면안","주겠","줄수","주면안"]

elif [x for x in indirect_str_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "indirect speech act"
    datapoint["rule"] = "request"
    datapoint["rule_order"] = "55"
```

Abbildung 13: Die Regel anhand der Muster für indirekte Aufforderungen

Zunächst wird eine Liste von sieben Strings erstellt, die aus der Kombination der wichtigsten linguistischen Mittel der Muster in Tabelle 20 bestehen. Die Liste enthält die drei Strings "jugess", "julsu" und "jumeanan", die die Kombination des Hilfsverbs *-a/eo juda*, der periphrastischen Konstruktion *-/su-iss/eobsda* oder *-myeon andeoda* und des Verbalinfixes *-gess-* sowie ihre Varianten in Kombination mit dem Honorativinfix *si* darstellen. Wenn ein Prädikat einen der Strings in der Liste "indirect_str_list" enthält und für den Datenpunkt "zp_mood" den Wert "ques" hat, der für Interrogativsätze steht, dann weist die Regel dem Prädikat die Adressatenrolle "2sh+" zu. Es wird erwartet, dass diese Regel den Nullsubjekten der Prädikate in den folgenden Beispielsätzen aus dem Originalkorpus die Adressatenrolle richtig zuweist.

- (26) a. *igeos-eul jagseongha-e ju-si-gess-seubnikka-?*
 das-OBJ ausfüllen-AUX-HON-IFX-SFS.INT-SF
 'Würden (Sie) das bitte ausfüllen?'
- b. *yeogwon-eul boyeoju-si-I su iss-seubnikka-?*
 Reisepass-OBJ zeigen-HON-PPC-SFS.INT-SF
 'Können (Sie) (mir) (Ihren) Reisepass zeigen?'

In (26)a wird dem Prädikat durch die Regel in Abbildung 13 der Referent in der Adressatenrolle zugeordnet, weil im Prädikat der String "jusigess" vorkommt und das Prädikat in einem Interrogativsatz steht, so dass es in den Zusatzdaten des Originalkorpus den Wert "ques" beim Merkmal "mood" hat. (26)a ist eine Äußerung, in der der Sprecher den Adressaten auffordert, ein Dokument auszufüllen, um ein Paket zu versenden. Daher wird der durch die Regel zugewiesene Adressat als korrekte Vorhersage der Regel bewertet. In (26)b wird ebenfalls der Referent in der Adressatenrolle zugewiesen, weil der String "jusilsu" im Prädikat vorhanden ist und das Prädikat den Wert "ques" beim Merkmal "mood" in den Zusatzdaten hat. Bei Beispiel (26)b handelt es sich um eine Äußerung, in der der Sprecher den Adressaten auffordert, seinen Pass zu zeigen, so dass der zugewiesene Adressat als richtige Vorhersage der Regel bewertet wird.

Bisher wurden heuristische Regeln eingeführt, die sich auf linguistische Mittel auf morphosyntaktischer Ebene stützen, wie beispielsweise Verbalsuffixe zur Markierung des Imperativs und Muster, um den Prädikaten einen Adressaten als Referenten zuzuweisen. Im Folgenden werden nun die linguistischen Mittel vorgestellt, die zur Wiederherstellung von Referenten in der Adressatenrolle verwendet werden können, weil sie zum Ausdruck der Kategorie Respekt gegenüber dem Adressaten dienen.

4.1.3. Linguistische Mittel zum Ausdruck von Respekt

Eine Besonderheit der koreanischen Sprache ist, dass es Honorifika in Form von Verbalsuffixen, Infixen und Pronomen gibt, die Respekt und Bescheidenheit ausdrücken. Honorifika werden meist benutzt, um soziale Beziehungen auszudrücken, bei denen der Kommunikationspartner entweder einen höheren sozialen Status hat oder ein Fremder ist (vgl. Cho Yongkil 2005: 49–50). Indem man sich die Tatsache zunutze macht, dass linguistische Mittel zum Ausdruck der Kategorie Respekt bereits die höfliche Einstellung zum Gesprächspartner enthalten, können sie dazu verwendet werden, heuristische Regeln für die Wiederherstellung der Referenten in der Adressatenrolle aufzustellen. Die linguistischen Mittel zum Ausdruck von Bescheidenheit werden wiederum genutzt, um heuristische Regeln für die Wiederherstellung von Referenten in der Sprecherrolle zu erstellen. Denn ein Sprecher kann durch diese linguistischen Mittel Respekt gegenüber seinen Gesprächspartnern zeigen, indem er bescheiden über sich selbst spricht (vgl. Lee Nam-Seok 1994: 128). Hierauf wird im Kapitel 4.2.3.3 näher eingegangen. Im Allgemeinen handelt es sich bei einem Prädikat mit dem Honorativinfix *si* um eine Höflichkeitsform. Wie bereits in Kapitel 1.2 erwähnt, ist das Auftreten des Honorativinfixes in einem Prädikat kein Garant für die Zuordnung der Adressatenrolle. Dies liegt daran, dass das Honorativinfix primär als Subjekt-Honorifikum fungiert, jedoch in der Literatur darauf hingewiesen wird, dass es auch direkt Respekt gegenüber dem Adressaten signalisieren kann. Bei den Äußerungen im Originalkorpus ist es sehr unwahrscheinlich, dass über einen Dritten gesprochen wird, da es sich meist um Gespräche zwischen zwei Personen handelt, die sich zum ersten Mal in einer Reisesituation begegnen. In der vorliegenden Arbeit wird das Honorativinfix *si* als ein sogenannter Shifter betrachtet, der sprecherorientierte in adressatenorientierte Äußerungen umwandelt. Daher wird das Honorativinfix in den Regeln zur Zuweisung von Referenten in der Adressatenrolle verwendet, da es

einer der entscheidenden Anhaltspunkte dafür ist, dass Nullsubjekte als Adressaten interpretiert werden können. Wie sie in heuristischen Regeln eingesetzt wird, wird in Kapitel 4.2.3.2 erläutert.

Es gibt zudem Hinweise auf lexikalischer Ebene, die dabei helfen können, die Referenz adressatenbezogener Nullsubjekte wiederherzustellen. Honorativverben gehören zu denjenigen, die nicht einfach durch eine Kombination des Honorativinfixes *si* mit einem Verbstamm gebildet werden. Da sie ihre eigene Höflichkeitsform haben, können sie nicht von den Regeln mit dem Honorativprefix *si* erfasst werden. Die folgende Tabelle zeigt die Verben in ihren Neutralformen und ihren Honorativformen, die in dieser Studie zur Erstellung einer Regel für die Zuweisung der Adressatenrolle an einen Referenten herangezogen werden. Piao (2012: 53) stellt Honorativformen für insgesamt zehn Verben zusammen, von denen die drei Verben in Tabelle 21 häufig in Alltagsgesprächen genutzt werden.

Neutralform	Honorativform
<i>iss-da</i> 'sein/bleiben'	<i>gyesi-da</i>
<i>meog-da</i> 'essen'	<i>deusi-da</i>
<i>malha-da</i> 'reden'	<i>malsseumha-da</i>

Tabelle 21: Drei Verben in ihrer Neutral- und Honorativform von Piao (2012: 53)

Betrachtet man die Honorativformen dieser drei Verben, ist festzustellen, dass die Stämme, die dem Verbalsuffix *da* der Neutralform vorausgehen, in den Honorativformen ihre Form verändert haben. Im Falle der Honorativformen der ersten beiden Verben ist das Honorativprefix *si* im Stamm enthalten, aber da ein POS-Tagger Verbstämme nicht auftrennt, sondern als eigenständige Verben erkennt, können heuristische Regeln zum Honorativprefix solche Honorativverben nicht abdecken.

Darum muss eine Regel aufgestellt werden, die diese drei Honorativverben einschließlich ihrer Varianten für die Zuweisung der Adressatenrolle an einen Referenten nutzt. Dies ist deshalb notwendig, da ein Prädikat mit einem Honorativverb aufgrund der Merkmale des Analysekorpus mit hoher Wahrscheinlichkeit benutzt wird, um Respekt gegenüber dem Adressaten auszudrücken, auch wenn Honorativverben zur Beschreibung von Handlungen, Vorgängen und Zuständen eines Dritten verwendet werden können. Da linguistische Mittel zum Ausdruck von Höflichkeit oder Respekt stark auf den Adressaten bezogen sind, beschreibt Piao (2012: 53) Honorativverben als ein Honorifikum in Bezug auf den Adressaten.

```
elif re_morphmatch(morphs, morph="^ (드시|안드|계시|말씀하고계시) $", pos="^(V|X)$") :
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "114"
```

Abbildung 14: Regel bei Verwendung von Honorativverben

Die Regel in Abbildung 14 enthält die Stämme der Honorativverben *deusida* 'essen.HON', *malsseumhada* 'reden.HON' und *gyesida* 'sein/bleiben.HON'. Darüber hinaus wurde der Stamm der Negationsform von *deusida* in die Regel aufgenommen, um zu berücksichtigen, dass seine Negationsform im Korpus nicht als Negationspräfix *an* und Verb *deusida* segmentiert wird, sondern als der Stamm *andeu*. Außerdem wird der Stamm des Hilfsverbs *go gyesida*, die Verlaufsform des Honorativverbs *gyesida*, in die Regel aufgenommen, da auch er eine Respektbezeugung an den Adressaten darstellen kann. Wenn eine der oben genannten Strings in einem Prädikat vorkommt,

ordnet die Regel dem Prädikat die Adressatenrolle als Referenten zu. Die folgenden Beispiele (27)a und (27)b wurden dem Originalkorpus entnommen, um zu illustrieren, wie die Honorativverben in Äußerungen benutzt werden und wie ihre Subjekte interpretiert werden können.

- (27) a. *jigeum eodi gyesi-nyayo-?*
 jetzt wo sein.HON-SFS.INT-SF
 'Wo sind (Sie) jetzt?'
- a'. *gyesi/V[hono=plus,mood=ques, pos=V,subpos=base,tense=pres,morphpos=VB]+*
ngayo/E[hono=plus,mood=ques, pos=E,subpos=send,morphpos=EE]
- b. *eoneu bang-e mug-go gyesi-jyo-?*
 welch Zimmer-LOC sich aufhalten-AUX.PROG.HON-SFS-SF
 'In welchem Zimmer halten (Sie) (sich) auf?'
- b'. *mug/V[mood=ques, pos=V,subpos=base,tense=pres,morphpos=VB]+*
gogyesi/X[pos=X,subpos=base,morphpos=VX]+
jyo/E[mood=ques, pos=E,subpos=send,morphpos=EE]

Beim Beispiel (27)a handelt es sich um eine Äußerung, in der ein Angestellter eines Taxiunternehmens nach dem Standort eines Kunden fragt. In (27)b fragt ein Hotelangestellter einen Gast, der seinen Zimmerschlüssel verloren hat, in welchem Zimmer er sich aufhält. Durch die Verwendung der Honorativverben in (27) zeigt der Sprecher seinen Respekt gegenüber dem Adressaten. Die Prädikate in (27)a und (27)b enthalten den Stamm des Honorativverbs *gyesida* und des Hilfsverbs *go gyesida*. Ihre POS sind in den Zusatzdaten als Verb ("V") und Hilfsverb ("X") getaggt, wie in (27)a' und (27)b' gezeigt wird. Da die POS-Informationen in der in Abbildung 14 abgebildeten Regel sowohl "V" als auch "X" enthalten, können die Prädikate in (27)a und (27)b durch die Regel der Adressatenrolle zugeordnet werden. Die von der Regel zugewiesenen Referenten werden als richtige Vorhersage der Regeln bewertet, weil die Subjekte der Prädikate in den beiden Äußerungen in (27) tatsächlich die Adressaten sind.

In Kapitel 4.1 wurde dargestellt, welche linguistischen Mittel zur Resolution von adressatenbezogenen Nullsubjekten verwendet werden können und wie diese als heuristische Regeln aufgebaut sind. Linguistische Mittel auf morphosyntaktischer Ebene, von denen erwartet wird, dass sie eine wichtige Rolle bei der Wiederherstellung von Referenten in der Adressatenrolle spielen, waren Imperativmarker und Muster für indirekte Aufforderungen. Außerdem wurden Honorativverben eingeführt, die auf der lexikalischen Ebene zum Ausdruck von Respekt gegenüber dem Adressaten genutzt werden.

Im nächsten Kapitel 4.2 wird der Fokus auf die Resolutionsoptionen von sprecherbezogenen Nullsubjekten liegen und darauf, wie die entsprechenden heuristische Regeln formuliert werden müssen.

4.2. Resolution von sprecherbezogenen Nullsubjekten

4.2.1. Exhortativmarker

Verbalsuffixe, die den Exhortativ markieren, können als ein Hinweis für die Wiederherstellung der Referenten sprecherbezogener Nullsubjekte genutzt werden. Der Exhortativ ist eine Art des Satzmodus, bei dem ein Sprecher einen Adressaten ermutigt, etwas gemeinsam zu tun, während er sich selbst in der vorgeschlagenen Handlung inkludiert. Ähnlich wie beim Imperativ wird das Subjekt eines Satzes im Exhortativ in einer unmarkierten Situation nicht realisiert. Häufig wird das Nullsubjekt in einem Exhortativsatz als Subjekt in der 1. Person Plural interpretiert, weil die vom Sprecher vorgeschlagene Handlung von Sprecher und Adressat gemeinsam ausgeführt wird.⁷²

Eine Besonderheit der koreanischen Sprache besteht darin, dass Exhortativsätze durch Exhortativmarker erkennbar sind. Die beiden Verbalsuffixe *-bsida* und *-ja* gelten im Allgemeinen als Exhortativmarker und werden häufiger als alle anderen Verbalsuffixe für den Satzmodus Exhortativ verwendet (vgl. Kwon Jae Il 2004: 40). Um eine Regel für Prädikate mit Exhortativmarkern zu erstellen, wurde in dieser Studie die Tagging-Information in den Zusatzdaten des Originalkorpus genutzt. In den Daten wurde einem Prädikat mit einem Exhortativmarker vom Parser der Wert "sugg" für das Merkmal "mood" zugewiesen.

```
elif datapoint["zp_mood"] == "sugg":  
    datapoint["subject_by_rule"] = "1ph-"  
    datapoint["group"] = "mood derivable"  
    datapoint["rule"] = "sugg"  
    datapoint["rule_order"] = "2"
```

Abbildung 15: Regel für Prädikate im Exhortativ

Die Regel in Abbildung 15 weist einem Prädikat den Referenten "1ph-" (1. Person Plural –Honorativ⁷³) zu, wenn das Prädikat den Wert "sugg" im Datenpunkt "zp_mood" hat. Exhortativmarker gelten als klare Hinweise für die Wiederherstellung eines spezifischen Referenten, weshalb die Regel in Abbildung 15 im Resolutionssystem an zweiter Stelle angewandt wird, wie aus dem Datenpunkt "rule_order" ersichtlich ist.

⁷² In der Forschungsliteratur wird darauf hingewiesen, dass Exhortativsätze nicht nur genutzt werden, um den Sprecher und den Adressaten zu einer gemeinsamen Handlung aufzufordern, sondern auch, um eine Einzelhandlung entweder des Sprechers oder des Adressaten vorzuschlagen (Ko Yong-Kun 1976; Yang In-Seok 1976). Das bedeutet, dass je nach Kontext das Subjekt, das die in einem Exhortativsatz vorgeschlagene Handlung ausführen soll, entweder als Sprecher oder als Adressat interpretiert werden kann. Es ist eine der Strategien, um die Wirkung eines direkten Befehls an einen Adressaten abzuschwächen oder die Kooperation des Gesprächspartners zu erlangen (vgl. Lim Dong Hoon 2011: 354). In dieser Studie wurde die entsprechende Regel jedoch aufgrund der Tendenz erstellt, dass ein Nullsubjekt in einem Prädikat in unmarkierten Situationen generell in der 1. Person Plural im Exhortativ steht.

⁷³ *uli* und *jeohui* werden im Koreanischen als Pronomen der 1. Person Plural klassifiziert. Das Pronomen *jeohui* drückt Bescheidenheit aus, hat also die Eigenschaft [+Honorativ], während das Pronomen *uli* diese Eigenschaft nicht besitzt. Sie unterscheiden sich auch darin, dass *uli* benutzt wird, um sich auf den Sprecher und den Adressaten gemeinsam zu beziehen, während *jeohui* verwendet wird, um sich auf den Sprecher und verwandte Personen zu beziehen, während der Adressat ausgeschlossen wird (vgl. Kim Jung-Nam 2003: 266–267). Daher wird das Subjekt *uli* einbezogen, das den Sprecher und den Adressaten einschließt, so dass in dieser Studie der durch die entsprechende Regel zugewiesene Referent als Pronomen der 1. Person Plural mit der Eigenschaft [–Honorativ] festgelegt wurde.

In den folgenden Beispielen aus dem Originalkorpus wird erwartet, dass die Subjekte der 1. Person Plural den Prädikaten nach der Regel in Abbildung 15 korrekt zugeordnet werden.

- (28) a. *ne-, syoping gyesog ha-leo ga-bsida-*.
ja-NF Shopping weiter machen-zu gehen-SFS.EXH-SF
'Ja, lass uns shoppen gehen.'
- b. *[uli sajin-do mani jjig-go-,]s1 [igeo minihompi-e olli-ja-!]s2*
wir Foto-auch viel schießen-und-NF das Minihompi-auf stellen-SFS.EXH-SF
'Lass uns auch viele Fotos machen und sie auf die Minihompy stellen.'

Die fett gedruckten Prädikate in (28) enthalten die Verbalsuffixe *-bsida* in (28)a und *-ja* in (28)b, die typischerweise den Exhortativ markieren. Aufgrund dieser Suffixe wird den Prädikaten vom Parser in den Zusatzdaten des Originalkorpus der Wert "sugg" für das Merkmal "mood" zugewiesen. In (28)a handelt es sich um eine Äußerung, in der der Sprecher dem Adressaten vorschlägt, gemeinsam shoppen zu gehen, während (28)b eine Äußerung ist, in der der Sprecher dem Adressaten vorschlägt, das gemeinsame Foto auf die Website Cyworld hochzuladen. Da in den beiden Äußerungen in (28) erwartet wird, dass der Sprecher und der Adressat die vorgeschlagene Handlung gemeinsam ausführen, werden die von der Regel in Abbildung 15 zugewiesenen Referenten "1ph-" als eine richtige Vorhersage bewertet. Die Exhortativmarker zeichnen sich somit als gute Indikatoren für die Nullsubjektresolution aus.

4.2.2. Sprecher- und subjektbezogene Modalität

Es wurde bereits in Kapitel 2.1.3.2 erläutert, dass verschiedene linguistische Mittel in Verbalkomplexen eine Beschränkung auf Subjekte der 1. Person aufweisen, wenn sie in Deklarativsätzen verwendet werden. Die folgenden Beispiele (zitiert aus Jang Chaerin 2017: 5) zeigen, welche linguistischen Mittel im Koreanischen diese Art von Einschränkung aufweisen.

- (29) a. *(nae-ga/*ne-ga/*geu salam-i) geogi anj-eulge-*.
ich-SBJ/*du-SBJ/*die Person-SBJ dort sitzen-SFS(Absicht/Plan)-SF
'Ich werde dort sitzen.'
- b. *(nae-ga/*ne-ga/*geu salam-i) geogi anj-eullaे⁷⁴-*.
ich-SBJ/*du-SBJ/*die Person-SBJ dort sitzen-SFS(Wille)-SF
'Ich will dort sitzen.'
- c. *(nae-ga/*ne-ga/*geu salam-i) geogi anj-eul geo-ya-*.
ich-SBJ/*du-SBJ/*die Person-SBJ dort sitzen-PPC(Absicht/Plan)-SFS-SF
'Ich werde dort sitzen.'
- d. *(nae-ga/*ne-ga/*geu salam-i) geogi anj-gess-eo-*.
ich-SBJ/*du-SBJ/*die Person-SBJ dort sitzen-IFX(Wille)-SFS-SF
'Ich werde dort sitzen.'

⁷⁴ Das Verbalsuffix (*eu*)llaе(*yo*) kann sowohl in Deklarativ- als auch in Interrogativsätzen genutzt werden. Nur anhand des Verbalsuffixes kann also nicht erkannt werden, ob die Äußerung in einem deklarativen oder interrogrativen Kontext genutzt wird. Für diesen Fall muss generell das Satzzeichen berücksichtigt werden. Steht ein Schlusspunkt, gilt der Satz als Deklarativsatz, wird ein Fragezeichen gesetzt, gilt er als Interrogativsatz.

Die fettgedruckten linguistischen Mittel in den Prädikaten in (29) können von einem Sprecher verwendet werden, um eine Absicht oder den eigenen Willen auszudrücken. Dafür kommen in (29)a und (29)b Verbalsuffixe (sentence-final suffix, SFS) in (29)c eine periphrastische Konstruktion (periphrastic construction, PPC) und in (29)d ein Infix (infix, IFX) zum Einsatz. Anhand dieser Beispiele lässt sich erkennen, dass die Subjekte dieser Deklarativsätze nur als die Subjekte in der 1. Person Singular interpretiert werden können, was in früheren Studien damit erklärt wurde, dass diese linguistischen Mittel eine sprecherorientierte Modalität ausdrücken (Park Jae-yon 2003; 2006; Jang Chaerin 2017). Verschiedene Studien haben sich dem Begriff der Modalität aus unterschiedlichen Perspektiven angenähert, was zu einer Vielfalt und einem breiten Sinn des Begriffs führt. In der vorliegenden Arbeit wird der Begriff *Modalität* aus semantisch ausgerichteten Ansätzen übernommen, die sich auf einen Modalitätsbegriff konzentrieren, der die Einstellung des Sprechers zu den in der Proposition geäußerten Informationen anzeigt (Palmer 1979; Coates 1983; 1995; Facchinetto et al. 2003; Frawley 2008). Die Kriterien zur Unterteilung der Modalität variieren ebenso wie die Definition von Modalität selbst von Forscher zu Forscher. Das Konzept der sprecherorientierten Modalität findet sich bei Bybee et al. (1994) und Bybee & Fleischman (1995) zur Unterscheidung von agensorientierter und sprecherorientierter Modalität. Die agensorientierte Modalität umfasst alle modalen Bedeutungen wie Verpflichtung, Wunsch, Fähigkeit, Erlaubnis, die Bedingungen an ein Agens stellen. Die sprecherorientierte Modalität umfasst Direktive wie Imperativ, d. h. Sprechakte, mit denen ein Sprecher versucht, einen Adressaten zu einer Handlung zu bewegen (vgl. Bybee & Fleischman 1995: 6). Park Jae-yon (2003: 258) wiederum argumentiert, dass das wichtigste Kriterium für die Unterscheidung zwischen sprecherorientierter und subjektorientierter Modalität⁷⁵ im Koreanischen mit der Unterscheidung zwischen konstativen und performativen Äußerungen in Austins (1962) früherer Formulierung zusammenhängt. Sie erklärt, dass konstative Äußerungen, die als wahr oder falsch beurteilt werden können, mit der subjektorientierten Modalität zusammenhängen, während performativ Äußerungen, die zur Darstellung des Vollzugs von Handlungen verwendet werden und deren Wahrheitswert nicht beurteilt werden kann, zur sprecherorientierten Modalität gehören (vgl. Park Jae-yon 2003: 257–258).

Die folgenden Beispiele zeigen, dass linguistische Mittel mit derselben modalen Semantik bei ihrer Orientierung zwischen Sprecher oder Subjekt unterschiedlich sein können.

- (30) a. *jajeongeo-leul deo ta-llaeyo-*.
Fahrrad-OBJ mehr fahren-SFS(Wille)-SF
'(Ich/Sie*/Er*) werde mehr Fahrrad fahren.'
- b. *jajeongeo-leul deo ta-lyeogo ha-eyo-*.
Fahrrad-OBJ mehr fahren-PPC(Wille)-SFS-SF
'(Ich/Sie/Er) werde mehr Fahrrad fahren.'

Der Ausdruck von Willen wird durch das Verbalsuffix *llaeyo* in (30)a und durch die periphrastische Konstruktion *lyeogo hada* in (30)b ausgedrückt. Außerdem wurden in den beiden Äußerungen die Subjekte der Prädikate nicht realisiert. Das Nullsubjekt in (30)a kann nur als Sprecher, das Nullsubjekt in (30)b dagegen als jede belebte Entität mit dem Willen, mehr Fahrrad fahren zu wollen, interpretiert werden.

Der Unterschied zwischen dem Verbalsuffix *llaeyo* und der periphrastischen Konstruktion *lyeogo hada* wird anhand von Subjektbeschränkungen erklärt. Im Allgemeinen kann *llaeyo* nur mit einem Subjekt

⁷⁵ Anstelle des Terminus *agensorientierte Modalität* von Bybee et al. (1994) und Bybee & Fleischman (1995) verwendet Park Jae-yon (2003) den Begriff *subjektorientierte Modalität* von Palmer (vgl. Palmer 1979: 36). Der Grund dafür ist, dass *Subjekt* ein umfassenderer Begriff ist als *Agens* und sowohl im Zusammenhang mit epistemischer Modalität als auch mit deontischer oder dynamischer Modalität anwendbar ist (vgl. Park Jae-yon 2003: 253).

der 1. Person genutzt werden, so dass es den Willen des Sprechers ausdrückt, während *lyeogo hada* keine solche Subjektbeschränkung aufweist (Park Jae-yon 2003; Yang Ji Hyeon 2013; Jang Chaerin 2017). Da es keine Beschränkung für ein Subjekt eines Prädikats einschließlich periphrastischer Konstruktionen wie *lyeogo hada* gibt, werden prinzipiell alle Subjekte in der 1., 2. und 3. Person als mögliche Optionen in Betracht gezogen. Als Grund für diesen Unterschied wird der Grammatikalisierungsgrad der Konstruktion *llaeyo* angeführt, die ausgehend von der periphrastischen Konstruktion *lyeogo hada* zu einem Verbalsuffix geworden ist (Park Son-Ja 1992; Park Jae-yon 1998). Infolgedessen kann das Verbalsuffix *llaeyo* von einem Sprecher nur dazu verwendet werden, seinen eigenen Willen zu äußern, etwas zu tun.

Park Jae-yon (2003) erklärt, dass die Subjektbeschränkung von *llaeyo* aus seiner Sprecherorientierung resultiert, die sich in typischen modalen Ausdrücken zeigt, und dass *lyeogo hada* aufgrund seiner Subjektorientierung als Prädikat fungiert, um Sachverhalte in Bezug auf ein Subjekt zu beschreiben. Sie führt weiter aus, dass die Sprecherorientierung im Allgemeinen eine Eigenschaft von Modalausdrücken mit Flexionsformen ist, deren Grammatikalisierung abgeschlossen ist, während die Subjektorientierung eine Eigenschaft von periphrastischen Konstruktionen ist, deren Grammatikalisierung noch nicht beendet ist.

Aus den vorangegangenen Erörterungen lässt sich schließen, dass sich die in (29) dargestellten linguistischen Mittel zum Ausdruck von Willen in Bezug auf ihre Subjektbeschränkungen unterscheiden. Im Fall von (29)a und (29)b wird die Modalität durch Verbalsuffixe ausgedrückt, deren Grammatikalisierungsprozess zum Ausdruck des Sprecherwillens in Aussagen abgeschlossen ist. Daher wird erwartet, dass die Genauigkeit der Nullsubjektresolution einer Regel sehr hoch ist, wenn diese Verbalsuffixe verwendet werden. Die folgende Abbildung zeigt beispielhaft, wie das Verbalsuffix *eulggeo* als Regel zur Wiederherstellung von Referenten in der Sprecherposition erstellt wurde.

```
elif re_morphmatch(morphs,morph="^(을게요|ㄹ게요|게요|을께요|ㄹ께요|ㄹ게|ㄹ께|ㄹ건데요|ㄹ건가|
=거나|을건데요|는건데요)?$", pos="^E$") and datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "80"
```

Abbildung 16: Regel beim Verbalsuffix *eulggeo* in Deklarativsätzen

Die Regel in Abbildung 16 enthält Varianten des Verbalsuffixes *eulggeo*. Zu den Varianten gehört zum Beispiel auch ein Verbalsuffix *eulggeo*, das als grammatisch falsch gilt, aber in der Umgangssprache häufig vorkommt (Yoon Eun Kyung 2006). Darüber hinaus wurden auch informelle Formen wie *eulge* ohne *yo*⁷⁶ in die Liste der Varianten aufgenommen. Es ist jedoch zu erwarten, dass diese Art von informellen Formen aufgrund der Natur des Korpus in dieser Studie nur selten zu finden ist. Daher ordnet die Regel einem Prädikat den Sprecherreferenten "1sh+" (1. Person, Singular, +Honorativ) anstelle von "1sh—" (1. Person, Singular, -Honorativ) zu, wenn eine der Varianten des Verbalsuffixes *eulggeo* auftritt und im Datenpunkt "zp_mood" den Wert "decl" aufweist.

In Bezug auf die periphrastische Konstruktion (*eu*)*lyeogo hada* wird in der Literatur betont, dass sie keine Subjektbeschränkungen aufweist und daher in der vorliegenden Studie nicht als Grundlage für eine heuristische Regel dienen sollte. Dennoch wird eine solche Regel aus verschiedenen Gründen formuliert. Ein Beispiel hierfür ist die Verwendung modaler Bedeutungen, insbesondere solcher, die den Ausdruck von Willen beinhalten, um eine Regel für die Zuweisung eines Referenten in der Sprecherrolle zu einem Nullsubjekt aufzustellen, obwohl frühere Studien betonen, dass diese Konstruktion keine Subjektbeschränkungen aufweist.

⁷⁶ Die Unterschiede zwischen den Verbalsuffixen in Bezug auf das Respektsverhältnis und ihre Verwendung in den sechs Sprechweisen im Koreanischen wird in Kapitel 4.2.3.3 näher beschrieben (vgl. Tabelle 23).

Die Gründe für die Formulierung solcher Regeln sind vielfältig. Erstens weisen modale Bedeutungen, die durch periphrastische Konstruktionen ausgedrückt werden, nicht immer deskriptive Merkmale auf.

- (31) a. *na-neun jigeum bab-eul meog-eulyeogo ha-e-*.
Ich.SBJ-DP jetzt Reis-OBJ essen-PPC(Wille)-SFS-SF
'Ich werde/will jetzt essen'⁷⁷.
- b. *na-neun jigeum bab-eul meog-eullae-*.
Ich.SBJ-DP jetzt Reis-OBJ essen-SFS(Wille)-SF
'Ich werde/will jetzt essen.'
- c. *jigeum bab-eul meog-eulyeogo ha-e-*.
jetzt Reis-OBJ essen-PPC(Wille)-SFS-SF
'(Ich>Er>Du?) will jetzt essen.'

Park Jae-yon (2004: 39) erklärt, dass die Äußerung (31)a mit der periphrastischen Konstruktion *eulyeogo hada* je nach Kontext die gleiche illokutionäre Wirkung haben kann wie eine Äußerung mit dem Verbalsuffix *eullae* wie in (31)b. Dies hängt laut Palmer (2001) damit zusammen, dass die Unterschiede zwischen sprecherorientierter und subjektorientierter Modalität keineswegs eindeutig sind (vgl. Palmer 2001: 84S.). Die Tatsache, dass periphrastische Konstruktionen wie *eulyeogo hada* auch dazu verwendet werden können, den Willen des Sprechers wie bei einem voll grammatisierten Verbalsuffix zu äußern, weist auf die Möglichkeit hin, diese linguistischen Mittel als eine Regel zur Wiederherstellung des Referenten in der Sprecherrolle von Nullsubjekten anzuwenden.

Darüber hinaus gibt es intuitive Lesarten von Prädikaten, auch wenn ein Subjekt nicht realisiert wird und kein Kontext gegeben ist. So argumentieren frühere Studien, dass das Subjekt des Prädikats mit *eulyeogo hada* als jedes Subjekt in allen grammatischen Personen interpretiert werden kann, was aber im Detail bisher nicht behandelt wurde. Für das Nullsubjekt in (31)c ist eine Wiederherstellung des Referenten als Sprecher die intuitivste, da er als Subjekt des Prädikats interpretiert werden kann, wenn keine anderen Äußerungskontexte gegeben sind. Die intuitivste Interpretation der Äußerung in (31)c ist dementsprechend, dass der Sprecher seinen Willen äußert, jetzt essen zu wollen. Einer der Gründe, warum das Nullsubjekt in (31)c bevorzugt als Sprecher interpretiert wird, ist, dass das Pronomen der 1. Person Singular in allgemeinen Gesprächssituationen häufig weggelassen wird. Imo (2014) veranschaulicht den Grund dafür mit Hilfe des Konzepts der *Ich-Jetzt-Hier-Origo* von Bühler (1934). Die Default-Origo beim Sprechen kann in den meisten Fällen aktiviert werden, so dass der Adressat davon ausgehen kann, dass das fehlende Element das unmarkierte *Ich* ist. Die Tatsache, dass alle anderen Pronomen mit Ausnahme von *Ich* in der Regel realisiert werden, ist ein deutlicher Hinweis darauf, dass ihre Tilgung dem Adressaten Probleme bereiten könnte, während die Tilgung von *ich* unproblematisch ist (vgl. Imo 2014: 153–154).

Der Satzmodus spielt offenbar ebenfalls eine wichtige Rolle bei der Interpretation von ausgelassenen Subjekten mit einer bestimmten grammatischen Person. Thomas (1979: 46) stellt fest, dass subjektlose Deklarativsätze im Allgemeinen mit einem Subjekt der 1. Person interpretiert werden, während subjektlose Interrogativsätze mit einem Subjekt der 2. Person verstanden werden. Die Tatsache, dass ein nicht realisiertes Subjekt in Deklarativsätzen in einem Gespräch mit hoher Wahrscheinlichkeit auf den Sprecher selbst referiert, ist einer der Gründe, warum das Nullsubjekt in (31)c am ehesten als Sprecher interpretiert wird. Aus diesem Grund ist es am besten nachvollziehbar, die modale Bedeutung

⁷⁷ Die VP *bab-eul meog-da* besteht aus den Elementen aus *bab* 'Reis' als Objekt und *meog-da* 'essen' als Verb, wird aber auch häufig in der Bedeutung 'essen' ohne ein bestimmtes Objekt genutzt.

in (31)c, die durch die periphrastische Konstruktion ausgedrückt wird, als den Willen des Sprechers zu interpretieren.

Auch wenn für die Interpretation des Nullsubjekts in (31)c die Referenz auf den Sprecher am wahrscheinlichsten ist, ist es ebenfalls möglich, es als Drittsubjekt zu interpretieren, wenn in der Diskurssituation bereits eine andere Entität aktiviert wurde. Mit anderen Worten, die Interpretation des Nullsubjekts in (31)c als Subjekt der 3. Person ist nur möglich, wenn eine andere belebte Entität, die in der vorliegenden Gesprächssituation zu essen versucht, eingeführt wird. Haegeman (2013) argumentiert, dass die Subjektauslassung in der 1. Person der Default ist und „out of the blue“, also ohne Vorankündigung, erfolgen kann, während Subjekte in der 3. Person nur ausgelassen werden können, wenn ein Referent kontextuell salient gemacht wird (vgl. Haegeman 2013: 90). Zur Identifizierung eines ausgelassenen Elements argumentiert Thomas (1979), dass das Fehlen einer 3. Person nicht automatisch identifiziert werden kann, wie im Falle der Auslassung von Sprecher und Adressat, sondern aus dem Kontext ermittelt werden muss. Sigurðsson (2011) stellt wiederum fest, dass die Lesart als 3. Person ebenfalls eingeschränkter ist als die Lesart als 1. Person, die oft einen Gesprächskontext erfordert (vgl. Sigurðsson 2011: 279). Aus diesem Grund wird die Äußerung (31)c so interpretiert, dass sie den Willen eines Dritten, jetzt essen zu wollen, beschreibt, wenn bereits eine weitere Person im Bewusstsein der Gesprächsteilnehmer ist.

Die unpassendste Interpretation der Äußerung (31)c wäre als Beschreibung des Willen des Adressaten. Das liegt daran, dass es für einen Sprecher unnatürlich ist, den Willen eines Adressaten anstatt seines eigenen zu erklären. Ähnlich verhält es sich mit psychologischen Verben, bei denen nur der Sprecher seine inneren Gefühle mitteilen kann, wie in Kapitel 1 und 2 gezeigt wurde. Daher wäre es sehr unnatürlich, das Nullsubjekt von (31)c als Adressaten zu interpretieren, es sei denn, es handelt sich um eine sehr spezielle Situation, in der ein Sprecher die Fähigkeit des Gedankenlesens besitzt, um den Willen des Adressaten zu erkennen. Zu fragen, ob der Adressat jetzt essen will, ist dagegen eine völlig übliche Äußerung. Dies bedeutet, dass Änderungen im Satzmodus zu anderen Interpretationen der Nullsubjekte führen, wie am Ende dieses Kapitels näher erläutert wird.

Bisher wurde gezeigt, dass die periphrastische Konstruktion (*(eu)lyeogo hada* eine performative Kraft haben kann, um den Willen des Sprechers zum Ausdruck zu bringen, und die Interpretation des Nullsubjekts als ein Subjekt der 1. Person Singular in Aussagen intuitiv gegenüber den Interpretationen als 2. oder 3. Person bevorzugt wird. Daher werden in dieser Studie periphrastische Konstruktionen wie (*(eu)lyeogo ha*, die Willensäußerungen repräsentieren, auch in einer Regel abgebildet, die den Nullsubjekten den Sprecher als Referenten zuweisen können.

```
elif re_morphmatch(morphs,morph="^(려고하|으려고하|사려하|려|려하|으려하|으려고|려고|어보려고하|
    려고그리|려고요|으려고요|으려고요|려는것이|고자하|으려고하|기로하|으려구요|려구요|사려하) $",
    pos="^(X|E|V)$") and datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "66"
```

Abbildung 17: Regel bei periphrastischen Konstruktionen für die Willensäußerung in Deklarativsätze

Abbildung 17 veranschaulicht die Regel für die Zuweisung des Referenten "1sh+", d. h. die Sprecherrolle, wenn eine der Varianten der periphrastischen Konstruktionen wie (*(eu)lyeogo hada*, *goja hada* und *gilo hada* im Prädikat vorkommt und die Modusinformation den Wert "decl" hat. Die periphrastischen Konstruktionen, die für die Erstellung der Regeln verwendet wurden, wurden aus Grammatikbüchern und verwandten Arbeiten gezogen.⁷⁸

⁷⁸ Park Sunyoung (2020: 13–14) stellt Ausdrücke für Willensäußerungen im Koreanischen und die Literaturlisten zusammen.

Es ist zu erwarten, dass heuristische Regeln, die periphrastische Konstruktionen zur Grundlage haben, in hohem Maße Nullsubjekte abdecken können, wenn man bedenkt, dass sie in der koreanischen Grammatik weit verbreitet sind, weil sie für dieselbe Funktion benutzt werden können wie einzelne Morpheme (z. B. Verbalaffixe, Kim Sun-hye 2019: 135). Die folgende Tabelle zeigt die Anteile des Verbalinfixes *gess* im Vergleich mit periphrastischen Konstruktionen, die zum Ausdruck von Vermutungen in einem gesprochenen Korpus⁷⁹ identifiziert wurden.

Kategorie	Linguistisches Mittel	Anteil
Verbal infix	- <i>gess</i> -	52,70 % (n=5.655)
Periphrastische Konstruktionen	- <i>n/l geos gat-da</i>	29,91 % (n=3.210)
	- <i>l geosi-da</i>	14,37 % (n=1.542)
	- <i>l teoi-da</i>	2,47 % (n=265)
	- <i>n/l deusha-da</i>	0,48 % (n=51)
	- <i>n/l moyangi-da</i>	0,07 % (n=8)
	Total	100,00 % (N=10.731)

Tabelle 22: Anteile der linguistischen Mittel zum Ausdruck der Bedeutung [Vermutung] im Korpus *saeyeonsemalmungchi2* nach Kim Sun-hye (2019: 175)

Tabelle 22 zeigt, dass von den insgesamt 10.731 Fällen, in denen Vermutungen geäußert wurden, das Verbalinfix *gess* mit einem Anteil von 53 % am häufigsten benutzt wurde. In den übrigen Fällen wurden in der gleichen Funktion verschiedene periphrastische Konstruktionen verwendet. Auch wenn es Unterschiede in ihrer Auftretenshäufigkeit gibt, machen sie insgesamt etwa 47 % der Gesamtzahl aus. Die vorliegenden Häufigkeitsangaben lassen darauf schließen, dass im Koreanischen nicht nur Einzelmorpheme wie Verbalinfixe, sondern auch periphrastische Konstruktionen häufig verwendet werden, um eine spezifische grammatische Bedeutung auszudrücken. Daher kann davon ausgegangen werden, dass die Verwendung dieser linguistischen Mittel in den heuristischen Regeln zu einer verbesserten Identifizierung der Nullsubjekte führen wird.

Anhand der Beispiele in (29) wurde bereits erläutert, dass Verbalsuffixe wie *eulge* und *eullae*, periphrastische Konstruktionen wie *eul geoda* und Verbalinfixe wie *gess* verwendet werden können, um den Sprecherwillen in Deklarativsätzen auszudrücken. Wie solche Verbalsuffixe und periphrastischen Konstruktionen als Regeln zur Resolution von Nullsubjekten mit Referenz auf den Sprecher erstellt werden können, ist bereits in Abbildung 16 und Abbildung 17 gezeigt worden. Diese Regeln weisen einem Prädikat den Sprecher als Referenten zu, wenn im Prädikat eines Deklarativsatz ein Verbalsuffix oder eine periphrastische Konstruktion gefunden wurde, das oder die in den Regeln enthalten ist.

Allerdings reichen Verbalinfixe wie *gess* allein nicht aus, um diese Regel aufzusetzen, selbst wenn die Information über den Satzmodus genutzt wird. Denn das Verbalinfix *gess* kann z. B. nicht nur für Willensäußerungen, sondern auch zum Ausdruck von Vermutungen verwendet werden (Na Jinseok 1965; Suh Cheong Soo 1990; Lim Dong Hoon 2001; Park Jae-yon 2004). Die folgenden Beispiele aus Park Jae-yon (2004: 71–72) und Jang Chaerin (2017: 4) geben Äußerungen wieder, in denen das Verbalinfix *gess* verwendet wird, um eine Vermutung und den Willen auszudrücken ((32)a vs. (32)c).

⁷⁹ Bei dem von Kim Sun-hye (2019) analysierten gesprochensprachlichen Korpus handelt es sich um "saeyeonsemalmungchi2", das aus 990.000 Wortformen besteht und auf der Grundlage natürlicher Äußerungen erstellt wurde, die zwischen 1990 und 2005 aufgezeichnet und transkribiert wurden (für weitere Informationen vgl. Kapitel 1.2.2 in Kim Sun-hye 2019).

Auch ein Vergleich mit der periphrastischen Konstruktion *I sudo issda* zur Illustration des Gewissheitsgrads in einer Vermutung findet sich in den Beispielen ((32)a vs. (32)b)

- (32) a. *jigeum bakk-e-neun bi-ga o-gess-da-*.
 jetzt draußen-LOC-DP Regen-SBJ regnen-IFX(Vermutung)-SFS.DECL-SF
 'Es kann jetzt draußen regnen.'
- b. *jigeum bakk-e-neun bi-ga o-I sudo iss-da-*.
 jetzt draußen-LOC-DP Regen-SBJ regnen-PPC(Möglichkeit)-SFS.DECL-SF
 'Es könnte jetzt draußen regnen.'
- c. *je-ga bogsaha-gess-seubnida/ bogsaha-Igeyo/ bogsaha-IIaeyo-*.
 ich-SBJ kopieren-IFX(Wille)-SFS.DECL-SF kopieren-SFS(Wille) kopieren-SFS(Wille)-SF
 'Ich werde/will (das Dokument) kopieren.'

In (32)a wird das Verbal infix *gess* zum Ausdruck einer Vermutung verwendet. Mit der Äußerung (32)a artikuliert der Sprecher seine Vermutung, dass es jetzt draußen regnet. Der Sprecher kann seine Vermutung mit der periphrastischen Konstruktion *I sudo issda* wie in (32)b ausdrücken, die auf eine Möglichkeit hinweist. Vergleicht man (32)a und (32)b, so zeigt sich ein Unterschied in der Gewissheit der Vermutung des Sprechers über die Proposition, was durch die Modalverben *kann* und *könnte* in den deutschen Übersetzungen zum Ausdruck kommt. (32)a impliziert eine gewisse Wahrscheinlichkeit, dass es tatsächlich regnet, während (32)b eine potenzielle Situation darstellt, in der es gerade draußen regnet, jedoch mit einer gewissen Unsicherheit. Laut Park Jae-yon (2004) drückt *gess* aus, dass der Sprecher mehr als 50 Prozent Gewissheit über die Proposition hat, während die Gewissheit bei *I sudo issda* bei unter 50 Prozent liegt (vgl. Park Jae-yon 2004: 71–72).

Das Beispiel (32)c zeigt drei Varianten mit unterschiedlichen linguistischen Mitteln, die den Willen des Sprechers ausdrücken, ein Dokument zu kopieren. Die erste Variante enthält das Verbal infix *gess* und die zweite und dritte Variante enthalten die Verbalsuffixe *Igeyo* und *IIaeyo*. Jang Chaerin (2017: 4) erklärt, dass diese drei Äußerungen austauschbar sind, um die Bereitschaft oder ein Versprechen des Sprechers auszudrücken. Beim Vergleich der drei Varianten gibt es allerdings Bedeutungsunterschiede, wobei die dritte Variante mit dem Verbalsuffix *IIaeyo* nicht verwendet werden kann, um ein Versprechen vonseiten des Sprechers zu artikulieren, sondern nur seine Absicht. Die ersten beiden Varianten können dagegen verwendet werden, um Versprechen zu verbalisieren (Lee Yun Jin & No Ji Ni 2003; Park Jae-yon 2006; Bak Jaehee 2011).

Wenn man jedoch davon ausgeht, dass das Subjekt in (32)c nicht realisiert ist, kann das Nullsubjekt in allen drei Varianten als Sprecher rekonstruiert werden. Insbesondere das Verbalsuffix *gess* in Kombination mit dem Verbalsuffix *seubnida*, das im gesprochenen Koreanisch üblicherweise verwendet wird, um den Adressaten höflich über eine Situation zu informieren (vgl. Kang Hyoun Hwa et al. 2016: 329), dient dazu, ein Versprechen des Sprechers bescheiden auszudrücken (vgl. Bak Jaehee 2011: 51). Daher wurde in dieser Studie auch die Kombination des Verbal infixes *gess* und des Verbalsuffixes *seubnida* bei der Erstellung einer Regel zur Wiederherstellung des Sprechers als Referenten bei diesen Prädikaten verwendet.

```
elif re_morphmatch(morphs, morph="^겠|것$", pos="^E$") and
    re_morphmatch(morphs, morph="^습니다$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "64"
```

Abbildung 18: Regel beim Verbal infix *gess* und beim Suffix *seubnida*

Die Regel in Abbildung 18 ist abgestimmt auf Prädikate, die eine der Varianten des Verbal infixes *gess* und das Verbalsuffix *seubnida* enthalten. Da beide linguistische Mittel die POS-Information "E" haben werden, was der Wert des POS-Taggings für verschiedene Verbalaffixe in den Zusatzdaten ist, wird dieser Wert auch in der Regel verwendet. Dem Prädikat, das die obige Bedingung erfüllt, wird durch die Regel der Sprecher als Referent "1sh+" zugewiesen, wenn das Prädikat für den Datenpunkt "zp_mood" den Wert "decl" hat.

Bei der Formulierung der bisher eingeführten Regeln spielt die Satzmodusinformation "decl", d. h. Deklarativ, eine bedeutende Rolle bei der Zuweisung des Sprecherreferenten zu einem Prädikat. Wie zuvor erwähnt, kann die Änderung des Satzmodus den Wechsel von der Interpretation des Nullsubjekts als Sprecher zum Adressaten bewirken. Insbesondere solche linguistischen Mittel, die in Deklarativsätzen eine Interpretation als Sprecher auslösen, verschieben die Interpretation gewöhnlich in Interrogativsätzen in Richtung des Adressaten. Diese Situation ist beispielsweise beim bereits in diesem Kapitel vorgestellten Verbalsuffix *eullae(yo)* zu beobachten, das in Deklarativsätzen üblicherweise den Willen des Sprechers zum Ausdruck bringt (vgl. Beispiel (29)b), während es in Interrogativsätzen den Willen des Adressaten thematisiert (vgl. u. a. Park Jae-yon 2003; Huh Kyunghang 2011; Kang Hyoun Hwa et al. 2016; Jang Chaerin 2017).⁸⁰

Die Beispiele (33)a und (33)b enthalten Prädikate mit dem Verbalsuffix *eullae* und veranschaulichen, wie ihre Nullsubjekte bei Veränderung des Satzmodus interpretiert werden.

- (33) a. *yeogi anj-eullae-*.
hier sitzen-SFS(Wille)-SF
'(Ich/*du/*er) will hier sitzen.'
- b. *yeogi anj-eullae-?*
hier sitzen-SFS(Wille)-SF
'Willst (du/*ich/*er) hier sitzen?'

In (33)a wird das Verbalsuffix *eullae* in einem Deklarativsatz verwendet, womit der Sprecher seinen Willen ausdrückt, „hier“ zu sitzen. Diese Interpretation steht in engem Zusammenhang mit der Tatsache, dass *eullae* das linguistische Mittel ist, dessen Funktion mit einer sprecherorientierten Modalität verbunden ist. Aus diesem Grund ist es unmöglich, das Nullsubjekt des Prädikats in (33)a als etwas anderes als den Sprecher zu interpretieren, wie in der deutschen Übersetzung zu sehen ist. In (33)b (zitiert aus Jang Chaerin 2017: 6) wird *eullae* in einem Interrogativsatz benutzt, womit der Sprecher nach dem Willen des Adressaten fragt, ob er „hier“ sitzen möchte. Diese Interpretation ist darauf zurückzuführen, dass der Wechsel vom Ausdruck des Sprecherwillen zum Adressatenwillen in Interrogativsätzen üblich ist. Park Jae-yon (2003: 250) bezeichnet die Veränderung in der Interpretation des Referenten zwischen Deklarativ- und Interrogativsatz als *sprecher-adressaten-orientierte Modalität*. Dies zeigt, dass die Satzmodusinformation bei der Resolution von Nullsubjekten eine wichtige Rolle spielt, wie unschwer an der vorgestellten Situation des Verbalsuffixes *eullae* zu erkennen ist.

Aufgrund der Tatsache, dass in einem Fragesatz die modale Semantik auf den Adressaten bezogen ist, kann (*eu*)*llae(yo)* nicht nur als Frage nach dem Willen des Adressaten verstanden, sondern auch als indirekte Aufforderung an den Adressaten verwendet werden. Die folgenden Beispiele stellen Äußerungen mit dem Verbalsuffix (*eu*)*llae(yo)* in Prädikaten dar, die als Bitte des Sprechers an den Adressaten interpretiert werden können.

⁸⁰ Dies gilt nicht für das Verbalsuffix (*eu*)*lge(yo)*, das nur in Deklarativsätzen verwendet werden kann (vgl. Jang Chaerin 2017: 42).

- (34) a. *i ttukkeong jom yeol-eo ju-IIae-?*
 dies Deckel bitte öffnen-AUX-SFS(Wille)-SF
 'Kannst (du) bitte diesen Deckel öffnen?'
- b. *i ttukkeong jom yeol-eo ju-si-IIaeyo-?*
 dies Deckel bitte öffnen-AUX-HON-SFS(Wille)-SF
 'Können (Sie) bitte diesen Deckel öffnen?'

Das Beispiel (34)a (zitiert aus Park Jae-yon 2003: 260) wird als Aufforderung an den Adressaten verstanden, einen Deckel zu öffnen, und nicht als eine Frage an den Adressaten, ob er den Willen hat, den Deckel zu öffnen. Genauso das Beispiel (34)b, bei dem zusätzlich das Honorativinfix *si* und das Verbalsuffix *yo* hinzugefügt werden. Wenn das Verbalsuffix *yo* am Ende des Prädikats hinzugefügt wird, wird der Satz von der informellen Sprechweise in die hohe Sprechweise verschoben. Aus diesem Grund wird auch das Honorativinfix *si* in (34)b, das optional an ein Prädikat angehängt werden kann und zur hohen Sprechweise dazugehört, zusammen mit dem Verbalsuffix *IIaeyo* als natürlich angesehen.⁸¹ Aus der Perspektive der Nullsubjektresolution kann ein Nullsubjekt eines Prädikats mit (*eu*)*IIae(yo)* in einem Interrogativsatz also als adressatenbezogen interpretiert werden, unabhängig davon, ob es sich um eine Frage nach dem Willen eines Adressaten oder um eine an den Adressaten gerichtete Bitte handelt. In Abbildung 19 und Abbildung 20 wird dargestellt, wie zwei Regeln bei der Verwendung des Verbalsuffixes (*eu*)*IIae(yo)* in Deklarativ- bzw. Interrogativsätzen erstellt wurden.

```
elif re_morphmatch(morphs, morph="^(을래요|ㄹ래요)$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "81"
```

Abbildung 19: Regel für die Verwendung des Verbalsuffixes *eullaeyo* in Deklarativsätzen

```
elif re_morphmatch(morphs, morph="^(을래요|ㄹ래요|ㄹ걸니까)$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "82"
```

Abbildung 20: Regel für die Verwendung des Verbalsuffixes *eullaeyo* in Interrogativsätzen

Die Regel in Abbildung 19 weist einem Prädikat den Referenten "1sh+", sprich die Sprecherrolle zu, wenn das Prädikat mit dem Verbalsuffix (*eu*)*IIaeyo* beim POS-Tagging den Marker "E" enthält und sein Satzmoduswert "decl" ist. Die Regel in Abbildung 20 findet ein Prädikat in einem Interrogativsatz, das eines der Verbalsuffixe (*eu*)*IIaeyo* oder *Igeobnikka* mit der POS-Information "E" enthält. Das letztere linguistische Mittel setzt sich aus der periphrastischen Konstruktion */ geoda* (mit der Funktion Willensäußerung) und dem Verbalsuffix *bnikka* zusammen, das als Verbalsuffix für Interrogativsätze fungiert, die benutzt werden, um dem Adressaten Respekt zu erweisen (vgl. Kang Hyoun Hwa et al.

⁸¹ Die Unterschiede in der Sprechweise zwischen (34)a und (34)b werden sinngemäß durch die Unterschiede zwischen den Personalpronomen *du* und *Sie* in den deutschen Übersetzungen dargestellt, auch wenn sich die Pronomen der 2. Person im Koreanischen *neo* und *dangsin* von der Verwendung der deutschen Pronomen *du* und *Sie* unterscheiden (vgl. O Sang Yi 2005: 247). Zum Unterschied zwischen koreanischen und deutschen Pronomen in der 2. Person siehe O Sang Yi (2005: 256).

2016: 328). Die kombinierte Phrase *lgeobnikka* wird also benutzt, um nach dem Willen des Adressaten zu fragen. Wenn ein Prädikat gefunden wird, das eines der in der Regel festgehaltenen linguistischen Mittel enthält und den Wert "ques" als Satzmodusinformation hat, weist die Regel in Abbildung 20 dem Prädikat "2sh+", sprich die Adressatenrolle zu.

Bisher wurde analysiert, wie verschiedene linguistische Mittel auf der morphosyntaktischen Ebene, die die modale Bedeutung des Sprecherwillens ausdrücken können, verwendet wurden, um Regeln für die Wiederherstellung von Referenten in der Sprecherposition aus den Nullsubjekten aufzustellen. Darüber hinaus wurde der Einfluss des Wechsels des Satzmodus gezeigt, der zu einem Wechsel von der Interpretation des Nullsubjekts als Sprecher zum Adressaten führt. Im Folgenden wird untersucht, welche Anhaltspunkte auf lexikalischer Ebene zur Wiederherstellung von Referenten in der Sprecherposition verwendet und für heuristische Regeln genutzt werden können.

4.2.3. Lexikalische Indikatoren für die Identifizierung von Sprecherreferenten

Als lexikalische Indikatoren gelten Hinweise, die sich auf der lexikalischen Ebene des Verbalkomplexes befinden. Diese Indikatoren umfassen performative Verben, psychologische Verben und Verben zur Bescheidenheitsbekundung und signalisieren, dass das Nullsubjekt als Sprecher interpretiert werden kann. Besonders die Berücksichtigung des ersten und dritten Indikators ist von Bedeutung, da ihnen in der Forschung zu koreanischen Nullsubjekten bisher wenig Aufmerksamkeit geschenkt wurde. Zudem wird die Bedeutung der Berücksichtigung von Kombinationen linguistischer Mittel hervorgehoben.

4.2.3.1. Performative Verben

Auf lexikalischer Ebene eignen sich zunächst performative Verben als Basis einer Regel zur Nullsubjektreolution. Performative Verben wie z. B. *myonglyeonghada* 'befehlen' oder *butaghada* 'bitten' sind eines der linguistischen Mittel, mit denen der Sprechakt der Aufforderung ausgeführt werden kann. Blum-Kulka & Olshtain (1984) erklären, dass die Ausführung einer Aufforderung mit performativen Verben einer sehr direkten und expliziten Strategie („explicit performatives“) entspricht. Damit ist gemeint, dass die illokutionäre Kraft der Äußerung durch entsprechende performative Verben wie „bitten“ in „Ich bitte Sie den Platz sofort freizumachen.“ wirkt (vgl. Blum-Kulka & Olshtain 1984: 202).

Kim Kanghee (2019) analysiert linguistische Strategien zur Ausführung von direktiven Sprechakten in einem semikolloquialen koreanischen Korpus. Dieses Analysekorpus wurde aus Drama-Skripten extrahiert und besteht aus 14.363 Turns und 160.800 Wortformen (*Eojeol* auf Koreanisch), die mit allgemeinen Spacing Units übereinstimmen. Laut ihrer Korpusanalyse können im Koreanischen Aufforderungen mithilfe performativer Verben ausgeführt werden, wobei es Unterschiede im Grad der Direktheit gibt. Performative Verben wie *myonglyeonghada* 'befehlen', die für direkte Befehle verwendet werden, werden selten für Aufforderungen benutzt. In den meisten Fällen wurden performative Verben wie *butaghada* 'bitten' und *pilyohada* 'brauchen' gefunden, die den Sprechakt der Aufforderung euphemistisch ausführen. Insgesamt am häufigsten wurde im Korpus *butaghada* 'bitten' verwendet. Kim Kanghee (2019) erklärt, dass die Verwendung solcher performativer Verben Teil einer Strategie zur Abschwächung von direktiven Sprechakten ist, da die Verwendung von Befehlsverben wie *myonglyeonghada* 'befehlen' als gesichtsbedrohender Akt an die Adressaten angesehen werden können.

Aus syntaktischer Sicht entspricht der Sprecher im Falle der performativen Verben dem Subjekt. Ein performatives Verb wird in einer Äußerung verwendet, um einen Adressaten um etwas zu bitten, wobei der Sprecher die Bitte vorbringt und nicht der Adressat (vgl. Beispiel (35)).

- (35) a. *kolla du jan butagha-bnida-*.
 Cola zwei Glas bitten-SFS.DECL-SF
 '(Ich) bitte (Sie) um zwei Gläser Cola.'
- b. *naeil-buteo 2-bag 3-il-gan pilyoha-bnida-*.
 Morgen-ab 2-Nächte 3-Tage-für brauchen-SFS.DECL-SF
 '(Ich) brauche ab morgen (einen Reiseführer) für 2 Nächte und 3 Tage.'
- c. *joheun yeohaeng-doe-si-gi bala-bnida-*.
 Gut Reise-haben-HON-NMZ hoffen-SFS.DECL-SF
 '(Ich) hoffe, (Sie) haben eine gute Reise.'

In (35) dienen die performativen Verben dazu, eine Aufforderung zu äußern. In (35)a bittet der Sprecher mit dem Verb *butaghada* 'bitten' um zwei Gläser Cola, wobei er das Subjekt des Verbs ist, da er die Bitte äußert. In (35)b bittet der Sprecher mit dem Verb *pilyohada* 'brauchen' um einen Reiseführer und ist auch damit wieder in der Subjektposition. In (35)c verwendet der Sprecher das Prädikat *gi balada* 'hoffen', um dem Adressaten eine gute Reise zu wünschen. Das Nullsubjekt in (35)c wird dementsprechend als Sprecher interpretiert, weil es der Sprecher ist, der eine gute Reise wünscht. Da das Subjekt von performativen Verben in Deklarativsätze also dem Sprecher entspricht, der um etwas bittet oder etwas braucht, wurden zwei Regeln formuliert, die diese Eigenschaft nutzen.

```
performatives_head_list = ["부탁", "권유"]

elif [x for x in performatives_head_list if x in datapoint["verb"]] and
      datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "performative"
    datapoint["rule"] = "performative predicate"
    datapoint["rule_order"] = "140"
```

Abbildung 21: Erste Regel zur Verwendung performativer Verben in Deklarativsätze

```
elif re_morphmatch(morphs,morph="^ (기?원하|바라|기바라|추천하|필요하|필요없|필요는없|빌|권하)$",
                   pos="^(V|X|J)$") and datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "performative"
    datapoint["rule"] = "performative predicate"
    datapoint["rule_order"] = "141"
```

Abbildung 22: Zweite Regel zur Verwendung performativer Verben in Deklarativsätze

Die Regel in Abbildung 21 weist einem Prädikat die Sprecherrolle als Referenten "1sh+" zu, wenn Prädikate wie *butaghada* 'bitten' und *gwonyuhada* 'vorschlagen' den Wert "decl", sprich Nutzung in einem Deklarativsatz, als Satzmodusinformation haben. Bei der Suche nach diesen performativen Verben findet die Regel die Prädikate, die *butag* 'Bitte' oder *gwonyu* 'Vorschlag' enthalten, die Kopfnomen der Verben sind. Der Grund für diese Art der Suche ist, dass sich die Kopfnomen in ihrer Form nicht verändern, auch wenn verschiedene Verbalaffixe an den letzten Teil der Prädikate *-hada* angehängt werden.

Die Regel in Abbildung 22 enthält verschiedene Prädikate wie z. B. *wonhada* oder *balada* 'wollen', *pilyohada* 'brauchen' und *pilyoeobsda* 'nicht brauchen', die für Aufforderungen verwendet werden können. Da die Kernbestandteile dieser Prädikate beim POS-Tagging in den Zusatzdaten als verschiedene Wortarten kategorisiert wurden (Verb: V, Verbalaffix: X, Adjektiv: J), werden diese POS-Informationen auch zum Suchen dieser Prädikate verwendet, wie in Abbildung 22 dargestellt. Wenn

ein Prädikat gefunden wird, das die Bedingungen erfüllt und den Wert "decl" beim Merkmal Satzmodus aufweist, dann weist die Regel dem Prädikat den Referenten "1sh+", sprich die Sprecherrolle, zu.

4.2.3.2. Psychologische Verben

Neben den performativen Verben gehören auch die sogenannten psychologischen Verben zu den lexikalischen Hinweisen, die zur Wiederherstellung der Referenten in der Sprecherposition verwendet werden können. Psychologische Verben werden in drei Typen unterteilt, je nachdem, ob sie Empfindungen, Erkenntnisse oder Gefühle ausdrücken. Die Liste der subjektiven Verben und Adjektive in Yoo Hyun-kyung (1998), die auch bei Hong Munpyo (2011) und Park Arum & Hong Munpyo (2015) verwendet wurde (vgl. Kapitel 2.2), wurde bei der Erstellung der entsprechenden Regeln für psychologische Verben konsultiert.⁸²

In den folgenden Abbildungen wird für jeden Typ psychologischer Verben jeweils nur eine Regel dargestellt, um zu zeigen, wie die Regeln mit psychologischen Verben aufgebaut sind, obwohl zu jedem dieser Typen mehr als eine Regel erstellt wurde. Eine vollständige Liste der zugehörigen Regeln findet sich in Appendix 1. Diese Regeln gehören zur Regelgruppe "experience", die dem Gruppennamen der entsprechenden Regeln auf der Grundlage psychologischer Verben entspricht.

```
elif re_morphmatch(morphs,morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼) $",
pos="^(J|V)$") and datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "sensation_predicate"
    datapoint["rule_order"] = "146"
```

Abbildung 23: Regel für Verben des Empfinden in Deklarativsätze

Die Regel in Abbildung 23 ordnet einem Empfindungsverb den Referenten "1sh+", sprich den Sprecher, zu, wenn das Prädikat den Wert "decl" im Datenpunkt "zp_mood" hat. Die Regel bezieht die Stämme von Prädikaten wie z. B. *chubda* 'sich kalt fühlen', *deobda* 'sich heiß fühlen', *jjibjjibhada* 'sich unwohl fühlen' und *neukkida* 'sich fühlen' ein und verwendet die POS-Informationen in den Zusatzdaten, nachdem diese Prädikate als POS-Tag entweder Adjektive (J) oder Verben (V) haben.

```
cognition_head_list = ["예정", "생각", "기억", "확인", "깜박", "예상", "착각", "이해", "확실", "집중",
"계획", "숙지", "오해", "파악", "동의", "사과", "깜빡"]
if [x for x in cognition_head_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "cognition_predicate"
    datapoint["rule_order"] = "136"
```

Abbildung 24: Regel für Verben des Denkens in Deklarativsätze

⁸² Da im Koreanischen sowohl Verben als auch Adjektive verwendet werden, um Empfindungen, Erkenntnisse oder Gefühle zu beschreiben, finden sich beide Wortarten in den entsprechenden Regeln. Aus Gründen der Vereinfachung werden in dieser Studie alle diese Verben bzw. Adjektive als psychologische Verben bezeichnet. Eine umfassende Liste der subjektiven Verben und Adjektive, wie von Yoo Hyun-kyung (1998) zusammengestellt, ist im Appendix von Park Arum & Hong Munpyo (2015) verfügbar.

Die Regel in Abbildung 24 verwendet kognitive Verben, wie z. B. *saenggaghada* 'denken', *gieoghada* 'sich erinnern', *ohaehada* 'missverstehen' und *kkambaghada* 'vergessen'. Findet die Regel eines der Prädikate in der Liste, das darüber hinaus den Wert "decl" als Satzmodusinformation besitzt, dann ordnet die Regel dem Prädikat den Sprecher als Referenten zu. Bei der Suche nach kognitiven Verben findet die Regel die Kopfnomen der Prädikate, da sie sich in ihren Formen nicht verändern.

```
elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|
    즐거우|무섭|기쁘|행복하|난감하|반갑|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "152"
```

Abbildung 25: Regel für Verben des Fühlens in Deklarativsätze

Die Regel in Abbildung 25 weist solchen Verben den Referenten "1sh+" zu, die Emotionen ausdrücken, wie z. B. *gippeuda* 'sich freuen', *haengboghada* 'glücklich sein', *seulpeuda* 'traurig sein' und *danghwangseuleobda* 'sich ärgern', wenn sie im Deklarativsatz vorkommen, also den Wert "decl" als Satzmodusinformation haben.

Von den drei oben eingeführten Regeln wird erwartet, dass sie den Nullsubjekten in den folgenden Beispielen, die aus dem Originalkorpus extrahiert wurden, die Sprecherrolle als Referenten zuordnen.

- (36) a. *[mwonga-ga mud-eo iss-eo]_{S1} [jjibjjibha-neyo-.]_{S2}*
 etwas-SBJ bekleckt-sein-CNJ unangenehm sein-SFS-SF
 'Da ist etwas (auf dem Blatt), und (ich) fühle (mich) unangenehm.'
- b. *jeonghwaghi gieogna-ji-neun anh-seubnida-*.
 genau sich erinnern-NEG-SFS.DECL-SF
 '(Ich) erinnere (mich) nicht genau.'
- c. *[masissge deu-si-eoss-dani]_{S1} [gippeu-gunyo-.]_{S2}*
 genüsslich essen.HON-HON-PST-CNJ sich freuen-SFS.DECL-SF
 '(Ich) freue (mich), dass (Sie) (Ihr) (Essen) genossen haben.'

Bei der Person, die sich in (36)a unwohl fühlt, in (36)b nicht genau erinnern kann oder in (36)c glücklich fühlt, handelt es sich um den Sprecher, der über seine aktuelle Erfahrung spricht. Es wird erwartet, dass jedem fettgedruckten Prädikat in (36) durch die drei oben genannten Regeln der Sprecher als Referent zugewiesen wird, da die in den Regeln genannten Bedingungen erfüllt sind.

Der Satzmodus spielt bei den linguistischen Mitteln auf lexikalischer Ebene eine wichtige Rolle, insbesondere bei den psychologischen Verben. Die Beispiele (37)a und (37)b enthalten das Gefühlsverb *haengboghada* 'glücklich sein' und zeigen, dass die Interpretation der Nullsubjekte je nach Satzmodus variiert.

- (37) a. *neomu haengbogha-neyo-*.
 sehr glücklich sein-SFS.DECL-SF
 '(Ich) bin sehr glücklich.'
- b. *neomu haengbogha-(si)-jyo-?*
 sehr glücklich sein-(HON)-SFS-SF
 '(Sie) sind sehr glücklich, nicht wahr?'

Im Beispiel (37)a wird *haengboghada* in einem Deklarativsatz verwendet, wobei das Nullsubjekt nur als Sprecher interpretiert wird, während es in (37)b nur als Adressat interpretiert werden kann, da *haengboghada* in einem Interrogativsatz benutzt wurde. Das Honorativinfix *si* in (37)b kann optional hinzugefügt werden. Subjekte solcher Prädikate werden oft weggelassen, da sie gewöhnlich entweder als Sprecher oder Adressat interpretiert werden können, was für die Diskursteilnehmer:innen eine alte Information ist (vgl. Yoo Hyun-kyung 2015: 83).

Unter Berücksichtigung dieser Eigenschaften psychologischer Verben werden zusätzliche Regeln für die Zuordnung von Referenten als Adressaten formuliert, wenn sie in Interrogativsätzen auftreten. Die folgende Regel zeigt beispielsweise, wie sie für Verben des Fühlens formuliert wird. Diese Formulierung gilt für alle Regeln, die auf psychologischen Verben basieren und in Interrogativsätzen auftreten.

```
elif re_morphmatch(morphs,morph="^(의숙하|자부하|망설여지|망설이|아침|상관없|즐겁|재미있|재미없|
    즐거우|무섭|기쁘|행복하|난감하|반갑|환영하|신기하|당황스럽|막막하|안타깝|부럽)$",
    pos="^(J|V)$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "135"
```

Abbildung 26: Regel für Verben des Fühlens in Interrogativsätzen

Die Regel in Abbildung 26 identifiziert ein Prädikat, das den mit "J" (Adjektiv) oder mit "V" (Verb) getaggten Stamm eines Gefühlsverbs (z. B. *haengboghada*) enthält, und weist ihm beim Wert "ques" den Referenten "2sh+", sprich die Adressatenrolle zu.

Im Unterschied zu den meist unveränderlichen Subjektbeschränkungen von Verbalsuffixen zeigen jene, die sich aus der Kombination von lexikalischen Hinweisen und Satzmodusinformation ergeben, eine gewisse Variabilität. Das bedeutet, dass in einigen Fällen die Subjektbeschränkung nicht mehr gültig ist, zum Beispiel wenn weitere linguistische Mittel wie das Verbalinfix *gess* hinzugefügt werden (vgl. Kapitel 1.2). Daher sollte dieses Phänomen auch bei der Erstellung der entsprechenden Regeln berücksichtigt werden. Die Beispiele in (38) zeigen deutlich, wie sich die Subjektbeschränkungen zwischen dem Verbalsuffix *llaeyo* und dem Gefühlsverb *haengboghada* 'glücklich sein' oder *geogjeonghada* 'besorgt sein' in Bezug auf die Hinzufügung anderer linguistischer Mittel unterscheiden.

- (38) a. *i os-eun je-ga ib-eullaeyo-*. /**ib-eusi-llaeyo-*. /**ip-gess-eullaeyo-*.
 dies Kleidung- ich- tragen- /*tragen-HON- /*tragen-IFX(Wille)-
 DP SBJ SFS(Wille)-SF SFS(Wille)-SF SFS(Wille)-SF
 'Ich will diese Kleidung tragen.'
- b. *jeong-mal haengbogha-gess-eoyo-*.
 sehr glücklich sein-IFX(Vermutung)-SFS-SF
 '(Sie/Er/*ich) sind/ist bestimmt sehr glücklich.'
- c. *neomu geogjeongha-ji anh-eusi-eodo dwae-yo-*.
 viel besorgt sein-NEG-HON-PPC(Pflicht)-SFS-SF
 '(Sie/Er/*ich) müssen/muss sich nicht viele Sorgen machen.'

Das Prädikat in (38)a (zitiert aus Kang Hyoun Hwa et al. 2016: 374) enthält das Verbalsuffix *llaeyo*, das nicht mit linguistischen Mitteln wie dem Honorativinfix (*eu*)*si* oder dem Verbalinfix *gess* kombiniert werden kann, was im Beispiel als ungrammatisch markiert wird. Der erste Grund dafür ist, dass das

Verbalsuffix *llaeyo* fest mit der Funktion verknüpft ist, den Willen des Sprechers in Deklarativsätze auszudrücken, so dass es unpassend ist, es mit dem Honorativinfix (*eu*)*si* zu kombinieren, weil das Infix verwendet wird, um Respekt gegenüber anderen Personen außer dem Sprecher auszudrücken. Umgekehrt ist es in Interrogativsätzen allerdings möglich, es mit dem Honorativinfix (*eu*)*si* zu kombinieren, um eben diese Respektskategorie gegenüber dem Adressaten zum Ausdruck zu bringen. Der zweite Grund, warum die in (38)a gezeigte Kombination nicht möglich ist, ist, dass die Bedeutungen wiederholt werden, die durch das Verbalsuffix *llaeyo* und das Verbal infix *gess* ausgedrückt werden. Wie das Verbalsuffix *llaeyo* kann auch das Verbal infix *gess* genutzt werden, um eine Willensäußerung darzustellen, so dass sie nicht kombiniert werden können, da es sonst zu ungrammatischen Redundanzen kommt. Diese Kombinationsbeschränkung tritt nicht nur in Deklarativsätzen, sondern auch in Interrogativsätzen auf (vgl. Kang Hyoun Hwa et al. 2016: 374; Jang Chaerin 2017: 7–8).

Im Gegensatz zum Verbalsuffix *llaeyo* kann bei psychologischen Verben das Verbal infix *gess* oder das Honorativinfix (*eu*)*si* wie in (38)b und (38)c hinzugefügt werden. In diesem Fall ist es nicht mehr möglich, das Subjekt des Gefühlsverbs *haengboghada* 'glücklich sein' oder *geogjeonghada* 'besorgt sein' im Deklarativsatz als Sprecher zu interpretieren. Dies widerspricht der in früheren Studien vertretenen Argumentation, dass Subjekte von psychologischen Verben in Deklarativsätzen grundsätzlich als Sprecher interpretiert werden. Das Subjekt des Prädikats in (38)b kann sich nicht auf den Sprecher beziehen, da das Verbalsuffix *gess* zum Ausdruck von Vermutungen hinzugefügt wurde. Yoo Hyun-kyung (1998) erklärt, dass das Hinzufügen von Vermutungsausdrücken einer der Umstände ist, der die Subjektbeschränkung von psychologischen Verben tilgt. Wie in Kapitel 1.2 erwähnt, kann dies auch als Verstoß gegen die Grice'schen Konversationsmaximen betrachtet werden, wonach der Sprecher deutlich sagen muss, wovon er bereits weiß, dass es wahr ist (wie z. B. sein aktuelles Gefühl), und es nicht als unsichere Vermutung angeben darf. Das Nullsubjekt des Prädikats in (38)c kann ebenfalls nicht als Sprecher interpretiert werden, da das Honorativinfix (*eu*)*si* generell nicht verwendet wird, um Respekthaltung gegenüber sich selbst anzuseigen.

Dies ist der Grund, warum das Nullsubjekt in (38)b und (38)c nur als ein anderes Subjekt als der Sprecher interpretiert werden kann. Im Prinzip könnte der Sprecher das Gefühl eines Dritten vermuten, wie in der deutschen Übersetzung von (38)b und (38)c gezeigt wird. Jedoch ist es aufgrund der Charakteristika der Gesprächssituationen im Korpus dieser Studie sehr unwahrscheinlich, dass die Äußerung dazu dient, die Gefühle eines Dritten zu erraten oder anzugeben, da die Diskursteilnehmer:innen sich zum ersten Mal in einer Reisesituation begegnen. Daher wird in dieser Studie angenommen, dass es eher wahrscheinlich ist, dass die Äußerung dazu dient, die Gefühle des Adressaten zu erraten oder anzugeben. Die folgenden Regeln wurden erstellt, um widerzuspiegeln, dass ein Nullsubjekt eines psychologischen Verbs mit hoher Wahrscheinlichkeit den Adressaten als Referenten hat, wenn das Verbal infix *gess* oder (*eu*)*si* hinzugefügt wird.

```
elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|
    즐거워|무섭|기쁘|행복하|난감해|반갑|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
    re_morphmatch(morphs, morph="^겠$", pos="^E$"):
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "151"
```

Abbildung 27: Regel für Verben des Fühlens mit dem Verbal infix *gess*

```

elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하고민되|
고민하|놀라|당황하|놀래|놀랍|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이) $",
pos="^(V|N)$") and datapoint["zp_hono"] == "plus":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "109"

```

Abbildung 28: Regel für Verben des Fühlens mit dem Verbal infix (eu)si

Die Regel in Abbildung 27 ordnet einem Prädikat einen Referenten "2sh+", also die Adressatenrolle zu, wenn das Prädikat den Stamm des mit "J" oder "V" getagten psychologischen Verbs und das mit "E" getagte Verbal infix *gess* enthält. In Abbildung 28 wird einem Prädikat mit einem Kopfnomen oder Verbstamm psychologischer Verben ein Referent in der Adressatenrolle zugewiesen, wenn es das Honorativ infix (eu)si enthält, das in den Zusatzdaten mit dem Wert "plus" markiert ist.

Wichtig bei den Regeln zu den psychologischen Verben ist die Reihenfolge der Regelanwendung. Zum Beispiel müssen die Regeln in Abbildung 27 und Abbildung 28 vor der Regel in Abbildung 25 angewandt werden, damit ein adressatenbezogenes Nullsubjekt nicht durch die Regel in Abbildung 25 bereits als Sprecher rekonstruiert wird. Dieser Punkt wurde bei der Festlegung der Reihenfolge der einzelnen Regeln berücksichtigt, wie aus dem Datenpunkt "rule_order" ersichtlich ist. So hat der Datenpunkt "rule_order" der Regel in Abbildung 27 und Abbildung 28 einen kleineren Wert (151, 109) als der Datenpunkt "rule_order" der Regel in Abbildung 25 (152). Das bedeutet, dass die ersten beiden Regeln früher angewandt werden als die dritte. Die Prinzipien für die Festlegung der Reihenfolge zur Anwendung der Regeln werden in Kapitel 4.3.3 ausführlicher behandelt.

Zusammenfassend kann der Wechsel des Satzmodus von Deklarativ- zu Interrogativsätzen bei Prädikaten wie psychologischen Verben eine Reinterpretation der Referenten von Nullsubjekten vom Sprecher hin zum Adressaten bewirken. Darüber hinaus kann das Hinzufügen eines Verbal infixes wie *gess* oder (eu)si eine Rolle dabei spielen, eine sprecherbezogene Äußerung in eine adressatenbezogene umzuwandeln.

Wie in den vorangegangenen Kapiteln gezeigt, wird in früheren Studien nur der Wechsel von der Zuweisung der Sprecher- zur Adressatenrolle beim Satzmoduswechsel erwähnt. Es wurde nicht darauf eingegangen, dass die Subjektbeschränkung auf den Sprecher mit der Hinzunahme von linguistischen Mitteln wie dem Verbal infix *gess* oder (eu)si nicht mehr gültig ist. In der vorliegenden Studie wird die Tendenz herausgearbeitet, dass Nullsubjekte der Prädikate bevorzugt als Adressaten interpretiert werden, wenn das Verbal infix *gess* als Marker für eine Vermutung oder das Honorativ infix (eu)si hinzugefügt wird. Darüber hinaus wird die Reihenfolge der Regelanwendung im NRS angepasst, um Konflikte zwischen den Regeln zu vermeiden. Aus diesem Grund wird erwartet, dass die Genauigkeit der in dieser Studie vorgeschlagenen heuristischen Regeln für die Resolution von Nullsubjekten hoch ist.

4.2.3.3. Honorifika

Die letzten linguistischen Mittel auf lexikalischer Ebene, die dazu beitragen können, die die Sprecher als Referenten von Nullsubjekten wiederherzustellen, sind mit dem Konzept der Honorifika verbunden. In Kapitel 4.1.3 wurde erläutert, dass Honorifika im Koreanischen mit den Kategorien Respekt und Bescheidenheit zusammenhängen. Linguistische Mittel, die die Respektkategorie ausdrücken, könnten als Honorifika für den Adressaten angesehen werden, da sie Respekt vor dem Adressaten ausdrücken, so dass die entsprechenden Hinweise als heuristische Regeln für die Zuweisung der Adressatenrolle formuliert wurden. Bei den linguistischen Mitteln, die Bescheidenheit ausdrücken können, zeigt der Sprecher eine höfliche Haltung gegenüber dem Gesprächspartner, indem er bescheiden auf sich selbst

referiert oder indem er seine Handlungen oder Absichten bescheiden ausdrückt. Diese Bescheidenheit wird auch über die Wahl des Pronomens der 1. Person Singular zum Ausdruck gebracht, bei der sich der Sprecher zwischen der Eigenreferenz mit *na* und *jeo* entscheiden kann. Dass der Sprecher das Personalpronomen *jeo* anstelle von *na* verwendet, um auf sich selbst zu referieren, beweist, dass er bescheiden über sich spricht und damit dem Adressaten gegenüber Höflichkeit zum Ausdruck bringt. Anhand der Beispielsätze in (39) (zitiert aus Piao 2012: 40) wird deutlich, inwiefern sich die Wahl der Personalpronomina unterscheidet, wenn sie sich auf den Sprecher selbst beziehen.

- (39) a. *na-do naeil haggyo-e ga-gess-eo-*.
 ich-auch morgen Schule-zu gehen-IFX-SFS-SF
 'Ich werde morgen auch zur Schule gehen.'
- b. *jeo-do naeil haggyo-e ga-gess-seubnida-*.
 ich.HUM-auch morgen Schule-zu gehen-IFX-SFS.DECL-SF
 'Ich werde morgen auch zur Schule gehen.'

Sätze wie in den Beispielen (39)a und (39)b werden verwendet, um den Willen des Sprechers auszudrücken, morgen zur Schule zu gehen. (39)b enthält einen Bescheidenheitsmarker in Form des Personalpronomens *jeo* und des Verbalsuffixes *seubnida*, die in (39)a fehlen. In Beispiel (39)b wird anstelle von *na* das Personalpronomen *jeo* verwendet, das Bescheidenheit kodiert, die als "HUM" (Humble) glossiert wird. Das Personalpronomen *jeo* wird verwendet, wenn der Gesprächspartner älter ist oder einen höheren sozialen Status hat als der Sprecher, oder wenn die Diskursteilnehmer:innen in einer Situation höflich zueinander sein müssen, z. B. wenn sie sich zum ersten Mal treffen (vgl. Piao 2012: 41). Neben der Wahl der Personalpronomina unterscheidet sich auch die Wahl der Verbalsuffixe in den beiden Äußerungen in (39). So kann das Personalpronomen *jeo* in (39)b nicht mit dem Verbalsuffix *eo* wie in (39)a verwendet werden. Das liegt daran, dass es erstens je nach Verbalsuffix Unterschiede in der Sprechweise gibt und zweitens steht für jede Sprechweise ein passendes Personalpronomen zur Verfügung. Die folgende Tabelle zeigt die repräsentativen Verbalsuffixe für den Modus Indikativ auf der Grundlage der sechs Sprechweisen im Koreanischen nach Seo Cheong-Soo (1997: 39).⁸³ Darüber hinaus wurden die Verbalsuffixe weitgehend danach unterteilt, ob sie als Honorifika gelten können oder nicht.

	Sprechweise	Verbalsuffix im Indikativ	Respekt
1	<i>aju-nopim</i> 'sehr hohe Sprechweise'	-bnida	[+honorativ]
2	<i>yesa-nopim</i> 'hohe Sprechweise'	-(s)o	
3	<i>dulu-nopim</i> 'allgemein verwendbare hohe Sprechweise'	-(eo)yo	
4	<i>yesa-najchum</i> 'informelle Sprechweise'	-ne	[-honorativ]
5	<i>aju-najchum</i> 'sehr informelle Sprechweise'	-da	
6	<i>dulu-najchum</i> 'allgemein verwendbare informelle Sprechweise'	-eo	

Tabelle 23: Einstufung der Verbalsuffixe im Indikativ nach Sprechweise (nach Seo Cheong-Soo 1997: 39)

⁸³ Die Gesamtliste der Verbalsuffixe für andere Satzmodi wie Interrogativ, Imperativ und Hortativ und ihre Klassifizierung nach Respektverhältnis, Sprechweise und Formalität ist in Tabelle 1 von Cho Youngkil (2005: 52) aufgeführt.

Wie in Tabelle 23 zu sehen ist, gibt es im Koreanischen sechs Sprechweisen, die sich weitgehend in hohe Sprechweisen (die sog. *nopim*-Sprechweisen) und informelle Sprechweisen (*najchum*-Sprechweisen) unterteilen lassen. Die ersten drei Sprechweisen gehören zu den hohen Sprechweisen, die eine honorative Funktion ausüben, während die übrigen drei Sprechweisen zu den informellen Sprechweisen gehören, d. h. ohne honorative Funktion sind. In jeder der beiden Sprechweisengruppen gibt es einen Unterschied im Grad der honorativen Funktion, wie aus den deutschen Übersetzungen ersichtlich ist. In Tabelle 23 sind die sechs repräsentativen Verbalsuffixe, die für den Indikativ verwendet werden können, nach ihren Sprechweisen sortiert. Es ist wichtig zu wissen, ob ein Verbalsuffix die Semantik [honorativ] ausdrücken kann, weil diese Information eine wichtige Rolle bei der Bestimmung der geeigneten Personalpronomen spielt. Zum Beispiel gehört das Verbalsuffix *eo* in (39)a zur sechsten Sprechweise *dulu-najchum*, in der die Personalpronomen *na* 'ich' und *neo* 'du' als Subjekt des Prädikats mit diesem Verbalsuffix geeignet sind. Das Verbalsuffix *seubnida* in (39)b gehört zur ersten Sprechweise *aju-nopim*, in der die Personalpronomen *jeo* 'ich' oder *dangsin* 'Sie' als Subjekt des Prädikats geeignet sind, weil sie die höfliche Variante von *na* 'ich' oder *neo* 'du' sind. Das bedeutet, dass selbst wenn ein Subjekt nicht overt realisiert wird, sein Personalpronomen vorhersagbar ist, je nachdem, welches Verbalsuffix mit welcher honorativen Funktion verwendet wird. Aus diesem Grund wird in dieser Studie der Referent eines Nullsubjekts als "1sh+" oder "2sh+" bezeichnet, wenn ein Verbalsuffix, das zu einer hohen Sprechweise gehört, im Prädikat enthalten ist, das die Eigenschaft [+Honorativ] hat. Die Angabe zur Sprechweise bei der Wahl eines Verbalsuffixes lässt allein jedoch keinerlei Einschränkungen in Bezug auf die grammatische Person zu. Vielmehr hilft die Klassifikation der Sprechweise dabei, festzustellen, ob die Honorativform der Subjektpersonen verwendet werden sollte.

Die Tatsache, dass einige Verben zum Ausdruck von Bescheidenheit genutzt werden, ermöglicht es, sie zu verwenden, um ein Nullsubjekt als Sprecher zu interpretieren, was zum Beispiel beim Hilfsverb *a/eo deulida* der Fall ist, wo die Höflichkeitsform *a/eo juda* wäre.⁸⁴ Die Grundbedeutung von *a/eo juda* wird als "beneficiary" bezeichnet, d. h. die durch das Hauptverb ausgedrückte Handlung wird als für jemanden nützlich angesehen (vgl. Seo Cheong-Soo 1996: 638). Im Allgemeinen geht man davon aus, dass diese Grundbedeutung von *a/eo juda* auch in der Honorativform *a/eo deulida* enthalten ist. Es ist bekannt, dass es keine besondere Beschränkung der grammatischen Personen beim Subjekt für ein Prädikat mit *a/eo juda* gibt. Es wird jedoch erklärt, dass es eine Beschränkung gibt, dass das Subjekt des Hilfsverbs *a/eo juda* immer ein Agens sein muss, das mit Bewusstsein und Absicht tätig werden kann (Bang Woonkyu 1994; Roh Chae Hwan 2006). Ebenso sollte das Subjekt eines Prädikats mit *a/eo deulida* das Attribut [+belebt] haben (vgl. Bae Soo-Ja 2007: 206; Kang Hyoun Hwa et al. 2016: 529). Obwohl in der Literatur bisher keine Einschränkungen bezüglich des Subjekts bei Prädikaten mit *a/eo deulida* für eine bestimmte grammatische Person erwähnt wurden, wird in dieser Studie erwartet, dass ein Nullsubjekt dieses Prädikats mit hoher Wahrscheinlichkeit als Sprecher interpretiert wird. Der Grund dafür ist, dass das Prädikat verwendet wird, um Bescheidenheit in Bezug auf das Verhalten des Sprechers zu signalisieren (vgl. Cho Yongkil 2005: 51). Mit anderen Worten: *a/eo deulida* kann benutzt werden, um die Bescheidenheit des Sprechers selbst auszudrücken, wenn er einer anderen Person einen Gefallen tut. Die folgenden Beispiele sind dem Originalkorpus entnommen, um zu zeigen, wie das Hilfsverb *a/eo deulida* in einer Äußerung verwendet wird und wie das Nullsubjekt interpretiert werden kann.

⁸⁴ Das Hilfsverb *a/eo juda* ist eine der Komponenten der in Kapitel 4.1.2 eingeführten Muster. Die Muster können dazu verwendet werden, höflich eine Bitte zu äußern, indem gefragt wird, ob der Adressat bereit oder in der Lage ist, etwas für den Sprecher zu tun (siehe Tabelle 20). Im Gegensatz zu *a/eo juda* kann *a/eo deulida* nicht für eine Bitte oder einen Befehl verwendet werden, da es eine Funktion in der Kategorie Höflichkeit hat (vgl. Bae Soo-Ja 2007: 207).

- (40) a. [deo-usi-myeon]_{s1} [dangyeonhi **teul-eo deuli-eoyajyo-**]_{s2}
 heiß sein-HON-wenn definitiv einschalten-AUX.HUM-SFS.DECL-SF
 'Wenn es Ihnen heiß ist, sollte (ich) sie definitiv einschalten.'
- b. **seonmul pojang ha-e deuli-Ikkayo-?**
 Geschenk Einpacken machen-AUX.HUM-SFS.INT-SF
 'Soll (ich) (es) (Ihnen) als Geschenk einpacken?'
- c. [deoisang **dowadeuli-ji mosha-e**]_{s1} [joisongha-bnida-]_{s2}
 nicht mehr helfen.HUM-NEG-CNJ tut jm. Leid-SFS.DECL-SF
 'Tut mir leid, dass (ich) (Ihnen) nicht weiterhelfen konnte.'

Der Satz in Beispiel (40)a ist eine Äußerung, in der ein Hotelangestellter sagt, dass er die Klimaanlage für einen Hotelkunden einschalten wird. In (40)b fragt ein Verkäufer, ob er einen von einem Kunden gekauften Artikel als Geschenk einpacken soll. Mit der Äußerung in (40)c bringt der Sprecher zum Ausdruck, dass es ihm leidtut, dass er keine weitere Hilfe geben kann. Betrachtet man die fettgedruckten Prädikate, so werden alle Nullsubjekte als Sprecher interpretiert, da sie die Handlungen darstellen, die der Sprecher für den Adressaten zu tun bereit ist, indem er das Hilfsverb *a/eo deulida* verwendet. Es ist auch erwähnenswert, dass die Interpretation des Nullsubjekts als Sprecher unabhängig von den Satztypen der Prädikate möglich ist, so ist die Interpretation die gleiche in einem Deklarativsatz (40)a, in einem Interrogativsatz (40)b und sogar in einem Nebensatz (40)c. Das bedeutet, dass die Präsenz des Hilfsverbs *a/eo deulida* allein ausreicht, um die Interpretation als Sprecher bei Nullsubjekten zu ermöglichen. Dies ist der Grund, warum die Regeln, die auf Basis dieses Hilfsverbs formuliert wurden, keine Informationen zum Satzmodus spezifizieren.

Die folgenden Regeln wurden aufgestellt, um einem Prädikat, das das Hilfsverb *a/eo deulida* enthält, den Sprecher als Referenten zuzuweisen.

```
elif re_morphmatch(morphs, morph="^ (아드리|어드리)$", pos="^X$") :
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness auxiliary verb"
    datapoint["rule_order"] = "156"
```

Abbildung 29: Regel für das Hilfsverb *-a/eo deulida*

Die Regel in Abbildung 29 weist einem Prädikat den Referenten "1sh+", also den Sprecher, zu, wenn das Prädikat einen Stamm des Hilfsverbs *a/eo deulida* mit dem Wert "X" für seine POS-Tagging-Information enthält, die einem Hilfsverb zugeordnet ist.

```
elif re_morphmatch(morphs, morph="^ (가져다드리|알려드리|보여드리|해드리|연락드리|도와드리|
  둘려드리|전해드리|전화드리|갖다드리)$", pos="^V$") :
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "157"
```

Abbildung 30: Ergänzende Regel bei Verwendung des Hilfsverbs *-a/eo deulida*

Die Regel in Abbildung 30 berücksichtigt die Tatsache, dass das Hilfsverb *a/eo deulida* bei bestimmten Verben vom POS-Tagger nicht als solches erkannt wird, sondern mit ihm zusammen als ein einziges Verb betrachtet wird. Zum Beispiel enthält die Form *dowa deulida* 'helfen.HUM' das Hilfsverb *a deulida*, wird aber als nur ein Verb mit dem Stamm *dowa deuli* erkannt. Daher kann mit der Regel in Abbildung

29 einem Prädikat, das diese Art von Verb enthält, kein Sprecher als Referent zugewiesen werden. Aus diesem Grund umfasst die Regel in Abbildung 30 Stämme dieser Art von Verben als String und ordnet dem Prädikat, das einen dieser Stämme enthält und die Tagging-Information "V" hat, einen den Referenten "1sh+" zu.

4.3. Ergänzende heuristische Regeln

An diesem Punkt wird eine detaillierte Erklärung der ergänzenden Regeln und ihrer Vorteile für die Nullsubjektresolution präsentiert. Zunächst werden die Regeln für die Resolution von Nullsubjekten in Nebensätzen eingehend betrachtet und mit einer früheren Resolutionsstrategie verglichen, um die Vorteile der hier entwickelten Regeln zu verdeutlichen, die eine morphologisch bedingte Zuweisung ohne die Notwendigkeit einer syntaktischen Analyse des Nebensatzes ermöglichen (Kapitel 4.3.1). Danach werden vier Regeln im NRS eingeführt, die Prädikate ausschließen, deren Subjekte weniger wahrscheinlich als Diskursteilnehmer interpretiert werden. Diese Regeln wurden entwickelt, um die Leistung der Nullsubjektresolution des NRS zu verbessern, indem sie bestimmte Prädikatstypen identifizieren und ausschließen (Kapitel 4.3.2). Zum Abschluss werden die Prinzipien für die Anpassung der Regelreihenfolge im NRS beschrieben. Dies ist von entscheidender Bedeutung, da die Reihenfolge der Regelanwendung in einigen Fällen kritisch für die Leistung der Regeln ist, und um sicherzustellen, dass Referenten nicht fälschlicherweise anderen Prädikaten zugeordnet werden (Kapitel 4.3.3).

4.3.1. Regeln zur Resolution von Nullsubjekten in Nebensätzen

Heuristische Regeln auf Basis von Satzmodusinformationen können keine Nullsubjekte über die Satzgrenze hinaus erfassen und rekonstruieren. Dies liegt daran, dass die Information über den Satzmodus nur aus einem Hauptsatz zu ziehen sind, der durch Verbalsuffixe, die den Satzmodus markieren, und Satzzeichen gekennzeichnet ist, falls Verbalsuffixe wie *-a-yo/eoyo/yeoyo* gebraucht werden, die für verschiedene Satztypen genutzt werden. Im Koreanischen wird ein Satz mit einem Verbalsuffix am Ende eines zusammengesetzten Satz als Hauptsatz betrachtet, während ein Satz, der durch Konjunktionen wie z. B. *go* 'und' und *eoseo* 'weil' vor dem Hauptsatz eingeleitet wird, als Nebensatz gilt. Die Unterscheidung zwischen koordinierenden und subordinierenden Strukturen im Koreanischen erfolgt im Allgemeinen anhand der Art der verwendeten Konjunktionen. Beispielsweise werden Sätze, die durch die Konjunktion *go* 'und' verbunden sind, als koordinierende Struktur charakterisiert, während Sätze, die durch die Konjunktion *eoseo* 'weil' verbunden sind, als subordinierende Struktur betrachtet werden. In dieser Arbeit werden Konjunktionen als gute Indikatoren für die Nullsubjektresolution in beiden Arten der Satzverknüpfung angesehen.

Die Ansichten über die syntaktische Struktur komplexer koreanischer Sätze sind noch nicht einheitlich (vgl. Lee Chung Hoon 2015). Das Beispiel (41) aus Mok Joungsoo & You Hyunjo (2007: 281) veranschaulicht, wie im Koreanischen zwei Sätze zu einem zusammengesetzten Satz verbunden werden, und wird in dieser Arbeit analysiert. In dieser Studie werden Sätze, die durch eine Konjunktion verbunden sind, der Einfachheit halber, wie in (41) beschrieben, ohne Unterscheidung zwischen koordinierten und subordinierten Strukturen behandelt. Bei der Analyse der syntaktischen Struktur eines zusammengesetzten Satzes wird die Konjunktion innerhalb des Satzes mit Lee Chung Hoon (2015) als Kopf eines Nebensatzes S1 betrachtet, nicht als Kopf des Gesamtsatzes. Ein weiterer Grund für die Beschreibung der Struktur des komplexen Satzes wie in (41) liegt darin, dass die Konjunktion gemeinsam mit anderen linguistischen Mitteln als Indikator für die Resolution der Nullsubjekte in Nebensätzen dienen kann.

- (41) [[*bom-i* ***ga-go***]_{S1} [*yeoleum-i* *o-nda-*.]S₂]S
 Frühling-SBJ gehen-CNJ Sommer-SBJ kommen-SFS.DECL-SF
 'Der Frühling geht und der Sommer kommt.'

In (41) verbindet die Konjunktion *-go* die erste Proposition, dass der Frühling geht, mit der zweiten Proposition, dass der Sommer kommt. Diese Konjunktion hat mehrere Funktionen und kann benutzt werden, wenn die Handlungen, die in den damit verbundenen Sätzen ausgedrückt werden, in zeitlicher Abfolge stattfinden (vgl. Kang Hyoun Hwa et al. 2016: 158).

Wenn ein Satzglied, insbesondere das Subjekt, in zwei aufeinanderfolgenden Satzbestandteilen eines komplexen Satzes gleich ist, erfolgt häufig eine Auslassung dieses Satzglieds. Park Cheong-hee (2013: 101) präsentiert Statistiken über die Auslassung von Satzgliedern in komplexen Sätzen in ihrem Korpus bestehend aus Drama-Skripten, Drehbüchern für Filmszenen und dem "hyeondae gueo wonsi malmungchi"⁸⁵. Sie stellt bei ihrer Auswertung fest, dass die Auslassungsrate von Subjekten bei etwa 90 % liegt, was sehr hoch ist im Vergleich mit der Auslassungsrate anderer Satzglieder wie Prädikaten, Objekten und Ergänzungen, die unter 10 % liegt.

Diese Auslassung betrifft also bevorzugt Subjekte komplexer Sätze, wenn sie für beide Satzbestandteile gleich sind (vgl. Kwon Jae Il 1985: 113; Park Cheong-hee 2013: 220). Ob das Subjekt des vorangehenden oder des folgenden Satzes ausgelassen wird, hängt mit der hierarchischen Struktur des Gesamtsatzes zusammen. Es ist üblich, dass in koordinierten komplexen Sätzen das Subjekt eines folgenden Satzes weggelassen wird, während die Subjekte wie in subordinierten Nebensätzen üblicherweise ausfallen (Kwon Jae Il 1985; Kim Ok 1998; Park Cheong-hee 2013).

Beispiel (42) (zitiert nach Kim Ok 1998: 19) illustriert die Situation bei komplexen Sätzen mit koordinierten Satzbestandteilen und Beispiel (43) zeigt die Auswirkungen auf Sätze mit subordinierten Satzbestandteilen. Anhand dieser Beispiele ist es möglich, einen Unterschied bei der Subjektauslassung je nach Art der komplexen Sätze zu erkennen.

- (42) a. [?][*yeonghui-ga* *aleubdab-go*]S₁ [*yeonghui-ga* *seulgilob-da-*.]S₂
 Younghee-SBJ schön sein-CNJ Younghee-SBJ weise sein-SFS.DECL-SF
 'Younghee ist schön und Younghee ist weise.'
- b. [*yeonghui-ga* *aleumdab-go*]S₁ [Ø_i *seulgilob-da-*.]S₂
 Younghee-SBJ schön sein-CNJ Ø.SBJ weise sein-SFS.DECL-SF
 'Younghee ist schön und ist weise.'
- c. *[Ø_i *aleumdab-go*]S₁ [*yeonghui-ga* *seulgilob-da-*.]S₂
 Ø.SBJ schön sein-CNJ Younghee-SBJ weise sein-SFS.DECL-SF
 'Ist schön und Younghee ist weise'

Die Konjunktion *-go* verbindet S1 und S2 in (42) und die verbundenen Sätze besagen, dass Younghee schön und weise ist. (42)a wird als eher unnatürlich angesehen, da in beiden Sätzen das gleiche Subjekt *Younghee* overt realisiert wird. Deshalb ist dieser Satz als potentiell ungrammatisch markiert. Die Beurteilung der Grammatikalität von Satz (42)a kann je nach Standpunkt des Forschers variieren. Dies liegt daran, dass nach einigen Auffassungen der Satz als ungrammatisch betrachtet werden könnte. Dieser Standpunkt basiert darauf, dass in der Tiefenstruktur, wenn es zwei oder mehr identische Subjekte gibt, eines davon weggelassen werden muss (vgl. Park Cheong-hee 2013: 220). In

⁸⁵ Das Korpus "hyeondae gueo wonsi malmungchi", übersetzt als 'modern gesprochenes primitives Korpus', besteht aus transkribierten Daten des gesprochenen Koreanischen. Es setzt sich aus fünf Transkriptionsdateien zusammen, die Äußerungen in Standardkoreanisch enthalten. Diese wurden von Park Cheong-hee (2013) aus 200 gesprochenen Transkriptionsdateien im Rahmen des 21st Century Sejong Project ausgewählt. Weitere Informationen zu den Daten sind bei Park Cheong-hee (2013: 67–68) verfügbar.

koordinierten komplexen Sätzen mit demselben Subjekt wird es als am natürlichssten angesehen, das Subjekt des folgenden Satzes S2 wie in (42)b auszulassen, während es unnatürlich wäre, das Subjekt des vorangehenden Satzes S1 wie in (42)c wegzulassen. In den deutschen Übersetzungen wurde versucht, die koreanische Struktur zu spiegeln, damit die Probleme des koreanischen Satzes illustriert werden können. In (42)a ist es redundant, dass das Subjekt *Younghée* wiederholt wird, in (42)b werden die beiden Eigenschaften *schön* und *weise* von *Younghée* natürlicherweise durch *und* verbunden, und in (42)c wird *Younghée*, das Subjekt von *schön*, nicht zuerst in S1 präsentiert, was den Satz ebenso wie im Koreanischen ungrammatisch macht.

In (43) (zitiert aus Kim Ok 1998: 19–20) wird die Konjunktion *-seo* verwendet, um einen subordinierenden komplexen Satz zu bilden, in dem das Subjekt *Younghée* auf den Markt geht und dort Obst kauft.

- (43) a. ²[*yeonghui-ga* *sijang-e* *ga-seo*]_{S1} [*yeonghui-ga* *gwa-il-eul* *sa-ass-da-*]_{S2}
 Younghee-SBJ Markt-auf gehen-CNJ Younghee-SBJ Obst-OBJ kaufen-PST-SFS.DECL-SF
 'Younghée ging auf den Markt und Younghée kaufte Obst.'
- b. [*yeonghui-ga* *sijang-e* *ga-seo*]_{S1} [\emptyset_i *gwail-eul* *sa-ass-da-*]_{S2}
 Younghee-SBJ Markt-auf gehen-CNJ \emptyset .SBJ Obst-OBJ kaufen-PST-SFS.DECL-SF
 'Younghée ging auf den Markt und kaufte Obst.'
- c. [\emptyset_i *sijang-e* *ga-seo*]_{S1} [*yeonghui-ga* *gwail-eul* *sa-ass-da-*]_{S2}
 \emptyset .SBJ Markt-auf gehen-CNJ Younghee-SBJ Obst-OBJ kaufen-PST-SFS.DECL-SF
 'Auf dem Markt kaufte Younghée Obst.'

Wie im Beispiel (42)a wird auch (43)a als redundant angesehen, da in beiden Sätzen die gleichen Subjekte realisiert werden. Im Gegensatz zu koordinierten komplexen Sätzen kann in einem subordierenden komplexen Satz eines der Subjekte entweder in S1 oder S2 ohne Einschränkung weggelassen werden, wie in (43)b oder (43)c gezeigt wird. Allerdings gibt es einen leichten Bedeutungsunterschied zwischen den beiden letztgenannten Sätzen, wie die deutsche Übersetzung zeigt. Hier kann der Sprecher entscheiden, welches Subjekt er entsprechend seiner Absicht nicht overt realisiert.

Wie in Kapitel 2 dargestellt, nutzt Hong Munpyo (2011) die Subject-Sharing-Eigenschaften koreanischer Konjunktionen, um die Referenten von Nullsubjekten in Nebensätzen wiederherzustellen. Er stellt insgesamt sieben Konjunktionen vor, die Subject-Sharing-Eigenschaften aufweisen:

1. *-go*,
2. *-eoseo*,
3. *-myonseo*,
4. *-myeo*,
5. *-lyeogo*,
6. *-lyeomyeon* und
7. *-neula(go)*.

Die ersten beiden Konjunktionen werden zur Bildung komplexer Sätze wie in den Beispielen (42) und (43) gebraucht. Die von Hong Munpyo (2011) vorgeschlagene Strategie zur Resolution der Nullsubjekte in Nebensätzen besteht darin, die Nullsubjekte in S2, die durch eine der sieben Konjunktionen mit S1 verbunden sind, zu rekonstruieren, indem angenommen wird, dass sie dieselben sind wie die expliziten Subjekte in S1. In dieser Strategie wird z. B. in (42)b oder (43)b das explizite Subjekt *Younghée* in S1 als Antezedenz des Nullsubjekts in S2 bestimmt. Auf diese Weise kann das Nullsubjekt in S1 in (43)c jedoch nicht rekonstruiert werden, da die Resolutionsstrategie auf Basis der Subject-Sharing-Eigenschaften

nur die Fälle behandelt, in denen die Antezedenzen vor den Nullsubjekten stehen, nicht aber die Nullsubjekte mit kataphorischen Referenten.

In der vorliegenden Arbeit werden die Subject-Sharing-Eigenschaften bei der Wiederherstellung der Referenz von Nullsubjekten in Nebensätzen nicht verwendet, da es bei ihrer Anwendung verschiedene Probleme geben kann. Erstens weisen die sieben von Hong Munpyo (2011) vorgeschlagenen Konjunktionen je nach Verwendung nicht immer Subject-Sharing-Eigenschaften auf. Die folgende Tabelle zeigt zum Beispiel die Konjunktionen *-go* und *-eoseo* je nach Verwendungsart in Anlehnung an Kang Hyoun Hwa et al. (2016).

Nr.	Konjunktion	Nr.	Verwendung	Subject-Sharing
1	<i>-(eo)seo</i>	1.1	Kausale Beziehung	–
		1.2	Temporale Abfolge	+
2	<i>-go</i>	2.1	Einfache Auflistung	–
		2.2	Temporale Abfolge	–
		2.3	Konsekutivsatz	+

Tabelle 24: Verwendung der Konjunktionen *-eoseo* und *-go* und ihre Subject-Sharing-Eigenschaften

Die Konjunktion *-(eo)seo* hat nach Kang Hyoun Hwa et al. (2016) zwei Hauptverwendungen: In einem durch *-(eo)seo* verbundenen Satz kann S1 die Ursache oder den Grund von S2 angeben (1.1) oder S1 und S2 treten in zeitlicher Reihenfolge auf (1.2). Im Fall von 1.1 brauchen die Subjekte von S1 und S2 nicht die gleichen zu sein, während sie im Verwendungskontext 1.2 gleich sein müssen (vgl. Kang Hyoun Hwa et al. 2016: 203). Daher kann es Fälle geben, in denen das rekonstruierte Nullsubjekt falsch ist, wenn das explizite Subjekt als das Antezedens des Nullsubjekts ohne Berücksichtigung der Verwendung gewählt wird.

Ahnliches gilt für die Konjunktion *-go*: Laut Kang Hyoun Hwa et al. (2016) hat die Konjunktion drei Hauptverwendungsarten: S1 und S2 können mithilfe von *-go* einfach ungeachtet der Reihenfolge aufgelistet werden (2.1) oder in zeitlicher Reihenfolge verbunden sein (2.2). Hier gibt es keine Einschränkung, dass die Subjekte von S1 und S2 gleich sein müssen. Allerdings müssen die Subjekte von S1 und S2 gleich sein, wenn die Handlung oder das Ergebnis von S2 eintritt, während die Handlung oder das Ergebnis von S1 weiterläuft (Kang Hyoun Hwa et al. 2016: 160), d. h., wenn es sich um einen konsekutiven Satz handelt. Die Beispiele (44)a und (44)b mit den Konjunktionen *-eoseo* und *-go* wurden aus dem Originalkorpus extrahiert, allerdings in den Verwendungskontexten 1.1 und 2.1 ohne Subject-Sharing-Eigenschaften.

- (44) a. [Ø *alaecheung-eoseo* *dambae-leul* *pyi-eoseo*_{S1} *[changmun-eulo* *yeongi-ga* *deuleo-bnida-*]_{S2}
Ø.SBJ unten-von Zigarette- OBJ rauchen- CNJ Fenster- durch Rauch- hierinkommen-
hereinkommen-SF

'Weil (man) unten Zigarette raucht, kommt der Rauch oben durch das Fenster herein.'

- b. [Ø *jinanju-e* *yeyagha-ess-go*_{S1} *[yeogwon-eun* *yeogi iss-seubnida-*]_{S2}
Ø.SBJ letzte Woche-ADV reservieren-PST-CNJ Reisepass.SBJ-DP hier sein-SFS.DECL-SF
'(Ich) habe letzte Woche reserviert, und mein Reisepass ist hier.'

S1 in (44)a gibt die Ursache für den Rauch in S2 an, der durch das Fenster eintritt, womit es sich in S1 um die Angabe der Wirkung der Handlung in S2 handelt. Das Nullsubjekt in S1 referiert auf eine Person, die raucht, während das Subjekt von S2 *yeonji* 'Rauch' ist. Dies zeigt, dass die Subjekte von S1 und S2 unterschiedlich sind, auch wenn sie durch *-eoseo* verbunden sind. Die Äußerung in (44)b erfolgte beim Check-in in einer Unterkunft. S1 und S2 in (44)b werden ohne Einhaltung einer bestimmten Reihenfolge aufgelistet, wobei der Sprecher in S1 angibt, dass er die Unterkunft letzte Woche gebucht

habe. Im Anschluss sagt er in S2, dass sein Reisepass „hier“ sei. In (44)b referiert das Nullsubjekt in S1 auf den Sprecher und das Subjekt in S2 ist *yeogwon* 'Reisepass'. Das bedeutet, dass ein durch *-go* verbundener Satz nicht garantiert, dass die Subjekte der beiden Sätze identisch sind.

Die von Hong Munpyo (2011) vorgeschlagene Strategie kann auch nicht mit Fällen umgehen, in denen das Nullsubjekt in S1 vorkommt und das potenzielle Postzedens in S2 steht, wie in (43)c dargestellt. Selbst wenn man annimmt, dass das explizite Subjekt in S2 dasselbe ist wie das Nullsubjekt in S1, kann die Vorhersage einer solchen Regel falsch sein, da diese Strategie die Verwendung von Konjunktionen nicht berücksichtigt, die möglicherweise kein Subject-Sharing zulassen. Auch wenn der Verwendungskontext einer Konjunktion in zusammenhängenden Sätzen Subject-Sharing zulässt, kann die Strategie von Hong Munpyo (2011) bei der Resolution eines Nullsubjekts im Nebensatz nicht benutzt werden, wenn beide Subjekte in S1 und S2 nicht realisiert sind. Bei den Beispielen in (45) handelt es sich um Fälle, in denen in keinem Satz mit den Konjunktionen *-seo* und *-go* ein Subjekt overt realisiert ist, und die den Verwendungskontexten in 1.2 und 2.3 entsprechen, in denen die Eigenschaft des Subject-Sharing vorliegt.

- (45) a. *[ppalli jib-e ga-seo]S1 [swi-go sip-da-.]S2*
schnell Hause-nach gehen-CNJ sich ausruhen-PPC-SFS.DECL-SF
'(Ich) möchte schnell nach Hause gehen und mich ausruhen.'
- b. *[sueob-e neuj-eoseo]S1 [taegsi-leul ta-go]S2 [haggyo-e ga-neun jungi-ya-.]S3*
Unterricht-ADV spät sein-CNJ Taxi-OBJ nehmen-CNJ Schule-zu fahren-PPC-SFS-SF
'(Ich) nehme ein Taxi und (ich) fahre zur Schule, weil (ich) für den Unterricht spät dran bin.'

Im Beispiel (45)a (zitiert aus Kang Hyoun Hwa et al. 2016: 203) liegt eines der von Kang Hyoun Hwa et al. (2016) vorgestellten Beispiele für die Konjunktion *-seo* mit der Eigenschaft des Subject-Sharing vor. S1 und S2 sind in einer zeitlichen Reihenfolge verbunden, d. h. zuerst erfolgt das Nachhause-Gehen und später das Ausruhen. In (45)a besitzt die Konjunktion *-seo* die Eigenschaft des Subject-Sharing, weil die Person, die schnell nach Hause gehen und sich zu Hause ausruhen möchte, der Sprecher ist. Der Rückgriff auf die Möglichkeit des Subject-Sharing ist dennoch nicht hilfreich, um die Nullsubjekte in (45)a zu rekonstruieren, da kein Subjekt realisiert wird. Beim Beispiel (45)b (zitiert aus Kang Hyoun Hwa et al. 2016: 160) handelt es sich um ein Beispiel von Kang Hyoun Hwa et al. (2016), in dem die Konjunktion *-go* in einem Verwendungskontext steht, in dem die Eigenschaft des Subject-Sharing benutzt wurde. So wird die Konjunktion *-go* in (45)b verwendet, um die Sätze S2 und S3 zu verbinden. Die Handlung in S2 (ein Taxi nehmen) wird fortgesetzt, und infolgedessen kann die Person, die das Taxi nimmt, zur Schule gehen (S3). Da die Person, die ein Taxi nimmt und zur Schule geht, mit dem Sprecher identisch ist, wird die Subjektbeschränkung der Konjunktion *-go* eingehalten, die besagt, dass die Subjekte der beiden verbundenen Sätze gleich sein müssen. Wie bei (45)a ist das Subject-Sharing jedoch kein Indikator, wie die Nullsubjekte in (45)b zu rekonstruieren sind, da es kein realisiertes Subjekt gibt, das als Referenzpunkt genutzt werden kann.

Ein weiterer Grund, der das Potenzial einer Konjunktion zum Subject-Sharing zur Wiederherstellung der Referenz eines in einem Nebensatz gefundenen Nullsubjekts erschwert, ist die Tatsache, dass ein zusammenhängender Satz, der aus mehr als drei Sätzen besteht, eine hierarchische Struktur und nicht immer eine flache Struktur aufweist. Dies bedeutet, dass zuerst eine Analyse der syntaktischen Struktur der verbundenen Sätze notwendig ist. Das Beispiel (46) ist Teil eines Diskurses aus dem Originalkorpus, in dem A, der ein Volleyballstadion besucht, B nach den Einrichtungen des Stadions fragt.

- (46) A. *hwajangsil-do geugos-e iss-nayo?*
 Toiletten-auch dort-ADV es gibt-SFS.INT-SF
 'Gibt es dort auch Toiletten?'
- B. *[deuleoga-seo]_{S1} [balo oenjjog-e iss-euni]_{S2} [geujjog-eulo ga-seyo-.]_{S3}*
 reingehen-CNJ gleich linke Seite-auf es gibt-CNJ die Seite-auf gehen-SFS.IMP-SF
 'Wenn (Sie) hineingehen, befinden (die Toiletten) sich direkt auf der linken Seite, also gehen (Sie)
 auf diese Seite.'

In (46) stellt Sprecher A die Frage an Adressat B, ob es im Stadion Toiletten gibt. B antwortet, dass es auf der linken Seite Toiletten gebe, wenn A das Stadion betrete und auf die linke Seite gehe. Der Antwortsatz von B besteht aus insgesamt drei Sätzen, die durch die Konjunktion *-seo* in S1 und die Konjunktion *-euni* in S2 verbunden sind. Wie zuvor erläutert, weist die Konjunktion *-seo* die Eigenschaft des Subject-Sharing auf, wenn die durch sie verbundenen Sätze in einer zeitlichen Reihenfolge stehen. Es ist wichtig zu beachten, dass die Sätze im Antwortsatz von B nicht linear, sondern hierarchisch strukturiert sind. Das bedeutet, dass S1 mit der Konjunktion *-seo* nicht direkt mit dem unmittelbar folgenden S2 verbunden ist, sondern mit dem letzten Satz S3. Diese syntaktische Struktur wird deutlich, wenn die Bedeutung dieser Äußerung betrachtet wird: Hier teilt der Sprecher B dem Adressaten A mit, dass er nach dem Betreten des Stadions (S1) auf die (linke) Seite gehen soll (S3), um dort Toiletten zu finden (S2). Die Reihenfolge der Sätze in der Äußerung von B könnte entsprechend seiner kommunikativen Absicht wie folgt umgestellt werden: *[balo oenjjog-e iss-euni]_{S2} [deuleoga-seo]_{S1} [geujjog-eulo ga-seyo.]_{S3}*. Durch diese Umstellung wird die hierarchische Struktur deutlicher, und es wird ersichtlich, dass die Konjunktion *-seo* verwendet wird, um S1 und S3 in eine zeitliche Abfolge zu bringen, wobei die Eigenschaft des Subject-Sharing zum Tragen kommt. S1 und S3 haben ein gemeinsames Subjekt, da die Person, die zuerst hineingeht (S1) und dann auf die (linke) Seite geht (S3), der Adressat ist. Hier kann verdeutlicht werden, dass die syntaktischen Strukturen eines verbundenen Satzes mit mehr als drei Sätzen genauer analysiert werden müssen, um die Eigenschaften des Subject-Sharing von Konjunktionen zu erfassen.

Zusammenfassend lässt sich sagen, dass die hierarchischen Strukturen der Teilsätze in einem komplexen Satz zunächst analysiert werden müssen, damit die Eigenschaften des Subject-Sharing von Konjunktionen genutzt werden können, die Hong Munpyo (2011) zur Wiederherstellung der Referenz des koreanischen Nullsubjekts in Nebensätzen verwendet werden. Darüber hinaus ist es notwendig, den Verwendungskontext der Konjunktion im Nebensatz zu bestimmen, um festzustellen, ob Subject-Sharing im vorliegenden Fall überhaupt aus grammatischer Sicht möglich ist. Selbst wenn die Verwendung der Konjunktion feststeht, kann das Merkmal des Subject-Sharing nicht genutzt werden, wenn keine Subjekte in den verbundenen Sätzen realisiert sind.

In der vorliegenden Arbeit werden auch heuristische Regeln vorgeschlagen, die morphosyntaktische Hinweise in Verbalkomplexen von Nebensätzen nutzen. Da es nicht notwendig ist, die Struktur von Nebensätzen zu analysieren und die Verwendung einer Konjunktion zu bestimmen, sind die heuristischen Regeln relativ einfach zu implementieren. Die Effektivität der vorgeschlagenen Regeln zur Resolution von Nullsubjekten in Nebensätzen wird in Kapitel 5 anhand der Leistung des BS basierend auf der Resolutionsstrategie von Hong Munpyo bewertet.

Die Regeln, die auf Nullsubjekte in Nebensätzen abzielen, verwenden linguistische Mittel mit sprecherorientierter Modalität in Kombination mit einigen Konjunktionen, um den entsprechenden Prädikaten die Sprecherrolle als Referenten zuzuweisen. Das folgende Beispiel aus dem Originalkorpus verdeutlicht, welche Kombinationen von Konjunktionen und linguistischen Mitteln zu dieser Interpretation der Nullsubjekte in Nebensätzen führen können.

- (47) a. [gwaja-leul sa-lyeogo ha-neunde]_{S1} [eoneu jjog-eulo ga-ya ha-jyo-?]_{S2}
 Kekse-OBJ kaufen-PPC(Absicht)-CNJ welche Richtung-in gehen-PPC(Pflicht)-SFS-SF
 '(Ich) möchte Kekse kaufen. In welche Richtung muss ich gehen?'
- b. [gicha-sigan-eul byeongyeongha-go sip-eunde]_{S1} [byeongyeongha-l su iss-nayo-?]_{S2}
 Bahn-Fahrzeit-OBJ ändern-PPC(Wille)-CNJ ändern-PPC(Möglichkeit)-SFS-SF
 '(Ich) möchte meine Reisezeit verschieben. Ist die Änderung möglich?'

In (47) wird S2 mit S1 durch die Konjunktion -(n)eunde verbunden, die im Allgemeinen zur Beschreibung des Hintergrunds oder der Situation benutzt wird (Son Jae Eun 2007; Lee Heeja & Lee Jonghee 2010; Choi Ju Hee 2013 u. a.).⁸⁶ In (47)a beschreibt S1 eine Situation, in der der Sprecher Kekse kaufen möchte, bevor er in S2 fragt, in welche Richtung er gehen soll, um Kekse zu kaufen. In (47)b erklärt der Sprecher in S1 seine Situation, dass er seine Reisezeit ändern möchte, bevor er in S2 fragt, ob diese Änderung möglich ist. Wie in S1 in (47) gezeigt, dient der Nebensatz mit -(n)eunde dazu, die Situation zu erklären, was die Aufmerksamkeit des Adressaten auf die Informationen des folgenden Satzes lenken soll (vgl. Song Dae-heon 2015: 173).

Es ist zu beachten, dass das Auftreten der Konjunktion -(n)eunde allein nicht dazu ausreicht, die Nullsubjekte von S1 in (47) als Sprecher zu interpretieren, aber die Kombination mit periphrastischen Konstruktionen wie *lyeogo hada* oder *go sipda* kann zu dieser Interpretation führen. Die periphrastische Konstruktion *lyeogo hada* in S1 in (47)a ist bereits zuvor als linguistisches Mittel beschrieben worden, das ebenso wie das Verbalsuffix *llaeyo* den Willen des Sprechers in Deklarativsätzen wiedergeben kann (vgl. (31)a & (31)b in Kapitel 4.2.2). Die in S1 in (47)b gebrauchte periphrastische Konstruktion *go sipda* kann ebenfalls benutzt werden, um den Willen oder Wunsch des Sprechers wiederzugeben. Anders als die Verbalsuffixe, die nur in Hauptsätzen gebraucht werden können, können periphrastische Konstruktionen auch in Nebensätzen benutzt werden. Abgesehen von der Verwendung in Interrogativsätzen ist die durch diese periphrastischen Konstruktionen ausgedrückte Bedeutung grundsätzlich an den Sprecher gerichtet. Ihre Kombination mit der Konjunktion -(n)eunde kann vor allem dann benutzt werden, wenn der Sprecher seine Situation erklären will, bevor er die nächsten Sätze äußert. Deshalb wird die folgende Regel aufgestellt, die auf der Tendenz basiert, dass Nullsubjekte von Prädikaten, die diese Kombination von linguistischen Mitteln enthalten, natürlicherweise als Sprecher interpretiert werden.

```
elif re_morphmatch(morphs, morph="^(☞려고?하|으?려고?하|고자하|☞까하|[아어]?드리|고싶) $",
pos="^([E|X]$") and
  re_morphmatch(morphs,morph="^([ㄴ 는]데|는데|는건데|은데|[ㄴ 은는]건데) $", pos="^E$") and
  datapoint["verb_syn"] == "clau":
  datapoint["subject_by_rule"] = "1sh+"
  datapoint["group"] = "clau"
  datapoint["rule"] = "clau_neunde"
  datapoint["rule_order"] = "26"
```

Abbildung 31: Vereinfachter Codeblock für die Regel zur Konjunktion -(n)eunde

Zur Vereinfachung der Erklärung zeigt Abbildung 31 nur einige der in der Regel verwendeten linguistischen Mittel, deren Bedeutung grundsätzlich auf den Sprecher ausgerichtet ist, darunter periphrastische Konstruktionen wie *lyeogo hada* oder *go sipda*, die den Willen des Sprechers zum Ausdruck bringen. Darüber hinaus wird in der Regel das Hilfsverb *eo deulida* eingesetzt, das zuvor als eines der linguistischen Mittel beschrieben wurde, die zur Wiederherstellung des Sprechers als Referenten führen können, da es zum Ausdruck der Bescheidenheit des Sprechers in Deklarativ- und Interrogativsätzen sowie in Nebensätzen verwendet werden kann (siehe Beispiel (40) in Kapitel 4.2.3.3). Wenn ein Prädikat mit einem dieser linguistischen Mittel in einem Nebensatz vorkommt, der

⁸⁶ Für eine Literaturübersicht über die semantischen und syntaktischen Eigenschaften der Konjunktion -(n)eunde siehe Son Kyeong-ae (2015: 103–106).

die Konjunktion -(n)eunde enthält, weist die Regel dem Prädikat den Referenten "1sh+", d. h. die Sprecherrolle zu. Um festzustellen, ob ein Prädikat in einem Nebensatz vorkommt, wird der Wert "clau" (als Abkürzung für "clause") verwendet, den der Parser einem Prädikat in einem Nebensatz in den Zusatzdaten zuweist, während ein Prädikat in einem Hauptsatz den Wert "root" erhält. Da -(n)eunde im gesprochenen Koreanisch tendenziell häufig verwendet wird (vgl. Kang Hyoun Hwa et al. 2016: 233), ist anzunehmen, dass Regeln mit der Konjunktion -(n)eunde eine hohe Identifikationsleistung für die Nullsubjekte in Nebensätzen aufweisen.

Periphrastische Konstruktionen wie *go sipda* in Nebensätzen garantieren nicht, dass die entsprechenden Nullsubjekte als Sprecher interpretiert werden. Beim Beispiel (48)a handelt es sich um eine Variante von (47)b, bei der die Konjunktion -eumyeon in S1 anstelle der Konjunktion -eunde gebraucht wird. Im Allgemeinen wird die Konjunktion -(eu)myeon verwendet, um zwei Sätze in einer kitionalen Beziehung zu verbinden (Lee Eunkyung 1996; Koo Hyun Jung 2001; Kang Hyoun Hwa et al. 2016).

- (48) a. *[gicha-sigan-eul byeongyeongha-go sip-eumyeon]_{S1} [byeongyeongha-l su iss-eyo-.]_{S2}*
 Bahn-Fahrzeit-OBJ ändern-PPC(Wille)-CNJ ändern-PPC(Möglichkeit)-SFS-SF
 'Wenn (Sie/man/ich) die Reisezeit ändern möchte(n), können/kann (Sie/man/ich) (sie) ändern.'
- b. *[gicha-sigan-eul byeongyeongha-go sip-eusi-eumyeon]_{S1} [byeongyeongha-l su iss-eyo-.]_{S2}*
 Bahn-Fahrzeit-OBJ ändern-PPC(Wille)-HON-CNJ ändern-PPC(Möglichkeit)-SFS-SF
 'Wenn (Sie/?man/*ich) die Reisezeit ändern möchte(n), können/kann (Sie/?man/*ich) (sie) ändern.'

Im Zusammenhang mit S1 in (48)a kann jede Person, die den Wunsch hat, die Reisezeit mit dem Zug zu ändern, als Subjekt interpretiert werden. Somit lässt sich feststellen, dass nicht alle Konjunktionen zu einer Zuordnung von Nullsubjekten zur Sprecherrolle führen, selbst wenn sie mit linguistischen Mitteln mit sprecherorientierter Modalität kombiniert werden. Wenn jedoch das Honorativinfix (*eul*)*si* an das Prädikat in S1 angehängt wird, wie in (48)b gezeigt, wird das Nullsubjekt von S1 mit großer Wahrscheinlichkeit als Adressat interpretiert. Die Person, die die Reisezeit ändern möchte, wird also aufgrund des Honorativinfixes bevorzugt als Adressat interpretiert. Diese Tendenz gilt für alle Nebensätze unabhängig von der Konjunktion. Aus diesem Grund wird die folgende Regel aufgestellt.

```
elif datapoint["zp_hono"] == "plus" and datapoint["verb_syn"] == "clau":  

    datapoint["subject_by_rule"] = "2sh+"  

    datapoint["group"] = "clau"  

    datapoint["rule"] = "clau_plus"  

    datapoint["rule_order"] = "23"
```

Abbildung 32: Regel für Nullsubjekte in Nebensätzen mit dem Honorativinfix -(eu)si

Die Regel in Abbildung 32 weist einem Prädikat den Referenten "2sh+", d. h. die Adressatenrolle zu, wenn es in den Zusatzdaten des Originalkorpus den Wert "plus" für das Vorkommen des Honorativinfixes (*eul*)*si* und den Wert "clau" für die Zugehörigkeit zu einem Nebensatz hat. Die Regel in Abbildung 32 wird, wie im Datenpunkt "rule_order" angegeben, vor der Regel in Abbildung 31 angewandt, da die Adressatenrolle unabhängig von der Konjunktion zugewiesen werden kann, wenn ein Prädikat in einem Nebensatz das Honorativinfix enthält.

Es gibt zudem Hinweise, die dabei helfen können, Nullsubjekte als Sprecher zu interpretieren, unabhängig davon, welche Konjunktion in einem Nebensatz benutzt wird. Im Beispiel (49) finden sich

Ausschnitte aus dem Originalkorpus, in denen die Konjunktion *-myoen* in S1 von (49)a und die Konjunktion *-go* in S1 von (49)b gebraucht wird.

- (49) a. [deulai ppalli **ha-eo deuli-myeon**_{S1}] [sib bun i-myeon doe-l geo-eyo-.]S₂
 Föhnen.OBJ schnell machen-AUX.HUM-CNJ zehn Minuten sein-PPC-PPC-SFS-SF
 'Wenn (ich) (Ihnen) schnell (Ihr Haar) föhne, kann das zehn Minuten dauern.'
- b. [geuleom yeyagha-si-n pyo-leul **chwisoha-eo deuli-go**_{S1}] [eeo paeseu-lo bakkwu-eo lkkayo-?]S₂
 dann reservieren- Ticket- stornieren- Airpass-in unwandeln-
 HON-ADN OBJ AUX.HUM-CNJ AUX.HUM-SFS.INT-SF
 'Soll (ich) dann (Ihr) reserviertes Ticket stornieren und in einen Airpass umwandeln?'

Auch wenn in S1 von (49)a und (49)b unterschiedliche Konjunktionen benutzt werden, werden ihre Nullsubjekte am ehesten als Sprecher interpretiert, da sie das Hilfsverb *eo deulida* enthalten. Dieses wurde zuvor als eines der linguistischen Mittel beschrieben, die zur Resolution von sprecherbezogenen Nullsubjekten führen können, da es in allen Satztypen zum Ausdruck der Bescheidenheit des Sprechers genutzt wird. Daher kann die folgende Regel aufgestellt werden, um den Nullsubjekten in Nebensätzen, die *eo deulida* enthalten, den Sprecher als Referenten zuzuordnen.

```
elif re_morphmatch(morphs, morph="^(아드리|어드리|드리|축하드리)$", pos="^(x|v)$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_various"
    datapoint["rule_order"] = "36"
```

Abbildung 33: Regel für Nullsubjekte in Nebensätzen mit dem Hilfsverb *-a/eo deulida*

Die Regel in Abbildung 33 weist einem Prädikat den Referenten in der Sprecherrolle zu, wenn es eine Variante des Hilfsverbs *a/eo deulida* enthält und den Wert "clau" hat, sprich in einem Nebensatz auftritt.

Es wurden insgesamt 47 Regeln vorgeschlagen, die speziell für die Resolution von Nullsubjekten in Nebensätzen aufgestellt wurden. Diese Regeln benutzen den Wert "clau" in den Zusatzdaten, um zu bestimmen, ob ein Prädikat zu einem Nebensatz gehört. Sie sind mit dem Wert "clau" für den Datenpunkt "group" in Appendix 1 zu finden. Da der Wert "clau" einem Prädikat zugewiesen wird, das eine Konjunktion enthält, ist der zugewiesene Wert auch morphologisch bedingt, ohne dass eine syntaktische Analyse des Nebensatzes erforderlich wäre. Darüber hinaus ist keine syntaktische Analyse oder Disambiguierung der Verwendung von Konjunktionen erforderlich, um die oben vorgestellten Regeln anzuwenden, da die in den Regeln beschriebenen linguistischen Mittel und Konjunktionen leicht in Verbalkomplexen in Nebensätzen zu finden sind. Außerdem nutzen einige Regeln Konjunktionen wie -(n)eunde, von denen erwartet wird, dass sie in der gesprochenen Sprache häufig verwendet werden, so dass die Identifikationsleistung der Regeln, die auf Nullsubjekte in Nebensätzen abzielen, relativ höher sein dürfte als die von Hong Munpyo (2011), der diese Konjunktion in seiner Resolutionsstrategie ausschließt, da sie kein Subject-Sharing ermöglicht (vgl. Son Kyeong-ae 2015). In Kapitel 5 wird die Eignung der Resolutionsstrategien zwischen dem NRS und dem BS zur Identifizierung und Resolution von Nullsubjekten in Nebensätzen im selben Analysekorpus verglichen.

4.3.2. Regeln zum Ausschluss der Subjekte in der 3. Person

Eine der Methoden zur Erhöhung der Aussagegenauigkeit der heuristischen Regeln ist der Ausschluss von Prädikaten, deren Subjekte mit geringer Wahrscheinlichkeit als Diskursteilnehmer:innen interpretiert werden, indem vier Regeln erstellt werden, die im NRS den Wert "none" zuweisen. Diese Regeln dienen dazu, Prädikate zu finden, deren Nullsubjekte eher in der 3. Person interpretiert werden.

Der erste Hinweis, der in den entsprechenden vier Regeln herangezogen wird, sind objektive Adjektive. Dabei handelt es sich um Adjektive, die kontinuierliche oder temporäre Eigenschaften eines Gegenstands beschreiben. Laut einer Liste objektiver Adjektive bei Yoo Hyun-kyung (2000: 255) gehören Adjektive wie *balgda* 'hell sein', *jjalbda* 'kurz sein', *ssada* 'billig sein' und *ttateushada* 'warm sein' zu den objektiven Adjektiven. Da es sehr wahrscheinlich ist, dass das Subjekt solcher Prädikate ein Gegenstand mit der semantischen Eigenschaft [–belebt] ist, wird eine Regel aufgestellt, die einige objektive Adjektive einschließt, denen der Wert "none" zugewiesen wird.

Die optimale Methode für die Verwendung objektiver Adjektive bei der Erstellung von Regeln ist die Einbeziehung aller verfügbaren Listen, was jedoch zu praktischen Problemen führt. Zunächst ist die Zahl der objektiven Adjektive sehr groß. Yoo Hyun-kyung (2000) nennt insgesamt 2.926 koreanische Adjektive, klassifiziert sie nach ihrem Typ und untersucht ihre Verteilung. Nach ihrer Zählung machen die subjektiven Adjektive 13,5 % aller Adjektive aus, während die restlichen 86,5 % objektive Adjektive sind. Unter den objektiven Adjektiven beträgt der Anteil der Eigenschaftsadjektive („state-property adjectives“) etwa 77,5 % (2.267 Adjektive), die für die entsprechenden heuristischen Regeln der vorliegenden Arbeit benutzt werden sollen. Aufgrund der langen Liste von Eigenschaftsadjektiven wird bei Yoo Hyun-kyung (2000: 255–256) nur eine Teilliste mit 218 Adjektiven dieser Art vorgestellt. Da die in der Literatur keine vollständige Liste objektiver Adjektive zu finden ist, können sie nicht in die entsprechende Regel aufgenommen werden. Daher wurden in der vorliegenden Arbeit zunächst alle mit "J" (Adjektiv) gekennzeichneten Prädikate aus den Zusatzdaten des Originalkorpus extrahiert. Anschließend wurden 196 Adjektive manuell ausgewählt, die zur Beschreibung von Eigenschaften oder Zuständen eines Gegenstands gebraucht werden können. Die folgende Abbildung zeigt die modifizierte Regel, die aus Platzgründen so überarbeitet wurde, dass sie nur einige der in der Regel benutzten objektiven Adjektive enthält. Die gesamte Liste der für diese Regel verwendeten Adjektive ist in Appendix 1 zu sehen.

```
elif re_morphmatch(morphs, morph="^(싱싱하|저렴하|뿌옇|매콤하|가볍|가깝|₩₩+되|₩₩+어지|₩₩+한)$",
    pos="^(J|V|N)$"):
    datapoint["subject_by_rule"] = "none"
    datapoint["group"] = "exclude"
    datapoint["rule"] = "exclude predicate"
    datapoint["rule_order"] = "85"
```

Abbildung 34: Vereinfachter Codeblock für die Regel bei Eigenschaftsadjektiven

Die Regel in Abbildung 34 enthält eine Liste der Stämme von Eigenschaftsadjektive wie z. B. *gakkabda* 'nah sein', *jeolyeomhada* 'günstig sein' und *maekomhada* 'scharf sein'. Wenn ein Prädikat einen der gelisteten Adjektivstämme beinhaltet, weist die Regel dem Prädikat den Wert "none" zu. Die Regel in Abbildung 34 nutzt nicht nur die Stämme der objektiven Adjektive, sondern auch einige linguistische Mittel, durch die im Koreanischen transitive Verben ins Passiv gesetzt werden können. Im Koreanischen können Passivausdrücke durch Hinzufügung des Verbal infixes *hi*, der periphrastischen Konstruktion *eo jida* oder des Funktionsverbs *doe da* gebildet werden (vgl. Mok Jungsoo & Kim Yeong-jung 2006: 375). Die folgenden Beispiele zeigen, wie Passivausdrücke für das transitive Verb *ilgda* 'lesen' im Koreanischen gebildet werden.

- (50) a. *X-ga chaeg-eul ilg-neunda-*.
 X-SBJ Buch-OBJ lesen-SFS.DECL-SF
 'X liest ein Buch.'
- b. *chaeg-i jal ilg-hi-nda-*.
 Buch-SBJ gut lesen-IFX.PASS-SFS.DECL-SF
 'Das Buch liest sich gut.'
- c. *chaeg-i jal ilg-eo ji-nda-*.
 Buch-SBJ gut lesen-PPC.PASS-SFS.DECL-SF
 'Das Buch liest sich gut.'

(50)a zeigt die Verwendung von *ilgda* als transitives Verb. Dieses Prädikat erfordert zwei Argumente, ein Subjekt und ein Objekt, wobei das Subjekt muss eine belebte Entität sein muss, die über die Fähigkeit verfügt, etwas zu lesen. In (50)b und (50)c wurde das Infix *hi* bzw. die periphrastische Konstruktion *eo jida* an den Stamm des Prädikats *ilgda* angehängt, so dass die Prädikate ins Passiv überführt werden. In diesem Fall handelt es sich um einstellige Prädikate, deren Subjekt auf einen lesbaren Gegenstand referiert, z. B. ein Buch. Wenn Verben im Passiv als einstellige Prädikate verwendet werden, wird für ihre Subjekte gewöhnlich eine NP [-menschlich] benutzt, wobei sie nicht die semantische Rolle *Experiencer* erhalten können (vgl. Yoo Hyun-kyung 2017: 84). Daher werden Bestandteile, die zu Passivausdrücken gehören, in der Regel in Abbildung 34 aufgenommen, weil ihr (Null-)Subjekt weniger wahrscheinlich als Diskursteilnehmer:innen interpretiert werden kann.⁸⁷

Eine weitere Regel für die Zuweisung des Werts "none" umfasst Hilfsverben oder periphrastische Konstruktionen, deren Subjekte weniger wahrscheinlich die semantische Eigenschaft [+menschlich] haben. Die folgenden Beispiele legen offen, welche linguistischen Mittel im Koreanischen solche Prädikate bilden.

- (51) a. *yeonghui-ui baji-ga jjalb-a iss-da-*.
 Younghhee-POSS Hose-SBJ kurz sein-AUX-SFS.DECL-SF
 'Younghhees Hose ist kurz geworden.'
- b. *nalssi-ga galsulog deo-eo ji-nda-*.
 Wetter-SBJ immer mehr heiß sein-PPC-SFS.DECL-SF
 'Das Wetter wird immer heißer.'
- c. *ip-i nolah-ge doe-eo iss-da-*.
 Blätter-SBJ gelb sein-PPC-AUX-SFS.DECL-SF
 'Die Blätter sind gelb geworden.'
- d. *eoje bo-n yeonghwa-neun bo-I manha-ess-eoyo-?*
 gestern sehen-ADN Film.SBJ-TOP sehen-PPC-PST-SFS-SF
 'War der Film, den Sie gestern gesehen haben, sehenswert?'

Das Beispiel (51)a (zitiert aus Bae Soo-Ja 2007: 244) gibt zum einen die Dauer des Zustands an, nach dem Younghhees Hose kürzer geworden ist und besagt, dass der Zustand immer noch gilt, indem das Hilfsverb *a issda* zum Adjektiv *jjalbda* 'kurz sein' hinzugefügt wird. Bisherige Studien haben das

⁸⁷ In einigen Kontexten können objektive Adjektive oder Verben im Passiv als zweistellige Prädikate benutzt werden, deren Subjekte das semantische Merkmal [+menschlich] und die Theta-Rolle *Experiencer* haben. Weitere Informationen dazu finden sich in Kapitel 5 von Yoo Hyun-kyung (2017). Aufgrund dieser Eigenschaft solcher Prädikate können einige falsche Vorhersagen durch die Regel in Abbildung 34 verursacht werden.

Hilfsverb *a issda* als ein linguistisches Mittel beschrieben, das die Dauer eines Zustands beschreibt (Seo Cheong-Soo 1996; Son Se-mo-dol 1996; Park Sun Ok 2003; Bae Soo-Ja 2007). Im Beispiel (51)b (zitiert aus Liu 2017: 37) wird die Veränderung des Wetters angemerkt (es wird heißer), indem die periphrastische Konstruktion *eo jida* an das Adjektiv *deobda* 'heiß sein' angefügt wird. In früheren Studien wurde erklärt, dass die Phrase *eo jida* Zustandsveränderungen markiert, wenn sie mit Adjektiven kombiniert wird (Son Se-mo-dol 1996; Ko Yong-Kun 2004; Jun Youngchul 2008). Im Beispiel (51)c (zitiert aus Liu 2017: 131) wird das Adjektiv *nolahda* 'gelb sein' mit zwei linguistischen Mitteln kombiniert: der periphrastischen Konstruktion *ge doeda* und dem Hilfsverb *eo issda*, das auch in (51)a verwendet wird. Es wird angenommen, dass die Phrase *ge doeda* eine Zustandsveränderung ausdrückt, wenn sie mit einem Adjektiv kombiniert wird (Jun Youngchul 2008; Kim Jung-Nam 2009; Kim Sunyoung 2015), ähnlich wie bei Verwendung der periphrastischen Konstruktion *eo jida* in (51)b.⁸⁸ Das Beispiel (51)c illustriert eine Zustandsveränderung, bei der das Subjekt (die Blätter) gelb geworden sind, und drückt durch Hinzufügen des Hilfsverbs *eo issda* aus, dass dieser veränderte Zustand fortgesetzt wird. Da die Prädikate in (51)a, (51)b und (51)c objektive Adjektive enthalten, kann ihnen durch die Regel in Abbildung 34 der Wert "none" zugewiesen werden, falls ihre Subjekte nicht realisiert sind. Es wird jedoch eine Regel erstellt, die auf die zuvor eingeführten linguistischen Mittel zurückgreift, nämlich das Hilfsverb *a issda* und die periphrastischen Konstruktionen *eo jida* und *ge doeda*. Dies wird als notwendig angesehen, da die Wahrscheinlichkeit als sehr gering eingestuft wird, dass ein Subjekt des Prädikats mit diesen linguistischen Merkmalen als menschliches Subjekt interpretiert wird. Zudem soll die Regel in Abbildung 34 ausgeweitet werden, um weitere Fälle abzudecken. Im Beispiel (51)d (zitiert aus Park Hami 2014: 91) wird die periphrastische Konstruktion *I manhada* an das Verb *boda* 'sehen' angehängt, um zu fragen, ob der Film, den der Adressat gesehen hat, sehenswert war. Die Phrase *I manhada* ist ein linguistisches Mittel, das anzeigen kann, dass die durch das vorangehende Verb ausgedrückte Handlung lohnenswert ist (Park Hami 2014; Kang Hyoun Hwa et al. 2016). Wenn diese Phrase in einem Prädikat enthalten ist, ist es sehr wahrscheinlich, dass das, worauf das Subjekt verweist, etwas Lohnenswertes ist, und es ist sehr unwahrscheinlich, dass das (Null-)Subjekt des Prädikats als menschliches Subjekt interpretiert wird. Aus diesem Grund wird diese periphrastische Konstruktion in die entsprechende Regel aufgenommen. Die folgende Abbildung zeigt, wie die Regel unter Verwendung der oben genannten Hilfsverben oder periphrastischen Konstruktionen aufgebaut ist.

```
elif re_morphmatch(morphs, morph="^(어있|아있|아지|어지|지|제만하게되어있)$", pos="^X$") :
    datapoint["subject_by_rule"] = "none"
    datapoint["group"] = "exclude"
    datapoint["rule"] = "exclude_predicate"
    datapoint["rule_order"] = "86"
```

Abbildung 35: Regel bei Hilfsverben und periphrastischen Konstruktionen für die Zuordnung des Wertes "none"

Die Regel in Abbildung 35 weist dem Prädikat den Wert "none" zu, wenn es eine Variante des Hilfsverbs *a/eo issda* oder der periphrastischen Konstruktion *a/eo jida* bzw. *I manhada* oder *ge deoeo issda* enthält. Die Regel basiert auf dem Wert des POS-Taggings dieser linguistischen Mittel, die in den Zusatzdaten des Originalkorpus als X markiert sind. Betrachtet man die Regelreihenfolge der Regeln in Abbildung 34 und Abbildung 35, so wird die Regel in Abbildung 35 unmittelbar nach der Regel in Abbildung 34 angewandt, um die von der Regel in Abbildung 34 nicht erfassten Fälle abzudecken.

Die bisher eingeführten Regeln für die Zuweisung des Werts "none" basieren darauf, dass die Subjekte der Prädikate, die die Bedingungen der Regeln erfüllen, mit größerer Wahrscheinlichkeit nichtmenschlich sind. Die folgende Regel bezieht sich auf den Fall, dass die Subjekte der Prädikate ein

⁸⁸ Siehe Kapitel 4.2.3 von Liu (2017) für semantische Unterschiede zwischen den beiden linguistischen Mitteln.

menschliches Subjekt sein können, aber die Wahrscheinlichkeit, dass sie in der 1. oder 2. Person interpretiert werden, gering ist. Die Regel fußt auf periphrastischen Konstruktionen wie *dago hada* oder *lago hada*, die im Koreanischen indirekte Rede wiedergeben können. Indirekte Rede fungiert als mittelbare Wiedergabe von Aussagen und Gedanken, während die direkte Rede der unmittelbaren Wiedergabe einer Äußerung dient (vgl. Ulrich 1975: 117). Die folgenden Beispiele (zitiert aus Lee Sangbok 1974: 147) zeigen, wie die indirekte Rede im Koreanischen aufgebaut ist. Die Beispiele (52)a und (52)b umfassen Äußerungen des Sprechers A und die Äußerungen in den Beispielen (52)a' und (52)b' stammen von Sprecher B, der das von A Gehörte an C weitergibt.

(52) a. Sprecher A

„*naeil-eun jib-eseo swi-gess-seubnida-*.“
 Morgen-DP Hause-zu sich ausruhen-IFX-SFS.DECL-SF
 'Morgen werde ich mich zu Hause ausruhen.'

a'. Sprecher B → Adressat C

*A-ga [naeil-eun jib-eseo swi-gess-**da]go ha**-ess-eoyo-.*
 A-SBJ morgen-DP Hause-zu sich ausruhen-IFX-PPC-PST-SFS-SF
 'A hat mir gesagt, dass er sich morgen zu Hause ausruhen werde.'

b. Sprecher A

„*igeos-i cheolsu gabang-i-bnida-*.“
 das-SBJ Cheolsu Tasche-sein-SFS.DECL-SF
 'Das ist die Tasche von Cheolsu.'

b'. Sprecher B → Adressat C

*A-ga na-ege [igeos-i cheolsu gabang-i-**la]go ha**-ess-eoyo-.*
 A-SBJ ich-IO das-SBJ Choelsu Tasche-sein-PPC-PST-SFS-SF
 'A hat mir gesagt, dass das die Tasche von Cheolsu ist.'

In den Beispielen (52)a' und (52)b' ist anhand der periphrastischen Konstruktionen *dago hada* und *lago hada* zu erkennen, dass es sich bei diesen Äußerungen um eine indirekte Redewiedergabe handelt. Die Umwandlung von direkter Rede in indirekte Rede im Koreanischen beinhaltet normalerweise die Verwendung von berichtenden periphrastischen Konstruktionen wie *da/lago hada* 'sagen, dass', die die Phrase *-go (mal)hada* beinhalten, die als typische Zitierphrase im Koreanischen gilt (Jang Mira 2008; Koh Kyoung-tae 2014). Zum Beispiel wird das Verbalsuffix *(seu)bnida* in der ursprünglichen Äußerungen (52)a und (52)b durch die Kombination der periphrastischen Konstruktionen *da/lago hada* in indirekte Rede umgewandelt, wie in (52)a' und (52)b', wenn B die Worte von A wiedergibt. Wenn zum Beispiel das Subjekt *A-ga* in (52)a' nicht overt ausgedrückt wird, kann dennoch anhand der Zitierphrase erkannt werden, dass die Person, die sich morgen zu Hause ausruhen wird, nicht der Sprecher B oder der Adressat C ist, sondern die Person A, von der die ursprüngliche Aussage stammt. Da Ausdrücke in indirekter Rede oft benutzt werden, um zu vermitteln, was eine dritte Person gesagt hat, ist es sehr wahrscheinlich, dass die Subjekte in diesen Ausdrücken in der 3. Person stehen. Aus diesem Grund wird die folgende Regel aufgestellt, um einem Prädikat den Wert "none" zuzuweisen, wenn es eine der Phrasen *dago hada* oder *lago hada* enthält, die die indirekte Rede kennzeichnen.

```

    elif re_morphmatch(morphs, morph="^(w+다고하|w+라고하|다고하)$", pos="^X$") :
        datapoint["subject_by_rule"] = "none"
        datapoint["group"] = "exclude"
        datapoint["rule"] = "exclude predicate"
        datapoint["rule_order"] = "89"

```

Abbildung 36: Regel für periphrastische Konstruktionen in indirekter Rede

Zusammenfassend lässt sich sagen, dass einige Nullsubjekt-Prädikate allerdings ausgeschlossen werden, wenn es unwahrscheinlich ist, dass ihre Nullsubjekte Diskursteilnehmer:innen als Referenten haben. Es wird erwartet, dass die Leistung des NRS durch den Ausschluss solcher Prädikate erhöht wird, da es auf diskursteilnehmerbezogene Nullsubjekte zugeschnitten ist. In Kapitel 5 wird bewertet, ob diese vier Regeln tatsächlich zu einer Verbesserung der Gesamtleistung des NRS führen. Die linguistischen Mittel und der Wert der syntaktischen Funktionen, die in den in diesem Kapitel vorgestellten Regeln eingesetzt wurden, beruhen auf morphosyntaktischen Hinweisen, die in Verbalkomplexen leicht zu finden sind. Daher haben diese Regeln den Vorteil, dass sie einfach zu implementieren sind, wie auch die zuvor vorgestellten Regeln zur Zuweisung des Sprechers oder des Adressaten.

4.3.3. Fazit: Reihenfolge der Regelanwendung

Insgesamt wurden in dieser Studie 212 Regeln für das NRS aufgestellt, um primär diskursteilnehmerbezogene Nullsubjekte zu rekonstruieren. Die heuristischen Regeln können vier verschiedene Werte zuweisen:

1. "1ph-" (wir),
2. "1sh+" (ich),
3. "2sh+" (Sie),
4. "none".

Die ersten drei Werte beziehen sich auf Diskursteilnehmer:innen, während der vierte Wert auf Prädikate verweist, bei denen die Nullsubjekte mit geringerer Wahrscheinlichkeit als Diskursteilnehmer:innen interpretiert werden. Die folgende Tabelle präsentiert die Anzahl der Regeln nach den vier Wertzuweisungen aus einer Gesamtanzahl von 212 heuristischen Regeln.

Nr.	Zugewiesener Wert	Anteil (Anzahl der Regeln)	
1	"1ph-"	0,47 % (1)	98,11 % (208)
2	"1sh+"	44,60 % (95)	
3	"2sh+"	52,58 % (112)	
4	"none"	1,89 % (4)	
Total		100,00 %	(212)

Tabelle 25: Verteilung der 212 Regeln nach den vier Wertzuweisungen

Von den insgesamt 212 Regeln weisen 208 Regeln diskursteilnehmerbezogene Referenten zu, was 98,11 % entspricht. Die verbleibenden vier Regeln können den Wert "none" zuweisen. Die Anzahl der Regeln legt nahe, dass in koreanischen Verbalkomplexen 208 Kombinationen linguistischer Mittel existieren können, um Nullsubjekte zu identifizieren, die entweder den Sprecher oder den Adressaten als Referenten haben. In Kapitel 5 wird eine umfassende Analyse der Anzahl der verwendeten und

ungenutzten Regeln im NRS durchgeführt, um die Anzahl der Parameter abzuschätzen, die für die Bestimmbarkeit von diskursteilnehmerbezogenen Nullsubjekten im Koreanischen relevant sind.

In der vorliegenden Arbeit hat sich die Anzahl der Regeln durch die Einführung von bis dato nicht auf den Themenbereich angewandten linguistischen Mitteln wie periphrastischen Konstruktionen und die Berücksichtigung verschiedener Kombinationen von linguistischen Mitteln erhöht. Bei dieser großen Anzahl von Regeln sind Konflikte zwischen den Regeln bei der Integration in das NRS wahrscheinlich, müssen aber vermieden werden, um ihre Referenten korrekt zuordnen zu können. Deshalb werden erstens die Regeln, die weit grammatisierte linguistische Mittel nutzen, früher im NRS angewandt als die Regeln, die linguistische Mittel mit einem niedrigeren Grammatikalisierungsgrad beinhalten. Zu den stärker grammatisierten Mitteln gehören Prädikate mit Imperativ- oder Exhortativmarkern in Form von Verbalsuffixen ebenso wie Verbalmuster, die üblicherweise in Fragesätzen für eine Aufforderung an den Adressaten benutzt werden. Tabelle 26 zeigt Beispiele für Regeln, die linguistische Mittel nutzen, die eindeutige Hinweise auf einen bestimmten Referenten geben können, und legt deren Rangordnung unter den insgesamt 212 Regeln im NRS fest.

Nr.	Reihenfolge der Regeln	Top	Kerninformation		Referent
1	1	0,47 %	Imperativmarker		"2sh+"
2	2	0,94 %	Exhortativmarker		"1ph-"
3	55	25,94 %	Muster für die indirekte Aufforderung	Interrogativsatz	"2sh+"
4	81	38,21 %	Verbalsuffix <i>eullaeyo</i>	Deklarativsatz	"1sh+"
5	82	38,68 %	Verbalsuffix <i>eullaeyo</i>	Interrogativsatz	"2sh+"

Tabelle 26: Beispiele für früh im NRS angewandte Regeln

Von den in Tabelle 26 dargestellten Regeln basieren die ersten beiden Regeln auf Verbalsuffixen zur Markierung von Imperativ und Exhortativ. Da diese linguistischen Mittel eindeutige Anhaltspunkte für die Bestimmung des Subjekts der 2. Person Singular oder der 1. Person Plural sind, werden sie im NRS global an erster und zweiter Stelle eingesetzt. Die dritte Regel in Tabelle 26 betrifft die Muster und die Satzmodusinformation zu Interrogativsätzen und gibt ebenfalls eindeutige Anhaltspunkte für die Wiederherstellung der Adressatenrolle als Referent, da die Kombination zeigt, dass sie für indirekte Aufforderungen gebraucht werden. Die letzten beiden Regeln in Tabelle 26 setzen das Verbalsuffix *eullaeyo* voraus, das den Sprecherwillen ausdrückt. Diese Regeln liefern klare Hinweise für die Bestimmung der Referenten anhand der Satzmodusinformation, so dass sie in den Top 38 % der Regeln eingestuft sind.

Bei linguistischen Mitteln mit einem hohen Grammatikalisierungsgrad sind die Kombinationsmöglichkeiten mit anderen linguistischen Mitteln oft eingeschränkt (siehe Beispiel (38)a in Kapitel 4.2.3.2). So ist es zum Beispiel unmöglich, das Honorativinfix *si* oder das Infix *gess* mit dem Verbalsuffix *eullaeyo* in Deklarativsätzen zu kombinieren (vgl. Kang Hyoun Hwa et al. 2016: 374). Dies hat den Vorteil, dass keine Änderungen in der Subjektinterpretation berücksichtigt werden müssen. Da für die Regel mit dem Verbalsuffix *eullaeyo* in Deklarativsätzen keine Regel hinzugefügt werden muss, ist die Reihenfolge der Regelanwendung an dieser Stelle unerheblich. Das Honorativinfix *si* kann optional zum Verbalsuffix *eullaeyo* in Interrogativsätzen hinzugefügt werden. Die Regel, die das Verbalsuffix *eullaeyo* mit der Satzmodusinformation zum Interrogativsatz nutzt, inkludiert jedoch nicht das Honorativinfix *si*, weil es nicht als Shifter für die Wiederherstellung der Adressaten als Referent fungiert, sondern nur optional an Prädikate angehängt werden kann. Es besteht daher keine Notwendigkeit, die Reihenfolge der Regelanwendung anzupassen, da in diesem Fall auch keine Regel hinzugefügt wird.

Im Gegensatz zu den Verbalsuffixen können periphrastische Konstruktionen oder psychologische Verben mit verschiedenen linguistischen Mitteln kombiniert werden und dementsprechend auch in Nebensätzen auftreten. Das bedeutet, dass die Anzahl verwandter Regeln größer ist als die der Verbalsuffixe. Hier muss die Reihenfolge der Regelanwendung sorgfältig abgestimmt werden, da sonst die durch die entsprechenden Regeln zugewiesenen Referenten möglicherweise falsch sind. In den Beispielen (53) werden die Gefühlsverben *buleobda* 'neidisch sein', *jeulgeobda* 'Spaß haben' und *asuibda* 'enttäuscht sein' in Deklarativ- und Interrogativsätze benutzt. Bisherige Studien konzentrieren sich auf die Subjektbeschränkung von psychologischen Verben, die besagt, dass in Deklarativsätzen nur der Sprecher und in Interrogativsätzen nur der Adressat als Subjekt interpretiert werden kann. Wie bereits gezeigt, kann die Interpretationsrichtung des Subjekts jedoch durch die Hinzufügung linguistischer Mittel wie dem Honorativ infix *si* oder dem Infix *gess* geändert werden (siehe Beispiel (38)b und (38)c in Kapitel 4.2.3.2). Anhand der folgenden Beispiele wird untersucht, was bei der Priorisierung der Anwendung der entsprechenden Regeln berücksichtigt werden muss.

- (53) a. *jeongmal buleob-gunyo-*.
so neidisch sein-SFS.DECL-SF
'(Ich) bin so neidisch.'
- b. *jeongmal buleo-eusi-gunyo-*.
so neidisch sein-HON-SFS.DECL-SF
'(Sie) sind vielleicht eifersüchtig.'
- c. *jeulgeo-(eus)i-eoss-nayo-?*
Spaß haben-(HON)-PST-SFS.INT-SF
'Hatten (Sie) Spaß?'
- d. *asuib-(eus)i-gess-gunyo-*.
enttäuschen sein-(HON)-IFX-SFS.DECL-SF
'(Sie) sind vermutlich enttäuscht.'

Mit der Äußerung (53)a gibt der Sprecher sein derzeitiges Neidgefühl an. Da der Stamm des Gefühlsverbs *bulebda* in einem Deklarativsatz ohne zusätzliche linguistische Mittel steht, ist es am wahrscheinlichsten, dass das Nullsubjekt in (53)a als Sprecher interpretiert wird. Wenn wie in (53)b das Honorativ infix -(eu)si an (53)a angehängt wird, wird das Subjekt eher als Adressat identifiziert, auch wenn das Gefühlsverb *buleobda* in einem Deklarativsatz gebraucht wurde. Dies liegt daran, dass das Honorativ infix -(eu)si einer der Ausdrücke ist, mit denen in Gesprächen Respekt gegenüber dem Adressaten angezeigt wird. Deshalb wird (53)b als Aussage interpretiert, in der der Sprecher das Neidgefühl des Adressaten vermutet.

Es gibt zwei Regeln, die den Prädikaten in (53)a und (53)b Referenten in der Sprecher- oder Adressatenrolle zuordnen können. Die erste Regel nutzt die Kombination aus Gefühlsverb und Deklarativinformation und die zweite Regel nutzt die Kombination aus Gefühlsverb und Honorativ infix. Wird die erste Regel vor der zweiten Regel angewandt, wird das Nullsubjekt von (53)b fälschlicherweise dem Sprecher zugeordnet, da das Gefühlsverb von (53)b in einem Deklarativsatz steht. Deshalb sollte die zweite Regel zuerst angewandt werden, um in (53)b korrekt den Adressaten als Referent zuzuordnen. Im Fall von (53)c fragt der Sprecher den Adressaten, ob er Spaß hatte. Da der Stamm des Gefühlsverbs *jeulgeobda* 'Spaß haben' in einem Interrogativsatz ohne zusätzliche linguistische Mittel verwendet wurde, ist es am naheliegendsten, dass das Nullsubjekt als Adressat interpretiert wird. Es ist auch möglich, dass das Honorativ infix an dieser Stelle vom Sprecher hinzugefügt wird, wenn er mehr Respekt gegenüber dem Adressaten ausdrücken möchte. Wenn das Honorativ infix an (53)c angehängt wird, kann die zweite Regel angewandt werden, um den Adressaten als Referent wie in (53)b zu identifizieren. Aber auch wenn das Honorativ infix nicht hinzugefügt wird,

kann der Referent immer noch durch die dritte Regel zugewiesen werden, die eine Kombination aus Gefühlsverb und Interrogativinformation nutzt. Da die zweite Regel, die das Honorativinfix benutzt, unabhängig von der Satzmodusinformation den Adressaten als Referent zuordnen kann, wird sie vor den beiden anderen Regeln angewandt, die auf der Satzmodusinformation aufbauen.

Es wirkt am natürlichsten, das Beispiel (53)d als eine Äußerung zu interpretieren, mit der der Sprecher eine Vermutung über die enttäuschten Gefühle des Adressaten angibt, da das Infix *gess* dem Gefühlsverb im Deklarativsatz hinzugefügt wurde. Um einem Prädikat wie in (53)d den Adressaten als Referent zuzuordnen, wird die vierte Regel zur Verwendung eines Gefühlsverbs mit dem Infix *gess* aufgestellt. Wenn die erste Regel auf Grundlage der Kombination aus Gefühlsverb und Deklarativinformation vor der vierten Regel angewandt wird, wird der Referent in (53)d falsch als Sprecher identifiziert, weil das Gefühlsprädikat in (53)d in einem Deklarativsatz auftritt. Daher sollte die vierte Regel, die (53)d den Adressaten als Referent zuweist, früher angewandt werden als die erste Regel, weil das Infix *gess* als Shifter fungieren kann, um ein Nullsubjekt eines Gefühlsverbs in einem Deklarativsatz wie beim Honorativinfix als Adressaten zu interpretieren. Wenn das Honorativinfix zu (53)d hinzugefügt wird, ordnet die zweite Regel mit dem Honorativinfix den Adressaten korrekt als Referent zu.

Die folgende Tabelle zeigt die Reihenfolge der Anwendung der vier Regeln mit Gefühlsverben, ihre Kerninformationen, ihre Referenten und wann sie im NRS zur Anwendung kommen.

Nr.	Reihenfolge der Regeln	Top	Kerninformation		Referent
1	112	52,83 %	Gefühlsverb	Honorativinfix (<i>eu</i>) <i>si</i>	"2sh+"
2	135	63,68 %		Interrogativsatz	"2sh+"
3	151	71,23 %		Infix <i>gess</i>	"2sh+"
4	152	71,70 %		Deklarativsatz	"1sh+"

Tabelle 27: Reihenfolge und Charakteristika von Regel-Anwendungen bei Gefühlsverben

Betrachtet man die Reihenfolge der vier Regeln, die den Prädikaten in (53) entweder die Sprecher- oder Adressatenrolle zuweisen, so wird die erste Regel mit dem Honorativinfix (*eu*)*si* zuerst angewandt, weil das Honorativinfix der eindeutigste Anhaltspunkt ist, um das Nullsubjekt eines Gefühlsverbs unabhängig vom Satzmodus als Adressaten zu interpretieren. Darüber hinaus ist die dritte Regel mit dem Infix *gess* früher anzuwenden als die vierte Regel mit der Deklarativinformation, weil das Infix hilft, das Nullsubjekt eines Gefühlsverbs als Adressaten zu interpretieren, auch wenn es in einem Deklarativsatz vorkommt. Ein weiterer Punkt in Tabelle 27 ist, dass diese Regeln nach der Hälfte der gesamten Regeln angewandt werden. Zum Beispiel sollen alle Regeln in Tabelle 26 früher angewandt werden als die in Tabelle 27. Auf diese Weise wird erwartet, dass die Leistung des NRS für die Aufgabe der Nullsubjektresolution bis zu einem gewissen Grad durch die Anwendung der Regeln, die definitive Anhaltspunkte für bestimmte Referenten verwenden, gewährleistet wird. Darüber hinaus werden die Regeln inklusive solcher Anhaltspunkte, die als Shifter fungieren können, früher als andere verwandte Regeln angewandt, um Konflikte zwischen den Regeln zu vermeiden, gerade wenn unterschiedliche Referenten zugewiesen werden. All diese Anpassungsversuche der Regelanwendung im NRS zielen darauf ab, die Genauigkeit der heuristischen Regeln bei der Zuweisung ihrer Zielreferenten zu erhöhen.

4.4. Output-Dataframe von Nullsubjektresolutionssystem

Der Output des NRS wird in Form eines Dataframes erstellt, der verschiedene Informationen über die annotierten Prädikate mit Nullsubjekten enthält, die dem Input des NRS entsprechen. Diese Informationen werden aus drei Quellen zusammengetragen:

1. Zusatzdaten,
2. Annotierte Daten,
3. Nullsubjektresolutionssystem.

Eine ausführliche Beschreibung der im Output-Dataframe des NRS enthaltenen Informationen wird in Tabelle 28 wiedergegeben.

Information	Spaltenname	Wert
1. aus Zusatzdaten		
1.1 Satz-ID	"etri_sent"	
1.2 Satz-String	"sent"	
1.3 Prädikat-String	"verb"	"plus"
1.4 Vorhandensein von Honorativinfix	"zp_hono"	"decl", "ques", "impe", "sugg", "excl"
1.5 Satzmodus	"zp_mood"	"root", "clau", "modi", "subj", "scom", "dobj"
1.6 Syntaktische Funktion	"verb_syn"	"pres", "futu", "past", "ipast", "past_futu", "past_ipast"
1.7 Tempus	"tense_zp"	"plus"
2. aus annotierten Daten		
2.1 Markable-ID	"zp_markable_ID"	
2.2 Nullsubjekttyp	"anaphora_type"	"extra", "intra", "inter", "other"
2.2.1 für Prädikat mit "extra" Referent	"zero_referent"	"1ph-", "1ph+", "1sh-", "1sh+", "2sh-", "2sh+", "2ph-", "2ph+", "3s", "3p", "other"(String)
2.2.2 für Prädikat mit "intra" oder "inter" Antezedenz	"antecedent"	verknüpfte NP mit dem Prädikat
3. aus NRS		
3.1 durch Regeln zugewiesener Referent	"subject_by_rule"	"1ph+", "1sh+", "2sh+", "none", "N/A"
3.2 angewandte Regel	"rule"	
3.3 Regelgruppe	"group"	
3.4 Regelreihenfolge	"rule_order"	

Tabelle 28: Die im Output-Dataframe enthaltenen Informationen über die Prädikate mit Nullsubjekt

Die Anzahl der aufgeführten Informationen in Tabelle 28 entspricht der Anzahl der Spalten im Dataframe. Demnach besteht dieses Dataframe aus 15.883 Zeilen und 15 Spalten, da es insgesamt 15 Informationen für 15.883 Prädikate bereithält.⁸⁹

Die Informationen aus den Zusatzdaten umfassen zunächst verschiedene morphosyntaktische Informationen zu den Input-Prädikaten (1.1–1.7). Die Informationen aus den annotierten Daten beinhalten die verschiedenen annotierten Werte für die Nullsubjekte der Input-Prädikate, einschließlich ihrer Nullsubjekttypen, Referenten oder Antezedenzen (2.1–2.2). Die Informationen aus dem NRS enthalten die Werte, die den Prädikaten durch die heuristischen Regeln des NRS zugewiesen wurden, sowie Informationen über die angewandte Regel (3.1–3.4). Sollte keine der heuristischen Regeln auf ein Prädikat anwendbar sein, wird der Wert "N/A" zugeordnet.

Die Informationen im generierten Dataframe, das als Ergebnis des NRS vorliegt, erweisen sich als nützlich für die Leistungsbewertung der heuristischen Regeln im NRS. Insbesondere die Informationen über die durch die heuristischen Regeln des NRS zugewiesenen Referenten (3.1) und die annotierten Referenten oder Antezedenzen aus den annotierten Daten (2.2.1 und 2.2.2) sind von besonderer Bedeutung. Die Leistung der Regeln kann effektiv durch den Vergleich dieser beiden Werte bewertet

⁸⁹ Falls zusätzliche Informationen zu den Prädikaten benötigt werden, kann sich die Anzahl der Spalten im entsprechenden Dataframe weiter erhöhen.

werden. Eine umfassende Methodik zur Evaluierung der Korrektheit der Regelvorhersagen und zur Leistung des entwickelten NRS zur Resolution des koreanischen Nullsubjekts wird in Kapitel 5 präsentiert.

5. Bewertung der Leistung des Nullsubjektresolutionssystems

In Kapitel 5 liegt das Hauptaugenmerk auf einer umfassenden Bewertung des entwickelten NRS. Das NRS basiert auf 212 heuristischen Regeln, die auf einer Kombination von linguistischen Mitteln in den Verbalkomplexen auf lexikalischer und morphosyntaktischer Ebene beruhen. 208 Regeln sind darauf ausgerichtet, Nullsubjekte im Zusammenhang mit Diskursteilnehmer:innen zu rekonstruieren, indem sie Prädikaten mit Nullsubjekten den Referenten in der Sprecher- oder Adressatenrolle zuweisen. Vier zusätzliche Regeln sollen Prädikate mit Nullsubjekten der 3. Person ausschließen. Die Hauptmotivation für die Entwicklung dieser vier Regeln liegt in der Annahme, dass sie zur Verbesserung der Genauigkeit des NRS bei der Resolution von diskurstteilnehmerbezogenen Nullsubjekten beitragen.

Die Leistung der 208 Regeln im NRS steht in Kapitel 5.1 im Vordergrund. Dieses Kapitel konzentriert sich darauf, die Hypothese dieser Arbeit zu überprüfen, dass die Rekonstruierbarkeit von diskurstteilnehmerbezogenen Nullsubjekten auf den expliziten linguistischen Mitteln in den Verbalkomplexen beruht. Die Bewertung der Resolutions- und Identifizierungsfähigkeit der Regeln wird zeigen, inwiefern die expliziten Hinweise für die Referenten in der Sprecher- oder Adressatenrolle bei der erfolgreichen Identifizierung und Resolution der Nullsubjekte mit Bezug auf die Diskurstteilnehmer:innen beitragen (Kapitel 5.1.1). Anschließend wird analysiert, welche Regel wie viel zur Resolutionsgenauigkeit beiträgt. Diese Analyse bietet wertvolle Einblicke in robuste Indikatoren für die korrekte Rekonstruktion von Nullsubjekten, die auf Sprecher- oder Adressatenreferenten zurückgehen. Es lässt sich vermuten, dass die linguistischen Mittel, die diesen ausgewählten Regeln zugrunde liegen, Schlüsselparameter für die Rekonstruierbarkeit von diskurstteilnehmerbezogenen Nullsubjekten im Koreanischen darstellen. Damit wird die in dieser Studie aufgestellte Hypothese überprüft, dass nicht nur einzelne Parameter, sondern auch deren Kombination die Interpretation von diskurstteilnehmerbezogenen Nullsubjekten im Koreanischen beeinflussen (Kapitel 5.1.2 und Kapitel 5.1.3). Schließlich erfolgt eine Analyse der Anzahl der genutzten und ungenutzten Regeln im NRS, um eine Schätzung der Anzahl der Parameter zu ermöglichen, die für die Identifizierung von Nullsubjekten mit Referenz auf Diskurstteilnehmer:innen im Koreanischen relevant sind (Kapitel 5.1.4).

In Kapitel 5.2 steht die Leistungsanalyse des in dieser Studie entwickelten NRS im Mittelpunkt. Das Hauptziel besteht darin zu bewerten, wie effektiv das NRS bei der Nullsubjektresolution arbeitet. Zunächst wird ein Vergleich zwischen der Vorhersagegenauigkeit der ursprünglichen 208 und der gesamten 212 Regeln durchgeführt, um festzustellen, ob die zusätzlichen vier Regeln tatsächlich zu einer Verbesserung der Gesamtleistung des NRS führen, wie angenommen. Des Weiteren wird analysiert, welche Regeln im NRS besonders hohe Resolutions- und Identifizierungsfähigkeiten aufweisen (Kapitel 5.2.1). Der Fokus liegt weiterhin auf dem Leistungsvergleich zwischen dem NRS und dem BS (Kapitel 5.2.2). Es wurde bereits betont, dass die Implementierung des BS notwendig ist, um die Leistung des NRS objektiv zu bewerten (siehe Kapitel 1.2). Zunächst werden die Hintergründe und die Leistungsanalyse des BS dargestellt (Kapitel 5.2.2.1), gefolgt von einem Leistungsvergleich zwischen dem NRS und dem BS (Kapitel 5.2.2.2). Der Leistungsvergleich der beiden Systeme zielt auf zwei Aspekte ab: Zunächst wird die Leistung der beiden Systeme für die drei Typen von Nullsubjekten untersucht. Hier zeigt sich ein signifikanter Unterschied in den Resolutionsstrategien, da das BS für jeden Typ spezifische Strategien verwendet, während das NRS sich auf diskurstteilnehmerbezogene Nullsubjekte unabhängig von ihrem Typ konzentriert. Anschließend wird die Leistung von NRS und BS bei der Resolution von Nullsubjekten in Nebensätzen verglichen, um den Ansatz des NRS zu bewerten, bei dem neue heuristische Regeln vorgeschlagen werden, um die Resolution von Nullsubjekten in Nebensätzen zu verbessern (siehe Kapitel 4.3.1). Diese Analyse zieht Rückschlüsse auf die Stärken und Schwächen des NRS.

5.1. Überprüfung der Hypothese dieser Arbeit

5.1.1. Analyse der Leistung der Regeln

Um die Vorhersagegenauigkeit einer Regel zu bewerten, werden die von der Regel zugewiesenen Referenten mit den annotierten Referenten aus dem vom NRS generierten Output-Dataframe verglichen. Dabei werden die Werte in der Spalte "subject_by_rule", die von den Regeln im NRS zugewiesen wurden, den Werten in den Spalten "zero_referents" und "antecedent" gegenübergestellt, die aus den annotierten Daten extrahiert wurden.

Das folgende Dataframe veranschaulicht die korrekten Vorhersagen durch die Regeln im NRS, bei denen dem Input-Prädikat der Wert "1sh+" zugewiesen wird, die dem Pronomen der 1. Person Singular in Honorativform entsprechen. Dieses Beispiel verdeutlicht, welche Fälle in dieser Studie als korrekte Vorhersagen der Regeln bewertet werden.

	etri_sent	sent	verb	subject_by_rule	zero_referents	antecedent
4449	15479	네, 객실로 가져다드립니다.	가져다드립니다	1sh+	1sh+	
2952	13055	픽업은 해드리지않습니다.	해드리지않습니다	1sh+	1ph+	
2662	11605	제가 알고있는 걸 알려드린것뿐인걸요 .	알려드린것뿐인걸요	1sh+		제가
10039	23410	저희가 초별구이를해서 가져다 해서 가져나드릴까하는데 어떤신지 ?	해서 가져다드릴까하는데	1sh+		저희가

Abbildung 37: Output-Dataframe-Schnitt: Schlüsselinformationen zur Regelbewertung

Abbildung 37 präsentiert vier Input-Prädikate und die zugehörigen Informationen aus dem Output-Dataframe. Es werden nur 7 von insgesamt 15 Spalten des Output-Dataframes gezeigt, die allgemeine Informationen über das Input-Prädikat und die wichtigsten Informationen zur Bewertung der Regelgenauigkeit enthalten. Die erste Spalte im Dataframe fungiert als eindeutige Indexnummer, die jedem Input-Prädikat zugewiesen wird. Die zweite Spalte "etri_sent" enthält die Satz-ID, während die dritte Spalte "sent" den Satz selbst zeigt, der das Input-Prädikat enthält. In der vierten Spalte "verb" steht das Input-Prädikat, das ein Nullsubjekt aufweist. Die Werte in den letzten drei Spalten, nämlich "subject_by_rule", "zero_referents" und "antecedent", sind entscheidend für die Bewertung der Korrektheit einer Regelvorhersage. In Abbildung 37 ist für alle vier Input-Prädikate der Wert "1sh+" in der "subject_by_rule"-Spalte angegeben, was dem zugewiesenen Referenten in der Sprecherrolle entspricht. Bei den ersten beiden Prädikaten gibt es in der "zero_referents"-Spalte einen Wert, der die Referenten ihrer Nullsubjekte zeigt. Diese stammen aus den annotierten Daten für die Annotationsoption "referent_type". Für das dritte und vierte Prädikat steht in der Spalte "antecedent" der Wert *je-ga* 'ich' bzw. *jeohi-ga* 'wir', was die annotierten Antezedenzen für die Prädikate darstellt.

Für das erste Prädikat wird der zugewiesene Referent durch eine Regel in der "subject_by_rule"-Spalte als "1sh+" identifiziert, während sein Referent in der "zero_referents"-Spalte ebenfalls als "1sh+" annotiert wird. Da sowohl der Wert in der Spalte "subject_by_rule" als auch bei "zero_referents" "1sh+" ist, wird diese Zuweisung als korrekte Vorhersage der Regel bewertet.

Im Fall des zweiten Prädikats wird deutlich, dass nur die grammatische Person, nicht aber der Numerus eine entscheidende Rolle bei der Bewertung der Korrektheit der Regelvorhersage spielt. Hier wird der zugewiesene Referent der 1. Person Singular in Honorativform ("1sh+") in der "subject_by_rule"-Spalte als korrekt bewertet, obwohl der annotierte Referent der 1. Person Plural in Honorativform ("1ph+") in der "zero_referents"-Spalte steht. Diese Bewertung erfolgt, da die grammatische Person übereinstimmt und beide Referenten die Sprecherrolle innehaben. Die Entscheidung, solche Fälle als korrekte Vorhersage einer Regel zu bewerten, basiert darauf, dass die linguistischen Mittel, die bei der

Formulierung der Regeln für die Zuweisung eines Referenten in der Diskursteilnehmerrolle verwendet wurden, nicht zwischen Singular und Plural unterscheiden, sondern vielmehr der Unterscheidung zwischen Sprecher- und Adressatenrolle eines Referenten dienen. Diese Bewertungsmethode wurde daher unter der Annahme gewählt, dass eine Vorhersage als korrekt klassifiziert werden sollte, wenn eine korrekte Zuweisung zwischen Sprecher oder Adressat, d. h. zwischen 1. oder 2. Person, besteht. Zusätzlich ist die Bewertung als korrekte Vorhersage in diesem Fall mit linguistischen und kulturellen Spezifika in der koreanischen Kommunikation zu erklären. So ist die Betonung der Gruppenzugehörigkeit und kollektiven Identität eines Sprechers v. a. bei Servicemitarbeitenden üblich im Gegensatz zur Hervorhebung der Individualität in einem solchen Kontext. Sprecher neigen möglicherweise dazu, mit dem Pronomen *jeohui* in der Pluralform auf sich selbst zu referieren und dabei die Gruppe zu unterstreichen, zu der sie gehören. Dies steht im Kontrast zur Verwendung des Pronomens *jeo* in der Singularform, das eher die individuelle Identität hervorheben würde (vgl. Kim Hye Ryoung 2016: 284). Aufgrund der Beschaffenheit des Forschungskorpus dieser Studie, das Gespräche in Dienstleistungssituationen während Reisen umfasst, ist also zu erwarten, dass die Pluralform *jeohui* 'wir' häufiger verwendet wird als die Singularform *jeo* 'ich', besonders dann, wenn das Servicepersonal auf sich selbst referiert. Daher wurde der Referent der 1. Person Singular, der dem zweiten Prädikat in Abbildung 37 von der angewandten Regel zugewiesen wurde, als korrekte Vorhersage der Regel betrachtet, weil der annotierte Referent mit der 1. Person Plural ebenso verwendet wird, um sich auf den Sprecher zu beziehen.⁹⁰ Für das dritte und vierte Prädikat in Abbildung 37 entsprechen die annotierten Antezedenzen für die Prädikate den Referenten der Sprecherrolle. Daher werden auch diese beiden Fälle aus denselben Gründen wie die ersten beiden Fälle als korrekte Vorhersagen der Regeln bewertet.

Im Folgenden wird die Genauigkeit der Regeln anhand der oben beschriebenen Methode zur Bewertung der Korrektheit der Regelvorhersage dargestellt. Von den 208 Regeln, die in dieser Studie zur Zuweisung der Referenten in der Sprecher- oder Adressatenrolle erstellt wurden, wurden 172 Regeln auf Input-Prädikate angewandt, um einen Referenten zuzuweisen.⁹¹ Im Folgenden wird daher die Genauigkeit der 172 Regeln dargestellt. Die Genauigkeit wird durch das Verhältnis der korrekt rekonstruierten Nullsubjekte zur Gesamtzahl der identifizierten Nullsubjekte berechnet. Die 172 Regeln identifizierten insgesamt 12.530 Nullsubjekte, von denen 11.950 Fälle als korrekte Vorhersagen dieser Regeln bewertet wurden. Dies führt zu einer Genauigkeit von rund 0,95, wie in der folgenden Formel dargestellt:

$$\text{Genauigkeit der Regeln} = \frac{11.950}{12.530} = 0,95$$

Eine Genauigkeit von 0,95 bedeutet, dass die Regeln in etwa 95 % der identifizierten Fälle die Nullsubjekte korrekt rekonstruiert haben. Dies deutet darauf hin, dass die linguistischen Mittel, auf denen die angewandten Regeln basieren, mit einer Trefferquote von 95 % erfolgreich dazu beigetragen haben, die diskursteilnehmerbezogenen Nullsubjekte zu rekonstruieren.

Neben der Genauigkeit der 172 Regeln ist auch die Anzahl der korrekt identifizierten diskursteilnehmerbezogenen Nullsubjekte durch diese Regeln im Verhältnis zur Gesamtzahl der diskursteilnehmerbezogenen Nullsubjekte im Analysekörper von Interesse, um die Effektivität der

⁹⁰ Die Fälle, die nur aufgrund der Übereinstimmung der grammatischen Person als korrekte Vorhersagen der Regeln eingestuft wurden, stellen weniger als 1 % aller korrekten Vorhersagen dar.

⁹¹ Die Anzahl der angewandten Regeln in diesem Abschnitt beträgt 172, wenn die zusätzlichen vier Regeln zum Ausschluss der Subjekte in der 3. Person nicht einbezogen werden. Bei Hinzuziehung der vier zusätzlichen Regeln wurden insgesamt 169 Regeln im NRS angewandt, um den Referenten in der Sprecher- oder Adressatenrolle zuzuweisen.

linguistischen Merkmale in den Verbalkomplexen für die Identifizierung und Resolution der diskursteilnehmerbezogenen Nullsubjekte genauer zu überprüfen. Im Analysekorpus sind insgesamt 14.049 Nullsubjekte in der Sprecher- oder Adressatenrolle enthalten, von denen 12.083 durch 172 Regeln identifiziert wurden. Das bedeutet, dass allein mithilfe der potenziellen Hinweise für die Referenten in der Sprecher- oder Adressatenrolle im Verbalkomplex etwa 86 % der diskursteilnehmerbezogenen Nullsubjekte erfolgreich identifiziert werden können. Dieses Ergebnis stimmt teilweise mit der Analyse der Eckdaten der annotierten Daten überein, die besagt, dass bis zu 90 % der Nullsubjekte mit den heuristischen Regeln im NRS identifiziert werden können, da sie als Extra-NS basierend auf linguistischen Merkmalen in den Verbalkomplexen annotiert wurden (vgl. Kapitel 3.3.1).

Von den insgesamt 12.083 identifizierten Nullsubjekten wurden 11.950 Nullsubjekte durch die Regeln korrekt rekonstruiert. Dies resultiert in einer Genauigkeit von 0,99, wie in der folgenden Formel verdeutlicht wird:

$$\text{Genauigkeit der Regeln in Bezug auf Diskurteilnehmer:innen} = \frac{11.950}{12.083} = 0,99$$

Die Genauigkeit von 0,99 zeigt, dass die linguistischen Mittel, die den angewandten Regeln zugrunde liegen, nahezu fehlerfrei zur korrekten Resolution der diskursteilnehmerbezogenen Nullsubjekte beitragen, sobald die Regeln ein Prädikat mit dem Nullsubjekt in der Sprecher- oder Adressatenposition identifizieren. Dieses Ergebnis bestätigt die Hypothese dieser Arbeit, dass im Koreanischen explizite linguistische Merkmale im Verbalkomplex eine entscheidende Rolle bei der Rekonstruktion von Nullsubjekten mit Referenz auf die Diskurteilnehmer:innen spielen. Allerdings können nur 86 % der diskursteilnehmerbezogenen Nullsubjekte im Analysekorpus durch diese expliziten Hinweise im Verbalkomplex erklärt werden. Dies lässt vermuten, dass bei der Interpretation der restlichen 14 % der Nullsubjekte weitere Parameter eine Rolle spielen könnten. Es ist also möglich, dass ihre Rekonstruierbarkeit nicht ausschließlich von den expliziten linguistischen Mitteln im Verbalkomplex abhängt. Daher sollten in zukünftigen Studien auch andere potenzielle Parameter für diese Rekonstruierbarkeit untersucht werden.

In den kommenden Kapiteln liegt der Fokus auf der Leistung der einzelnen Regeln, um verlässliche Indikatoren für Nullsubjekte mit Referenz auf Sprecher oder Adressaten zu ermitteln. Diese Indikatoren sollen wichtige Parameter für die Rekonstruierbarkeit von Nullsubjekten im Zusammenhang mit den Diskurteilnehmer:innen aufzeigen. Die Analysen in den Kapiteln 5.1.2, 5.1.3 und 5.1.4 basieren auf der Leistung der 208 Regeln unter Einbeziehung der vier zusätzlichen Regeln. Dadurch variieren die angewandten Regeln sowie die Anzahl der identifizierten und rekonstruierten Nullsubjekte leicht im Vergleich zu den Ergebnissen in Kapitel 5.1.1, wo die Analyse ausschließlich auf den 208 ursprünglichen Regeln basierte. Die Entscheidung, die Leistung der 208 Regeln unter Hinzunahme der vier zusätzlichen Regeln zu analysieren, wird als vorteilhaft erachtet, da die Einbeziehung der vier Regeln eine genauere Resolution der diskursteilnehmerbezogenen Nullsubjekte durch die 208 Regeln ermöglicht, wie in Kapitel 5.2.1 näher erläutert wird. Dies ist entscheidend für die Untersuchung zuverlässiger Indikatoren für die Resolution von sprecher- oder adressatenbezogenen Nullsubjekten.

5.1.2. Zuverlässige Indikatoren für die korrekte Rekonstruktion sprecherbezogener Nullsubjekte

Für die Zuweisung des Referenten der 1. Person Plural ist lediglich eine Regel zuständig, die sich auf Exhortativmarker wie zum Beispiel *bsida* oder *ja* stützt. Diese Regel trug maßgeblich zur korrekten Resolution von 43 Nullsubjekten bei, wobei zuvor 44 passende Prädikate identifiziert wurden, was einem Genauigkeitswert von 0,98 entspricht. Nur eine Vorhersage durch diese Regel wurde als

inkorrekt bewertet, da in diesem Fall das Nullsubjekt als Adressat interpretiert werden sollte statt als Sprecher. Das folgende Beispiel illustriert diese fehlerhafte Vorhersage der Regel.

- (54) *geunyang jeo-rang gachi ga-bsida-*.
 einfach ich-mit zusammen gehen-SFS.EXH-SF
 'Gehen (**Sie**) einfach mit mir zusammen.'

Die Regel wies aufgrund des Exhortativmarkers *bsida* dem Prädikat in (54) den Referenten der 1. Person Plural zu. Aufgrund des Satzbestandteils *jeo-rang* 'mit mir' wurde der Satz eher als Imperativsatz interpretiert, in dem der Sprecher den Adressaten auffordert, mit ihm zusammen zu gehen. Dies könnte auf das Phänomen zurückzuführen sein, dass Exhortativmarker im Koreanischen nicht nur für Anregungen, sondern auch für Aufforderungen verwendet werden können. Eine eingehende Korpusanalyse von Kim Kanghee (2019)⁹² legt offen, dass das Verbalsuffix *bsida* nicht ausschließlich für Vorschläge steht, sondern auch für andere Sprechakte wie Befehle und Bitten genutzt werden kann. Dies geschieht oft aus Gründen der Höflichkeit, um gesichtsbedrohende Akte zu vermeiden, die durch direkte Befehle mit Imperativmarkern ausgelöst werden können (vgl. Lee Ji Soo 2016: 79).

Dass nur ein Nullsubjekt in einem Prädikat mit Exhortativmarker als Adressat interpretiert wurde und 98 % der identifizierten Nullsubjekte dem Referenten der 1. Person Plural korrekt zugeordnet wurden, untermauert trotz des kleinen vorliegenden Datensamples die Beobachtung, dass ein einzelner Parameter wie der Exhortativmarker als zuverlässige Kennzeichnung für Nullsubjekte mit dem Referenten der 1. Person Plural dienen kann.

Für die Zuweisung eines Referenten in der 1. Person Singular wurden 81 von 95 Regeln im NRS angewandt, wobei sie einen Genauigkeitswert von 0,96 aufweisen. Diese Regeln haben 5.179 Nullsubjekte unter den identifizierten 5.382 korrekt rekonstruiert. Abbildung 38 illustriert die kumulierte Anzahl der korrekten und inkorrekten Vorhersagen durch die 81 angewandten Regeln. Die Tendenz der Grafik ermöglicht die Bewertung des individuellen Beitrags einer Regel bei der schrittweisen Anwendung zur Identifizierung und korrekten Resolution eines sprecherbezogenen Nullsubjekts. Wenn eine Regel eine hohe Identifizierungs- und Resolutionsfähigkeit aufweist, lassen sich die zugrunde liegenden linguistischen Mittel dieser Regel als Schlüsselparameter für die Rekonstruierbarkeit von sprecherbezogenen Nullsubjekten interpretieren.

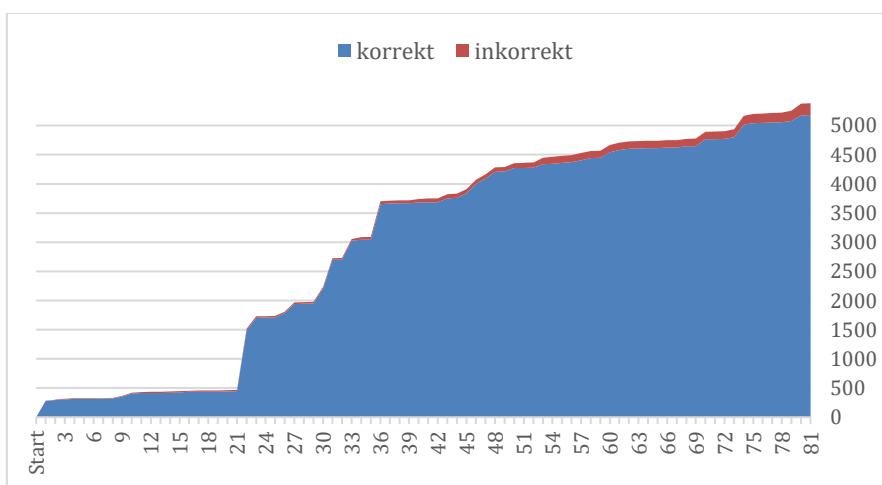


Abbildung 38: Vorhersageentwicklung: Referent der 1. Person Singular mit 81 Regeln

⁹² Für Informationen zum Analysekörper von Kim Kanghee (2019) siehe Kapitel 4.2.3.1.

In Abbildung 38 repräsentiert die x-Achse die Anzahl der angewandten Regeln, während die y-Achse die Anzahl der Prädikate zeigt, die von den Regeln identifiziert wurden und einen Referenten erhalten haben. Der blaue Bereich in der Grafik zeigt den kumulativen Wert der korrekten Vorhersagen der Regeln, während der rote Bereich den kumulativen Wert der inkorrekteten Vorhersagen der Regeln darstellt, wenn die Regeln nacheinander angewandt werden. Auffällig ist, dass die ersten 21 Regeln vergleichsweise wenige Nullsubjekte identifizieren, während es sechs deutliche Anstiege der identifizierten Fälle gibt, die mit der Anwendung der Regeln 22 bis 36 zusammenhängen. Konkret haben die Regeln 22, 23, 27, 31, 33 und 36 zu diesen Anstiegen beigetragen. Diese sechs Regeln werden bei der Vorstellung der individuellen Ergebnisse genauer erläutert (siehe Tabelle 30). Ab der Anwendung von Regel 37 nimmt der Anteil des roten Bereichs deutlich zu, was bedeutet, dass ein Anstieg der inkorrekteten Vorhersagen zu beobachten ist. Es lassen sich anhand der Grafik entsprechend drei distinkte Abschnitte mit eindeutigen Tendenzen einteilen, die im Folgenden einzeln betrachtet werden sollen.

Abschnitt	Anzahl korrekter Vorhersagen	Anzahl identifizierter Fälle	Genauigkeit
1. Regel 1–21 (−25,93 %)	445	467	0,95
2. Regel 22–36 (27,16–44,44 %)	3.206	3.235	0,99
3. Regel 37–81 (45,68–100,00 %)	1.528	1.680	0,91
Total	5.179	5.382	0,96

Tabelle 29: Leistung der Resolution von sprecherbezogenen Nullsubjekten in den Regelabschnitten (prozentualer Anteil an Gesamtwerten in Klammern)

Die ersten 21 Regeln in Abschnitt 1 sind auf die Resolution von Nullsubjekten in Nebensätzen ausgerichtet. Mit den 21 Regeln wurden 467 Nullsubjekte identifiziert, was etwa 9 % der insgesamt 5.382 mit den 81 Regeln identifizierten Nullsubjekte entspricht. Dies erklärt die schwache Tendenz im linken Teil der Grafik in Abbildung 38. Trotz der vergleichsweise geringen Identifizierungsfähigkeit dieser Regeln ist der rote Bereich klein, da sie einen Genauigkeitswert von 0,95 aufweisen. Diese Regeln beziehen sich auf eine in dieser Studie neu eingeführte Resolutionsstrategie für die Nullsubjekte in Nebensätzen, für die einige linguistische Mittel zum Ausdruck einer sprecherorientierten Modalität in Kombination mit bestimmten Konjunktionen herangezogen wurden, während in früheren Studien wie bei Hong Munpyo (2011) auf das Konzept des Subject-Sharing von Konjunktionen zurückgegriffen wurde (siehe Kapitel 4.3.1). Um die linguistischen Mittel zu verdeutlichen, auf die die Regeln in Abschnitt 1 zurückgreifen, wird ein Beispiel von Input-Prädikaten gezeigt, bei denen die Vorhersage der Regel 1 in Abschnitt 1 als korrekt bewertet wurde.

- (55) [i os-eul sa-go sip-eunde-,]s₁ [saijeu-ga iss-nayo-?]s₂
 dies Kleidungsstück-OBJ kaufen-PPC(Wille)-CNJ-NF Größe-SBJ haben-SFS.INT-SF
 'Ich möchte dieses Kleidungsstück kaufen. Welche Größen haben (Sie)?'

Regel 1 zielt darauf ab, Prädikate zu identifizieren, die bestimmte periphrastische Konstruktionen wie *go sipda* enthalten, die eine sprecherorientierte Modalität ausdrücken, zusammen mit der Konjunktion -(n)eunde. Diese Kombination von linguistischen Mitteln wird bei der Erstellung dieser Regel herangezogen, weil angenommen wird, dass sie vor allem verwendet wird, wenn der Sprecher seine Situation erklären möchte, bevor er die nächsten Sätze äußert. Die Regel identifiziert 285 Nullsubjekte, von denen 279 Vorhersagen als korrekt bewertet werden, was zu einem Genauigkeitswert von 0,98 führt. Zusätzlich identifiziert diese Regel allein fast 61 % aller Nullsubjekte, die von den Regeln in Abschnitt 1 identifiziert wurden. Diese Ergebnisse zeigen, dass diese Regel eine

relativ hohe Identifizierungs- und Resolutionsleistung aufweist und dass die Kombination dieser linguistischen Mittel besonders effektiv bei der korrekten Identifizierung von sprecherbezogenen Nullsubjekten in Nebensätzen ist. Darüber hinaus unterstreicht der Genauigkeitswert von 0,95 für die 21 Regeln in Abschnitt 1, die ebenfalls auf die Kombination linguistischer Mittel für die sprecherorientierte Modalität und bestimmter Konjunktionen in Nebensätzen zurückgreifen, die entscheidende Bedeutung der Parameterkombination für die korrekte Identifizierung von sprecherbezogenen Nullsubjekten in Nebensätzen.

Lediglich 15 Regeln aus Abschnitt 2 können etwa 60 % aller Nullsubjekte identifizieren, da sie 3.235 der insgesamt 5.382 identifizierten Nullsubjekte abdecken. Dies unterstreicht die hohe Identifizierungsfähigkeit der Regeln in Abschnitt 2. Fünf dieser Regeln verwenden Verbalsuffixe als Hauptmerkmal, die dazu dienen, eine sprecherorientierte Modalität auszudrücken. Die übrigen zehn Regeln basieren auf periphrastischen Konstruktionen, die den Willen des Sprechers ähnlich wie vollständig grammatikalisierte Verbalsuffixe ausdrücken. Um die Verbalsuffixe zu veranschaulichen, auf denen die fünf Regeln basieren, werden Beispiele von Input-Prädikaten präsentiert, bei denen die Regelvorhersage als korrekt bewertet wird.

- (56) a. *eolleun ga bo-Igeyo-*.
gleich gehen-SFS(Wille).DECL-SF
'(Ich) werde gleich gehen.'
- b. *geuleom damnyo-leul junbiha-e deuli-Ikkayo-?*
dann Decke-OBJ bereitlegen-AUX-SFS.INT-SF
'Dann soll (ich) eine Decke bereitlegen?'

Vier der fünf Regeln identifizieren ein Prädikat in einem Deklarativsatz mit einem Verbalsuffix wie z. B. *Igeyo*, das den Willen des Sprechers ausdrückt. Eine weitere Regel weist einem Prädikat in einem Interrogativsatz mit einem Verbalsuffix wie z. B. *Ikkayo* einen Referenten in der Sprecherrolle zu. Dieses Verbalsuffix wird verwendet, um die Meinung des Adressaten über die Handlungen des Sprechers zu erfragen. In der Äußerung (56)a und (56)b ist die Person, die gleich gehen wird oder die ankündigt, eine Decke bereitzulegen, der Sprecher. Daher wird die von der Regel zugewiesene sprecherbezogene Referenz als korrekt erachtet.

Ein vollständig grammatikalisiertes Verbalsuffix wird als aussagekräftigeres Signal für die Wiederherstellung eines Referenten in der Sprecherrolle betrachtet im Vergleich zu periphrastischen Konstruktionen, deren Grammatikalisierung noch nicht abgeschlossen ist. Im Leistungsvergleich der Regeln in Abschnitt 2 nach ihren beiden linguistischen Mitteln (Verbalsuffixe vs. Periphrasen) weisen die auf Verbalsuffixen basierenden fünf Regeln einen Genauigkeitswert von 0,996 auf, da sie 2.149 von 2.157 identifizierten Nullsubjekte korrekt rekonstruierten konnten. Der Genauigkeitswert von 0,996 dieser fünf Regeln bestätigt, dass die Verbalsuffixe, die verwendet werden, um den Willen des Sprechers in Aussagen auszudrücken oder um nach der Handlung des Sprechers zu fragen, signifikante Hinweise auf sprecherbezogene Nullsubjekte liefern. Die hohe Identifizierungsfähigkeit dieser Regeln wird durch die Analysedaten gestützt, da etwa 40 % der mit den 81 Regeln identifizierten Nullsubjekte allein durch diese fünf Regeln abgedeckt werden können. Die Formulierung aller fünf Regeln basiert auf einer Kombination von Verbalsuffixen mit Satzmodusinformationen, und diese Regeln zeigen eine bemerkenswerte Identifizierungs- und Resolutionsleistung. Dies legt nahe, dass diese Kombination der Parameter entscheidend für die Identifizierung und korrekte Resolution von Nullsubjekten mit dem Referenten in der Sprecherrolle ist.

Die zehn Regeln in Abschnitt 2, die auf periphrastischen Konstruktionen basieren, weisen ebenfalls eine gute Leistung auf. Ihr Genauigkeitswert beträgt 0,98, da sie 1.057 von 1.078 identifizierten Nullsubjekten korrekt rekonstruiert haben. Das folgende Beispiel von Input-Prädikaten veranschaulicht die linguistischen Mittel, auf die die zehn Regeln zurückgreifen.

- (57) a. *yag-eul sa-lyeogo ha-neundeyo-*.
 Medikamente-OBJ kaufen-PPC(Wille)-SFS.DECL-SF
 '(Ich) möchte gerne Medikamente kaufen.'
- b. *bodeu-leul billi-go sip-seubnida-*.
 Snowboard-OBJ mieten-PPC(Wille)-SFS.DECL-SF
 '(Ich) würde gerne ein Snowboard mieten.'
- c. *yeogiseo sajin-eul jjig-eodo doe-nayo-?*
 hier Foto-OBJ machen-PPC(Erlaubnis)-SFS.INT-SF
 'Darf (ich) hier ein Foto machen?'

Sechs der zehn Regeln stützen sich auf periphrastische Konstruktionen wie z. B. *lyeog hada* oder *go sipda*, die den Willen des Sprechers in einem Deklarativsatz ausdrücken können. Die restlichen vier Regeln weisen einem Prädikat einen Referenten in der Sprecherrolle zu, wenn es eine periphrastische Konstruktion wie *eodo doeda* in einem Interrogativsatz verwendet. Diese Kombination wird verwendet, um den Adressaten um Erlaubnis für die Handlung des Sprechers zu bitten. Die Zuweisung der Sprecherrolle durch die in den Beispielsätzen in (57) angewandten Regeln ist korrekt, da der Sprecher in Beispiel (57)a mit *lyeog hada* und in Beispiel (57)b mit *go sipda* seinen Willen zum Ausdruck bringt, Medikamente zu kaufen oder ein Snowboard zu mieten. Und der Sprecher in Beispiel (57)c fragt mit *eodo doeda*, ob er „hier“ ein Foto machen darf.

Der Genauigkeitswert von 0,98 dieser zehn Regeln deutet darauf hin, dass periphrastische Konstruktionen, die verwendet werden, um den Sprecherwillen in Aussagen auszudrücken oder um nach Erlaubnis für Handlungen des Sprechers zu fragen, als ein neuer zuverlässiger Indikator für sprecherbezogene Nullsubjekte betrachtet werden können. Diese Regeln basieren auf der Kombination von periphrastischen Konstruktionen und Satzmodusinformationen, wodurch die entscheidende Rolle der Parameterkombination für die Rekonstruierbarkeit von Nullsubjekten mit Referenz auf den Sprecher deutlich wird.

Die sechs Regeln, die den stärksten Anstieg der Anzahl richtiger Vorhersagen bewirken, wurden ebenfalls im Abschnitt 2 gefunden und sind in Abbildung 38 als deutliche Sprünge erkennbar. Tabelle 30 präsentiert diese Regeln in der Reihenfolge ihres Beitrags zur Steigerung der korrekten Vorhersagen:

Rang		Regel	Genaugigkeit
1	Regel 22	want verbal suffix ("rule_order": 64 ⁹³)	0,996 ⁹⁴
2	Regel 36	suggestory verbal suffix ("rule_order": 92)	0,993
3	Regel 31	want verbal suffix ("rule_order": 80)	1,000
4	Regel 33	want periphrastic construction ("rule_order": 87)	1,000
5	Regel 23	want periphrastic construction ("rule_order": 66)	1,000
6	Regel 27	permission periphrastic construction ("rule_order": 71)	1,000

Tabelle 30: Genaugigkeit der sechs wichtigsten Regeln für die korrekte Resolution sprecherbezogener Nullsubjekte

Eine Hälfte der einflussreichsten sechs Regeln basiert auf Verbalsuffixen, während die andere Hälfte auf periphrastischen Konstruktionen fußt. Lediglich sechs der insgesamt 81 Regeln decken etwa 55 % der korrekten Vorhersagen (2.837 von 5.179) ab und können etwa 53 % der identifizierten Nullsubjekte abdecken (2.845 von 5.832). Die ersten beiden dieser sechs Regeln zeigen nahezu fehlerfreie Resolutionsraten von 0,996 und 0,993, während die Vorhersagen der restlichen vier Regeln sogar gänzlich frei von Fehlern sind. Die Analyse verdeutlicht, dass diese sechs Regeln maßgeblich zur Steigerung der Identifizierung von sprecherbezogenen Nullsubjekten und zu ihrer korrekten Resolution beitragen. Dies unterstreicht die Erkenntnis, dass nicht nur Verbalsuffixe, sondern auch periphrastische Konstruktionen als zuverlässige Indikatoren für die Wiederherstellung sprecherbezogener Referenten betrachtet werden können. Es fällt dabei auf, dass linguistische Mittel, die den Willen des Sprechers ausdrücken, eine bedeutende Rolle bei der Identifizierung und Resolution sprecherbezogener Nullsubjekte spielen, da vier der sechs wichtigsten Regeln darauf Bezug nehmen.

Die 45 Regeln in Abschnitt 3 führten zu 152 inkorrektten Vorhersagen, was rund 75 % der insgesamt 203 falschen Vorhersagen ausmacht. Dies ist in Abbildung 38 deutlich zu erkennen, wo ein Anstieg des Anteils im roten Bereich festzustellen ist. Darüber hinaus weisen die Regeln in Abschnitt 3 im Vergleich zu den 36 Regeln in den Abschnitten 1 und 2 eine relativ geringe Identifizierungsfähigkeit auf. Obwohl Abschnitt 3 mehr Regeln enthält als die jeweiligen Abschnitte 1 und 2, wurden nur halb so viele Nullsubjekte identifiziert (1.680 vs. 3.702). Dennoch zeigen selbst die geringe Identifizierungsfähigkeit der Regeln in Abschnitt 3 einen Genauigkeitswert von 0,91, was bedeutet, dass etwa 91 % der identifizierten Nullsubjekte korrekt rekonstruiert wurden. Die 45 Regeln in Abschnitt 3 basieren auf periphrastischen Konstruktionen, psychologischen Verben, Hilfsverben, Infixen, performativen Verben und einigen Konstruktionen zur Bezeichnung der Verlaufs- und der Passivform. Um die linguistischen Mittel zu veranschaulichen, auf die sich die Regeln in Abschnitt 3 stützen, werden Beispiele von Input-Prädikaten vorgestellt, bei denen die Vorhersage der Regel als korrekt bewertet wurde.

⁹³ Die Angaben zu "rule_order" werden hier dargestellt, da die Nummer in "rule_order" hilft, die entsprechende Regel in Appendix 1 zu finden. In diesem Appendix sind alle Regeln im NRS aufgeführt, was es erleichtert, die Formulierung einer bestimmten Regel nachzuschlagen, falls erforderlich.

⁹⁴ In diesem Abschnitt wird der Genauigkeitswert auf die dritte Dezimalstelle gerundet im Gegensatz zu anderen Abschnitten dieser Arbeit, in denen lediglich auf die zweite Dezimalstelle gerundet wird. Diese Entscheidung wurde getroffen, um subtile Unterschiede in den Genauigkeitswerten besser herausarbeiten zu können, insbesondere dann, wenn die zweite Dezimalstelle identisch ist. Dieser Ansatz wird auch in den folgenden Abschnitten für entsprechende Fälle beibehalten.

- (58) a. *[olaenmane deungsanha-neun geo-la]_{S1} [himdeu-neyo-.]_{S2}*
 nach langer Zeit wandern-NMZ-CNJ müde sein-SFS-SF
 'Es ist lange her, dass (ich) wandern war, deshalb bin (**ich**) müde.'
- b. *hogsi inteones jamkkan sseu-I su iss-nayo-?*
 vielleicht Internet kurz nutzen-PPC(Möglichkeit)-SFS.INT-SF
 'Kann (**ich**) das Internet kurz nutzen?'
- c. *hyangsu-leul chaj-go iss-seubnida-*
 Parfüm-OBJ suchen-AUX-SFS.DECL-SF
 '(**Ich**) suche ein Parfüm.'

Eine Regel in Abschnitt 3 zielt darauf ab, psychologische Verben in Deklarativsätze wie z. B. *himdeulda* 'müde sein' zu identifizieren und weist solche Prädikaten einen Referenten in der Sprecherposition zu. Eine weitere Regel identifiziert in Interrogativsätzen Prädikate, die eine Variante der periphrastischen Konstruktion wie z. B. *I su issda* enthalten, die eine modale Bedeutung im Sinne von Möglichkeit ausdrückt. Obwohl für diese periphrastische Konstruktion keine bekannten Subjektbeschränkungen vorliegen, wurde in dieser Studie die Annahme getroffen, dass Prädikate mit solchen periphrastischen Konstruktionen in Interrogativsätzen im Allgemeinen verwendet werden, um die mögliche Handlung des Sprechers zu erfragen. Eine andere Regel weist einem Prädikat einen Referenten in der Sprecherrolle zu, wenn es das Hilfsverb wie z. B. *go issda* enthält, das eine Verlaufsform bezeichnet und sich in einem Deklarativsatz befindet. Diese Regel basiert auf der Annahme, dass Handlungen, die von einem Sprecher in einem Gespräch mit Fremden beschrieben werden, höchstwahrscheinlich vom Sprecher selbst ausgeführt werden. In Äußerung (58)a wird die Person, die ihre Müdigkeit aufgrund des Wanderns ausdrückt, als Sprecher identifiziert. In Äußerung (58)b verwendet der Sprecher einen Fragesatz und die periphrastische Konstruktion *I su issda*, um zu fragen, ob er das Internet nutzen darf. In der Äußerung (58)c wird die Person, die nach Parfüm sucht, als Sprecher identifiziert. Daher wird der von der Regel zugewiesene Referent in allen Äußerungen in (58) als korrekte Vorhersage bewertet. Mit Ausnahme der psychologischen Verben wurden in der vorliegenden Studie erstmals alle anderen linguistischen Mittel als Hinweise auf den Referenten in der Sprecherrolle berücksichtigt. 13 Regeln nutzen psychologische Verben als Hauptmerkmal, die restlichen 32 stützen sich auch weitere linguistische Mittel. Mithilfe der Anwendung der 32 Regeln auf Grundlage verschiedener linguistischer Mittel wurden insgesamt 1.238 Nullsubjekte identifiziert. Davon wurden 1.140 als korrekte Vorhersagen der Regeln bewertet, was einer Genauigkeit von 0,92 entspricht. Bei Anwendung der 13 Regeln zu psychologischen Verben wurde eine Genauigkeit von 0,88 erreicht, da 388 von 442 identifizierten Nullsubjekten korrekt rekonstruiert wurden. Dies deutet darauf hin, dass die neu eingeführten linguistischen Mittel präzisere Vorhersagen ermöglichen im Vergleich zu den oft erwähnten psychologischen Verben, die zur Resolution der sprecherbezogenen Nullsubjekte beitragen. Von den 45 Regeln in Abschnitt 3 basieren 43 auf einer Kombination der oben genannten linguistischen Mittel mit Satzmodusinformationen. Dies unterstreicht erneut die entscheidende Bedeutung der Parameterkombination für die korrekte Identifizierung von sprecherbezogenen Nullsubjekten.

Im Folgenden werden die Regeln analysiert, die zu einem Anstieg falscher Vorhersagen bei der Resolution von sprecherbezogenen Nullsubjekten geführt haben. Das Ziel dieser Untersuchung ist es, festzustellen, ob die linguistischen Mittel, die diesen Regeln zugrunde liegen, tatsächlich unzuverlässige Indikatoren für sprecherbezogene Nullsubjekte darstellen und gegebenenfalls die Gründe hierfür zu ergründen.

Bei der Zuordnung der Sprecherrolle verursachen 38 von 81 Regeln 203 inkorrekte Vorhersagen. Fünf dieser Regeln tragen erheblich zu diesem Anstieg inkorrechter Vorhersagen bei; in Tabelle 31 sind sie in aufsteigender Reihenfolge des größten Anteils an falschen Vorhersagen zusammengefasst.

Rang	Regel	Genauigkeit
1	Regel 53 feeling predicate ("rule_order": 147)	0,69
2	Regel 80 progress aspect ("rule_order": 209)	0,78
3	want Regel 34 periphrastic construction ("rule_order": 88)	0,47
4	possibility Regel 74 periphrastic construction ("rule_order": 190)	0,93
5	want Regel 40 periphrastic construction ("rule_order": 118)	0,43

Tabelle 31: Genauigkeit der fünf Regeln mit dem größten Einfluss auf den Anstieg fehlerhafter Prognosen für sprecherbezogene Nullsubjekte

Die fünf Regeln verursachen insgesamt 101 falsche Vorhersagen, was fast die Hälfte aller fehlerhaften Vorhersagen ausmacht (101 von 203). Diese Regeln basieren auf der Verwendung von Gefühlsverben, der Konstruktion der Verlaufsform und periphrastischen Konstruktionen, kombiniert mit Informationen über den Satzmodus. Mit Ausnahme der viertplatzierten Regel weisen die Genauigkeitswerte niedrige Werte zwischen 0,43 und 0,78 auf. Im Vergleich dazu erreichen die sechs Regeln mit dem größten Einfluss in Tabelle 30 Werte von fast 1,00. Die niedrigen Genauigkeitswerte legen nahe, dass die linguistischen Mittel, die den Regeln zugrunde liegen, möglicherweise keine zuverlässigen Indikatoren für die Rekonstruktion von sprecherbezogenen Nullsubjekten sind.

Die Ursachen für die große Anzahl inkorrekt er Vorhersagen durch die vier Regeln (Regel 53, 80, 34 und 40) mit relativ geringem Genauigkeitswert sollen anhand eines Beispiels aufgezeigt werden. Im Anschluss wird analysiert, ob die linguistischen Mittel, auf denen diese Regeln basieren, tatsächlich als unzuverlässige Marker für ein sprecherbezogenes Nullsubjekt angesehen werden müssen.

- (59) [pyeonha-go]_{S1} [joh-seubnida-.]_{S2}
 bequem-und gut-SFS.DECL-SF
 '(Der Rock) ist bequem und (er ist) gut.'

Regel 53 weist dem Prädikat von S2 in (59) einen Referenten in der Sprecherrolle zu, weil es das Gefühlsverb *johda* 'mögen' enthält und in einem Deklarativsatz steht. Das Verb in dem Beispiel wurde jedoch nicht als Gefühlsverb verwendet, um einen psychologischen Zustand des Sprechers zu beschreiben, sondern als Eigenschaftsadjektiv, um zu beschreiben, dass der Rock, den der Sprecher anprobiert hat, gut ist. Dies verdeutlicht einen der Gründe, warum die auf Gefühlsverben basierte Regel zu dem größten Anstieg falscher Vorhersagen führt, nämlich, dass viele Gefühlsverben auch andere Verwendungskontexte haben. Im Koreanischen gehört das Prädikat *johda* zu verschiedenen Prädikattypen. Wenn *johda* als einstelliges Eigenschaftsadjektiv verwendet wird, benötigt es ein Argument in der Subjektposition, dessen Eigenschaften vom Prädikat beschrieben werden. Aus diesem Grund bezieht sich das Nullsubjekt in (59) auf 'Rock (er)', dessen Eigenschaften vom Prädikat *johda* als 'gut' beschrieben werden. Von den 26 falschen Vorhersagen durch Regel 53 enthalten 13 Fälle das Prädikat *johda* und 11 Fälle das Prädikat *gwaenchanhda*, das als Gefühlsverb mit der Bedeutung 'sich gut fühlen' oder als Eigenschaftsadjektiv mit der Bedeutung 'gut sein' verwendet werden kann. Da diese beiden Prädikate als Eigenschaftsadjektive verwendet wurden, die die im jeweiligen Kontext erwähnte Entität beschreiben, und nicht als Gefühlsverben, wurden die von der Regel zugewiesenen Referenten in der Sprecherrolle als falsche Vorhersagen bewertet. Dies deutet darauf hin, dass die Kombination von Gefühlsverben und Satzmodusinformationen nicht als unzuverlässiger Indikator angesehen werden kann, da die meisten falschen Vorhersagen von Regel 53 darauf zurückzuführen

sind, dass die identifizierten Prädikate kein Gefühlsverb, sondern ein Eigenschaftsadjektiv enthalten. Es gibt vier weitere Regeln, die auf dieser Parameterkombination basieren und einen Genauigkeitswert von 0,91 erreichen, indem sie 70 der 77 identifizierten Prädikate korrekt einem Referenten in der Sprecherrolle zuordnen. Dieses Ergebnis bestätigt, dass die Kombination von Gefühlsverb und Satzmodusinformationen als guter Hinweis für die Identifikation von sprecherbezogenen Nullsubjekten interpretiert werden kann.

Die restlichen zwei falschen Vorhersagen der Regel 53 beinhalteten die Prädikate *geobughada* 'sich schlecht fühlen' und *heungmilobda* 'interessant finden'. Die folgenden Beispiele zeigen falsche Vorhersagen mit beiden Prädikaten. Diese Beispiele verdeutlichen weitere Gründe, warum Regeln, die auf Gefühlsverben basieren, zu falschen Vorhersagen führen können.

- (60) a. *[hyang-i neomu gangha-eseo]_{S1} [geobughada-gunyo-]_{S2}*
 Geruch-SBJ zu stark sein-CNJ sich schlecht fühlen-SFS.DECL-SF
 'Der (Knoblauch-)Geruch ist so stark, dass mir schlecht wird.'
- b. ***heungmilob-gunyo-***.
 interessant ist-SFS.DECL-SF
 '(Die Geschichte) ist (für mich) interessant.'

In (60)a ist das Nullsubjekt von S2 mit der NP *sog-i* 'Magen-SBJ' annotiert. Diese Annotation erfolgt aufgrund der häufigen Verwendung der NP *sog-i* mit einer VP wie *geobughada*, *ullonggeolida*, *meseukkeobda* oder *an johda* mit der Bedeutung 'sich schlecht fühlen'. Daher wird in dieser Studie die NP *sog-i* als das nicht realisierte Subjekt des Prädikats von S2 betrachtet, da sie eine Kollokation mit dem Verb *geobughada* bildet. Im Prädikat (60)b wird *jeonseol* 'die Geschichte' als Subjekt des Prädikats *heungmilobda* annotiert. In diesem Kontext wird das Prädikat als 'ist interessant' interpretiert, da der Sprecher die Geschichte als interessant bezeichnet, nachdem er sie gehört hat. In (60)a und (60)b kann das Nullsubjekt jedoch je nach Annotator als Sprecher interpretiert werden, der sich krank fühlt oder der etwas interessant findet. In diesem Fall werden die Referenten in der Sprecherposition, die durch Regel 53 zugewiesen wurde, als korrekte Vorhersagen gewertet. Dies hängt mit einem ungelösten Problem in der koreanischen Sprache bei der Subjektbestimmung von psychologischen Verben zusammen. Es gibt unterschiedliche Meinungen dazu, ob die NP in der Theta-Rolle Experiencer oder die NP mit der Theta-Rolle Stimulus-Causer oder Thema das Subjekt eines psychologischen Verbs ist. Um diese Frage zu klären, wird das folgende Beispiel herangezogen (zitiert aus Yoo Hyun-kyung 2015: 71).

- (61) *[na-neun]_{NP1} [i bang-i]_{NP2} museob-da-*.
 ich-DP dies Raum-SBJ Angst haben-SFS.DECL-SF
 'Ich habe Angst vor diesem Raum.'
 'Dieser Raum macht mir Angst.'

Einige Forscher argumentieren dafür, dass die NP1 mit der Theta-Rolle Experiencer das Subjekt des psychologischen Verbs ist (vgl. Lee Eunkyung 1996; Choi Hyoung Kang 2004; Yoo Hyun-kyung 1997; 1998; Kim Jung-Nam 2005 u. a.). Die Begründung dafür liegt in der Kompatibilität von NP1 mit dem Honorativinfix *si* oder dem Subjekt-Marker *kkeseo*, die das Subjekt-Honorifikum repräsentieren.⁹⁵ In dieser Perspektive wird die Satzstruktur von (61) als ähnlich der ersten deutschen Übersetzung 'Ich habe Angst vor diesem Raum' betrachtet. Auf der anderen Seite gibt es Ansichten, die die NP2 als Subjekt des psychologischen Verbs betrachten (vgl. Im Hongbin 1972; 2007; Woo Soon-jo 1994; Kim Young Ha 1996; Yi Seon-ung 2012 u. a.). Es wird argumentiert, dass die NP1 *na-neun* 'ich-DP' durch *na-*

⁹⁵ Für eine eingehende Erläuterung dieser Argumentation mit entsprechenden Beispielen vgl. Yoo Hyun-kyung (2015: 71–72).

ege 'mir' ersetzt werden kann, sodass die NP2 *bang-i* 'Raum-SBJ' mit dem Subjekt-Marker das Subjekt des Satzes ist. Unter diesem Gesichtspunkt ähnelt die Struktur von Satz (61) der zweiten deutschen Übersetzung 'Der Raum macht mir Angst'. Die bisherigen Diskussionen konzentrieren sich ausschließlich darauf, welches der beiden Argumente das Subjekt ist, wenn zwei Argumente für psychologische Verben realisiert werden. In realen Gesprächen kann es jedoch vorkommen, dass keine Argumente realisiert werden, wie in den Beispielsätzen von (60) zu sehen ist. In solchen Fällen können andere Kriterien für die Bestimmung des Referenten von Nullsubjekten eine Rolle spielen, die bisher nicht untersucht wurden. Dies könnte die Übereinstimmung zwischen den Annotatoren beeinflussen, wenn der Referent eines Nullsubjekts annotiert wird und nur Prädikate realisiert werden. Für zukünftige Forschungsarbeiten ist eine detaillierte Analyse und Bewertung der Intuition koreanischer Muttersprachler zu diesem Thema notwendig.

Die dritt- und fünftplatzierte Regel (Regel 34 und 40) in Tabelle 31 weisen niedrige Genauigkeitswerte von 0,47 bzw. 0,43 auf, und somit weniger korrekte Zuweisungen als über eine Zufallsauswahl erreicht werden würden. Beachtenswert ist, dass diese beiden Regeln auf periphrastischen Konstruktionen basieren, die den Sprecherwillen ausdrücken. Es wäre von Interesse zu untersuchen, auf welche periphrastischen Konstruktionen sich diese beiden Regeln stützen. Dies ist insbesondere relevant, da die zwei periphrastischen Konstruktionen *go sipda* und *lyeogo hada* in anderem Kontext als eindeutige Indikatoren für sprecherbezogene Referenten fungiert haben und eine Genauigkeit von 1,00 aufwiesen (siehe Tabelle 30 und Beispiele (57)a und (57)b).

- (62) a. *ama* *geumbang dochagha-I geo-bnida-*.
vielleicht bald ankommen-PPC(Vermutung)-SFS.DECL-SF
'Vielleicht kommt (**der Krankenwagen**) bald an.'
- b. *[jom iljjig o-si-eoss-deolamyeon]_{S1} [bo-I su iss-eoss-eul geo-bnida-.]_{S2}*
etwas früher kommen-HON-PST-CNJ sehen-PPC(Möglichkeit)-PST-PPC(Vermutung)-SFS.DECL-SF
'Wenn (Sie) etwas früher gekommen wären, hätten (**Sie**) (das Konzert) sehen können.'

Die Ursache für die falschen Vorhersagen der Regeln 34 und 40 liegt in der Tatsache begründet, dass die periphrastische Konstruktion (*eu*) *geoda*, auf der beide Regeln basieren, nicht nur für Willensäußerungen, sondern auch zum Ausdruck von Vermutungen verwendet wird (vgl. Kang Hyoun Hwa et al. 2016: 573–578). In den Beispielen in (62)a und (62)b, die die falschen Vorhersagen beider Regeln darstellen, wurde die periphrastische Konstruktion dazu verwendet, die Vermutung auszudrücken, dass der Krankenwagen gleich ankommt und dass der Adressat das Konzert sehen könnte.

Der Genauigkeitswert von 0,47 und 0,43 für diese beiden Regeln deutet darauf hin, dass die periphrastische Konstruktion in Deklarativsätzen in den Analysedaten eher als Vermutung denn als Wiedergabe des Sprecherwillen verwendet wurde. Dies lässt darauf schließen, dass der Grammatikalisierungsgrad auch bei periphrastischen Konstruktionen variiert kann. Denn die periphrastischen Konstruktionen *go sipda* und *lyeogo hada* wurden in den Analysedaten ausschließlich zum Ausdruck des Sprecherwillens in Deklarativsätzen verwendet (vgl. Tabelle 30). Die falschen Vorhersagen dieser beiden Regeln beruhen auf den Verwendungskontexten, weshalb die Kombination der periphrastischen Konstruktionen mit der Satzmodusinformation nicht als unzuverlässiger Indikator interpretiert werden kann. Denn die periphrastischen Konstruktionen mit Satzmodusinformation können als eindeutiger Indikator für die Identifikation von sprecherbezogenen Nullsubjekten dienen, wenn die Verwendungskontexte dieser Konstruktionen als Ausdruck der Sprecherabsicht disambiguieren werden.

Als Disambiguierungsstrategie kann die Einführung eines neuen Elements, wie zum Beispiel eines Adverbs im Satz für modale Vermutungen, z. B. *ama* 'vielleicht' wie in (62)a, in Betracht gezogen werden. Wenn in einem Satz ein Adverb vorhanden ist, das auf Unsicherheit hinweist, wird die

periphrastische Konstruktion (*eu*)*l geoda* wahrscheinlich verwendet, um eine Vermutung auszudrücken und nicht die Absicht des Sprechers. Daher kann das Prädikat mit dieser Kombination von periphrastischen Konstruktionen bei Vorhandensein eines solchen Adverbs von der Markierung für sprecherbezogene Nullsubjekte ausgeschlossen werden.

Eine zweite Möglichkeit wäre die Berücksichtigung einer Kombination von periphrastischen Konstruktionen. Wenn die periphrastische Konstruktion (*eu*)*l geoda* zusammen mit einer anderen periphrastischen Konstruktion wie *su issda* in der Funktion, eine Möglichkeit zu kennzeichnen, auftritt, wird die Kombination dieser beiden periphrastischen Konstruktionen verwendet, um eine Vermutung, aber keine Gewissheit, anzuzeigen. Beispielsweise wird in (62)b angedeutet, dass der Adressat die Möglichkeit hat, das Konzert zu sehen, aber es ist ungewiss, ob es tatsächlich stattfinden wird. Da es wahrscheinlicher ist, dass die periphrastische Konstruktion (*eu*)*l geoda* in diesem Fall dazu verwendet wird, eine Vermutung statt der Sprecherabsicht auszudrücken, kann ein Prädikat mit dieser Kombination von periphrastischen Konstruktionen nicht als Hinweis für die Identifizierung von sprecherbezogenen Nullsubjekten angesehen werden.

Das folgende Beispiel repräsentiert eine falsche Vorhersage der letzten Regel (Regel 80).

- (63) *susanmul-, yachae-, gwail deung-eul pal-go iss-seubnida-*.
Meeresfrüchte-NF Gemüse-NF Obst usw.-OBJ verkaufen-AUX-SFS.DECL-SF
'(Leute in dem Markt) verkaufen Meeresfrüchte, Gemüse und Obst usw.'

Das Nullsubjekt in (63) kann als Sprecher wiederhergestellt werden, wenn kein Äußerungskontext gegeben ist, was dazu führt, dass die zugewiesenen Referenten von den Regeln als korrekt bewertet werden. Die Interpretation der Nullsubjekte als Sprecher könnte sogar die intuitivste Lesart für diese Nullsubjekte sein, wenn nur die einzelne Äußerung angezeigt wird. Bei der Zuweisung eines Referenten können die Regeln den Äußerungskontext nicht berücksichtigen, da sie sich ausschließlich auf explizite linguistische Marker in den Prädikaten stützen. Dies zeigt, dass nicht nur explizite Hinweise in Verbalkomplexen, sondern auch der Kontext einer Äußerung eine Rolle bei der Resolution von sprecherbezogenen Nullsubjekten im Koreanischen spielen kann.

5.1.3. Zuverlässige Indikatoren für die korrekte Rekonstruktion adressatenbezogener Nullsubjekte

Für die Zuweisung des Referenten in der 2. Person Singular wurden insgesamt 112 Regeln entwickelt, von denen 87 tatsächlich im NRS verwendet wurden. Diese 87 Regeln zeichnen sich durch eine Genauigkeit von 0,97 aus, da sie erfolgreich 6.566 der identifizierten 6.739 Nullsubjekte korrekt rekonstruiert haben. Abbildung 39 veranschaulicht die kumulierte Anzahl der korrekten und inkorrekt vorhergesagten Referenten durch die Anwendung dieser 87 Regeln.

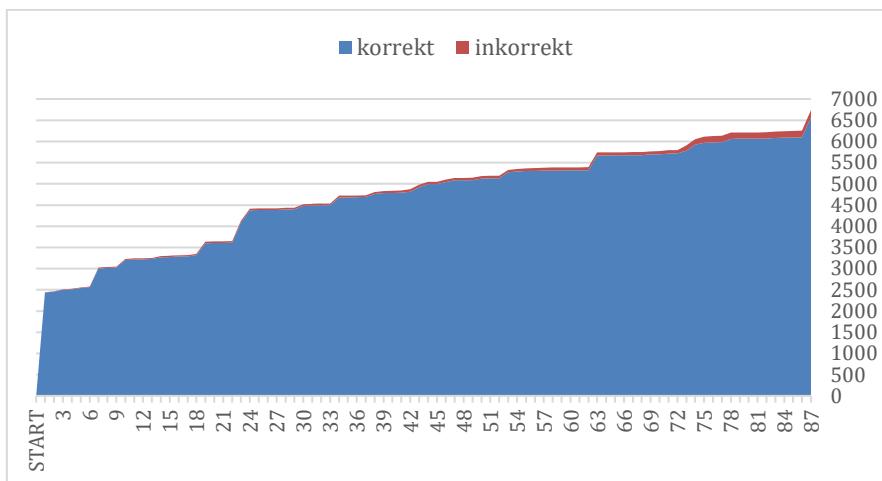


Abbildung 39: Vorhersagenentwicklung: Referenten der 2. Person Singular anhand von 87 Regeln

Auffällig ist, dass schon die erste Regel einen erheblichen Beitrag zur Steigerung der identifizierten Fälle leistet, da etwa 36 % der Nullsubjekte identifiziert werden können. Weitere bemerkenswerte Anstiege der identifizierten Fälle sind auf die erfolgreiche Anwendung der Regeln 7, 19, 23 und 24 zurückzuführen. Die ersten 24 Regeln haben eine hohe Identifizierungsfähigkeit, da sie bereits etwa 65 % aller mit den 87 Regeln identifizierten Nullsubjekte identifizieren. Der sehr geringe Anteil des roten Bereichs in Abbildung 39 zeigt, dass die Vorhersagen dieser Regeln in hohem Maße korrekt sind. Bei der Anwendung der Regeln 25 bis 62 ist ein gradueller Anstieg der identifizierten Fälle in der Grafik zu beobachten. Regel 63 zeigt einen weiteren Anstieg der identifizierten Fälle, gefolgt von zwei weiteren durch Regel 74 und die abschließende Regel 87. Dieser letzte Abschnitt der Grafik weist eine deutliche Zunahme des roten Bereichs auf, was auf eine erhöhte Anzahl fehlerhafter Vorhersagen durch die angewandten Regeln hinweist. Aus der Grafik lassen sich drei unterschiedliche Abschnitte mit eindeutigen Tendenzen ableiten, die im Folgenden einzeln betrachtet werden. Eine eingehende Analyse der Regeln in diesen drei Abschnitten verdeutlicht die Rolle linguistischer Merkmale bei der Identifizierung und Resolution von adressatenbezogenen Nullsubjekten. Dabei lassen sich jene linguistischen Merkmale identifizieren, die eine hohe Identifizierungs- und Resolutionsfähigkeit aufweisen und somit als entscheidende Parameter für die erfolgreiche Rekonstruktion adressatenbezogener Nullsubjekte interpretiert werden können.

Abschnitt	Anzahl korrekter Vorhersagen	Anzahl identifizierter Fälle	Genauigkeit
1. Regel 1–24 (–27,59 %)	4.372	4.411	0,99
2. Regel 25–62 (28,74–71,26 %)	946	980	0,97
3. Regel 63–87 (72,41–100,00 %)	1.248	1.348	0,93
Total	6.566	6.739	0,97

Tabelle 32: Leistung der Resolution von adressatenbezogenen Nullsubjekten in den Regelabschnitten

Im ersten Abschnitt erzielen die 24 Regeln einen bemerkenswert hohen Genauigkeitswert von 0,99, d. h. nahezu fehlerfreie Vorhersagen. Von diesen Regeln basieren vier auf linguistischen Mitteln, die oft als eindeutige Hinweise für eine erfolgreiche Resolution von Nullsubjekten angeführt werden. Dazu gehören Imperativmarker, feste Wendungen für indirekte Aufforderungen und Verbalsuffixe in Interrogativsätzen zum Ausdruck des Adressatenwillens. Das folgende Beispiel illustriert die korrekte Vorhersage der vier Regeln und verdeutlicht damit die zugrundeliegenden linguistischen Mittel dieser Regeln.

- (64) a. *mul jom gajyeoda ju-seyo-*.
Wasser bitte bringen-SFS.IMP-SF
'Bringen (**Sie**) (mir) bitte Wasser.'
- b. *damyo jom gajyeoda ju-si-gess-eoyo-?*
Decke bitte holen-HON-IFX-SFS-SF
'Könnten (**Sie**) (mir) bitte eine Decke bringen?'
- c. *gyesanha-si-gess-seubnikka-?*
bezahlen-HON-IFX-SFS.INT-SF
'Möchten (**Sie**) bezahlen?'

Zwei der vier Regeln weisen einem Prädikat mit einem Imperativmarker wie *seyo* in (64)a einen Referenten in der Adressatenrolle zu, da der Marker einen Imperativsatz kennzeichnet. Der zugewiesene Referent wird als korrekt bewertet, da der Sprecher in (64)a mit einem Imperativsatz den Adressaten auffordert, Wasser zu bringen. Eine weitere Regel identifiziert Prädikate mit einem Ausdruck wie *jusigesseoyo* in (64)b in einem Interrogativsatz, um einem Referenten in der Adressatenrolle zuzuweisen, da es für indirekte Aufforderungen verwendet wird. Diese Zuordnung wird für (64)b als korrekt bewertet, da der Sprecher den Ausdruck verwendet, um höflich zu bitten, dass der Adressat ihm eine Decke bringen soll. Eine andere Regel ordnet dem Prädikat in (64)c einen Referenten in der Adressatenrolle zu, weil es eine Kombination des Honorativinfixes mit dem Verbalsuffix *gessseubnikka* in einem Interrogativsatz enthält, die verwendet wird, um den Willen des Adressaten zu erfragen. Der zugewiesene Referent ist korrekt, da der Adressat gefragt wird, ob er bezahlen möchte.

Im Leistungsvergleich der 24 Regeln in Abschnitt 1 erreichen diese vier Regeln, die auf eindeutigen Hinweisen für die Referenten in der Adressatenrolle basieren, eine Genauigkeit von 0,997, da sie 3.217 der 3.227 identifizierten Nullsubjekte korrekt rekonstruieren. Bemerkenswert ist, dass fast die Hälfte der durch die 87 Regeln identifizierten Nullsubjekte allein durch diese vier Regeln identifiziert wurden (3.227 von 6.739), und dass die identifizierten Nullsubjekte nahezu fehlerfrei rekonstruiert wurden. Dies unterstreicht die Effektivität nicht nur einzelner Parameter wie Imperativmarker, sondern auch einer Kombination von Parametern wie festen Wendungen für indirekte Aufforderungen und

Verbalsuffixen in Verbindung mit Informationen zum Satzmodus für die präzise Rekonstruktion von Nullsubjekten mit Referenz auf Adressaten.

Die übrigen vier Regeln in Abschnitt 1 basieren auf den in dieser Studie neu eingeführten linguistischen Mitteln wie periphrastischen Konstruktionen, die eine modale Bedeutung im Sinne von Vermutungen ausdrücken, sowie Formulierungen zur Darstellung von Vorschlägen oder Erfahrungen.

- (65) a. *ama cheken-e mulli-si-n geos gat-neyo-*.
wahrscheinlich Zecke-von gebissen werden-HON-PPC(Vermutung)-SFS-SF
'(Ich) glaube, (**Sie**) wurden wahrscheinlich von einer Zecke gebissen.'
- b. *hyudaejeonhwa-neun kkeo noh-neun ge joh-gess-seubnida-*.
Handy-DP ausschalten-PPC-IFX-SFS.DECL-SF
'(**Sie**) sollten (Ihr) Handy besser ausschalten.'
- c. *choegeune baileoseu geomsa-leul ha-si-n jeogi iss-seubnikka-?*
kürzlich Virus Test-OBJ machen-HON-PPC-SFS.INT-SF
'Wurden (**Sie**) kürzlich auf das Virus getestet?'

Zwei der vier Regeln identifizieren Prädikate, die entweder eine Variante der periphrastischen Konstruktionen wie *n geos gatda* in einem Deklarativsatz enthalten, die für modale Vermutungen verwendet werden, oder die eine Kombination einer solchen Konstruktion mit dem Honorativprefix aufweisen. Diese periphrastischen Konstruktionen wurden bei der Erstellung dieser Regeln berücksichtigt, da es äußerst unwahrscheinlich ist, dass ein Prädikat, das zum Ausdruck von Vermutungen verwendet wird, auf Handlungen oder Zustände eines Subjekts in der Sprecherrolle verweist. Dem Prädikat in (65)a wird ein Referent in der Adressatenrolle zugewiesen, da die Kombination der periphrastischen Konstruktion *n geos gatda* und des Honorativprefixes gefunden wird. Eine weitere Regel identifiziert in einem Deklarativsatz Prädikate mit einem Ausdruck wie *neun ge johgessda* in (64)b, um einem Referenten in der Adressatenrolle zuzuweisen, da er einen Vorschlag an den Adressaten signalisiert. Eine andere Regel ordnet dem Prädikat in (64)c einen Referenten in der Adressatenrolle zu, weil es in einem Interrogativsatz einen Ausdruck wie *n jeogi issda* enthält, der verwendet wird, um die Erfahrung von Adressaten zu erfragen.

Die vier Regeln, die auf neu eingeführten linguistischen Mitteln basieren, erreichen einen Genauigkeitswert von 0,925 und rekonstruieren 86 von 93 identifizierten Nullsubjekten korrekt. Obwohl ihre Identifizierungsleistung im Vergleich zu den ersten vier Regeln geringer ist, zeigt sich doch, dass die Hinzunahme periphrastischer Konstruktionen oder Formulierungen zur Analyse präzise Vorhersagen bezüglich adressatenbezogener Nullsubjekte zulässt, auch wenn nur wenige Nullsubjekte davon betroffen sind. Es besteht die Möglichkeit, dass die Kombination der linguistischen Mittel, auf die diese Regeln zurückgreifen, als ein entscheidender Parameter für die Rekonstruierbarkeit adressatenbezogener Nullsubjekte bestätigt werden kann, wenn die Regeln in anderen Analysedaten eine höhere Anzahl von Nullsubjekten identifizieren und korrekt rekonstruieren können.

Die restlichen 16 Regeln in Abschnitt 1 zielen darauf ab, Nullsubjekte in Nebensätzen zu rekonstruieren. Diese Regeln identifizierten insgesamt 1.091 Nullsubjekte in Nebensätzen. Davon wurden 1.069 als korrekte Vorhersagen bewertet, was einer Genauigkeit von 0,98 entspricht. Dies lässt darauf schließen, dass die speziell für die Resolution von Nullsubjekten in Nebensätzen entwickelten Regeln ihren Zweck genau erfüllen. Das folgende Beispiel veranschaulicht die korrekten Vorhersagen von drei dieser 16 Regeln, die eine hohe Identifizierungsfähigkeit aufweisen und somit die Hauptmerkmale dieser Regeln verdeutlichen. Durch diese drei Regeln allein werden etwa 84 % der von den insgesamt 16 Regeln identifizierten Nullsubjekte abgedeckt.

- (66) a. *[jigeum bangmunha-eo ju-si-myeon]_{S1} [jasehi alli-eo deuli-gess-seubnida-.]_{S2}*
jetzt besuchen-AUX-HON-CNJ ausführlich informieren-AUX-IFX-SFS.DECL-SF
'Wenn (**Sie**) (uns) jetzt besuchen, werde (ich) (Sie) ausführlich informieren.'
- b. *[ijjogeulo o-si-eoseo]_{S1} [gyesanha-si-myeon doe-bnida-.]_{S2}*
hierher kommen-HON-CNJ bezahlen-HON-PPC(Erlaubnis)-SFS.DECL-SF
'(**Sie**) können hierher kommen und bezahlen.'
- c. *[pyo-neun gaji-si-go]_{S1} [ijjogeulo nao-seyo-.]_{S2}*
Ticket-DP behalten-HON-CNJ hierher rauskommen-SFS.IMP-SF
'(**Sie**) können Ihr Ticket behalten und hier herauskommen.'

Die drei Regeln suchen nach einem Prädikat, das bestimmte Verben in ihrer Honorativform oder mit einem Honorativinfix enthält, um Respekt gegenüber dem Adressaten auszudrücken, in Kombination mit den Konjunktionen *myeon*, *eoseo* oder *go*. Dem Prädikat wird dann ein Referent in der Adressatenrolle zugewiesen. Die Vorhersage dieser Regeln ist in den Beispielsätzen von (66) korrekt, da die Person, die zu Besuch kommt bzw. die Möglichkeit hat, hierher zu kommen oder das Ticket zu behalten, der Adressat ist. Das wichtigste linguistische Mittel für die drei Regeln ist das Honorativinfix, das es ermöglicht, Nullsubjekte als Adressaten zu interpretieren. Diese Regeln wurden basierend auf der Annahme formuliert, dass das Honorativinfix nicht nur in Hauptsätzen, sondern auch in Nebensätzen eine wichtige Rolle spielen kann, da es zum Ausdruck von Respekt gegenüber dem Adressaten genutzt wird. Von den 16 Regeln nutzen 14 das Honorativinfix als Hauptmerkmal. Sie identifizieren etwa 84 % aller identifizierten Nullsubjekte, die durch die 16 Regeln identifiziert wurden, und erreichen einen Genauigkeitswert von 0,98, während die restlichen zwei Regeln einen Genauigkeitswert von 0,82 aufweisen. Dies zeigt, dass das Honorativinfix ein starker Indikator für adressatenbezogene Nullsubjekte in Nebensätzen ist.

Abschnitt 2 in Tabelle 32 umfasst 38 Regeln, die etwa 15 % der mit den 87 Regeln identifizierten Nullsubjekte abdecken. Im Vergleich zu Abschnitt 1 weisen die Regeln in Abschnitt 2 eine geringere Identifizierungsfähigkeit auf, da sie etwa 4,5-mal weniger Nullsubjekte identifizieren als die 24 Regeln in Abschnitt 1, die 4.411 Nullsubjekte identifiziert haben. Jedoch deutet ihr hoher Genauigkeitswert von 0,97 darauf hin, dass das identifizierte Nullsubjekt höchstwahrscheinlich als Adressat interpretiert wird. Die Regeln in Abschnitt 2 basieren auf verschiedenen Arten von linguistischen Mitteln wie Verbalsuffixen, psychologischen Verben, peripherastischen Konstruktionen, performativen Verben und Hilfsverben. Bemerkenswert ist, dass bei 25 dieser Regeln das Vorhandensein des Honorativinfixes genutzt wird, um einen adressatenbezogenen Referenten zuzuweisen. Insgesamt konnten anhand dieser Regeln 850 Nullsubjekte identifiziert werden, was etwa 87 % der durch die Regeln in Abschnitt 2 identifizierten Nullsubjekte entspricht. Dabei erreichen die Regeln einen Genauigkeitswert von 0,97, wobei 827 Nullsubjekte korrekt als Adressat interpretiert werden konnten. Diese Ergebnisse legen nahe, dass das Vorhandensein von Honorativinfixen ein aussagekräftiges Signal für adressatenbezogene Nullsubjekte darstellt. Um die Auswirkungen des Honorativinfixes auf die Resolution adressatenbezogener Nullsubjekte zu bewerten, muss analysiert werden, wie viele Nullsubjekte der insgesamt 15.883 Input-Prädikate allein durch die Verwendung des Honorativinfixes korrekt rekonstruiert werden können. Von diesen Prädikaten enthalten 6.422 das Honorativprefix (*eu*)*si*, was durch den Wert "hono" in den Zusatzdaten gekennzeichnet ist. Von diesen Prädikaten wurden 6.314 Nullsubjekte als Adressat interpretiert, was einen Genauigkeitswert von 0,983 ergibt. Den 6.422 Prädikaten wurden im NRS durch 64 Regeln ein Referent zugewiesen, wobei 6.291 den korrekten Referenten in der Adressatenrolle erhalten, was einen Genauigkeitswert von 0,980 ergibt. Dies zeigt, dass allein die Verwendung des Honorativinfixes einen leicht höheren Genauigkeitswert erzielt als die Anwendung der 64 Regeln. Folglich unterstreicht dies die signifikante Rolle des Honorativinfixes als Indikator für adressatenbezogene Nullsubjekte und deutet darauf hin, dass dieser einzelne Parameter besonders wichtig für die Resolution der Nullsubjekte mit Referenz auf Adressaten ist.

Die 25 Regeln in Abschnitt 3 identifizierten 20 % aller durch die 87 Regeln identifizierten Nullsubjekte und wiesen den niedrigsten Genauigkeitswert von 0,93 auf. Sie führten zu 100 falschen Vorhersagen, was etwa 58 % der insgesamt 173 falschen Vorhersagen ausmacht und damit den großen Anteil der roten Bereiche in Abbildung 39 erklärt. Drei Regeln sind für den signifikanten Anstieg der identifizierten Fälle in Abbildung 39 verantwortlich. Die Regeln 63 und 74 basieren auf der Kombination des Honorativinfixes mit periphrastischen Konstruktionen, die eine modale Bedeutung im Sinne von Erlaubnis oder Möglichkeit haben. Regel 87 basiert lediglich auf dem Honorativprefix und wird als letzte Regel angewandt. Diese drei Regeln identifizierten 978 Nullsubjekte, was etwa 73 % der durch die Regeln in Abschnitt 3 identifizierten Nullsubjekte entspricht. Von diesen wurden 960 korrekt vorhergesagt, was einer Genauigkeit von 0,98 entspricht. Dies bestätigt erneut, dass das Honorativprefix ein zuverlässiger Indikator für adressatenbezogene Nullsubjekte ist.

Den größten Einfluss auf die erfolgreiche Identifizierung und korrekte Resolution adressatenbezogener Nullsubjekte haben insgesamt sieben Regeln, die im Folgenden detailliert analysiert werden. Tabelle 33 zeigt eben diese Regeln in der Reihenfolge ihres größten Beitrags zur Steigerung des Anteils korrekter Vorhersagen (vgl. Abbildung 39).

Rang	Regel	Genauigkeit
1	Regel 1 impe ("rule_order": 1)	0,996
2	Regel 23 request ("rule_order": 55)	0,998
3	Regel 87 hono total ("rule_order": 212)	0,971
4	Regel 7 clau_myeon ("rule_order": 8)	0,984
5	Regel 63 permission periphrastic construction/plus ("rule_order": 137)	1,000
6	Regel 24 want verbal suffix ("rule_order": 56)	1,000
7	Regel 19 clau_plus ("rule_order": 23)	0,962

Tabelle 33: Genauigkeit der sieben wichtigsten Regeln für die korrekte Resolution adressatenbezogener Nullsubjekte

Die ersten beiden Regeln basieren auf Imperativmarkern und festen Ausdrücken für indirekte Aufforderungen, die in früheren Studien oft als starke Indikatoren für die Resolution von adressatenbezogenen Nullsubjekten angesehen wurden. Diese beiden Regeln konnten etwa 43 % der von den 87 Regeln identifizierten Nullsubjekte nahezu fehlerfrei rekonstruieren, was ihre hohe Identifizierungs- und Resolutionsfähigkeit zeigt. Die restlichen fünf Regeln in Tabelle 33 basieren auf einer Kombination verschiedener linguistischer Mittel wie periphrastischen Konstruktionen, Verbalsuffixe, Konjunktionen, des Honorativinfixes und Informationen über den Satztyp. Das gemeinsame linguistische Merkmal aller fünf Regeln ist das Honorativprefix. Da ihre Vorhersagen mit einer Wahrscheinlichkeit von mindestens 96 % bis maximal 100 % korrekt sind, bestätigt dies erneut, dass das Honorativprefix eine wichtige Rolle bei der korrekten Zuweisung der Referenten in der Adressatenrolle spielt. Die hier genannten fünf Regeln decken etwa 27 % der von den 87 Regeln identifizierten Nullsubjekte ab, wobei zusammen mit der ersten Regel bereits 71 % der Nullsubjekte identifiziert werden können. Diese Ergebnisse unterstreichen die hohe Leistung der Imperativmarker, der festen Wendungen für indirekte Aufforderungen und des Honorativinfixes bei der Identifizierung und Resolution der adressatenbezogenen Nullsubjekte. Sie deuten darauf hin, dass einzelne Parameter wie Imperativmarker oder Honorativprefix eine besonders entscheidende Rolle bei der korrekten Rekonstruktion von adressatenbezogenen Nullsubjekten spielen.

Im nächsten Schritt wird eine Untersuchung durchgeführt, um die Regeln zu analysieren, die zu einem Anstieg fehlerhafter Vorhersagen bei der Resolution von adressatenbezogenen Nullsubjekten geführt haben. Das Ziel dieser Analyse ist es festzustellen, ob die linguistischen Merkmale, auf denen diese Regeln basieren, tatsächlich als unzuverlässige Anzeichen für die adressatenbezogenen Nullsubjekten angesehen werden müssen, und gegebenenfalls die Gründe dafür zu ermitteln. Aus den 87 Regeln, die einem Prädikat einen Referenten in der Adressatenrolle zuordnen, resultieren 173 falsche Vorhersagen, die aus 34 Regeln stammen. Fünf dieser Regeln tragen wesentlich zum Anstieg der inkorrekt Vorhersagen bei. Tabelle 34 stellt diese fünf Regeln in aufsteigender Reihenfolge des größten Anteils an falschen Vorhersagen dar.

Rang	Regel	Genauigkeit
1	Regel 73 permission periphrastic construction ("rule_order": 180)	0,609
2	Regel 75 possibility periphrastic construction ("rule_order": 185)	0,746
3	Regel 87 hono total ("rule_order": 212)	0,971
4	Regel 19 clau_plus ("rule_order": 23)	0,962
5	Regel 1 impe ("rule_order": 1)	0,996

Tabelle 34: Genauigkeit der fünf Regeln mit dem größten Einfluss auf den Anstieg fehlerhafter Prognosen für adressatenbezogene Nullsubjekte

Diese fünf Regeln verursachen insgesamt 92 falsche Vorhersagen, was mehr als die Hälfte aller falschen Vorhersagen ausmacht. Bemerkenswert ist, dass die Regeln auf dem dritten bis fünften Platz zusätzlich zu den Top-7-Regeln in Tabelle 33 gehören, was bedeutet, dass sie den größten Anstieg sowohl korrekter als auch falscher Vorhersagen verursacht haben. Im Gegensatz zu den ersten beiden Regeln in Tabelle 34 weisen diese drei Regeln höhere Genauigkeitswerte auf. Dies legt nahe, dass das Honorativinfix und der Imperativmarker, auf denen diese Regeln basieren, immer noch verlässliche Indikatoren für adressatenbezogene Nullsubjekte sind. Die fehlerhaften Vorhersagen, die durch die drei Regeln verursacht wurden, betrafen stets Nullsubjekte mit Referenten in der 3. Person. Der Grund für diese Fehlvorhersage liegt darin, dass das Honorativinfix, das diesen Regeln zugrunde liegt,⁹⁶ als Subjekt-Honorifikum fungiert, um Respekt gegenüber Personen in der Subjektposition auszudrücken. Diese Regeln wurden dennoch entwickelt, da das Honorativinfix oft in Gesprächen genutzt wird, um Respekt gegenüber dem Adressaten zum Ausdruck zu bringen. Die Tatsache, dass diese Regeln mit einer Wahrscheinlichkeit von über 96 % der Nullsubjekte den korrekten Referenten in der Adressatenrolle zuweisen, deutet darauf hin, dass das Honorativinfix in den Analyseergebnissen in der Regel verwendet wird, um Respekt gegenüber dem Adressaten zu bekunden.

Die ersten beiden Regeln in Tabelle 34 basieren wiederum auf einer Kombination aus Satzmodusinformationen und periphrastischen Konstruktionen, die die modalen Bedeutungen im Sinne von Erlaubnis oder Möglichkeit ausdrücken. Warum diese Regeln relativ niedrige Genauigkeitswerte von 0,609 und 0,746 aufweisen, steht im Folgenden im Fokus. Um die Gründe für die falschen Vorhersagen dieser Regeln herausarbeiten zu können, werden Beispiele mit fehlerhafter Zuordnung herangezogen, die weiteren Aufschluss über mögliche Fehlerquellen geben können.

⁹⁶ Die falsche Vorhersage der Regel 1, die auf Imperativmarkern basiert, wird hauptsächlich auf Parsingfehler in den Zusatzdaten zurückgeführt. Die betroffenen Sätze sind keine Imperativsätze, enthalten jedoch das Honorativinfix, was eine weitere Ursache für den Fehler darstellt.

- (67) a. *don nae-myeon doe-janhayo-*.
 Geld bezahlen-PPC(Genügen)-SFS.DECL-SF
 '(Es ist ausreichend, wenn) (ich) dafür bezahlen kann.'
- b. *[geuli meol-ji anh-euni]s1 [geumbang dochagha-l su iss-eoyo-.]s2*
 nicht allzu weit sein-NEG-CNJ bald ankommen-PPC(Möglichkeit)-SFS-SF
 '(Es) ist nicht allzu weit weg, also können (wir) bald ankommen.'

Regel 73 wurde entwickelt, um in einem Deklarativsatz einem Prädikat mit periphrastischen Konstruktionen wie *myeon deoda*, d. h. mit der modalen Bedeutung von Pflicht, einen Referenten in der Adressatenrolle zuzuweisen. Ein Prädikat, das diese Bedingung erfüllt, wird oft genutzt, um eine Erlaubnis gegenüber dem Adressaten auszudrücken. Allerdings beträgt die Wahrscheinlichkeit, dass die zugeordneten Referenten durch diese Regel korrekt sind, nur etwa 61 %. Von den 43 fehlerhaften Vorhersagen dieser Regel wurden insgesamt 31 Nullsubjekte als Sprecher interpretiert, wie in (67)a, während die restlichen 12 Nullsubjekte als Subjekte der 3. Person interpretiert wurden. Insbesondere in (67)a wurde *myeon deoda* nicht im Sinne von Pflicht, sondern im Sinne von Genügen verwendet. Interessanterweise erreicht eine Regel, die unter denselben Bedingungen wie Regel 73 die Kombination der periphrastischen Konstruktionen mit dem Honorativinfix verwendet, einen Genauigkeitswert von 1, da alle Prädikate dafür verwendet wurden, eine Erlaubnis gegenüber dem Adressaten auszudrücken. Dies legt nahe, dass die Kombination aus periphrastischer Konstruktion und Satzmodusinformation allein ohne das Honorativinfix nicht immer eindeutig auf ein spezifisches Subjekt hinweist. Regel 75 wurde wiederum entwickelt, um einem Prädikat mit der periphrastischen Konstruktion *eul su issda* in einem Deklarativsatz einen Referenten in der Adressatenrolle zuzuweisen. Diese Regel wurde aufgestellt, weil Prädikate, die diese Bedingung erfüllen, oft verwendet werden, um mögliche Handlungen anzugeben, die vom Adressaten ausgeführt werden können. Der Genauigkeitswert dieser Regel von 0,756 deutet darauf hin, dass etwa 24 % der durch die Regel identifizierten Nullsubjekte nicht als Adressaten interpretiert werden. Von den 15 falschen Vorhersagen betreffen insgesamt 10 sprecherbezogene Nullsubjekte wie in (67)b, während die restlichen 5 Nullsubjekte Referenten in der 3. Person haben. Interessanterweise hat eine Regel, die zusätzlich das Vorhandensein des Honorativinfixes unter den gleichen Bedingungen wie Regel 75 verwendet, eine Genauigkeit von 0,972, da die meisten Nullsubjekte Referenten in der Adressatenrolle haben. Dies könnte darauf hindeuten, dass die Kombination aus periphrastischer Konstruktion und Satzmodusinformation alleine ohne das Honorativinfix ein relativ schwacher Hinweis für die Interpretation eines bestimmten Subjekts ist. Dies unterstreicht erneut die wichtige Rolle des Honorativinfixes für die korrekte Rekonstruktion von Nullsubjekten mit Bezug auf Adressaten.

5.1.4. Fazit: Anzahl der Parameter zur Rekonstruierbarkeit von Nullsubjekten

Im Fokus dieses Kapitels steht die Abschätzung der Anzahl der Parameter für die Rekonstruierbarkeit von Nullsubjekten mit Referenz auf Diskursteilnehmer:innen durch die Analyse der genutzten und ungenutzten Regeln im NRS. Der Begriff *Parameter* wird in diesem Abschnitt als Oberbegriff verwendet, um nicht nur einzelne Parameter, sondern auch eine Kombination von Parametern zu bezeichnen. Insgesamt wurden in dieser Studie 208 Regeln zur Zuweisung der Referenten in der Sprecher- oder Adressatenrolle aufgestellt. Basierend auf dieser Anzahl an aufgestellten Regeln wird vermutet, dass in koreanischen Verbalkomplexen bis zu 208 verschiedene Kombinationen linguistischer Merkmale existieren können, um Nullsubjekte zu identifizieren, die entweder den Sprecher oder den Adressaten als Referenten haben (siehe Tabelle 25 in Kapitel 4.3.3). Von diesen Regeln wurden 169, d.h. ca. 81 % für die Zuweisung eines Referenten verwendet.

In Kapitel 5.1.2 wurde festgestellt, dass eine einzelne Regel mit einer Trefferquote von 98 % erfolgreich Nullsubjekte mit Referenten in der 1. Person Plural wiederhergestellt hat, während 81 Regeln eine

Trefferquote von 96 % für die erfolgreiche Resolution der Nullsubjekte in der 1. Person Singular erzielten. Diese Ergebnisse legen nahe, dass etwa 82 Parameter existieren könnten, um Nullsubjekte als Sprecher zu identifizieren. Im Gegensatz dazu besteht die Möglichkeit, dass die Anzahl der Regeln zur Zuweisung der Referenten in der Adressatenrolle reduziert werden könnte. Wie in Kapitel 5.1.3 dargelegt, haben insgesamt 87 Regeln mit einer Trefferquote von 97 % Nullsubjekte mit Referenten in der Adressatenposition erfolgreich wiederhergestellt. Bei der Analyse der Leistung dieser Regeln wurde festgestellt, dass das Honorativ infix besonders effektiv zur korrekten Resolution der adressatenbezogenen Nullsubjekte beigetragen hat. Allein die Verwendung des Honorativ infixes konnte einen leicht höheren Genauigkeitswert erzielen als die Anwendung der 64 Regeln, die auf der Kombination verschiedener linguistischer Merkmale und des Honorativ infixes beruhen. Dies deutet darauf hin, dass die 64 im NRS verwendeten Regeln möglicherweise durch eine einzige Regel ersetzt werden könnten, die lediglich die Anwesenheit des Honorativ infixes berücksichtigt. Die Ergebnisse legen nahe, dass etwa 23 Parameter existieren könnten, um Nullsubjekte als Adressaten zu identifizieren, wenn man diese 64 Regeln aus den 87 Regeln ausschließt, die im NRS für die Zuweisung von Referenten in der Adressatenrolle verwendet wurden.

Eine weitere Analyse der 39 nicht im NRS verwendeten Regeln wird durchgeführt, um die Anzahl der potenziellen Parameter zu ermitteln, die wahrscheinlich zur Rekonstruierbarkeit der Nullsubjekte beitragen können, obwohl sie im NRS nicht zum Einsatz kamen. Zur Bewertung der Leistungsfähigkeit dieser ungenutzten Regeln wurde ein neues Resolutionssystem entwickelt, das ausschließlich aus diesen 39 Regeln besteht. Anschließend wurden diese Regeln auf dieselben Input-Prädikate angewandt, die bereits für das NRS verwendet wurden. Dieser Ansatz ermöglicht es, den tatsächlichen Beitrag der 39 unbenutzten Regeln zur Resolution von Nullsubjekten in den Input-Prädikaten zu bewerten, ohne die bereits verwendeten Regeln einzubeziehen. Die gewonnenen Ergebnisse liefern Erkenntnisse darüber, ob diese Regeln, obwohl zunächst ungenutzt, Potenzial für eine Verwendung in einem weiteren Nullsubjektresolutionssystem haben könnten. Auf diese Weise kann die Anzahl der potenziellen Parameter abgeschätzt werden.

Bei der Anwendung der 39 Regeln auf insgesamt 15.883 Input-Prädikate ergab sich eine weitgehende Zweiteilung: 26 Regeln konnten den Prädikaten entweder Referenten der Sprecher- oder Adressatenrolle zuordnen, während die restlichen 13 Regeln auf keine der 15.883 Input-Prädikate angewandt wurden.⁹⁷ Diese 26 Regeln wurden auf insgesamt 537 Prädikate angewandt, was etwa 3 % des Ausgangsmaterials ausmacht. Die Analyse dieser Fälle zeigt, dass die ermittelten Referenten bereits durch eine von 54 anderen Regeln im NRS wiederhergestellt werden konnten. Dies bedeutet, dass diese 54 Regeln im NRS vor den 26 Regeln zur Anwendung kamen und den Nullsubjekten der 537 Prädikate bereits einen der beiden Referenten der Diskursteilnehmerrolle zugewiesen haben, wodurch die Anwendung der entsprechenden 26 Regeln auf diese Prädikate im NRS unnötig war.

Die Tatsache, dass die restlichen 13 Regeln im neuen Resolutionssystem nicht verwendet wurden, legt nahe, dass unter den 15.883 Input-Prädikaten keine Prädikate existieren, die die Kombinationen der linguistischen Merkmale umfassen, auf denen diese 13 Regeln basieren. Es besteht also die Möglichkeit, dass diese Regeln dennoch verwendet werden könnten, wenn sie mit anderen Analysedaten getestet werden. Dies würde bedeuten, dass diese Regeln für die weitere Entwicklung eines Nullsubjektresolutionssystems nützlich sein und entsprechend integriert werden könnten. Diese Erkenntnis lässt vermuten, dass es weitere 13 Parameter geben könnte, die bei der Identifizierung der Nullsubjekte mit Bezug auf Diskursteilnehmer:innen eine Rolle spielen. Daraus lässt sich schließen, dass aus den ursprünglich 208 Regeln des NRS nach Ausschluss der 26 nicht verwendeten Regeln und der 64 verwendeten Regeln etwa 118 Parameter resultieren. Die in dieser Studie identifizierten

⁹⁷ Die "rule_order" für die offenbar redundanten übrigen 26 Regeln ist wie folgt: 9, 30, 44, 47, 50, 73, 83, 94, 96, 99, 115, 119, 121, 123, 126, 131, 135, 155, 164, 167, 168, 187, 193, 199, 200, 207. Die "rule_order" der 13 nicht angewandten Regeln lautet wie folgt: 10, 22, 51, 52, 110, 111, 116, 133, 134, 162, 165, 174, 196. Für eine Zuordnung zur Art der jeweiligen Regel vgl. Appendix 1 in der Spalte "rule_order".

Parameter sind etwa 5,6-mal so zahlreich wie diejenigen, die in früheren Studien intensiv diskutiert wurden, da 21 Regeln im NRS relevante Parameter aufweisen. Dieses Ergebnis verdeutlicht den wichtigen Beitrag dieser Studie zur Aufdeckung der vielfältigen Parameter für die Identifikation von Nullsubjekten mit Bezug auf Diskursteilnehmer:innen, was zu einem besseren Verständnis des weit verbreiteten Phänomens des Nullsubjekts im Koreanischen beiträgt.

5.2. Leistungsanalyse des Nullsubjektresolutionssystem

5.2.1. Bewertung der Leistung des Nullsubjektresolutionssystem

Der Fokus in diesem Kapitel liegt zunächst auf einem Vergleich der Genauigkeitswerte zwischen den initialen 208 Regeln und den 212 Regeln inklusive der vier ergänzenden Regeln (vgl. Tabelle 25 in Kapitel 4.3.3). Die vier zusätzlichen Regeln wurden entwickelt, um Prädikate auszuschließen, deren Subjekte weniger wahrscheinlich als Diskursteilnehmer:innen interpretiert werden. Die Integration dieser zusätzlichen Regeln zielt darauf ab, die Genauigkeit des NRS bei der Resolution von diskursteilnehmerbezogenen Nullsubjekten zu steigern. Die 208 Regeln ohne Einbeziehung der vier ergänzenden Regeln erreichen eine Genauigkeit von 0,954, da sie 11.950 der 12.530 identifizierten Nullsubjekte korrekt rekonstruieren (siehe Kapitel 5.1.1). Bei Hinzuziehung der vier zusätzlichen Regeln wurden insgesamt 13.715 Nullsubjekte identifiziert, von denen 13.027 Fälle als korrekte Vorhersagen der 212 Regeln bewertet wurden. Daraus resultiert eine Genauigkeit von 0,950:

$$\text{Genauigkeit des NRS} = \frac{13.027}{13.715} = 0,950$$

Diese Analyse der Genauigkeitswerte zeigt, dass die Einbeziehung der vier zusätzlichen Regeln zu einem um ca. 0,4 % niedrigeren Genauigkeitswert führt. Dies ist auf die relativ geringe Genauigkeit der vier Regeln zurückzuführen, die einen Genauigkeitswert von 0,80 aufweisen, da sie 1.239 von 1.550 identifizierten Nullsubjekten korrekt als Nullsubjekte in der 3. Person identifiziert haben. Im Vergleich zu den vier Regeln weisen die Regeln zur Zuweisung eines Referenten in der 1. Person Singular oder Plural und der 2. Person Singular einen Genauigkeitswert zwischen 0,96 und 0,98 auf. Diese Ergebnisse deuten darauf hin, dass diese zusätzlichen Regeln entgegen der ursprünglichen Annahme nicht zu einer signifikanten Verbesserung der Gesamtleistung des NRS beitragen können. Es ist jedoch erwähnenswert, dass die Regeln für die Zuweisung von diskursteilnehmerbezogenen Referenten im NRS unter Hinzunahme der vier zusätzlichen Regel einen Genauigkeitswert von 0,969 erreichen, da sie 11.788 von 12.165 Nullsubjekten korrekt rekonstruieren. Dies zeigt, dass diese Regeln aufgrund der vier zusätzlichen Regeln 365 Nullsubjekte weniger identifiziert haben, aber etwa 2 % genauere Vorhersagen treffen konnten, wenn man die Leistung der Regeln ohne Berücksichtigung der vier Regeln vergleicht.⁹⁸ Daraus kann gefolgert werden, dass die zusätzlichen Regeln zu besseren Vorhersagen der Regeln zur Resolution von diskursteilnehmerbezogenen Nullsubjekten beitragen, obwohl ihre Einbeziehung insgesamt die Genauigkeit des NRS nicht erhöhen konnte.

Im nächsten Schritt erfolgt eine eingehende Untersuchung der Identifizierungs- und Resolutionsleistung der insgesamt 169 im NRS regulär angewandten Regeln, indem sie sequenziell auf

⁹⁸ Es ist jedoch darauf hinzuweisen, dass der Vergleich mit Vorsicht zu genießen ist, da die Anzahl der angewandten Regeln je nach Anwendung der vier zusätzlichen Regeln variiert. Für die Zuordnung von Referenten in Sprecher- oder Adressatenrolle wurden im NRS mit Berücksichtigung der vier zusätzlichen Regeln insgesamt 169 Regeln angewandt, während ohne Berücksichtigung dieser vier Regeln 173 Regeln verwendet wurden.

die 12.165 Prädikate in der Reihenfolge ihrer Anwendung betrachtet werden. Die Analyse konzentriert sich dabei auf die kumulierte Anzahl korrekter und inkorrekt vorhergesagter Referenten. Diese detaillierte Untersuchung ermöglicht eine Einschätzung des jeweiligen Beitrags einer Regel zur korrekten Resolution der Nullsubjekte bei sukzessiver Anwendung der Regeln, was nicht durch die Genauigkeitswerte allein erkannt werden kann. Darüber hinaus erlaubt diese Analyse einen Einblick, wie viele der Nullsubjekte bei welcher Anzahl der angewandten Regeln bereits identifiziert werden können, was wiederum Rückschlüsse auf die Entwicklung der Identifizierungsfähigkeit der Regeln im NRS zulässt.

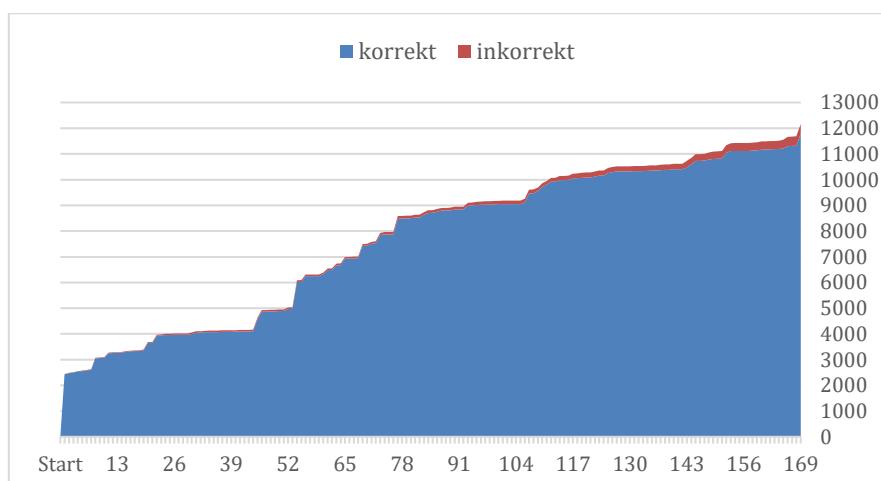


Abbildung 40: Entwicklung der korrekten und inkorrekt vorhergesagten Nullsubjekten der 169 Regeln des NRS

Die Grafik in Abbildung 40 veranschaulicht die Entwicklung des Anteils der bereits identifizierten Nullsubjekte während der schrittweisen Anwendung der Regeln. Die ersten 85 und damit 50 % der Regeln, können 8.814 Nullsubjekten von insgesamt 12.165 identifizierten Nullsubjekten einen Referenten zuordnen, was in etwa 72 % der Fälle entspricht. Während die ersten 25 % der Regeln lediglich etwa 34 % der Nullsubjekte identifizieren, steigt der Anteil der identifizierten Nullsubjekte durch die obersten 50 % der Regeln um etwa 38 %, wenn man die nächsten 25 % hinzuzählt. Dies deutet darauf hin, dass die Integration der restlichen 50 % der Regeln im NRS zu einer vergleichsweise moderaten Steigerung des Anteils der identifizierten Nullsubjekte um circa 28 % führt, da die ersten 50 % der Regeln bereits einen erheblichen Anteil der Nullsubjekte identifizieren. Dies könnte darauf hindeuten, dass die Identifizierungsfähigkeit der verbleibenden 50 % der Regeln im Vergleich zu den ersten 50 % nicht proportional ansteigt, was auf eine gewisse Sättigung oder abnehmende Qualität bei der Identifizierung der Nullsubjekte hindeuten könnte.

Die separate Berechnung der Genauigkeit von insgesamt vier Teilabschnitten zu je ca. 25 % bestätigt, dass die Genauigkeit der ersten beiden Abschnitte, die zu den oberen 50 % der Regeln in NRS gehören, höher ist als die der letzten beiden Abschnitte, die zu den unteren 50 % der Regeln in NRS gehören.

Abschnitt	Anzahl korrekter Vorhersagen	Anzahl identifizierter Fälle	Genauigkeit
1. Regel 1–43 (–25,4 %)	4.097	4.158	0,985
2. Regel 44–85 (26,03–50,39 %)	4.610	4.656	0,990
3. Regel 86–127 (50,88–75,15 %)	1.609	1.702	0,945
4. Regel 128–169 (75,74–100 %)	1.472	1.649	0,893

Tabelle 35: Regelgruppierung und entsprechende Genauigkeiten

Tabelle 35 veranschaulicht eine Veränderung in der Resolutionsqualität zwischen den Regelabschnitten. Der Vergleich zeigt, dass die Genauigkeit in Abschnitt 2 höher ist als in Abschnitt 3 und die Genauigkeit in Abschnitt 3 wiederum höher liegt als in Abschnitt 4. Dies könnte auf die Anwendungsreihenfolge der Regeln zurückzuführen sein, die auf eindeutigen Indikatoren für bestimmte Referenten basieren, bevor Regeln mit weniger klaren Indikatoren in NRS angewandt werden. Die Herausforderung besteht jedoch darin zu erklären, warum die Genauigkeit der Regeln in Abschnitt 1, die zuerst im NRS angewandt werden, geringer ist als die in Abschnitt 2.

In Abschnitt 1 der 43 Regeln basieren die ersten sechs Regeln auf dem Auftreten von Imperativ- oder Exhortativmarkern sowie spezifischen Ausdrücken, die einen Vorschlag an den Adressaten signalisieren. Die verbleibenden 37 Regeln dienen der Resolution von Nullsubjekten in Nebensätzen, die in dieser Studie neu eingeführt wurden (siehe Beispiel (55) in Kapitel 5.1.2 und Beispiel (66) in Kapitel 5.1.3). Die Platzierung von Regeln, die Nullsubjekte in Nebensätzen rekonstruieren können, im Bereich der ersten 25 % der Regeln erfolgt, um potenzielle Konflikte mit den Regeln zur Resolution von Nullsubjekten im Hauptsatz zu vermeiden. Dabei werden bevorzugt Regeln berücksichtigt, die auf Nebensatzprädikate angewandt werden, um Interferenzen mit den entsprechenden Regeln für Hauptsatzprädikate zu verhindern. Dieser Ansatz beruht auf den Ähnlichkeiten in der Verwendung linguistischer Mittel wie Verben des Empfindens, Denkens und Fühlens.

Die Analyse bestätigt, dass die 37 Regeln zur Resolution von Nullsubjekten in Nebensätzen die Gesamtgenauigkeit von Abschnitt 1 verringern. Die ersten sechs Regeln weisen einen Genauigkeitswert von 0,99 auf, während die 37 Regeln insgesamt einen Genauigkeitswert von 0,97 erreichen. Zudem identifizieren die ersten sechs Regeln etwa 1,7-mal mehr Nullsubjekte, indem sie 2.600 Nullsubjekte im Vergleich zu 1.558 Nullsubjekten identifizieren. Dies unterstreicht die Effizienz der ersten sechs Regeln nicht nur bei der Identifizierung, sondern auch bei der präzisen Resolution von diskursteilnehmerbezogenen Nullsubjekten.

Die Genauigkeitswerte der Regeln zeigen einen Abfall von Abschnitt 2 zu 3 und 4, die mit den unterschiedlichen linguistischen Mitteln, die für die Zuweisung genutzt wurden, erklärt werden können. Die Regeln in Abschnitt 2 stützen sich hauptsächlich auf Verbalsuffixe und periphrastische Konstruktionen, die die sprecher-adressaten-orientierte Modalität ausdrücken können (siehe Beispiel (56) und (57) in Kapitel 5.1.2 und Beispiel (64)c in Kapitel 5.1.3), sowie auf feste Wendungen, die üblicherweise verwendet werden, um einen Adressaten zu etwas aufzufordern (siehe Beispiel (64)b in Kapitel 5.1.3). Die meisten Regeln in Abschnitt 3 basieren auf performativen Verben, die verwendet werden, wenn der Sprecher dem Adressaten Befehle oder Aufforderungen erteilt (vgl. Kapitel 4.2.3.1), sowie auf psychologischen Verben, die das Empfinden, Denken und Fühlen des Sprechers oder Adressaten ausdrücken können (siehe Beispiel (58)a in Kapitel 5.1.2). Die Regeln in Abschnitt 4 stützen sich hauptsächlich auf periphrastische Konstruktionen, die eine Erlaubnis oder Möglichkeiten ausdrücken, oder sie enthalten Hilfsverben oder Infixe (siehe Beispiel (58)b und (58)c in Kapitel 5.1.2). Bei den Regeln in Abschnitt 4 handelt es sich um zur Nullsubjektresolution neu eingeführte

linguistische Mittel, die in früheren Studien nicht explizit genutzt wurden. Diese Regeln wurden in der Annahme formuliert, dass ihre Kombination mit anderen Informationen, wie z. B. dem Satzmodus, hilft, Nullsubjekte im Zusammenhang mit Diskursteilnehmer:innen zu rekonstruieren. Obwohl sie bei der Wiederherstellung der Referenten also hilfreich sein können, beruhen diese Hinweise nicht auf soliden Indikatoren. Daher sind sie für die Anwendung im letzten Teil des NRS vorgesehen. Die Genauigkeit dieser Regeln kann niedriger sein als die Genauigkeit der vorangehenden Regeln. Die obige Analyse hat bestätigt, dass die Regeln in Abschnitt 4 den niedrigsten Genauigkeitswert der vier Abschnitte aufweisen. Doch selbst in diesem Abschnitt liegt die Genauigkeit immer noch bei 89,3 %, was auf eine insgesamt solide Resolutionsleistung des Regelwerks hindeutet.

Tabelle 35 zeigt auch, dass die Anzahl der identifizierten Fälle in den Abschnitten 2, 3 und 4 abnimmt, wobei Abschnitt 2 die höchste Anzahl identifizierter Fälle aufweist, nämlich 4.656. Dies lässt darauf schließen, dass die Regeln in Abschnitt 2 im Vergleich zu den Regeln in den anderen Abschnitten eine besonders starke Identifizierungsfähigkeit für die Nullsubjekte in den Analysedaten haben. Im Vergleich dazu sinkt die Anzahl der Fälle, die von den Regeln in Abschnitt 3 und Abschnitt 4 identifiziert werden, um mehr als die Hälfte (1.702 bzw. 1.649). Dies deutet darauf hin, dass die Regeln in den Abschnitten 3 und 4 im Vergleich zu den Regeln in den Abschnitten 1 und 2 eine deutlich geringere Identifizierungsfähigkeit für die Nullsubjekte in den Analysedaten aufweisen. Beachtlich ist jedoch, dass rund 95 % bzw. 90 % dieser Fälle als korrekte Vorhersagen bewertet werden. Dies bedeutet, dass die zugeordneten Referenten unter Berücksichtigung ihrer Identifizierungsfähigkeit zu einem sehr hohen Prozentsatz korrekt sind.

In Abbildung 40 fallen im blauen Bereich markante Sprünge auf. Ein deutlicher Anstieg im blauen Teil der Grafik zeigt an, dass die Anzahl der korrekten Vorhersagen rasch ansteigt, sobald eine bestimmte Regel angewandt wird. Dieser Anstieg ist nicht nur durch die korrekte Wiederherstellung der Referenten der Nullsubjekte gekennzeichnet, sondern auch durch die erfolgreiche Identifizierung einer erheblichen Anzahl von Nullsubjekten. Den größten Einfluss auf den Anstieg korrekter Vorhersagen haben fünf Regeln, die im Folgenden näher untersucht werden. Diese Regeln (Regel 1, 45, 54, 69, und 77) gehören zu den ersten 50 % der insgesamt 169 Regeln, die auf die Analysedaten angewandt wurden. Dies deutet darauf hin, dass es sich höchstwahrscheinlich um Regeln handelt, die aufgrund relativ klarer Hinweise formuliert wurden. In der folgenden Tabelle sind diese fünf Regeln aufgeführt.

Rang	Regel	Zugewiesener Referent	Genauigkeit	
1	Regel 1	impe ("rule_order": 1)	"2sh+"	0,996
2	Regel 54	want verbal suffix ("rule_order": 64)	"1sh+"	0,996
3	Regel 77	suggestory verbal suffix ("rule_order": 92)	"1sh+"	0,993
4	Regel 69	want verbal suffix ("rule_order": 80)	"1sh+"	1,000
5	Regel 45	request ("rule_order": 55)	"2sh+"	0,998

Tabelle 36: Genauigkeit der fünf einflussreichsten Regeln auf die korrekten Vorhersagen

Die erste Regel in Tabelle 36 zeigt den markantesten Anstieg der korrekten Vorhersagen im NRS. Dies ist deutlich erkennbar an der steilen Anfangsneigung der Kurve in Abbildung 40. Diese Regel beruht auf Imperativmarkern, die dem Referenten die Adressatenrolle zuzuweisen, und erzielt den höchsten Zuwachs an korrekten Vorhersagen. Regel 1 allein konnte mit einer Genauigkeit von 0,996 rund 20 %

der durch die 169 angewandten Regeln insgesamt identifizierten Nullsubjekte identifizieren, was im Vergleich zur Identifizierungsfähigkeit und zum Zuwachs an richtigen Vorhersagen aller anderen Regeln eine unvergleichliche Leistung darstellt. Dieses Ergebnis unterstreicht die Effektivität des Imperativmarkers für die Resolution von adressatenbezogenen Nullsubjekten, was in früheren Studien häufig als ein starkes Indiz für ihre Resolution angesehen wurde.

Die Regeln auf den Plätzen 2–4 fokussieren auf Verbalsuffixe, wobei Regel 54 und 69 auf Prädikate in Deklarativsätzen mit Willensäußerung des Sprechers angewandt werden. Diese beiden Regeln wurden unabhängig voneinander formuliert, obwohl bei ihrer Erstellung ähnliche linguistische Mittel zum Einsatz kommen. Grund dafür ist, dass Regel 54 auf dem neu eingeführten Verbalsuffix *gessseunbnida* basiert, das speziell für die vorliegende Studie herangezogen wurde. Im Gegensatz dazu basiert Regel 69 auf den bereits in vorherigen Studien betrachteten Verbalsuffixen wie *eulggeo* oder *eullayo*, die als klare Indikatoren für einen Referenten in der Sprecherrolle gelten. Die Anzahl der identifizierten Fälle ist beim neu eingeführten Verbalsuffix *gessseunbnida* in Regel 54 höher als bei den bereits häufig analysierten Verbalsuffixen in Regel 69 (1.050 vs. 489). Es muss angemerkt werden, dass alle identifizierten Fälle von Regel 69 als korrekt bewertet wurden, was zu einer Genauigkeit von 1,0 führt, während die Genauigkeit bei Regel 54 0,996 beträgt. Dies verdeutlicht, dass die Verbalsuffixe, die bei der Formulierung beider Regeln verwendet wurden, äußerst zuverlässige Indikatoren für die Referenten in der Sprecherrolle darstellen.

Regel 77 auf dem 3. Platz basiert auf einer Variante des Verbalsuffixes *eulkkayo*. Diese Regel zielt auf ein Prädikat mit diesem Verbalsuffix in Interrogativsätzen, dem ein Referent in der Sprecherrolle zugewiesen wird. Das Verbalsuffix wurde in dieser Studie integriert, da es typischerweise in Interrogativsätzen verwendet wird, um den Adressaten zu gemeinsamen Aktivitäten einzuladen oder um nach seiner Meinung zu einer bestimmten Handlung des Sprechers zu fragen (vgl. Kang Hyoun Hwa et al. 2016: 372–373). Die Anzahl der durch diese Regel identifizierten Fälle beträgt 612, von denen 608 korrekte Vorhersagen sind, was einer Genauigkeit dieser Regel von 0,993 entspricht. Lediglich vier Fälle wurden als falsche Vorhersage dieser Regel bewertet. Der Grund hierfür liegt darin, dass es keine Beschränkung auf ein bestimmtes Subjekt gibt, wenn dieses Verbalsuffix verwendet wird, um den Adressaten nach seiner Meinung zur Vermutung des Sprechers zu fragen (vgl. Kang Hyoun Hwa et al. 2016: 371). Trotz dieser Ausnahme ordnet die Regel den Referenten die Sprecherrolle in etwa 99 % der Fälle korrekt zu, was damit zu erklären ist, dass die Prädikate mit diesem Verbalsuffix in den Analysedaten hauptsächlich mit einem Subjekt in der Sprecherposition verwendet werden.

Die fünfte Regel, die zu einem signifikanten Anstieg der Anzahl richtiger Vorhersagen in NRS beiträgt, soll Prädikate identifizieren, die feste Wendungen in Fragesätzen aufweisen, um einen Referenten in der Adressatenrolle zu bestimmen. Diese Regel macht sich dabei die Präferenz der Sprecher zunutze, bei höflichen Bitten an den Adressaten konventionell spezifische Ausdrücke in einer Frageform zu verwenden. Insgesamt konnten durch die Anwendung dieser Regel 478 Prädikate identifiziert werden, wobei 477 von ihnen korrekt vorhergesagt wurden,⁹⁹ was einer beeindruckenden Genauigkeit von 0,998 entspricht. Die nahezu fehlerfreie Vorhersage durch diese Regel verdeutlicht die hohe Leistung der Regel basierend auf den festen Wendungen zur Resolution adressatenbezogener Nullsubjekte in den Analysedaten.

Die fünf Regeln in Tabelle 36 allein identifizieren etwa 42 % der Nullsubjekte, die durch die Anwendung aller 169 Regeln identifiziert wurden, und spielen eine entscheidende Rolle bei der Steigerung der Gesamtleistung des NRS, indem sie für rund 43 % der korrekten Vorhersagen insgesamt verantwortlich sind. Es ist bemerkenswert, dass zwei dieser fünf Regeln (Regel 54 und 77) auf einer Kombination linguistischer Merkmale basieren, die zuvor nicht für die Resolution koreanischer Nullsubjekte in Betracht gezogen wurde. Diese Ergebnisse zeigen, dass sowohl die etablierten als auch die neu eingeführten linguistischen Merkmale eine hohe Identifizierungsfähigkeit besitzen und durch ihre

⁹⁹ Das Nullsubjekt in der einzigen Fehlvorhersage dieser Regel bezieht sich zwar auf den Sprecher, kann jedoch auch ohne Kontext als Adressat interpretiert werden.

höheren Genauigkeitswerte starke Hinweise auf die Identifizierung und korrekte Resolution diskursteilnehmerbezogener Nullsubjekte im Koreanischen liefern.

Die Analyse der Identifizierungs- und Resolutionsleistung der Regeln im NRS lassen sich dahingehend interpretieren, dass sowohl die zuvor oft erwähnten Hinweise für eine erfolgreiche Resolution von Nullsubjekten wie Imperativ- oder Exhortativmarker sowie feste Wendungen für indirekte Aufforderungen oder einige Verbalsuffixe als auch die in dieser Studie neu eingeführte Kombination linguistischer Mittel eine bemerkenswerte Leistung bei der Zuweisung des richtigen Referenten zeigen. Obwohl die Identifizierungsfähigkeit der Regeln mit den neu eingeführten linguistischen Mitteln in dieser Studie über die Analysedaten etwas geringer ist, hat sich die Methode, dass zuerst Regeln mit konkreten Hinweisen für bestimmte Referenten angewandt wurden, als wirksam erwiesen, die Leistung des gesamten NRS zu gewährleisten. Dies zeigt sich in der vergleichsweise geringeren Genauigkeit der später im NRS angewandten Regeln im Vergleich zu den zuerst angewandten Regeln.

5.2.2. Leistungsvergleich der beiden Nullsubjektresolutionssysteme

5.2.2.1. Hintergründe und Leistungsanalyse des Basissystems

Hong Munpyo (2011) verwendet in seiner Analyse zur Resolution koreanischer Nullsubjekte drei Resolutionsstrategien entsprechend den Typen Extra-, Intra- und Inter-NS, sowie eine zusätzliche Default-Strategie. Die Untersuchung basiert auf der Annahme, dass die Nullsubjektresolution im Koreanischen die Leistung des maschinellen Übersetzungssystems vom Koreanischen ins Englische oder Deutsche verbessert. Es wird vorgeschlagen, einen Resolutionsprozess anzuwenden, bei dem die Resolutionsstrategie für ein Nullsubjekt im Eingabesatz von der Umgebung, in der das Nullsubjekt gefunden wird, und nicht von seinem Typ abhängt. Dieser Vorschlag ergibt sich aus der Tatsache, dass der Typ des Nullsubjekts im Eingabesatz in einem maschinellen Übersetzungsprozess nicht bekannt sein kann (siehe Kapitel 2.2.1).

Die Implementierung eines Resolutionssystems basierend auf den von Hong Munpyo vorgeschlagenen Strategien und Prozessen birgt verschiedene Herausforderungen. So erwähnt Hong Munpyo nicht explizit die Anwendungsreihenfolge der Regeln in Bezug auf die Strategie für den Typ Extra-NS, die auf linguistischen Mitteln im Verbalkomplex basiert, um diskursteilnehmerbezogene Referenten zuzuweisen. Die genaue Reihenfolge dieser Regeln ist allerdings entscheidend, da sie die Gesamtleistung des Systems beeinflussen kann. Hong Munpyos Resolutionsstrategie für den Typ Intra-NS hat Defizite. Ursprünglich war diese Strategie darauf ausgerichtet, das Nullsubjekt von Hauptsätzen (S2) zu rekonstruieren, wenn es durch eine von sieben Konjunktionen mit direkt vorausgehenden Nebensätzen (S1) verbunden ist. Dies sollte erfolgen, indem das explizite Subjekt von S1 als der Referent der Nullsubjekte in S2 unter Verwendung der Subject-Sharing-Merkmale der Konjunktion identifiziert wird. Jedoch kann diese Strategie das Nullsubjekt von S2 nicht rekonstruieren, wenn das Subjekt von S1 nicht realisiert ist. Bei der Umsetzung dieser Resolutionsstrategie in einem Nullsubjektresolutionssystem sollte daher darüber nachgedacht werden, wie die Eigenschaften der Konjunktion effektiver genutzt werden können, indem solche Fälle ausgeschlossen werden, in denen das Subjekt von S1 nicht realisiert wird. Hong Munpyos auf dem CF-Ranking der CT basierende Resolutionsstrategie für den Typ Inter-NS weist einige Erklärungslücken auf, die ihre Anwendung in einem Nullsubjektresolutionssystem erschweren. Es wird nicht erklärt, welche NP im CF-Ranking salienter ist: eine NP mit einem postpositionalen Suffix oder eine NP ohne Suffix. Die Beantwortung dieser Frage ist wichtig, da das Ranking der Antezedens-Kandidaten je nach Perspektive verändert werden kann und somit Einfluss auf die Auswahl des passendsten Antezedens-NP hat. Zudem wurde nicht spezifiziert, wie vorgegangen werden soll, wenn mehrere Kandidaten-NP denselben höchsten

Status im CF-Ranking aufweisen. Ohne ein Kriterium kann in diesem Szenario keine NP als Antezedens gewählt werden. Daher ist es für die Entwicklung eines Nullsubjektresolutionssystems erforderlich, klare Kriterien für die Bestimmung der salienten NP im CF-Ranking sowie für die Auswahl einer NP als Antezedens zu definieren, insbesondere bei der Anwendung der Resolutionsstrategie für den Typ Interns von Hong Munpyo. Die Verwendung von Satzzeichen in der Default-Strategie als Indikator für einen bestimmten Satztyp ist mitunter zu stark vereinfacht und beeinträchtigt im schlechtesten Fall die Leistung. Da Satzzeichen keinen bestimmten Satztyp gewährleisten, ist es notwendig, zusätzliche Informationen über den Satztyp zu berücksichtigen, um die Default-Strategie angemessen zu implementieren. Der Resolutionsprozess von Hong, bei dem je nach Kontext des Nullsubjekts im Eingabesatz eine Resolutionsstrategie angewandt wird, erschwert die präzise Bewertung der Leistung der einzelnen Strategien. Dies liegt daran, dass die Umgebung, in der die Nullsubjekte auftreten, nicht immer eindeutig ihren genauen Typ widerspiegelt.

Bei der Implementierung des BS wird in dieser Studie der Typ der Nullsubjekte, d. h. Extra-, Intra-, und Inter-NS direkt genutzt, um den Gegenstand der Resolution für die jeweilige Resolutionsstrategie zu identifizieren. Diese Vorgehensweise ist möglich, da diese Information bereits für die Eingabeprädikate des BS verfügbar ist, die auch von dem neu erarbeiteten NRS genutzt werden. Die Information über den Nullsubjekttyp wird in den annotierten Daten unter der Annotationsoption „anaphora_type“ mit den Werten „extra“, „intra“ bzw. „inter“ gespeichert (siehe Abbildung 10 in Kapitel 3.2.2). Da der Nullsubjekttyp für das BS bekannt ist, besteht das System aus insgesamt vier Resolutionsschritten, wie in der folgenden Abbildung dargestellt.

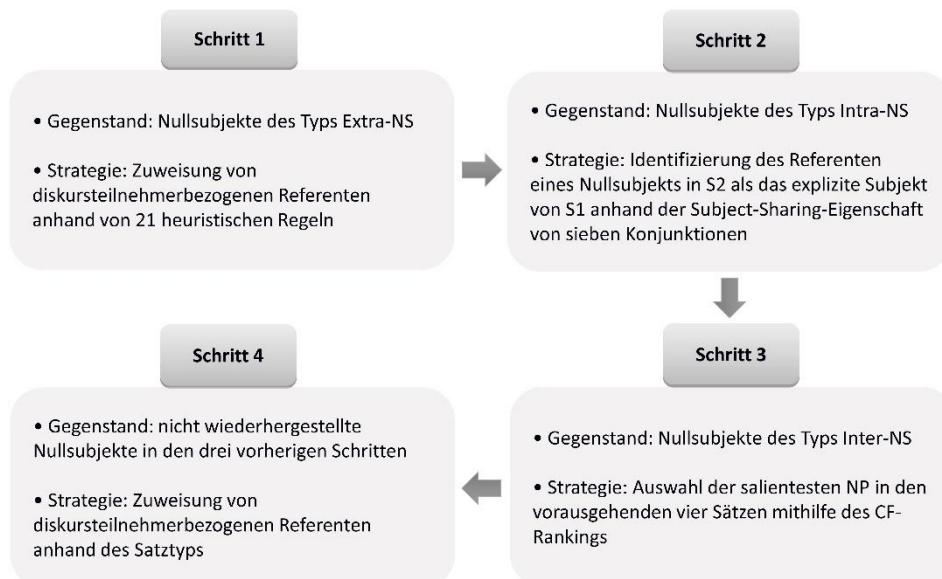


Abbildung 41: Vier Schritte zur Nullsubjektresolution im BS nach Nullsubjekttyp und Resolutionsstrategie

In den ersten drei Schritten des BS stehen die Input-Prädikate der Typen Extra-, Intra- oder Inter-NS im Zentrum der Resolution. Im letzten Schritt des BS wird die Default-Strategie auf die verbleibenden Fälle angewandt, die in den vorherigen Schritten nicht identifiziert werden konnten. Durch diese Implementierung des BS wird erwartet, dass die jeweilige Resolutionsstrategie genauer bewertet werden kann, indem die Effektivität der verschiedenen Resolutionsstrategien für die jeweiligen Typen von Nullsubjekten analysiert wird.

In einem ersten Schritt der Implementierung des BS werden 21 passende heuristische Regeln aus dem NRS übernommen,¹⁰⁰ weil Hong Munpyo nur solche linguistischen Mittel einbezieht, die in früheren

¹⁰⁰ Der Code für die 21 Regeln für den ersten Schritt des BS befindet sich in Appendix 2.

Studien häufig als starke Indikatoren für die Resolution von diskursteilnehmerbezogenen Nullsubjekten angesehen wurden, wie z.B. Imperativ- oder Exhortativmarker und subjektive Adjektive, die bereits bei der Erstellung der Regeln des NRS berücksichtigt wurden. Der Vorteil der Verwendung dieses Teils der Regeln aus dem NRS liegt darin, dass die Reihenfolge der Regelanwendung aus dem NRS wiederverwendet werden kann. Dabei werden Regeln mit eindeutigen Hinweisen für die Referenten der Sprecher- oder Adressatenrolle priorisiert gegenüber anderen Regeln mit weniger eindeutigen Hinweisen. Bei der Anwendung dieser Regeln im ersten Schritt des BS werden Fälle mit Übereinstimmungen in der grammatischen Person als korrekte Vorhersagen der Regeln gewertet, ähnlich wie bei der Auswertung der Regeln im NRS (vgl. Kapitel 5.1.1).

Im zweiten Schritt des BS sind Fälle des Typs Intra-NS der Resolutionsgegenstand. Die Annotation eines Nullsubjektes als Typ Intra-NS bedeutet, dass das Subjekt im direkt vorausgehenden Satz (S1) vor dem Satz mit dem Nullsubjekt (S2) explizit realisiert ist und als Antezedens des Nullsubjekts fungiert. Unter den zusammengesetzten Sätzen mit Intra-NS werden die Fälle extrahiert, in denen S1 eine der sieben Konjunktionen mit Subject-Sharing-Eigenschaften enthält. Auf diese Weise kann das explizite Subjekt in S1 als Referent des Nullsubjekts in S2 anhand der Subject-Sharing-Eigenschaft der Konjunktion wiedergegeben werden. Anschließend erfolgt die Leistungsbewertung der Resolutionsstrategie des zweiten Schritts durch manuelle Überprüfung, ob die Subjekte in S1 mit den annotierten Antezedenzen der Nullsubjekte in S2 identisch sind.

Das Ziel im dritten Schritt des BS ist die Resolution der Nullsubjekte des Typs Inter-NS. Diese Nullsubjekte sind bereits in den annotierten Daten mit ihren Antezedenzen annotiert. Die Resolutionsstrategie in diesem Schritt wählt aus den potenziellen Antezedens-Kandidaten in den vorherigen vier Sätzen die passendste Antezedens-NP auf Grundlage eines CF-Rankings aus. Bei der Bewertung der Resolutionsstrategie wird die ausgewählte NP, die am salientesten wirkt, mit dem annotierten Antezedens des Nullsubjekts verglichen. Die Implementierung und Bewertung des dritten Schritts im BS erfolgten manuell, da die Informationen aus den Zusatzdaten allein nicht präzise genug waren, um alle für die Resolutionsstrategie notwendigen Prozesse automatisieren zu können. Bei der Zählung der vier vorausgehenden Sätze vor den Nullsubjekten wurden kurze Antworten wie *ne* 'ja' oder *anio* 'nein' ausgeschlossen. Zur Bestimmung der syntaktischen Funktion der NP-Kandidaten wurden Informationen zur syntaktischen Funktion aus den Zusatzdaten von ETRI verwendet. Falls die syntaktische Funktion aus den Daten nicht korrekt war, wurde die tatsächliche syntaktische Funktion berücksichtigt. Wenn eine NP den Topik-Marker *eun/neun* enthielt, wurde sie unabhängig von der syntaktischen Funktion aus den Zusatzdaten als Topik-NP betrachtet. Da die im BS verwendeten CF-Rankings auf den von den Zusatzdaten bereitgestellten syntaktischen Funktionen basieren, erfolgt die Rangfolge wie folgt: [Topik-NP > Subjekt-NP > Direktes Objekt-NP > Indirektes Objekt-NP > Komplement-NP > Modifikator-NP]. Wenn es mehr als zwei NP-Kandidaten mit dem höchsten CF-Ranking gibt, wird derjenige NP-Kandidat als Antezedens ausgewählt, der näher am Nullsubjekt liegt, was Erkenntnissen aus früheren Studien folgt, bei denen die näher am Nullsubjekt liegende NP am ehesten als Antezedens dient (vgl. Kapitel 2.1.3.1).

Schließlich wird bei der Implementierung der Default-Strategie im letzten Schritt des BS die Satztypinformation aus den Zusatzdaten verwendet, um den Satztyp des betreffenden Satzes zu bestimmen, anstatt wie in Hong Munpyo (2011) auf Satzzeichen zurückzugreifen. Von allen Fällen, die in den vorangegangenen drei Schritten des BS nicht erfasst werden konnten, können in diesem Schritt nur Prädikate, die in Hauptsätzen gefunden wurden, als Resolutionsgegenstand dienen. Dies liegt daran, dass Prädikate in Nebensätzen keine Information über den Satztyp enthalten, da die Satztypinformation im Koreanischen in der Regel durch Verbalsuffixe am Satzende von Hauptsätzen gekennzeichnet ist (vgl. Lim Dong Hoon 2011: 323). Die Default-Strategie weist den Prädikaten in Deklarativsätzen den Referenten in der Sprecherrolle zu, während den Prädikaten in Imperativ- und Interrogativsätzen die Referenten in der Adressatenrolle zugewiesen werden. Bei der

Leistungsbewertung der Default-Strategie wird untersucht, ob der durch die Default-Strategie zugewiesene Referent mit dem annotierten Referenten des Nullsubjekts in den annotierten Daten übereinstimmt.

Anhand des BS konnten 12.894 Prädikate identifiziert werden, wobei 10.389 Nullsubjekte korrekt rekonstruiert wurden. Dies führt zu einem Genauigkeitswert von 0,81:

$$\text{Genauigkeit des BS} = \frac{10.389}{12.894} = 0,81$$

Neben des Genauigkeitswerts des BS ist auch die Genauigkeitswerte für jeden Schritt des BS von Interesse, weil diese Werte nicht direkt aus dem zuvor genannten Gesamtgenauigkeitswert des BS hervorgehen. Die folgende Tabelle bietet eine detaillierte Einsicht in die Genauigkeitswerte für jeden Schritt des BS und verdeutlicht, dass das BS in den verschiedenen Schritten unterschiedliche Genauigkeitswerte bzw. eine veränderliche Anzahl von identifizierten Nullsubjekten aufweist.

Schritt	Strategie	Anzahl korrekter Vorhersagen	Anzahl identifizierter Fälle	Genauigkeit
1.	Extra	3.919	3.949	0,99
2.	Intra	24	29	0,83
3.	Inter	687	1.143	0,60
4.	Default	5.759	7.773	0,74

Tabelle 37: Genauigkeitswert jedes Schritts des BS

Der erste Schritt, dessen Resolutionsziel Nullsubjekte des Typs Extra-NS sind, zeigt einen herausragenden Genauigkeitswert von 0,99, was eine nahezu fehlerfreie Vorhersage bedeutet. Dies könnte darauf hindeuten, dass die linguistischen Mittel, auf denen die Regeln basieren, zuverlässige Indikatoren sind, um die Nullsubjekte des Typs Extra-NS in den Analysedaten präzise zu rekonstruieren. Darüber hinaus wird erneut deutlich, dass die in früheren Studien intensiv diskutierten linguistischen Mittel tatsächlich einen wesentlichen Beitrag zur korrekten Resolution von diskursteilnehmerbezogenen Nullsubjekten leisten können. In diesem Schritt wurden 3.949 Nullsubjekte identifiziert, was etwa 31 % der insgesamt identifizierten Fälle im BS entspricht.

Im zweiten Schritt der Nullsubjektresolution im BS beträgt der Genauigkeitswert 0,83. Das bedeutet, dass die Subject-Sharing-Eigenschaft der Konjunktionen mit einer Wahrscheinlichkeit von etwa 83 % dazu beigetragen hat, die Nullsubjekte der folgenden Sätze korrekt zu rekonstruieren, indem die expliziten Subjekte der unmittelbar vorausgehenden Sätze als Referenten der Nullsubjekte angenommen wurden. Obwohl der Genauigkeitswert in diesem zweiten Schritt etwas niedriger ausfällt als im ersten Schritt, deutet er darauf hin, dass die Resolutionsstrategie in diesem Schritt des BS bei der Resolution von Nullsubjekten des Typs Intra-NS immer noch effektiv ist. Es ist jedoch zu beachten, dass die im Rahmen dieses Schritts identifizierten Nullsubjekte nur etwa 0,2 % der Gesamtzahl der identifizierten Fälle im BS ausmachen.

Der dritte Schritt der Nullsubjektresolution mithilfe des BS weist mit 0,60 den niedrigsten Genauigkeitswert der vier Schritte auf. Dies bedeutet, dass nur 60 % der salientesten NPs das Antezedenz der Nullsubjekte waren. In Anbetracht der Tatsache, dass die gleiche Resolutionsstrategie bei der Datengrundlage von Kim Mi-kyung (2003) für Volksmärchen und Zeitungstexte eine Erfolgsrate von ca. 77 % bzw. 51 % aufweist (vgl. Tabelle 6 in Kapitel 2.1.3.1), ist die Verwendung des CF-Rankings für die Resolution der Fälle des Typs Inter-NS auch bei den interaktiven Gesprächen in der simulierten Reisesituation nicht sehr erfolgreich. In diesem Schritt wurden etwa 9 % der identifizierten Fälle im BS abgedeckt.

Im letzten Schritt des BS beträgt der Genauigkeitswert 0,74, d. h. Vorhersagen der Default-Strategie anhand der Satztypinformation sind mit einer Wahrscheinlichkeit von 74 % korrekt. Angesichts der Tatsache, dass Hong Munpyos (2011) Default-Strategie auf Grundlage von Satzzeichen bei der Resolution von Nullsubjekten nicht effektiv war (siehe Tabelle 10 in Kapitel 2.2.1), führt die Verwendung der Satztypinformation anstelle der Satzzeichen zu einer besseren Leistung als die ursprüngliche Default-Strategie von Hong. Es ist bemerkenswert, dass etwa 74 % der identifizierten Nullsubjekte in diesem letzten Schritt des BS korrekt zugeordnet werden können.

5.2.2.2. Evaluation der Leistung beider Systeme im Vergleich

Bei der Bewertung der Leistung des BS und des NRS wird zunächst ein Vergleich der Gesamtleistung beider Systeme angestellt, gefolgt von einer Analyse ihrer Leistung in den Aspekten der Identifizierungs- und Resolutionsfähigkeit. Dabei zeigen sich deutliche Unterschiede in den Resolutionsstrategien.

Das BS weist einen Genauigkeitswert von 0,81 auf, da insgesamt 12.894 Nullsubjekte identifiziert wurden, von denen 10.389 Nullsubjekte ihren korrekten Referenten erhalten haben. Im Vergleich dazu liegt der Genauigkeitswert im NRS bei 0,97, da insgesamt 12.165 Nullsubjekte identifiziert wurden, von denen 11.788 Fälle als korrekte Vorhersagen bewertet wurden. Im Leistungsvergleich zwischen dem BS und dem NRS wurde die Leistung der 169 Regeln im NRS berücksichtigt, da diese direkt der Zuweisung eines Referenten in der Sprecher- oder Adressatenrolle dienten. Die Zahlen verdeutlichen, dass etwa 81 % der gesamten Input-Prädikate vom BS identifiziert werden konnten, während das NRS rund 77 % der gesamten Input-Prädikate identifizieren konnte. Das bedeutet, dass das BS circa 5 % der Nullsubjekten mehr identifiziert hat als das NRS. Obwohl die Anzahl der vom NRS identifizierten Nullsubjekte etwas geringer war als die vom BS, war der Gesamtgenauigkeitswert im NRS etwa 16 % höher als im BS. Dieser Vergleich legt nahe, dass die Resolution von Nullsubjekten in den Analysedaten unter Verwendung von Regeln, die lediglich auf der Kombination verschiedener linguistischer Mittel in Verbalkomplexen basieren, weniger Nullsubjekte identifizieren kann als die Anwendung verschiedener Resolutionsstrategien je nach Typ von Nullsubjekt. Dennoch zeigt die Tatsache, dass der Genauigkeitswert des NRS bei 0,97 liegt, dass die Nutzung von expliziten linguistischen Merkmalen in den Prädikaten allein ausreicht, um die identifizierten Nullsubjekte in den Analysedaten zu 97 % korrekt zu rekonstruieren.

Im Folgenden wird es untersucht, inwiefern das BS und das NRS die Nullsubjekte im Analysekörper nach Typen identifiziert haben und wie präzise die identifizierten Nullsubjekte in beiden Systemen korrekt rekonstruiert wurden. Die folgende Tabelle veranschaulicht den Anteil der identifizierten Nullsubjekte mithilfe von BS und NRS differenziert nach Nullsubjekttypen.

Nullsubjekttyp	BS	NRS	Total
1. Extra-NS	79,76 % (n=11.434)	82,88 % (n=11.881)	100,00 % (N=14.336)
	89,55 % (n=317)	45,20 % (n=160)	100,00 % (N=354)
	95,81 % (n=1.143)	10,39 % (n=124)	100,00 % (N=1.193)

Tabelle 38: Anteil der identifizierten Nullsubjekte nach Nullsubjekttyp in BS & NRS

In Bezug auf die Identifikation der Nullsubjekte mithilfe des BS zeigt sich, dass die Identifikationsrate über alle Typen zwischen ca. 80 % und 96 % liegt. Im Vergleich mit den Ergebnissen des NRS zeigt das BS eine herausragende Identifizierungsleistung für anaphorische Nullsubjekte in den Analysedaten, da die Nullsubjekte der Typen Intra- und Inter-NS als anaphorische Nullsubjekte eingestuft werden

können. Dabei identifiziert das BS Nullsubjekte des Typs Intra-NS etwa doppelt so häufig wie das NRS und solche des Typs Inter-NS sogar etwa neunmal so häufig. Selbst bei den Nullsubjekten des Typs Extra-NS beträgt der Unterschied in der Identifikationsrate zwischen dem BS und dem NRS lediglich 3 %. Mithilfe des NRS wurden hingegen vorwiegend Nullsubjekte des Typs Extra-NS identifiziert, und zwar zu 83 %. Das NRS weist eine geringe Identifizierungsfähigkeit für anaphorische Nullsubjekte auf, da über die Hälfte der Nullsubjekte des Typs Intra-NS und etwa 90 % der Nullsubjekte des Typs Inter-NS nicht identifiziert werden konnten. Dieses Ergebnis ist in gewissem Maße erwartbar, da das Hauptaugenmerk des NRS auf der Resolution diskursteilnehmerbezogener Nullsubjekte liegt. Darunter wurden solche Fälle gefasst, bei denen explizite Hinweise in den Verbalkomplexen das Hauptmotiv für die Nullform sein können, und die als Typ Extra-NS annotiert wurden.

Die Ergebnisse der Nullsubjektresolution mithilfe des NRS offenbaren jedoch Überraschungen, insbesondere in dem Kontext, dass Nullsubjekte der Typen Intra- und Inter-NS zu etwa 45 % bzw. 10 % allein durch die Verwendung von Hinweisen in den Verbalkomplexen identifiziert werden konnten. Die Hauptmotive für das Auftreten dieser Arten von Nullsubjekten sind ein gemeinsames Subjekt zwischen vorausgehendem und nachfolgendem Satz sowie die Realisierung der salientesten Diskursentität wie eines Topiks als Nullform, um die Kohärenz des Diskurses zu erhöhen (vgl. Kapitel 2.1.3.1). Die Beobachtung, dass 45 % der Fälle des Typs Intra-NS und 10 % der Nullsubjekte des Typs Inter-NS durch die Resolutionsstrategie des NRS identifiziert wurden, legt jedoch nahe, dass die Gründe für die erfolgreiche Resolution einiger dieser Nullsubjekttypen auch in den expliziten Hinweisen in den Verbalkomplexen liegen könnten, wenn die identifizierten Nullsubjekte korrekt rekonstruiert werden. Die folgende Tabelle zeigt die Resolutionsleistung des NRS nach Typen.

Nullsubjekttyp	Anzahl korrekte Resolution	Anzahl identifizierter Fälle	Genauigkeit
1. Extra-NS	11.656	11.881	0,98
2. Intra-NS	132	160	0,83
3. Inter-NS	0	124	0,00

Tabelle 39: Genauigkeitswert des NRS nach dem Nullsubjekttyp

Das NRS erreicht einen Genauigkeitswert von 0,98 für die Nullsubjekte des Typs Extra-NS, was die nahezu fehlerfreie Resolutionsfähigkeit des NRS für diesen Typ von Nullsubjekten unterstreicht. Die Ergebnisse zeigen, dass das NRS 83 % der Fälle des Typs Extra-NS mit einer Trefferquote von 98 % korrekt rekonstruiert, was darauf hindeutet, dass das NRS auf die Identifizierung und korrekte Resolution von Nullsubjekten des Typs Extra-NS spezialisiert ist. Die Stärke des NRS zeigt sich im Leistungsvergleich mit dem ersten Schritt des BS für die Nullsubjekte des Typs Extra-NS, bei dem das Hauptziel die Resolution dieses Typs ist. Im ersten Schritt des BS wurden 19 von 21 Regeln angewandt, um den Referenten der Diskursteilnehmerrolle zuzuweisen.¹⁰¹ Dadurch wurden 3.949 Nullsubjekte identifiziert, was lediglich etwa 28 % der 14.336 Nullsubjekte des Typs Extra-NS entspricht. Im Vergleich dazu wurden im NRS 169 Regeln angewandt, um 11.881 Nullsubjekte des Typs Extra-NS zu identifizieren, was etwa 83 % des Typs Extra-NS ausmacht. Dies verdeutlicht, dass das NRS etwa neunmal so viele Regeln wie der BS einsetzt und dass das NRS mehr als dreimal so viele Nullsubjekte des Typs Extra-NS identifizieren kann wie der BS. In diesem Kontext wird erneut deutlich, dass die neu eingeführten linguistischen Mittel in dieser Studie eine wichtige Rolle für die Identifizierung von mehr Nullsubjekten im NRS spielen. Die Genauigkeitswerte des NRS und des BS liegen bei 0,98 bzw. 0,99 für die Resolution der Nullsubjekte des Typs Extra-NS, was darauf hinweist, dass die von den angewandten

¹⁰¹ Die "rule_order" der drei nicht angewandten Regeln im BS lautet 6, 7 und 8. Diese Regeln ordnen Prädikaten einen Referenten in der Sprecherrolle zu, wenn sie linguistische Merkmale aufweisen, die Vermutungen, Urteile und Zweifel ausdrücken. Für die Zuordnung zur Art der jeweiligen Regel siehe Appendix 2 in der Spalte "rule_order".

Regeln in beiden Systemen genutzten Hinweise für die Referenten in der Sprecher- oder Adressatenrolle äußerst zuverlässige Indikatoren für die korrekte Resolution der diskursteilnehmerbezogenen Nullsubjekte sind. Dies verdeutlicht, dass nicht nur die in früheren Studien oft als starke Indikatoren für eine Resolution von Nullsubjekten mit Referenz auf Diskursteilnehmer:innen angesehenen Hinweise, sondern auch die Kombinationen linguistischer Mittel, die in dieser Arbeit neu eingeführt wurden, eine wichtige Rolle für die erfolgreiche Resolution von diskursteilnehmerbezogenen Nullsubjekten spielen.

Das NRS zeigt eine solide Leistung bei der Resolution von Nullsubjekten des Typs Intra-NS, da es für diesen Typ von Nullsubjekten einen Genauigkeitswert von 0,83 aufweist. Die Beobachtung, dass 45 % der identifizierten Nullsubjekte des Typs Intra-NS mit einer Trefferquote von 83 % korrekt rekonstruiert wurden, kann dahingehend interpretiert werden, dass die erfolgreiche Resolution einiger Nullsubjekte des Typs Intra-NS tatsächlich auf die expliziten linguistischen Merkmale in den Verbalkomplexen zurückgeführt werden kann.

Bei den Nullsubjekten des Typs Inter-NS identifizierte das NRS lediglich 10 % der gesamten Nullsubjekte dieses Typs im Analysekörper und konnte keinen der identifizierten Fälle korrekt rekonstruieren. Dies deutet auf eine Schwäche des NRS hin, nämlich seine geringe Fähigkeit, anaphorische Nullsubjekte zu identifizieren und zu rekonstruieren. Dies lässt darauf schließen, dass im Koreanischen ein anaphorisches Nullsubjekt nicht erfolgreich anhand expliziter linguistischer Merkmale im Verbalkomplex identifiziert und rekonstruiert werden kann.

Eine detaillierte Analyse der vom BS genutzten Schritte zur Identifikation von Nullsubjekten sowie der Genauigkeitswerte in jedem Schritt nach Typen ermöglicht es festzustellen, welche Resolutionsstrategien zur Identifizierung bestimmter Nullsubjekttypen im BS beigetragen haben, und wie präzise die identifizierten Nullsubjekte in jedem Schritt rekonstruiert wurden.

Nullsubjekttyp	SchrittStrategie	Anzahl korrekte Resolution	Anzahl identifizierter Fälle	Genauigkeit
Extra-NS	Schritt 1_{Extra}	3.919	3.949	0,99
	Schritt 4 _{Default}	5.633	7.485	0,75
Intra-NS	Schritt 2_{Intra}	24	29	0,83
	Schritt 4 _{Default}	162	288	0,56
Inter-NS	Schritt 3_{Inter}	687	1.143	0,60

Tabelle 40: Genauigkeitswert des BS in jedem Schritt nach dem Nullsubjekttyp

Die obige Tabelle zeigt, dass Nullsubjekte der Typen Extra- und Intra-NS nicht ausschließlich in den ersten beiden Schritten identifiziert wurden, in denen das Hauptziel die Resolution ebendieser Typen war. Vielmehr wurden etwa 65 % der Fälle des Typs Extra-NS und etwa 91 % der Fälle des Typs Intra-NS im vierten Schritt mit der Default-Strategie identifiziert. Nullsubjekte des Typs Inter-NS hingegen wurden ausschließlich im dritten Schritt des BS identifiziert, der speziell auf die Resolution ebendieses Typs abzielt. Diese Beobachtung hebt die Rolle der Default-Strategie im vierten Schritt bei der Identifikation von Nullsubjekten der Typen Extra- und Intra-NS hervor. Zusätzlich zeigt sich, dass ausschließlich die Resolutionsstrategie im dritten Schritt, die auf dem CF-Ranking basiert, für die Identifikation von Fällen des Typs Inter-NS von Bedeutung ist. Diese Analyse legt nahe, dass es möglich ist, den Anteil identifizierter Nullsubjekte aus allen drei Typen durch eine gezielte Anwendung des CF-Rankings und der Default-Strategie zu steigern.

Passend dazu, dass im BS das Resolutionsziel des ersten Schritts die Nullsubjekte des Typs Extra-NS und des zweiten Schritts Nullsubjekte des Typs Intra-NS sind, sind die Genauigkeitswerte auch in diesen Schritten höher als im vierten. Dies bedeutet, dass die Resolutionsstrategien im ersten und zweiten Schritt bei der Resolution der Fälle der Typen Extra- oder Intra-NS präziser als die Default-Strategie im

vierten Schritt funktionieren, auch wenn die Verwendung der Default-Strategie die Anzahl der identifizierten Nullsubjekte der Typen Extra- und Intra-NS erhöht. Interessant ist jedoch, dass der Genauigkeitswert der Default-Strategie für die Resolution der Nullsubjekte des Typs Extra-NS relativ höher ist als für die Resolution derjenigen des Typs Intra-NS (0,75 vs. 0,56). Bei den Inter-NS wurde der Genauigkeitswert im dritten Schritt als 0,60 berechnet, was zeigt, dass die Verwendung des CF-Rankings vergleichsweise ineffizient bei der Resolution der Nullsubjekte des Typs Inter-NS ist.

Hieraus lässt sich schlussfolgern, dass die Implementierung der Default-Strategie im BS dazu beitragen kann, mehr Nullsubjekte der Typen Extra- und Intra-NS zu identifizieren, jedoch auf Kosten der Gesamtgenauigkeit. Zusätzlich kann die Integration der Subject-Sharing-Eigenschaften von Konjunktionen dazu dienen, Intra-NS zu rekonstruieren, sie führt jedoch nicht zwangsläufig zu einer Erhöhung der Identifikationsrate dieser Nullsubjekte. Schließlich kann die Einbeziehung des CF-Rankings die Identifikationsrate von Fällen des Typs Inter-NS steigern, jedoch gleichzeitig zu einem erheblichen Rückgang der Gesamtgenauigkeit führen.

Im Folgenden wird die Effektivität der neu eingeführten Regeln im NRS zur Resolution von Nullsubjekten in Nebensätzen analysiert und mit der Leistung des BS verglichen. Diese Regeln im NRS beziehen sich auf eine in dieser Studie neu eingeführte Resolutionsstrategie für Nullsubjekte in Nebensätzen, die bestimmte linguistische Mittel zur Ausdrucksweise einer sprecherorientierten Modalität oder der Respektkategorie in Kombination mit bestimmten Konjunktionen einsetzt. Im Gegensatz dazu basiert das BS lediglich auf dem Konzept des Subject-Sharing von Konjunktionen, um Nullsubjekte in Nebensätzen zu rekonstruieren. Von den 208 heuristischen Regeln im NRS sind 105 in der Lage, Nullsubjekte in Nebensätzen zu identifizieren. Von diesen wurden 42 verwendet, um einen Referenten zuzuweisen. Die Leistung der Regeln zur Nullsubjektresolution in Nebensätzen wird durch die Identifikation von Nullsubjekten in Nebensätzen in den Input-Prädikaten bewertet. Die Informationen aus den Zusatzdaten geben Aufschluss darüber, ob ein Nullsubjekt in einem Haupt- oder Nebensatz steht, indem der Wert "root" für ein Prädikat in einem Hauptsatz und der Wert "clau" für ein Prädikat in einem Nebensatz angegeben wird. Insgesamt weisen 3.143 Prädikate den Wert "clau" auf, jedoch müssen nach manueller Prüfung 206 davon ausgeschlossen werden, da sie in Hauptsätzen auftreten. Somit werden insgesamt 2.937 Prädikate zur Leistungsbewertung der Regeln für die Resolution von Nullsubjekten in Nebensätzen herangezogen. Die folgende Tabelle veranschaulicht die Leistung des NRS im Vergleich zum BS in diesem Bereich.

System	Anzahl korrekte Resolution	Anzahl identifizierter Fälle	Genauigkeit
NRS	1.474	1.517	0,97
BS	151	251	0,60

Tabelle 41: Leistungsvergleich von NRS & BS bei der Nullsubjektresolution in Nebensätzen

Das NRS produziert einen höheren Genauigkeitswert von 0,97 als das BS mit einem Genauigkeitswert von 0,60. Darüber hinaus konnte das NRS sechsmal mehr Nullsubjekte identifizieren als das BS, da darüber 1.517 Nullsubjekte identifiziert wurden, im Vergleich zu 251 Nullsubjekten durch das BS. Dies unterstreicht die überlegene Leistung des NRS bei der Identifizierung und korrekten Resolution von Nullsubjekten in Nebensätzen in den Analysedaten.

Die Leistungsanalyse der Nullsubjektresolution in Nebensätzen zwischen dem NRS und dem BS verdeutlicht, dass das NRS eine fortschrittliche und verlässliche Resolutionsstrategie für die Resolution von Nullsubjekten in Nebensätzen nutzt, die die Einschränkungen des BS in Bezug auf eine geringe Identifizierungs- und Resolutionsfähigkeit für die Nullsubjekte in Nebensätzen übertrifft. Die Untersuchung offenbart, dass explizite Anhaltspunkte in Verbalkomplexen ebenso als überzeugende Indikatoren dienen können, um diskursteilnehmerbezogene Referenten in Nebensätzen angemessen zuzuweisen.

6. Schlussfolgerung

Kann das Phänomen der koreanischen Nullsubjekte nur durch den Topik-Drop-Parameter gekennzeichnet werden? Diese Frage bildete den Ausgangspunkt der vorliegenden Arbeit. Um ein vertieftes Verständnis des Phänomens der koreanischen Nullsubjekte zu erlangen, wurden in dieser Studie verschiedene potenzielle Parameter auf morphosyntaktischer und lexikalischer Ebene untersucht. Diese Parameter wurden als heuristische Regeln formuliert, um bestimmte Referenten zuzuweisen. Zudem wurden die Regeln als Nullsubjektresolutionssystem (NRS) integriert; ein Resolutionssystem mit den potenziellen Parametern zur Rekonstruktion koreanischer Nullsubjekte zu erstellen und seine Leistung an einer großen Anzahl authentischer Gespräche zu überprüfen, bietet eine wertvolle Möglichkeit, verschiedene theoretische Behauptungen und Beobachtungen empirisch zu belegen.

In dieser abschließenden Betrachtung werden die wichtigsten Ergebnisse und Erkenntnisse dieser Arbeit zusammengefasst. Hierbei geht es vor allem um die Frage, ob und inwieweit linguistische Merkmale in Verbalkomplexen eine wichtige Rolle bei der Resolution von Nullsubjekten spielen und wie effektiv die Resolutionsstrategie des NRS bei der Resolution des koreanischen Nullsubjekts im Vergleich zum Basissystem (BS) basierend auf den Resolutionsstrategien von Hong Munpyo (2011) ist (Kapitel 6.1). Abschließend werden zukünftige Forschungsrichtungen und praktische Implikationen der Ergebnisse der vorliegenden Arbeit diskutiert (Kapitel 6.2).

6.1. Fazit

Nullsubjekte sind im Koreanischen weit verbreitet und gelten in der gesprochenen Sprache als unmarkierte Option im Vergleich zu overten Subjekten (vgl. Kapitel 2.1.1). Im Gegensatz zu den prototypischen Pro-Drop-Sprachen wie Spanisch ist im Koreanischen der Wegfall von Pronomen auch ohne eine reiche Verbalflexion möglich, weshalb es als radikale Pro-Drop-Sprache klassifiziert wird. Während die explizite Subjekt-Verb-Kongruenz nicht nur Nullsubjekte in prototypischen Pro-Drop-Sprachen lizenziert, sondern auch ihre Referenten identifiziert, werden koreanische Nullsubjekte weithin als Diskurstopik identifiziert, indem der Topik-Drop-Parameter eingeführt wird. Die Klassifizierung koreanischer Nullsubjekte ausschließlich als Topik-Drop schränkt jedoch das Phänomen auf die Diskussion von Subjektauslassungen im Diskurskontext ein. Es gibt im Koreanischen jedoch eine weitere Art von Nullsubjekten, die nicht allein durch Topik-Drop erklärt werden kann, insbesondere wenn sie sich auf Sprecher oder Adressaten bezieht (vgl. Kapitel 1.1). Dies führt dazu, dass der Topik-Drop-Parameter für die Identifizierung von Referenten der Nullsubjekte in Bezug auf die Diskursteilnehmer:innen im Koreanischen an Bedeutung verliert.

Es stellt sich deshalb die Frage, anhand welcher anderer Parameter der Referent solcher Nullsubjekte im Koreanischen identifiziert werden kann. Die Hypothese dieser Studie besagt, dass linguistische Merkmale in Verbalkomplexen ein wichtiger Parameter für die Nullsubjektresolution sein können. Diese Hypothese basiert auf der Beobachtung, dass im Koreanischen der Satzmodus, der Sprechakt und die sprecher-adressaten-orientierte Modalität bis zu einem gewissem Maße durch linguistische Mittel in den Verbalkomplexen identifiziert werden können.

Um die Wirksamkeit linguistischer Merkmale in Verbalkomplexen bei der Rekonstruktion von Nullsubjekten in Bezug auf Sprecher oder Adressaten zu untersuchen, wurden zunächst verschiedene linguistische Merkmale in Verbalkomplexen untersucht. Diese wurden durch Literaturrecherche und empirische Befunde ermittelt und dienten als Hinweise auf einen Referenten in der Sprecher- oder

Adressatenrolle. Anschließend wurden die identifizierten linguistischen Merkmale in heuristische Regeln übertragen, die nach dem „wenn-dann“-Prinzip funktionieren. Hierbei bezieht sich die Bedingung auf spezifische Merkmale im Verbalkomplex, in Form von beispielsweise morphosyntaktischen Eigenschaften wie der Anwesenheit bestimmter Verbalsuffixe oder konventionalisierter Ausdrücke. Die Konsequenz besteht darin, dass spezifische Referenten in der Sprecher- oder Adressatenrolle einem Prädikat zugeordnet werden, um das Nullsubjekt des Prädikats zu rekonstruieren.

Im Koreanischen werden illokutionäre Indikatoren morphologisch realisiert, um Satztypen zu unterscheiden. Dadurch ist es möglich, auf der morphologischen Ebene zu erkennen, welche kommunikative Kraft der Sprecher mit seiner Äußerung dem Adressaten gegenüber ausdrückt. Diese Unterscheidung ermöglicht die Formulierung von Regeln, bei denen Verbalsuffixe wie die Imperativ- und Exhortativmarker *-(eu)seyo* oder *-bsida* verwendet werden. Die Verwendung von Imperativ- oder Exhortativmarkern in den Prädikaten kann für die Zuweisung von Referenten zur Adressaten- oder Sprecherrolle genutzt werden (vgl. Kapitel 4.1.1 und 4.2.1). Darüber hinaus wurde in Kapitel 4.1.2 festgestellt, dass bestimmte Ausdrücke in Interrogativsätzen wie z. B. *-ju(si)gesseoyo?* als Grundlage für eine heuristische Regel für die Zuweisung der Adressatenrolle dienen können, da sie konventionell verwendet werden, um höflich eine Aufforderung oder Bitte auszudrücken.

Linguistische Mittel wie das Honorativinfix *-(eu)si* und Honorativverben wie z. B. *deusida* 'essen.HON', die die Kategorie Respekt ausdrücken, wurden ebenfalls bei der Erstellung der Regeln genutzt, da sie bereits eine höfliche Einstellung gegenüber dem Gesprächspartner implizieren. Sie wurden daher verwendet, um heuristische Regeln für die Wiederherstellung der Referenten in der Adressatenrolle aufzustellen (vgl. Kapitel 4.1.3). Die linguistischen Mittel wie Hilfsverben zum Ausdruck von Bescheidenheit wie z. B. *-a/eo deulida* wurden wiederum dafür genutzt, heuristische Regeln für die Wiederherstellung von Referenten in der Sprecherrolle zu erstellen (vgl. Kapitel 4.2.3.3), da ein Sprecher durch diese linguistischen Mittel Respekt gegenüber seinen Gesprächspartnern zeigen kann, indem er bescheiden über sich selbst spricht (vgl. Beispiel (40) in Kapitel 4.2.3.3).

Lexikalische Indikatoren wie die Nutzung performativer und psychologischer Verben wurden ebenfalls bei der Erstellung der heuristischen Regeln berücksichtigt, da sie anzeigen, dass das Nullsubjekt jeweils als Sprecher oder Adressat interpretiert werden kann (vgl. Kapitel 4.2.3.1 und 4.2.3.2). Im Falle der performativen Verben wie z. B. *butaghada* 'biten' entspricht der Sprecher aus syntaktischer Sicht dem Subjekt, da er mit einem performativen Verb eine Bitte an einen Adressaten richtet (vgl. Beispiel (35) in Kapitel 4.2.3.1). Bei psychologischen Verben wie *haengboghada* 'glücklich sein' wurde die inhärente Subjektbeschränkung bei der Erstellung der heuristischen Regeln genutzt, da ihr Subjekt in Deklarativsätzen mit der Sprecherrolle und in Interrogativsätzen mit Adressatenrolle übereinstimmt (vgl. Beispiel (37) in Kapitel 4.2.3.2). Es wurde auch gezeigt, dass nicht nur ein Wechsel des Satzmodus von Deklarativ- zu Interrogativsätzen bei Prädikaten wie psychologischen Verben eine Reinterpretation der Referenten von Nullsubjekten vom Sprecher zum Adressaten bewirkt, sondern auch die Hinzufügung von Verbalinfixen wie *gess* oder *(eu)si* eine Rolle dabei spielt, eine Äußerung des Sprechers in eine Äußerung des Adressaten umzuwandeln (vgl. Beispiel (38) in Kapitel 4.2.3.2). In dieser Studie wurde die Tendenz herausgearbeitet, dass Nullsubjekte von Prädikaten bevorzugt als Adressaten interpretiert werden, wenn das Verbalprefix *gess* als Marker für eine Vermutung oder das Honorativinfix *(eu)si* hinzugefügt wird. Daher spiegelte sich der Einfluss des Satzmoduswechsels und der Hinzufügung von Verbalinfixen in der Erstellung der Regeln wider, was zu einem Wechsel von der Interpretation des Nullsubjekts als Sprecher zum Adressaten führte (vgl. Kapitel 4.2.3.2).

Einige Regeln wurden auf Basis verschiedener linguistischer Mittel auf morphosyntaktischer Ebene formuliert, wie z. B. Verbalsuffixe und periphrastische Konstruktionen, um die Referenten in der Sprecherrolle wiederherzustellen, da diese Mittel eine sprecherorientierte Modalität ausdrücken.

Verbalsuffixe wie z. B. *llaeyo* oder periphrastische Phrasen wie z. B. *lyeogo hada* dienten als Grundlage für einige heuristische Regeln, da sie den Sprecherwillen in Aussagen oder den Willen des Adressaten in Fragen ausdrücken (vgl. Kapitel 4.2.2).

Besonders die Berücksichtigung der Honorativverben, performativen Verben und periphrastischen Konstruktionen ist von Bedeutung, da diesen in der Forschung zu koreanischen Nullsubjekten bisher zu wenig Aufmerksamkeit geschenkt wurde. Insbesondere zeigten sich periphrastische Konstruktionen als vielversprechend für die Wiederherstellung der Referenten, da sie eine performative Kraft ähnlich zu einem voll grammatisierten Verbalsuffix haben können, um den Sprecherwillen auszudrücken. Zusätzlich können periphrastische Konstruktionen, wenn sie in den heuristischen Regeln berücksichtigt werden, insgesamt zu einer verbesserten Identifizierung von Nullsubjekten führen, da sie im Koreanischen häufig ähnlich wie Verbalinfixe verwendet werden, um eine spezifische grammatische Bedeutung auszudrücken (vgl. Tabelle 22 in Kapitel 4.2.2). Auch die Berücksichtigung der Kombination von lexikalischen Indikatoren und Verbalinfixen, die die Subjektbeschränkung aufheben, ist von Bedeutung, da in früheren Studien nur der Wechsel von der Zuweisung der Sprecher- zur Adressatenrolle beim Wechsel des Satzmodus erwähnt wurde.

Darüber hinaus wurden heuristische Regeln für die Resolution von Nullsubjekten in Nebensätzen formuliert, da Regeln auf Basis von Satzmodusinformationen keine Nullsubjekte über die Satzgrenze hinaus erfassen und somit auch solche Referenten nicht wiederherstellen können (vgl. Kapitel 4.3.1). Bei der Formulierung dieser Regeln wurden die morphosyntaktischen Merkmale in Verbalkomplexen von Nebensätzen berücksichtigt, wie beispielsweise die Kombination bestimmter Konjunktionen und linguistischer Mittel, deren Bedeutung grundsätzlich auf den Sprecher ausgerichtet ist. Diese heuristischen Regeln sind einfach zu implementieren, da sie eine morphologisch bedingte Zuweisung ermöglichen, ohne dass eine syntaktische Analyse des Nebensatzes erforderlich ist.

Die Leistungsanalyse der Nullsubjektresolution in Nebensätzen zwischen dem NRS und dem BS verdeutlichte zudem, dass explizite linguistische Merkmale in den Verbalkomplexen als überzeugende Indikatoren dienen können, um diskursteilnehmerbezogene Referenten in Nebensätzen angemessen zuzuweisen. Das BS wurde basierend auf den Resolutionsstrategien von Hong Munpyo (2011) implementiert (vgl. Kapitel 2.2.1). Ein direkter Leistungsvergleich seiner Resolutionsstrategien mit der Resolutionsstrategie des NRS gestaltete sich als herausfordernd, da die Art und Menge der verwendeten Daten zur Bewertung der Resolutionsstrategien unterschiedlich war (vgl. Kapitel 1.2). Das NRS verwendet eine Resolutionsstrategie, die sich ausschließlich auf linguistische Merkmale in Verbalkomplexen stützt, während das BS drei verschiedene Resolutionsstrategien je nach Typ des Nullsubjekts (in Form der Extra-NS, Intra-NS und Inter-NS sowie eine Default-Strategie) nutzt (vgl. Abbildung 41 in Kapitel 5.2.2.1). Die Leistung des NRS und des BS wurde deshalb bewertet, indem untersucht wurde, wie gut mithilfe beider Systeme 15.883 Nullsubjekte im gleichen Analysekörper rekonstruiert werden konnten. Dabei war zu beobachten, dass das NRS sechsmal mehr Nullsubjekte in Nebensätzen identifizieren konnte als das BS und dabei zusätzlich einen höheren Genauigkeitswert von 0,97 im Vergleich zum BS mit einem Genauigkeitswert von 0,60 erreichte (vgl. Tabelle 41 in Kapitel 5.2.2.2), obwohl das BS vier verschiedene Resolutionsstrategien zur Nullsubjektresolution nutzt.

Insgesamt wurden 208 Regeln zur Nullsubjektresolution entwickelt, die einen Referenten in der Sprecher- oder Adressatenrolle zuweisen können. In der vorliegenden Studie wurde ein regelbasierter Ansatz als vorteilhaft angesehen gegenüber nicht-deterministischer Ansätze zur Analyse der Nützlichkeit einzelner Parameter. Dieser Ansatz ermöglichte es, die Wirksamkeit bestimmter morphosyntaktischer und lexikalischer Hinweise in Verbalkomplexen auf bestimmte Referenten zur Nullsubjektresolution zu überprüfen. Um die Effektivität der entwickelten Regeln zu bewerten, wurden sie in das NRS integriert. Aufgrund der großen Anzahl von Regeln galten Konflikte zwischen den Regeln als wahrscheinlich, sie sollten aber vermieden werden, damit die Referenten korrekt zugeordnet

werden können. Daher wurde die Regelreihenfolge im NRS angepasst: Regeln, die auf weit verbreiteten grammatisierten linguistischen Mitteln wie Verbalsuffixen basieren, wurden relativ früher im NRS angewendet als Regeln, die auf linguistischen Mitteln mit einem niedrigeren Grammatikalisierungsgrad basieren wie beispielsweise periphrastische Konstruktionen. Eine detaillierte Analyse der Veränderung in der Resolutionsqualität zwischen den Regelabschnitten bestätigt die Beobachtung, dass die Regeln in relativ früheren Abschnitten einen höheren Genauigkeitswert aufweisen (vgl. Tabelle 35 in Kapitel 5.2.1). Dies könnte auf die Anwendungsreihenfolge der Regeln zurückzuführen sein, die auf eindeutigen Indikatoren für bestimmte Referenten basieren, bevor Regeln mit weniger klaren Indikatoren im NRS angewandt wurden.

Über die bereits erläuterten Regeln hinaus wurden vier ergänzende Regeln entwickelt und in das NRS integriert, um Prädikate auszuschließen, bei denen es unwahrscheinlich ist, dass ihre Nullsubjekte auf Diskursteilnehmer:innen verweisen (vgl. Kapitel 4.3.2). Diese Regeln wurden erstellt in der Erwartung, dass sie die Genauigkeit der anderen heuristischen Regeln erhöhen würden, da sie Prädikate wie Eigenschaftadjektive oder solche mit bestimmten Hilfsverben oder periphrastischen Konstruktionen identifizieren, die vom Resolutionsprozess ausgeschlossen werden konnten, wenn ihre Nullsubjekte eher als Referenten in der 3. Person interpretiert wurden. Die Analyse der Genauigkeitswerte des NRS zeigt, dass die Integration dieser vier zusätzlichen Regeln entgegen der ursprünglichen Annahme zu einem etwa 0,4 % niedrigeren Genauigkeitswert führt, was hauptsächlich auf die relativ geringe Genauigkeit dieser vier Regeln zurückzuführen ist, die einen Wert von 0,80 aufweisen. Trotz des Rückgangs der Gesamtgenauigkeit des NRS mit den zusätzlichen Regeln erreichten die Regeln zur Zuweisung der Referenten in der Sprecher- oder Adressatenrolle unter Einbezug der vier Ausschluss-Regeln einen Genauigkeitswert von 0,969, was etwa 1,5 % genauer ist im Vergleich zu den Vorhersagen ohne die zusätzlichen Regeln, bei denen die Genauigkeit bei 0,954 lag.

Die vorliegende Arbeit hat neben der Untersuchung potenzieller Parameter zur Rekonstruierbarkeit von Nullsubjekten im Koreanischen auch annotierte Daten generiert, die Informationen über Nullsubjekte und ihre Verweisstruktur enthalten. Dies erwies sich deshalb als notwendig, da keine geeigneten annotierten Daten verfügbar waren, die interaktive Alltagsgespräche abdeckten und somit viele Nullsubjekte in der Sprecher- oder Adressatenrolle enthalten könnten. Die Penn Korean Treebank war das einzige verfügbare Korpus mit annotierten Dialogsequenzen, jedoch basiert dieses Korpus auf militärischen Kontexten, die sich zu stark von alltäglichen Konversationen unterscheiden. Daher wurden die annotierten Daten basierend auf einem optimierten „zero mark-up“-Annotationsschema generiert, um die Herausforderungen bestehender Annotationsschemata bei der Identifizierung und Klassifizierung koreanischer Nullsubjekte zu überwinden. Mit dem optimierten Annotationsschema wurden drei Hauptfragen beantwortet:

1. Welche Elemente sollen annotiert werden?
2. Wie sollen diese Elemente annotiert werden?
3. Nach welchen Kriterien sollen diese Elemente klassifiziert werden?

Die erste Herausforderung bestand darin, dass es keine klare Richtlinie dafür gab, welche Elemente als Nullsubjekt annotiert werden sollten. Dies führte zu einer verringerten Zuverlässigkeit der Annotationsergebnisse in Bezug auf die Identifizierung von Nullsubjekten und beeinträchtigte folglich auch deren Genauigkeit. Die Hauptursache für die fehlende Identifizierung liegt in der Uneinheitlichkeit der grammatischen Beschreibungen darüber, was überhaupt als Nullsubjekt anzusehen ist. Wie in Kapitel 3.2.1 gezeigt wurde, führte diese Uneinheitlichkeit dazu, dass dieselben Sätze unterschiedlich klassifiziert wurden, entweder als Sätze mit Subjektellipse, subjektlose Konstruktionen oder sogar als Routineformeln (vgl. Beispiel (14) in Kapitel 3.2.1). In der vorliegenden Arbeit wurde diese Unterteilung nicht beachtet und stattdessen wurde der Fokus auf all diejenigen Fälle gelegt, in denen eine NP nicht in der Subjektposition eines Satzes realisiert wird und das implizite Subjekt durch ein overtes Subjekt ersetzt werden kann.

Die zweite Herausforderung bei der Annotation bezieht sich auf die Markierung von Nullsubjekten an bestimmten Stellen in den entsprechenden Sätzen. Im Koreanischen können Satzbestandteile, einschließlich der Subjekte, je nach Sprecherabsicht frei angeordnet werden. Dies macht die präzise Bestimmung der Position frei angeordneter Nullelemente bereits zu einer anspruchsvollen Aufgabe, was wiederum die Annotationsaufgabe komplexer gestaltet und mitunter zu Inkonsistenzen in den annotierten Daten führen kann. In dieser Studie wurde daher ein Annotationsschema vorgeschlagen, das die Annotation eines Verbalkomplexes ermöglicht, ohne dass das Nullsubjekt direkt an einer bestimmten Stelle im zugehörigen Satz eingefügt werden muss. Die reine Markierung sichtbarer Verbalkomplexe im vorgeschlagenen Annotationsschema zielt darauf ab, die Komplexität der Annotation zu verringern, indem Verbalkomplexe mit Nullsubjekten einfach gekennzeichnet werden können, ohne die exakte Position der Nullsubjekte im Auge behalten zu müssen. Es wurde angenommen, dass dieser Ansatz den Annotationsprozess beschleunigen wird, indem die Last der Annotation, die durch unsichtbare und flexible Elemente verursacht wird, reduziert wird.

Die letzte Herausforderung bei der Annotation bestand in der Kategorisierung der koreanischen Nullsubjekte in deiktische und anaphorische Typen, da diese Unterscheidung unscharf ist (vgl. Beispiel (13) in Kapitel 3.2.1). Anstatt also eine Zweiteilung der Nullsubjekttypen anzunehmen, folgte die vorliegende Studie der Einteilung von Hong Munpyo (2011) in drei Subtypen: Extra-NS, Inter-NS und Intra-NS (vgl. Kapitel 2.1.2). Es wurde erwartet, dass das in dieser Studie eingeführte Kategoriensystem im Vergleich zum binären Kategoriensystem für koreanische Nullsubjekte entscheidende Vorteile bietet. Dieses System zeichnet sich durch die Fähigkeit aus, die Motivationen hinter der Verwendung der Nullform im Koreanischen sowie das sprachliche Umfeld der Nullsubjekte präziser widerzuspiegeln. Ein bedeutender Fortschritt bestand darin, dass dieses Kategoriensystem den Nachteil der Unschärfe des bisherigen Systems überwinden konnte, indem es Nullsubjekten ausdrücklich erlaubt, lediglich einer Unterkategorie zugeordnet zu werden. Zudem ermöglicht die Übertragung der Ergebnisse aus dem neuen Kategoriensystem auf das binäre System eine verbesserte Vergleichbarkeit und Interpretation, insbesondere durch die klare Zuordnung von Nullsubjekten des Typs Extra-NS zu deiktischen Nullsubjekten und solchen des Typs Inter-NS oder Intra-NS zu anaphorischen Nullsubjekten.

Als passend für die Beantwortung der vorliegenden Forschungsfragen wurde das vom koreanischen Forschungsinstitut ETRI zur Verfügung gestellte Korpus ausgemacht (vgl. Kapitel 1.2), das zur Generierung der Annotationsdaten verwendet wurde (vgl. Kapitel 3.1.1). Die Datengrundlage besteht aus authentischen Gesprächen in simulierten Reiseszenarien und wurde speziell für die Entwicklung von Softwareanwendungen erstellt. Ein Subkorpus des Datensatzes mit insgesamt 29.506 Sätze in 2.349 Dialogen wurde exklusiv für diese Studie zur Verfügung gestellt, der im Folgenden annotiert werden konnte. Als Annotationstool wurde das Tool MMAX2 gewählt aufgrund seiner hohen Anpassungsfähigkeit und Flexibilität in Bezug auf Annotationsschemata. Die Annotation erfolgte gemäß dem in dieser Studie optimierten Annotationsschema. Wenn ein Subjekt eines Prädikats nicht realisiert war und durch ein overt Subjekt ersetzt werden konnte, wurde das Prädikat als ein neues Markable auf der "Zero"-Ebene im MMAX2-Projekt festgehalten. Auf dieser vordefinierten Ebene wurden alle Prädikate mit Nullsubjekten erfasst und gespeichert. Wenn ein Nullsubjekt eines Prädikats anaphorisch verwendet wurde und eine NP im Diskurs als sein Antezedenz fungierte, wurde das Nomen-Markable als sein Antezedenz markiert (vgl. Abbildung 9 in Kapitel 3.2.2). Wenn ein Nullsubjekt eines Prädikats in deiktischer Verwendung verwendet wurde, wurde sein Referent unter der Annotationsoption "referent_type" ausgewählt und gespeichert (vgl. Abbildung 10 in Kapitel 3.2.2). Als Ergebnis der Annotation wurden insgesamt 17.266 neue Markables auf der "Zero"-Ebene erstellt. In dieser Kohorte wurden 1.383 Prädikate von der Analyse manuell ausgeschlossen, entweder aufgrund von Fehlern oder weil sie als satzwertige Erweiterung von Satzgliedern dienten oder Satzglieder modifizierten, was nicht dem Fokus dieser Studie entsprach und somit keine Nullsubjekte

in finiten Sätzen darstellte. Zu Fehlern, die zum Ausschluss eines Prädikates führten, zählten Segmentierungs- und Parsingfehler sowie andere Ungenauigkeiten, die bereits in den Analysedaten vorhanden waren (vgl. Kapitel 3.3.2). Folglich verblieben insgesamt 15.883 Prädikate mit der gleichen Menge Nullsubjekte, die als Input für das NRS verwendet werden konnten.

Die annotierten Daten dieser Studie stellen eine wertvolle Ressource dar, um die erwähnte Annahme über koreanische Nullsubjekte zu überprüfen, ob ihre Verwendung in den linguistischen Merkmalen der Prädikate begründet liegt, die zur Interpretation der ausgelassenen Referenten führen können. Die Untersuchung der Verteilung der Anaphern-Typen in den annotierten Daten bestätigte, dass etwa 90 % der Nullsubjekte vom Typ Extra-NS am häufigsten auftraten, gefolgt von Inter-NS, während Nullsubjekte des Typs Intra-NS am seltensten zu finden waren (vgl. Tabelle 17 in Kapitel 3.3.1). Diese Beobachtung lässt darauf schließen, dass bis zu 90 % der Subjekte kein explizites Antezedens im Diskurs haben, sondern stattdessen Hinweise zur Identifizierung ihrer Referenten aufweisen, was wiederum ein Kriterium zur Identifizierung von Nullsubjekten in der Kategorie Extra-NS ist. Dies deutet darauf hin, dass bis zu 90 % der Nullsubjekte mit den entwickelten heuristischen Regeln erkannt werden können, da sie auf linguistischen Merkmalen basieren, die explizit in den Prädikaten vorhanden sind.

Die annotierten Daten dieser Arbeit wurden weiterhin zur Leistungsbewertung des NRS genutzt, indem die Vorhersagen der heuristischen Regeln im NRS mit den Referenten und Antezedenzien in den annotierten Daten verglichen wurden (vgl. Kapitel 5.1.1). Dabei wurden insgesamt 172 Regeln für die Zuweisung der Referenten in der Sprecher- oder Adressatenrolle angewandt, womit 12.083 Nullsubjekte in dieser Rolle identifiziert werden konnten, was etwa 86 % der im Korpus gefundenen diskursteilnehmerbezogenen Nullsubjekte entspricht. Diese Regeln wiesen einen beeindruckenden Genauigkeitswert von 0,99 auf, da sie bei 11.950 der 12.083 identifizierten Fälle den richtigen Referenten zuwiesen. Diese Ergebnisse bekräftigen die Hypothese der Arbeit, dass explizite linguistische Merkmale im Verbalkomplex eine entscheidende Rolle bei der Rekonstruierbarkeit von Nullsubjekten in Bezug auf die Diskursteilnehmer:innen im Koreanischen spielen.

Letztendlich bestand ein Forschungsinteresse darin, den Beitrag einzelner Parameter zur Zuweisungsgenauigkeit zu untersuchen, um die korrekte Rekonstruktion von Nullsubjekten zu unterstützen. Durch die Analyse der Leistung einzelner Regeln wurde festgestellt, welche Regel einen signifikanten Anstieg der korrekten Vorhersagen bewirkte. Diese Analyse lieferte wertvolle Einblicke in robuste Indikatoren für die präzise Rekonstruktion von Nullsubjekten mit Referenz auf Sprecher oder Adressaten. Es wurde vermutet, dass die linguistischen Merkmale, die diesen Regeln zugrunde liegen, Schlüsselparameter für die Rekonstruierbarkeit von diskursteilnehmerbezogenen Nullsubjekten im Koreanischen darstellen.

Ein wichtiges Ergebnis dieser Studie ist, dass die linguistischen Merkmale, die oft als entscheidend für die Rekonstruierbarkeit der Nullsubjekte diskutiert wurden, tatsächlich signifikante Parameter für die Resolution von Nullsubjekten in der Sprecher- oder Adressatenrolle sind. Dazu gehören Verbalsuffixe als Imperativmarker oder zum Ausdruck des Sprecherwillens, feste Wendungen für Aufforderungen und das Honorativ infix zur Kennzeichnung der Kategorie Respekt. Zudem wurde die wichtige Rolle periphrastischer Konstruktionen, die den Sprecherwillen ausdrücken oder eine modale Bedeutung im Sinne von Erlaubnis haben, hervorgehoben (vgl. Tabelle 30 in Kapitel 5.1.2 und Tabelle 33 in Kapitel 5.1.3). Diese Ergebnisse unterstreichen die Bedeutung der Verwendung periphrastischer Konstruktionen in den heuristischen Regeln, obwohl ihre Rolle in der Forschung zu koreanischen Nullsubjekten bisher wenig Aufmerksamkeit erhalten hat.

Die Effektivität der neu eingeführten linguistischen Merkmale zur Nullsubjektresolution in dieser Studie wurde durch einen Leistungsvergleich zwischen dem NRS und dem BS bestätigt. Bei der Implementierung des ersten Schritts des BS, der darauf abzielte, Nullsubjekte des Typs Extra-NS zu identifizieren, wurden 21 geeignete heuristische Regeln aus dem NRS übernommen. Dies geschah, weil Hong Munpyo (2011) nur solche linguistischen Merkmale in seiner eigenen Studie einbezieht, die bereits in früheren Studien als starke Indikatoren für die Resolution von diskursteilnehmerbezogenen Nullsubjekten betrachtet wurden, wie beispielsweise Imperativ- oder Exhortativmarkern sowie subjektive Adjektive, die bereits bei der Erstellung der Regeln des NRS berücksichtigt wurden. Von diesen 21 Regeln wurden 19 im BS angewandt, um den Referenten der Diskurteilnehmerrolle zuzuweisen. Dadurch konnten 3.949 Nullsubjekte identifiziert werden, was lediglich etwa 28 % der 14.336 Nullsubjekte des Typs Extra-NS in den Analysedaten entspricht. Im Vergleich dazu wurden im NRS 169 Regeln angewandt, um 11.881 Nullsubjekte des Typs Extra-NS zu identifizieren, was etwa 83 % aller Nullsubjekte dieses Typs ausmacht. Dies verdeutlicht, dass das NRS etwa neunmal so viele Regeln wie das BS einsetzt und dass das NRS mehr als dreimal so viele Nullsubjekte des Typs Extra-NS identifizieren konnte wie das BS. In diesem Kontext wurde deutlich, dass die neu eingeführten linguistischen Merkmale bei der Formulierung der heuristischen Regeln in dieser Studie eine wichtige Rolle bei der Identifizierung von mehr Nullsubjekten spielen. Sowohl das NRS als auch das BS erreichten einen Genauigkeitswert von 0,98 bzw. 0,99 für die Resolution von Nullsubjekten des Typs Extra-NS, was darauf hindeutet, dass die von den angewandten Regeln in beiden Systemen genutzten Hinweise auf die Referenten in der Sprecher- oder Adressatenrolle äußerst zuverlässige Indikatoren für die korrekte Rekonstruktion der diskursteilnehmerbezogenen Nullsubjekte sind. Dieses Ergebnis verdeutlicht den wichtigen Beitrag dieser Studie zur Aufdeckung der vielfältigen Parameter für die korrekte Bestimmung von Nullsubjekten in Bezug auf Diskurteilnehmer:innen, was insgesamt zu einem besseren Verständnis des weit verbreiteten Phänomens des Nullsubjekts im Koreanischen beiträgt.

Die Analyse der Leistung einzelner Regeln im NRS hebt die Rolle der morphologischen Parameter, insbesondere der Verbalsuffixe, hervor, die wesentlich zur Identifizierung und nahezu fehlerfreien Rekonstruktion von Nullsubjekten in Bezug auf Sprecher oder Adressaten beitragen. Von den sechs wichtigsten Zuweisungsregeln für die Sprecherrolle basieren drei auf Verbalsuffixen wie z. B. *eulgeyo* oder *eulkkayo*, die den Sprecherwillen in Aussagen reflektieren oder, im Fall von Fragen, die Meinung des Adressaten über die Handlungen des Sprechers erfragen (vgl. Tabelle 30 in Kapitel 5.1.2). Zwei der sieben wichtigsten Regeln zur Zuweisung der Adressatenrolle basieren ebenfalls auf Verbalsuffixen, wie z. B. *seyo* oder *gessseubnikka*, die als Imperativmarker fungieren oder in Interrogativsätzen verwendet werden, um den Willen des Adressaten zu erfragen (vgl. Tabelle 33 in Kapitel 5.1.3). Diese Ergebnisse zeigen, dass linguistische Merkmale auf der morphologischen Ebene einen bedeutenden Beitrag zur Identifizierung und Rekonstruktion von Nullsubjekten in Bezug auf Sprecher oder Adressat leisten. Dies legt nahe, dass das Phänomen der Nullsubjekte im Koreanischen in gewisser Weise der Situation in prototypischen Pro-Drop-Sprachen wie Spanisch ähnelt, obwohl das Koreanische als radikale Pro-Drop-Sprache klassifiziert wird. Diese Klassifizierung erfolgte aufgrund der Tatsache, dass das Koreanische trotz des Fehlens grammatischer Mittel zur Identifizierung der Referenten von Nullsubjekten den Gebrauch von *pro* erlaubt. Die Analyse innerhalb dieser Studie zeigt jedoch, dass im Koreanischen verschiedene voll grammatisierte Mittel existieren, die zur Rekonstruktion von Nullsubjekten beitragen können, insbesondere in Bezug auf die 1. und 2. Person, wenn es um Nullsubjekte in interaktiven Gesprächen geht. Die Analyseergebnisse dieser Studie können daher als empirische Evidenz betrachtet werden, die die Behauptung von Kim Mi-Young (2003) unterstützt, dass das Phänomen der Nullsubjekte im Koreanischen in Bezug auf Sprecher oder Adressat eher einem Phänomen ähnelt, das in prototypischen Pro-Drop-Sprachen wie Spanisch beobachtet wird, wo die Rekonstruierbarkeit von Nullsubjekten auf die reiche Verbalmorphologie zurückzuführen ist. Der Hauptunterschied besteht jedoch darin, dass die linguistischen Merkmale im Koreanischen keine

Unterscheidung zwischen Singular und Plural markieren, während im Spanischen die Verbalflexion nicht nur jede grammatische Person, sondern auch den Numerus markiert (vgl. Tabelle 1 in Kapitel 1.1).

6.2. Ausblick

Die vorliegende Arbeit trägt wesentlich zum Verständnis des Phänomens der koreanischen Nullsubjekte bei, indem sie eine umfassende Analyse dieses Phänomens bietet, insbesondere im Hinblick auf Nullsubjekte, die sich auf den Sprecher oder den Adressaten beziehen. Sie identifiziert und beschreibt potenzielle Parameter für die Resolution koreanischer Nullsubjekte detailliert, wodurch etwa 5,6-mal mehr potenzielle Parameter als in früheren Arbeiten diskutiert wurden (vgl. Kapitel 5.1.4). Des Weiteren wurde in dieser Studie ein funktionsfähiges System zur Resolution koreanischer Nullsubjekte entwickelt, das als erstes seiner Art gilt und in der Lage ist, einem Input-Prädikat einen Referenten für das Nullsubjekt zuzuweisen. Dieses System kann einfach zur Nullsubjektreolution weiterverwendet werden, solange Prädikate mit Nullsubjekten als Input eingegeben werden. Die generierten annotierten Daten stellen eine wichtige Ressource dar, um die Verteilung der Nullsubjekte nach Typen in interaktiven Gesprächen zu untersuchen und die Hypothese dieser Arbeit zu überprüfen. Die Analyse der annotierten Daten konnte bestätigen, dass diskursteilnehmerbezogene Nullsubjekte einen erheblichen Anteil an den gesamten koreanischen Nullsubjekten ausmachen und klare Hinweise zur Wiederherstellung ihrer Referenten liefern. Die annotierten Daten ermöglichen auch eine Leistungsbewertung des entwickelten NRS, was zur Überprüfung der Hypothese dieser Arbeit und zur Identifizierung der Schlüsselparameter für die erfolgreiche Nullsubjektreolution beitrug. Insgesamt stellt diese Arbeit den ersten Versuch dar, ein regelbasiertes Resolutionsverfahren anhand eines umfangreichen Datensatzes zu testen.

Es gibt vielversprechende Bereiche für weitere Forschung auf diesem Gebiet. Erstens könnte die Wirkung des optimierten Annotationsschemas dieser Studie mit anderen Annotationsprojekten über Nullsubjekte untersucht werden. Ein Inter-Annotator-Agreement könnte zeigen, ob das optimierte Schema zu konsistenteren annotierten Daten führt. Zweitens könnte die Leistung des NRS mit anderen Datensätzen, die natürliche Gespräche enthalten, überprüft werden. Die Validierung des NRS mit anderen Datensätzen könnte dazu beitragen, seine Anwendbarkeit auf verschiedene sprachliche Kontexte zu testen und mögliche Verbesserungen zu erarbeiten. Drittens könnte die Untersuchung weiterer Parameter zur Rekonstruierbarkeit von Nullsubjekten fortgesetzt werden, insbesondere in Bezug auf Sprecher oder Adressaten. Es wäre wichtig zu überprüfen, warum einige Nullsubjekte im NRS nicht rekonstruiert wurden und ob dies auf nicht erfasste linguistische Merkmale oder andere Faktoren zurückzuführen ist.

In künftigen Forschungsarbeiten könnte das NRS als integrierter Bestandteil eines umfassenden Systems zur Nullsubjektreolution weiterentwickelt werden. Dieses System sollte nicht nur in der Lage sein, Referenten für deiktische Nullsubjekte wiederherzustellen, sondern auch Koreferenzbeziehungen von anaphorischen Nullsubjekten zu erkennen. Da das NRS bereits eine bemerkenswerte Leistung bei der Resolution deiktischer Nullsubjekte gezeigt hat, könnte es als Grundlage für dieses System dienen. Damit könnten diejenigen Nullsubjekte, die nicht durch das NRS rekonstruiert werden können, mit größerer Wahrscheinlichkeit als anaphorische Nullsubjekte klassifiziert werden.

Die linguistischen Merkmale und ihre Kombination, die in dieser Studie für die Rekonstruktion von Nullsubjekten in Bezug auf Diskursteilnehmer eine Rolle spielen, könnten im Sprachunterricht für Nicht-Muttersprachler:innen praktische Anwendung finden. Indem verschiedene sprachliche Kontexte vermittelt werden, in denen ein ausgelassenes Subjekt natürlicher klingt als ein overtes, könnten Koreanisch-Lernende dazu befähigt werden, in authentischen sprachlichen Situationen angemessen

zu interagieren und zu kommunizieren. Insbesondere könnten die linguistischen Merkmale, die den wichtigsten Regeln mit dem größten Einfluss auf die korrekte Rekonstruktion von Nullsubjekten zugrunde liegen, gezielt im Lehrplan behandelt werden. Auf diese Weise erhalten Koreanisch-Lernende klare Hinweise auf bestimmte implizierte Subjekte und können besser verstehen, warum Nullsubjekte in verschiedenen Kontexten verwendet werden.

Literatur

- Androutsopoulos, Jannis & Gurly Schmidt. 2001. SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe. *Zeitschrift für angewandte Linguistik* 36. 49–80.
- Austin, John Langshaw. 1962. *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
- Bae Soo-Ja. 2007. *A study on Korean auxiliary verb*. Changwon National University PhD Thesis.
- Bak Jaehee. 2011. *The light verb construction in Korean*. University of Toronto PhD Thesis.
- Bang Woonkyu. 1994. bojodongsa -*judea* daehan yeongu [A study on the auxiliary verb -*juda*]. *Journal of Korean National Language and Literature* 17·18. 181–203.
- Bellotti, Adriana & Luigi Rizzi. 1988. Psych-verbs and θ-theory. *Natural Language & Linguistic Theory* 291–352.
- Blum-Kulka, S. & E. Olshtain. 1984. Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics* 5(3). 196–213.
- Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge University Press.
- Brown, Penelope & Stephen C. Levinson. 2007. Gesichtsbedrohende Akte. In Steffen K. Herrmann, Sybille Krämer & Hannes Kuch (eds.), *Verletzende Worte*, 59–88. Bielefeld: Transcript Verlag.
- Bühler, Karl. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. G. Fischer.
- Bybee, Joan L. & Suzanne Fleischman. 1995. *Modality in grammar and discourse*. John Benjamins Publishing Company.
- Bybee, Joan L., Revere Dale Perkins & William Pagliuca. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In *Subject and topic*. New York: Academic Press.
- Cho Choon-Hak. 1982. *A study of Korean pragmatics: Deixis and politeness*. University of Hawaii PhD Thesis.
- Cho Yongkil. 2005. *Grammatik und Höflichkeit im Sprachvergleich: Direktive Handlungsspiele des Bittens, Aufforderns und Anweisens im Deutschen und Koreanischen* (Beiträge zur Dialogforschung Bd. 32). Tübingen: M. Niemeyer.
- Choe Jae-Woong & Lee Min-Haeng. 1999. chojeom [Focus]. In *hyeongsig uimilongwa hangugeo gisul [Formal semantics and description of Korean]*, 157–205. H.S Media.
- Choi Hyoung Kang. 2004. *A study on case-marker phrase complements in Korean*. Seoul National University PhD Thesis.
- Choi Ju Hee. 2013. The research about the core-functions and restrictions of connective -neunde for Korean language education. *The Journal of Yeongju Language & Literature* 25. 133–163.
- Choi Suk-Jae. 2013. Survey the meaning of particle eun/nun and i/ka. *Urimal* 35. 82–111.
- Choi Yoonji. 2016. *A study on the information structure in Korean language*. Seoul National University PhD Thesis.
- Chomsky, Noam. 1981. *Lectures on government and binding*. Walter de Gruyter GmbH & Company KG.
- Chomsky, Noam. 1982. *Some concepts and consequences of the theory of government and binding*. Vol. 6. MIT press.
- Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In *Syntax: An international handbook of contemporary research*, vol. 1, 506–569. Berlin: De Gruyter.
- Chung Jin Woo & Park Jong Chul. 2009. Automated classification of sentential types in Korean with morphological analysis. *Eoneowa jeongbo (Language and information)* 13(2). 59–97.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.

- Coates, Jennifer. 1995. The expression of root and epistemic possibility in English. In *The verb in contemporary English: Theory and description*, 145–156. Cambridge: Cambridge University Press.
- Dryer, Matthew S. & Martin Haspelmath. 2013. The world atlas of language structures online.
- Ehrich, Veronika & Günter Saile. 1972. Über nicht-direkte Sprechakte. In *Linguistische Pragmatik*, 255–287. Wiesbaden: Athenaion.
- Eom Gyeong-ok. 2002. A study on the speech level of listeners-honorific for modern Korean. *語文論集 (The Journal of Lang. & Lit.)* 30. 79–98.
- Eom Iksang. 2013. Korean romanization system: The current status and suggestions for revision. *The Journal of Study on Language and Culture of Korea and China* (31). 3–30.
- Facchinetti, Roberta, Manfred G. Krug & Frank Robert Palmer. 2003. *Modality in contemporary English*. Berlin: Mouton de Gruyter.
- Franck, Dorothea. 1975. Zur Analyse indirekter Sprechakte. In *Beiträge zur Grammatik und Pragmatik*, 219–232. Kronberg: Scriptor.
- Frawley, William. 2008. *The expression of modality*. Germany: Walter de Gruyter.
- Frick, Karina. 2017. *Elliptische Strukturen in SMS: Eine korpusbasierte Untersuchung des Schweizerdeutschen*. De Gruyter.
- Gnanambal, S., M. Thangaraj, V. T. Meenatchi & V. Gayathri. 2018. Classification algorithms with attribute selection: An evaluation study using WEKA. *International Journal of Advanced Networking and Applications*. Eswar Publications 9(6). 3640–3644.
- Grosz, Barbara J. & Candace L. Sidner. 1986. Attention, intentions, and the structure of discourse. *Computational linguistics* 12(3). 175–204.
- Grosz, Barbara J., Scott Weinstein & Aravind K Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2). 203–225.
- Grosz, Barbara, Aravind Joshi & Scott Weinstein. 1983. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual meeting of the Association for Computational Linguistics*, 44–50. Association for Computational Linguistics.
- Gundel, Jeanette K. 1998. Centering theory and the givenness hierarchy: Towards a synthesis. In *Centering Theory in Discourse*, 183–198. Clarendon Press.
- Haegeman, Liliane. 2013. The syntax of registers: Diary subject omission and the privilege of the root. *Lingua* 130. 88–110.
- Hamann, Cornelia. 1992. *Late empty subjects in German child language*. Univ. de Genève, Faculté des Lettres.
- Han Kil. 2002. *hyeondae ulimalui nopimbeob yeongu [A Study on honorific system of the modern Korean language]*. Youkrack.
- Han Na-Rae. 2004. Korean null pronouns: Classification and annotation. In *Proceedings of the Workshop on Discourse Annotation*, 33–40.
- Han Na-Rae. 2006. *Korean zero pronouns: Analysis and resolution*. University of Pennsylvania PhD Thesis.
- Han Song Hwa. 2017. The meaning and function of Korean particle *un/nun* in spoken languages: For Korean pedagogic grammar. *Korean Semantics* 55. 81–111.
- Haspelmath, Martin, Matthew S. Dryer, David Gil & Bernard Comrie. 2005. *The world atlas of language structures*. New York: Oxford University Press.
- Herrmann, Wilfried. 1994. *Lehrbuch der modernen koreanischen Sprache*. Helmut Buske Verlag.
- Ho Jung-hwan. 2015. Politeness and indirect expressions in Korean. *韓國學研究論文集. 中國文化大學韓國語文學系* (4). 21–44.
- Holmberg, Anders. 2017. Linguistic typology. In *The Oxford Handbook of Universal Grammar*, 355–376. Oxford University Press.

- Hong Jaesung. 1997. suleomyeongsa sajeongwa -ida suleomyeongsa gumunui gisul [Dictionary of predicative noun and description of predicative noun construction -ida]. In *Proceedings of The Sixth International Conference on Korean Language*, 131–149.
- Hong Min-pyo. 2000. Centering theory and argument deletion in spoken Korean. *The Korean Journal of Cognitive Science* 11(1). 9–24.
- Hong Munpyo. 2011. Subjektellipse im gesprochenen Koreanischen und deren Behandlung für die maschinelle Übersetzung ins Deutsche. *Dokohak (Zeitschrift der Koreanischen Gesellschaft für Deutsche Sprachwissenschaft)* 24. 417–439.
- Huang, C. T. James. 1989. Pro-drop in Chinese: A generalized control theory. In Osvaldo A. Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*, vol. 15, 185–214. Dordrecht: Springer Netherlands.
- Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15. 531–574.
- Huh Kyunghang. 2011. A study on the meaning of -eulge and -eulrae. *The language and culture* 7(2). 215–233.
- Hwang Juck-Ryoon. 1990. ‘Deference’ versus ‘politeness’ in Korean speech. *International Journal of the Sociology of Language* 82. 42–55.
- Hyams, Nina M. 1986. *Language acquisition and the theory of parameters*. Dordrecht: Reidel Publishing Company.
- Im Hongbin. 1972. *gugeoui jujehwa yeongu* [A study on topicalization in Korean]. Seoul National University Master Thesis.
- Im Hongbin. 2007. *hangugeoui jujewa tongsa bunseog: juje gaenyeomui saeloun jeongae* [Syntax analysis of theme in Korean: A new development of the subject concept]. SNUP.
- Imo, Wolfgang. 2014. Elliptical structures as dialogical resources for the management of understanding. In Susanne Günthner, Wolfgang Imo & Jörg Bäcker (eds.), *Grammar and Dialogism*, 139–176. De Gruyter.
- Jaeggli, Osvaldo A. & Nina M. Hyams. 1987. Morphological uniformity and the setting of the null subject parameter. In *North East Linguistics Society*, vol. 18, 238–253.
- Jaeggli, Osvaldo & Kenneth J. Safir. 1989. The null subject parameter and parametric theory. In Osvaldo A. Jaeggli & Kenneth J. Safir (eds.), *The null subject parameter*, 1–44. Dordrecht: Springer Netherlands.
- Jang Chaerin. 2017. A comparative study of the meanings of intention expressions in Korean: Focused on -gess-, -eulgeosi-, -eulge(yo), -eulrae(yo). *Korean Semantics* 56. 1–34.
- Jang Mira. 2008. A study on selection and arrangement of education content of Korean quotations. *Bilingual Research* 38. 419–438.
- Jang Seokjin. 1986. *hwayonglon yeongu* [A study of pragmatics]. Topbooks.
- Jang Seokjin. 1994. *tonghab munbeoblon: damhwawa hwayong* [Unified grammar: Discourse and pragmatics]. Seoul National University.
- Jeon Hye Young. 1995. On the pragmatics of -kess- and politeness phenomena in Korean. 國語學 (*Journal of Korean Linguistics*) 26. 125–146.
- Jeong Yeon ju. 2011. How to deal with experiencer arguments of adjective predicates: On the conversions from the psychic to the figural and from the figural to the sensory. *Journal of Korealex* (17). 209–238.
- Jun Youngchul. 2008. The so-called double passive constructions in Korean. *Eoneohag* 0(52). 79–101.
- Kameyama, Megumi. 1998. Intrasentential centering: A case study. In *Centering Theory in Discourse*, 89–114. Clarendon Press.
- Kang Hyoun Hwa, Lee Hyungjung, Nam Sinhye, Jang Chaerin, Hong Yeonjung, & Kim Kanghee. 2016. *hangugeogyoyug munbeob: jalyopyeon* [Korean grammar for Korean language education: Materials]. Hanguelpark.

- Kaur, Harsurinder, Husanbir Singh Pannu & Avleen Kaur Malhi. 2019. A systematic review on imbalanced data challenges in machine learning: Applications and solutions. *ACM Computing Surveys* 52(4). 1–36.
- Kim Ahrim. 2018. Non-polite 3rd person pronominal uses of *yay/kyay/cyay* in spoken Korean. *Language and Linguistics* 80. 1–40.
- Kim Heungssoo. 1989. *hyeondaegugeo simlidongsa gumun [A study on psychological verb constructions in modern Korean]*. Topbooks.
- Kim Hung-gyu, Kang Beom-mo, & Hong Jungha. 2007. 21st century Sejong corpora (to be) completed. *The Korean Language in America* 12. 31–42.
- Kim Hye Ryoung. 2016. Semantic analysis of first-person pronouns: Based on analysis of meaning conditions. *民族文化研究 (Korean Classics Studies)* 70. 267–294.
- Kim Jung-Nam. 2003. The meaning and usage of the Korean pronoun *wuli*. *Korean Semantics* 13. 257–274.
- Kim Jung-Nam. 2005. *gugeo hyeongyongsui yeongu [A study on Korean adjectives]*. Youkrack.
- Kim Jung-Nam. 2009. Meaning and its distribution of *-gedoya*: As a part of basic study for developing Korean language education material. *Korean Semantics* 30. 59–88.
- Kim Kanghee. 2019. *A study on the discourse grammar of Korean directive speech act: Focused on a contextual analysis approach on the meaning, form, and usage*. Yonsei University PhD Thesis.
- Kim Mi-Kyung. 2002. Preferred argument structure in Korean discourse. *Korean Journal of Linguistics* 27(3). 315–343.
- Kim Mi-Kyung. 2003. A condition on the interpretation of zero objects in Korean discourse: A centering analysis. In *Proceedings of the Human and Language Technology (SIGHLT)*, 134–136.
- Kim Min-gook. 2016. Case markers and focus in Korean. *國語學 (Journal of Korean Linguistics)* 78. 165–219.
- Kim Minsun. 2005. *A study on the types and meaning of doeda constructions*. Seoul National University Master Thesis.
- Kim Mi-Young. 1994. *(The) centering of Korean discourse*. South Korea: Seoul National University Master Thesis.
- Kim Mi-Young. 2003. *An optimality approach to the referential interpretation of zero anaphora in Korean*. South Korea: Seoul National University PhD Thesis.
- Kim Ok. 1998. *A study on the ellipsis in conjunctive sentences in Korean*. Chonnam National University Master Thesis.
- Kim Sun-hye. 2019. *The study of Korean periphrastic construction*. Yonsei University PhD Thesis.
- Kim Sunyoung. 2015. The research of the construction “adjective+ge doeda.” *Korean Journal of Linguistics* 40(1). 19–39.
- Kim Young Ha. 1996. *hangugeo simli dongsui nonhang gujowa gyeobjueo [The argument structure and multiple subjects of the Korean psychological verb]*. *The Korean Language and Literature* 0(57). 73–107.
- Ko Eon-Suk. 2000. A discourse analysis of the realization of object NP forms in Korean. *Ohak Yonku* 36(1). 43–62.
- Ko Seok-Ju. 2011. A study on the types of subjectless sentences in Korean. *Discourse and Cognition* 18(2). 1–22.
- Ko Yong-Kun. 1976. Sentence-types and mood system in modern Korean. *語學研究 (Language research)* 12(1). 359–400.
- Ko Yong-Kun. 2004. *A study of tense, aspect and mood in Korean*. Taehagsa.
- Koh Kyoung-tae. 2014. A study on the learners' contents of the indirect quotation shown in the Korean language textbooks. *The Korean Language and Literature* (166). 141–163.
- Koo Hyun Jung. 2001. On discourse functions of conditionals. *Korean Language Research* (9). 27–52.

- Kwon Jae Il. 1985. *gugeoui boghabmun guseong yeongu* [A study on complex sentences in Korean]. Jipmoondang.
- Kwon Jae Il. 2004. *Realizing methods of the mood in spoken Korean*. Seoul National University.
- Lee Chung Hoon. 2015. The syntax of coordinating endings and structure building. *Korean Linguistics* 66. 49–85.
- Lee Chung-min. 1976. Cases for psychological verbs in Korean. *The Korean Journal of Linguistics* 1(1). 256–296.
- Lee Eunkyung. 1996. *A study on connective endings in Korean*. Seoul National University PhD Thesis.
- Lee Hanjung. 2006. Effects of focus and markedness hierarchies on object case ellipsis in Korean. *Discourse and Cognition* 13(2). 205–231.
- Lee Hansol. 1989. *Korean grammar*. London: Oxford University Press.
- Lee Heeja & Lee Jonghee. 2010. *eomi, josa sajeon* [Dictionary on suffix and particle]. Hankookmunhwasa.
- Lee Ik-Hwan & Lee Min-Haeng. 2000. Anaphora resolution and discourse structure: A controlled information packaging approach. *Language and Information* 4(1). 67–82.
- Lee Ji Soo. 2016. *A study on the grammar and speech acts of Korean imperative sentences*. Seoul National University PhD Thesis.
- Lee Ji Sun. 2014. *A study on imperative speech acts for Korean language education*. Yonsei University Master Thesis.
- Lee Nam-Seok. 1994. *Deixis und Honorifica: Allgemeine deiktische Phänomene und die pragmatische Komponente des Koreanischen*. Göttingen: Georg-August-Universität.
- Lee Namsoon. 1998. gyeogpyojiui bisilhyeongwa saenglyag [Non-realization and omission of case markers]. *國語學 (Journal of Korean Linguistics)* 31. 338–360.
- Lee Narah. 2014. A reconsideration of the omission of first and second person subjects in modern spoken Korean: Focusing on the pragmatic meanings of the overt subject expressions. *Discourse and Cognition* 21(3). 145–163.
- Lee Phil Young. 2003. Conditions and directionalities of the interpretation of indirect speech acts (ISAs). *tegseuteueoneohag (Text Linguistics)* 15. 313–339.
- Lee Sangbok. 1974. hangugeoui inyongmun yeongu [A study of quotations in Korean]. *eoneowa munhwa (The language and Culture)* 1. 131–154.
- Lee Seung-Wook. 1969. jueoui tongsae gwanhan gochal [A study on the syntax of subjects]. *國文學論集* 3. 25–37.
- Lee Sun-Hee, Donna K. Byron, & Jang Seok Bae. 2005. Why is zero marking important in Korean? In Robert Dale, Kam-Fai Wong, Jian Su & Oi Yee Kwong (eds.), *International Conference on Natural Language Processing*, vol. 3651, 588–599. Berlin, Heidelberg: Springer.
- Lee Yun Jin & No Ji Ni. 2003. A study on Korean modal expressions. *Journal of Korean Language and Education* 14(1). 173–209.
- Levinson, Stephen C. 1983. *Pragmatics*. New York: Cambridge University Press.
- Li, Charles & Sandra Thompson. 1976. Subject and topic: A new typology of language. *Subject and Topic* 457–489.
- Lim Dong Hoon. 2000. *hangugeo eomi siui munbeob* [The grammar of the Korean infix si]. Thae haksa.
- Lim Dong Hoon. 2001. The diachronic perspective on the meanings of Korean prefinal ending -gess-. *Journal of Korean Linguistics* 37. 115–147.
- Lim Dong Hoon. 2011. Sentence types in Korean. *國語學 (Journal of Korean Linguistics)*. The Society of Korean Linguistics 60. 323–359.
- Liu, Juan. 2017. *A study on change-of-state constructions: -eojida and -ge doeda in Korean language*. Seoul National University PhD Thesis.
- Mitkov, Ruslan. 1999. *Anaphora resolution: The state of the art*. School of Languages and European Studies, University of Wolverhampton.

- Mok Jungsoo. 2011. A spoken Korean grammar: For the integration with the written grammar. *Urimal* 28. 57–98.
- Mok Jungsoo & Kim Yeong-jung. 2006. A structure of passive constructions in Korean and their meaning “potential.” *Lingua Humanitatis* 8. 369–387.
- Mok Jungsoo & Yoo Hyunjo. 2003. A correlation between category of verbs/endings and person of subject. *語學研究 (Language research)* 39(3). 529–558.
- Mok Jungsoo & Yoo Hyunjo. 2007. On the pattern of connective verbal endings in multi-clause complex sentences in spoken Korean. *Korean Linguistics* 35. 275–303.
- Na Jinseok. 1965. *gugeo umjigssiui ttaemaegim yeongu* [A study on tense of Korean verbs]. *Han-Guel* (134). 50–147.
- Na Seung-Hoon. 2014. Third-order dependency parsing of Korean. In *Annual Conference on Human and Language Technology*, 78–80. Human and Language Technology.
- Na Seung-Hoon. 2015. Conditional random fields for Korean morpheme segmentation and POS tagging. *ACM Transactions on Asian and Low-Resource Language Information Processing* 14(3). 1–16.
- Nakaiwa, Hiromi & Satoru Ikehara. 1995. Intrasentential resolution of Japanese zero pronouns in a machine translation system using semantic and pragmatic constraints. In *Proceedings of the International Conference on Theoretical and Methodological Issues in Machine Translation (TM'95)*, 96–105. Leuven, Belgium.
- Nakaiwa, Hiromi, Satoshi Shirai, Satoru Ikehara & Tsukasa Kawaoka. 1995. Extrasentential resolution of Japanese zero pronouns using semantic and pragmatic constraints. In *Proceedings of the AAAI 1995 Spring Symposium Series: Empirical methods in discourse interpretation and generation*, 99–105.
- Nam Kisim. 2001. *hyeondae gugeo tongsalon* [Syntax of modern Korean]. Taehagsa.
- Nariyama, Shigeko. 2002. Grammar for ellipsis resolution in Japanese. In *Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation*, 135–145.
- Neleman, Ad & Kriszta Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38(4). 671–714.
- No Yongkyoon. 1991. A centering approach to the [CASE] [TOPIC] restriction in Korean. *Linguistics* 29(4). 653–668.
- Noh Dae Kyu. 1996. *hangugeoui ibmalgwa geulmal* [Spoken and written Korean]. Kookhakjalyowon.
- O Sang Yi. 2005. Anrede-und Hoflichkeitsformen im Koreanischen und im Deutschen: Aus kulturkontrastiver Perspektive. *Deutsche Sprache-und Literaturwissenschaft* 29. 243–262.
- Oh Sun-Young. 2007. Overt reference to speaker and recipient in Korean. *Discourse Studies* 9(4). 462–492.
- Okumura, Manabu & Kouji Tamura. 1996. Zero pronoun resolution in Japanese discourse based on centering theory. In *Proceedings of the 16th Annual Meeting of the Association for Computational Linguistics*, 871–876.
- Palmer, Frank Robert. 1979. *Modality and the English modals* (Longman Linguistics Library). 2. edition. London: Longman.
- Palmer, Frank Robert. 2001. *Mood and modality*. Cambridge university press.
- Panther, Klaus-Uwe & Linda Thornburg. 1998. A cognitive approach to inferencing in conversation. *Journal of Pragmatics* 30(6). 755–769.
- Park Arum & Hong Munpyo. 2015. Subjektellipse im Koreanischen und deren Behandlung für die maschinelle Übersetzung ins Deutsche. *Koreanische Gesellschaft für Germanistik* 56(1). 197–223.
- Park Chang-Young. 2014. *A study on subject in Korean*. Dongguk University Master Thesis.
- Park Cheong-hee. 2013. *A study of ellipsis in contemporary Korean*. Korea University PhD Thesis.
- Park Chul-woo. 2015. On the interrelation between the functions of auxiliary noun endings and the information structure in Korean. *國語學 (Journal of Korean Linguistics)* 73. 269–307.

- Park Hami. 2014. *A study on the teaching of adverbial dependent noun construction of Korean*. Seoul National University Master Thesis.
- Park Jae-yon. 1998. *A study on sentence final endings of Panmal style in contemporary Korean*. Seoul National University Master Thesis.
- Park Jae-yon. 2003. Speaker or hearer-oriented modality and subject-oriented modality in Korean. *Gugeohag (Journal of Korean Linguistics)* 41. 249–277.
- Park Jae-yon. 2004. *A study on the modal endings in Korean*. Seoul National University PhD Thesis.
- Park Jae-yon. 2006. *hangugeo yangtae eomi yeongu [Study on the modal endings in Korean]*. Taehagsa.
- Park Mae-Ran. 1990. Conflict avoidance in social interaction: A sociolinguistic comparison of the Korean and Japanese honorific systems. In *Japanese/Korean Linguistics*, 111–127. Cambridge University Press.
- Park Son-Ja. 1992. A Study on the Korean modality. *Cogito* 41. 55–91.
- Park Sun Ok. 2003. *A study on the Korean auxiliary Verb*. Chung-Ang University PhD Thesis.
- Park Sunyoung. 2020. *Korean volitive expressions: Focusing on mechanism of choosing grammar*. Seoul National University Master Thesis.
- Park Yang Kyu. 1980. Subjectless sentences in Korean. *國語學 (Journal of Korean Linguistics)* 9. 1–25.
- Park Young Me. 2005. Einige morpho-syntaktische Unterschiede zwischen den koreanischen und deutschen Imperativkonstruktionen. *Deutsche Sprache-und Literaturwissenschaft* 29. 181–199.
- Piao, Chengri. 2012. *A study of teaching Korean honorific system based on personal deixis: Focusing on Chinese learners of Korean language*. Seoul National University PhD Thesis.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic inquiry* 17(3). 501–557.
- Roh Chae Hwan. 2006. A study on the functional characteristic of auxiliary verb -a/-ə juda. *The language and culture* 2(3). 67–84.
- Rudin, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*. Nature Publishing Group UK London 1(5). 206–215.
- Ryu Byungryul. 2001. *Centering and zero anaphora in the Korean discourse*. Seoul National University Master Thesis.
- Searle, John R. 1976. A classification of illocutionary acts. *Language in Society*. Cambridge University Press 5(1). 1–23.
- Searle, John R. 1979. *Expression and meaning: Studies in the theory of speech acts*. Cambridge, England: Cambridge University Press.
- Seo Cheong-Soo. 1996. *hyeondae hangugeo munbeob yeonguui gaegwan [An overview of studies on grammar of modern Korean]*. Hankookmunhwasa.
- Seo Cheong-Soo. 1997. *jondaebeobui yeongu [A study of the Korean honorifics]*. H.S Media.
- Shin Hyon Sook. 1982. *mogjeoggyeog pyoji -leului uimiyeongu [A study on the meaning of object marker -leul]*. *Eoneo (The Korean Journal of Linguistics)* 7(1). 119–139.
- Sigurðsson, Halldór Árman. 2011. Conditions on argument drop. *Linguistic Inquiry* 42(2). 267–304.
- Son Jae Eun. 2007. A study on methods of teaching the Korean conjunctive ending -neunde. *Journal of Korean Language and Education* 18(1). 187–205.
- Son Kyeong-ae. 2015. *A study on the teaching device of connecting endings in Koran language: Based on the connecting ending -ado and -eunde*. Chung-Ang University PhD Thesis.
- Son Se-mo-dol. 1989. *haengdong bugwai ganjeob pyohyeon: gongsonhan pyohyeon bangsig [Indirect representation of imposition of action: A polite way of expression]*. *Journal of Korean Language and Culture* 7. 181–206.
- Son Se-mo-dol. 1996. *gugeo bojoyongeon yeongu [A study on auxiliary verbs in Korean]*. Hankookmunhwasa.

- Song Dae-heon. 2015. A study on meanings of connecting endings *-neunde/eunde* and discourse function in the Korean language. *Korean Language and Literature in International Context* 0(66). 167–191.
- Suh Cheong Soo. 1990. *-gess-e gwanhayeo* [About -gess-]. In *gugeo munbeobui yeongu* [A study of Korean grammar]. Hankookmunhwasa.
- Sung Kwang Soo. 1978. cheeon jeobsoggwa gonggyeog [Noun phrase structure conjunction and comitative case]. *Han-Guel* (162). 143–158.
- Taraldsen, Knut Tarald. 1978. On the NIC, vacuous application and the that-trace filter. MIT, Cambridge, Massachusetts, ms.
- Thomas, Andrew L. 1979. Ellipsis: The interplay of sentence structure and context. *Lingua* 47(1). 43–68.
- Tomioka, Satoshi. 2003. The semantics of Japanese null pronouns and its cross-linguistic implications. In *The interfaces: Deriving and interpreting omitted structures*, vol. 61, 321–339. Amsterdam: John Benjamins.
- Tsao, Feng-fu. 1977. *A functional study of topic in Chinese: The first step toward discourse analysis*. University of Southern California.
- Ulrich, Winfried. 1975. *Wörterbuch: Linguistische Grundbegriffe*. Verlag Ferdinand Hirt.
- Walker, Marilyn, Masayo Iida & Sharon Cote. 1994. Japanese discourse and the process of centering. *Computational linguistics* 20(2). 193–231.
- Wang Hahn-Sok. 1986. Toward describing the organization of Korean speech levels. *語學研究 (Language research)* 22(3). 351–373.
- Woo Soon-jo. 1994. *hangugeoui hyeongsangseonggwā gwangyepyojiui silhyeonyangsang* [Configuration of Korean and its realization patterns]. Seoul National University PhD Thesis.
- Yang In-Seok. 1972. *Korean syntax: Case markers, delimiters, complementation, and relativization*. University of Hawaii PhD Thesis.
- Yang In-Seok. 1976. hangugeo yangsangui hwayonglon (1): jeanmungwa myeonglyeongmun [Pragmatics in Korean (1): Exhortative and imperative sentences]. *Eoneo (The Korean Journal of Linguistics)* 1. 121–41.
- Yang Jaehyung. 1997. Centering theory and its application to Korean analysis. *Journal of industrial research institute* (3). 107–119.
- Yang Ji Hyeon. 2013. As to expression of intention: With focus on *-ryeo ha(da)*, *-ryeogo(yo)*, *-ryeogo ha(da)*. *Baedalmal (Korean Language)* 52. 51–74.
- Yang Jungseok. 1995. *gugeodongsawi uimibunseoggwa yeongyeol ilon* [Analysis of meaning and dependency theory of Korean verbs]. Pagijong Press.
- Yi Seon-ung. 2012. *hangugeo munbeoblonui gaenyeomeo yeongu* [A study on the conceptual words of Korean grammar]. Wolin.
- Yoo Hyun-kyung. 1997. *A classificatory study on Korean adjectives*. Yonsei University PhD Thesis.
- Yoo Hyun-kyung. 1998. *gugeo hyeongyongsa yeongu* [A study on Korean adjectives]. Hankookmunhwasa.
- Yoo Hyun-kyung. 2000. A study on the classification of Korean adjectives. *Journal of Korean Linguistics* 36. 221–258.
- Yoo Hyun-kyung. 2015. Issues on the subject of Korean adjective construction: Focusing on Kim (2005). *Morphology* 17(1). 69–91.
- Yoo Hyun-kyung. 2017. Person constraints on Korean predicates. *Journal of Korean Linguistics* 83. 65–92.
- Yoon Eun Kyung. 2006. A study on modality in Korean: Modal endings *-eulrae* and *eulge*. *eoneowa munhwa (The language and Culture)* 2(2). 41–63.
- Yun Sun Young. 2016. *Deutsch-koreanischer Sprachkontakt: Eine empirische Untersuchung des Sprachverhaltens bilingualer koreanischer Kinder in Deutschland*. Bonn: Rheinischen Friedrich-Wilhelms-Universität PhD Thesis.

Appendix

Appendix 1: Code für die gesamte Regel in Nullsubjektresolutionssystem

```
if datapoint["zp_mood"] == "impe":  
    datapoint["subject_by_rule"] = "2sh+"  
    datapoint["group"] = "mood derivable"  
    datapoint["rule"] = "impe"  
    datapoint["rule_order"] = "1"  
  
elif datapoint["zp_mood"] == "sugg":  
    datapoint["subject_by_rule"] = "1ph-"  
    datapoint["group"] = "mood derivable"  
    datapoint["rule"] = "sugg"  
    datapoint["rule_order"] = "2"  
  
elif re_morphmatch(morphs, morph="^(으?세요|으?세여|으?십시오|으?십시오) $", pos="^E$") and  
    datapoint["zp_mood"] == "decl":  
    datapoint["subject_by_rule"] = "2sh+"  
    datapoint["group"] = "mood derivable"  
    datapoint["rule"] = "impe"  
    datapoint["rule_order"] = "3"  
  
elif re_morphmatch(morphs, morph="^(는게좋|면좋|는게낫|는것이좋|는편이좋) $", pos="^X$") and  
    datapoint["zp_hono"] == "plus":  
    datapoint["subject_by_rule"] = "2sh+"  
    datapoint["group"] = "suggestory formula"  
    datapoint["rule"] = "suggestory periphrastic construction"  
    datapoint["rule_order"] = "4"  
  
elif re_morphmatch(morphs, morph="^(는게좋|면좋|는게낫|는것이좋|는편이좋) $", pos="^X$") and  
    datapoint["zp_mood"] == "decl":  
    datapoint["subject_by_rule"] = "2sh+"  
    datapoint["group"] = "suggestory formula"  
    datapoint["rule"] = "suggestory periphrastic construction"  
    datapoint["rule_order"] = "5"  
  
elif re_morphmatch(morphs, morph="^(을것같|ㄹ것같|는것같|ㄴ것같|은듯하) $", pos="^X$") and  
    datapoint["zp_hono"] == "plus":  
    datapoint["subject_by_rule"] = "2sh+"  
    datapoint["group"] = "guess"  
    datapoint["rule"] = "guess periphrastic construction"  
    datapoint["rule_order"] = "6"  
  
else:  
    datapoint["subject_by_rule"] = "N/A"  
  
try_str_list = ["보니까", "보니깐", "보니까는"]  
performative_head_list = ["부탁", "권유"]  
cognition_head_list = ["예정", "생각", "기억", "확인", "깜박", "예상", "착각", "이해", "확실", "집중", "계획", "숙지",  
    "오해", "파악", "동의", "사과", "깜빡"]  
sensation_head_list = ["시음", "민감", "구경", "시원"]  
feeling_head_list = ["실망", "공포", "영광", "장담", "한시름", "미안", "실례", "의아", "질색", "놀랍", "환영"]  
  
if datapoint["subject_by_rule"] == "N/A":  
    if re_morphmatch(morphs, morph="^(아?야하|어야하|아?야되|어야되|고계시|으?시|으셨) $", pos="^X|E$") and  
        re_morphmatch(morphs, morph="^(ㄴ[는]데|는데|는건데|은데|ㄴ[은]는데) $", pos="^E$") and  
        datapoint["zp_hono"] == "plus" and  
        datapoint["verb_syn"] == "clau":  
            datapoint["subject_by_rule"] = "2sh+"  
            datapoint["group"] = "clau"  
            datapoint["rule"] = "clau_neunde"  
            datapoint["rule_order"] = "7"  
  
    elif re_morphmatch(morphs, morph="^(으?시|아계시|고계시|ㄹ[을]줄알|가시|주무시|으셨) $", pos="^E|V$") and  
        re_morphmatch(morphs, morph="^(려?면|으면|다보면) $", pos="^E$") and  
        datapoint["verb_syn"] == "clau":  
            datapoint["subject_by_rule"] = "2sh+"            datapoint["group"] = "clau"  
            datapoint["rule"] = "clau_myeon"  
            datapoint["rule_order"] = "8"  
  
    elif re_morphmatch(morphs, morph="^으?시$", pos="^E$") and  
        re_morphmatch(morphs, morph="^있$", pos="^E$") and
```

```

re_morphmatch(morphs, morph="^으면$", pos="^E$") and datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_myeon"
    datapoint["rule_order"] = "9"
elif re_morphmatch(morphs, morph="^주무시$", pos="^V$") and
    re_morphmatch(morphs, morph="^고$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_go"
    datapoint["rule_order"] = "10"
elif re_morphmatch(morphs, morph="^(으?시|[아어]계시)$", pos="^E$") and
    re_morphmatch(morphs, morph="^다가?$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_daga"
    datapoint["rule_order"] = "11"
elif re_morphmatch(morphs, morph="^으?시$", pos="^E$") and
    re_morphmatch(morphs, morph="^으?면서$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_myeonseo"
    datapoint["rule_order"] = "12"
elif re_morphmatch(morphs, morph="^으?어서$", pos="^E$") and
    re_morphmatch(morphs, morph="^어서$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_eoseo"
    datapoint["rule_order"] = "13"
elif re_morphmatch(morphs, morph="^으?나$", pos="^E$") and
    re_morphmatch(morphs, morph="^ㄴ후에*$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_nhu"
    datapoint["rule_order"] = "14"
elif re_morphmatch(morphs, morph="^으?기전에$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_gijeon"
    datapoint["rule_order"] = "15"
elif re_morphmatch(morphs, morph="^으?나|$|", pos="^E$") and
    re_morphmatch(morphs, morph="^(ㄴ다음에*|ㄴ뒤에*|ㄴ뒤에도)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_ndam"
    datapoint["rule_order"] = "16"
elif re_morphmatch(morphs, morph="^으?나|$|", pos="^E$") and
    re_morphmatch(morphs, morph="^(ㄹ때|ㄹ때에?는|ㄹ때에?도)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_lttae"
    datapoint["rule_order"] = "17"
elif re_morphmatch(morphs, morph="^으?나|$|", pos="^E$") and
    re_morphmatch(morphs, morph="^다니$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_dani"
    datapoint["rule_order"] = "18"
elif re_morphmatch(morphs, morph="^으?나|$|", pos="^E$") and

```

```

re_morphmatch(morphs, morph="^어야지?$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_eoya"
datapoint["rule_order"] = "19"

elif re_morphmatch(morphs, morph="^으?시$", pos="^E$") and
re_morphmatch(morphs, morph="^더라도?", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_deolado"
datapoint["rule_order"] = "20"

elif re_morphmatch(morphs, morph="^으?시$", pos="^E$") and
re_morphmatch(morphs, morph="^ㄴ?다면?", pos="^E$") and datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_ndamyeon"
datapoint["rule_order"] = "21"

elif re_morphmatch(morphs, morph="^(아?야하|어야하|으?려하|[ㄹ을]터이|을테|[아어]보|[ㄹ을]것이)$",
pos="^X$") and
re_morphmatch(morphs, morph="^(니|니까|니깐)$", pos="^E$") and
datapoint["verb_syn"] == "clau" and
datapoint["zp_hono"] == "plus":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_nikka"
datapoint["rule_order"] = "22"

elif datapoint["zp_hono"] == "plus" and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_plus"
datapoint["rule_order"] = "23"

elif re_morphmatch(morphs, morph="^(ㄹ수있|을수있|ㄹ수없|ㄹ수도없|을수없)$", pos="^X$") and
re_morphmatch(morphs, morph="^으?ㄴ?", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_euni"
datapoint["rule_order"] = "24"

elif re_morphmatch(morphs, morph="^(걸리|멀쩡하|멀|나오)$", pos="^(J|V)$") and
re_morphmatch(morphs, morph="^(는데)$", pos="^E$"):
datapoint["subject_by_rule"] = "none"
datapoint["group"] = "exclude"
datapoint["rule"] = "exclude predicate"
datapoint["rule_order"] = "25"

elif re_morphmatch(morphs, morph="^(겠|았|었|아?야하|어야하|아?야되|어야되|ㄹ려고?하|으?려고?하|고자하|
ㄹ까하|[아어]?드리|고싶|[ㄹ을]생각이|[ㄹ을]지도모르)$", pos="^(E|X)$") and
re_morphmatch(morphs, morph="^(ㄴ|는)데|는데|는건데|은데|[ㄴ|은|는]건데)", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_neunde"
datapoint["rule_order"] = "26"

elif re_morphmatch(morphs, morph="^([ㄹ을]건데|려는데)$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_neunde"
datapoint["rule_order"] = "27"

elif re_morphmatch(morphs, morph="^(고있|어서그리|던중이|고있는중이|ㄴ것같|은것같|은듯하)$", pos="^X$") and
re_morphmatch(morphs, morph="^(ㄴ|데|는데|은데)$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_neunde"
datapoint["rule_order"] = "28"

elif re_morphmatch(morphs, morph="^(있었|았|었|었었)$", pos="^E$") and

```

```

re_morphmatch(morphs, morph="^는데$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_neunde"
datapoint["rule_order"] = "29"

elif re_morphmatch(morphs, morph="^(은건데|는건데)$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_neunde"
datapoint["rule_order"] = "30"

elif re_morphmatch(morphs, morph="^고싶$", pos="^X$") and
re_morphmatch(morphs, morph="^지만$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_jiman"
datapoint["rule_order"] = "31"

elif re_morphmatch(morphs, morph="^.*보$", pos="^V$") and
re_morphmatch(morphs, morph="^았$", pos="^E$") and
re_morphmatch(morphs, morph="^지만$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_jiman"
datapoint["rule_order"] = "32"

elif re_morphmatch(morphs, morph="^고싶$", pos="^X$") and
re_morphmatch(morphs, morph="^으니까?$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_nikka"
datapoint["rule_order"] = "33"

elif re_morphmatch(morphs, morph="^려니?$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_nikka"
datapoint["rule_order"] = "34"

elif re_morphmatch(morphs, morph="^(아|야하|어야하|으?려하|[ㄹ을]터이|을테|[아어]보|[ㄹ을]것이)$",
pos="^X$") and
re_morphmatch(morphs, morph="^(니|니까|니깐)$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_nikka"
datapoint["rule_order"] = "35"

elif re_morphmatch(morphs, morph="^(아드리|어드리|드리|축하드리)$", pos="^(X|V)$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_various"
datapoint["rule_order"] = "36"

elif re_morphmatch(morphs, morph="^자말$", pos="^X$") and
re_morphmatch(morphs, morph="^고$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_go"
datapoint["rule_order"] = "37"

elif re_morphmatch(morphs, morph="^(ㄹ까보|게되|고싶)$", pos="^X$") and
re_morphmatch(morphs, morph="^(아|어|어서|아서)$", pos="^E$") and
datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_aseo"
datapoint["rule_order"] = "38"

elif re_morphmatch(morphs, morph="^(게도|야하)$", pos="^X$") and
re_morphmatch(morphs, morph="^(다니|ㄴ다니)$", pos="^E$") and

```

```

datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "clau"
datapoint["rule"] = "clau_dani"
datapoint["rule_order"] = "39"
elif re_morphmatch(morphs, morph="^려고$", pos="^E$") and
    re_morphmatch(morphs, morph="^보$", pos="^V$") and
        re_morphmatch(morphs, morph="^(니|나까)$", pos="^E$") and
            datapoint["verb_syn"] == "clau":
            datapoint["subject_by_rule"] = "lsh+"
            datapoint["group"] = "clau"
            datapoint["rule"] = "clau_euni"
            datapoint["rule_order"] = "40"
elif [x for x in try_str_list if x in datapoint["verb"]] and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_euni"
    datapoint["rule_order"] = "41"
elif re_morphmatch(morphs, morph="^고있$", pos="^X$") and
    re_morphmatch(morphs, morph="^(으나|으니까|으니까)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_euni"
        datapoint["rule_order"] = "42"
elif re_morphmatch(morphs, morph="^(는줄알|ㄴ줄알|을줄알|ㄹ줄알|ㄹ줄아|잘알|알|아|알게되|잘아|모르|는줄모르|
    을줄모르|ㄹ줄모르|어야할지모르|는지모르|ㄴ지모르|을지모르|ㄹ지모르|을수있을지모르|알아들|궁금하|고집하)$",
    pos="^(X|V|J)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_cognition"
        datapoint["rule_order"] = "43"
elif re_morphmatch(morphs, morph="^(잊|잊어버리|믿|알아듣|외우|헷갈리)$", pos="^V$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_cognition"
        datapoint["rule_order"] = "44"
elif [x for x in cognition_head_list if x in datapoint["verb"]] and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_cognition"
        datapoint["rule_order"] = "45"
elif re_morphmatch(morphs, morph="^(보|못보|살펴보|둘러보|보려오|듣|들어서|들|들어보|맛보|배고프|싱거우|
    간지럽)$", pos="^(V|J)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_cognition"
        datapoint["rule_order"] = "46"
elif [x for x in sensation_head_list if x in datapoint["verb"]] and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"
        datapoint["group"] = "clau"
        datapoint["rule"] = "clau_cognition"
        datapoint["rule_order"] = "47"
elif re_morphmatch(morphs, morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼)$",
    pos="^(J|V)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으나깐)$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "lsh+"

```

```

datapoint["group"] = "clau"
datapoint["rule"] = "clau_cognition"
datapoint["rule_order"] = "48"
elif re_morphmatch(morphs, morph="^(괜찮|고맙|불편하|피곤하|지루하|흥미롭|곤란하|불쾌하|거북하|불안하|힘들|
힘들|기힘들|좋|어쩔수없|별수없)$", pos="^(J|X)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으니깐)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_cognition"
    datapoint["rule_order"] = "49"
elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하|고민되|고민하|놀라|
당황하|놀래|놀랍|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이)$", pos="^(V|N)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으니깐)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_cognition"
    datapoint["rule_order"] = "50"
elif re_morphmatch(morphs, morph="^(후회하|감사하|감사드리|감탄하|횡재하|들뜨|떨리|만족하|만족스럽|망설여지|
망설이|싫|싫어하)$", pos="^(V|J)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으니깐)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_cognition"
    datapoint["rule_order"] = "51"
elif [x for x in feeling_head_list if x in datapoint["verb"]] and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으니깐)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_cognition"
    datapoint["rule_order"] = "52"
elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
기쁘|행복하|난감하|반갑|환영하|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
    re_morphmatch(morphs, morph="^(ㄴ데|는데|은데|지만|다가|니까|으니까|나깐|으니깐)$", pos="^E$") and
    datapoint["verb_syn"] == "clau":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "clau"
    datapoint["rule"] = "clau_cognition"
    datapoint["rule_order"] = "53"
else:
    datapoint["subject_by_rule"] = "N/A"

indirect_str_list = ["주시겠", "주실수", "주시면안", "시면안", "주겠", "줄수", "주면안"]
want_str_list = ["시겠습", "시겠어", "시려고"]

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^ㄴ적이?있|ㄴ적이?없$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "experience periphrastic construction"
    datapoint["rule_order"] = "54"
elif [x for x in indirect_str_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "indirect speech act"
    datapoint["rule"] = "request"
    datapoint["rule_order"] = "55"
elif [x for x in want_str_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "56"
elif re_morphmatch(morphs, morph="^으?면도$", pos="^X$") and
    re_morphmatch(morphs, morph="^(겠|지않)$", pos="^(E|X)$") and

```

```

datapoint["zp_hono"] == "plus":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "suggestory formula"
datapoint["rule"] = "suggestory periphrastic construction"
datapoint["rule_order"] = "57"

elif re_morphmatch(morphs, morph="^(아?야하|[어여]야하)$", pos="^X$") and
    re_morphmatch(morphs, morph="^았$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "suggestory formula"
    datapoint["rule"] = "suggestory periphrastic construction"
    datapoint["rule_order"] = "58"

elif re_morphmatch(morphs, morph="^(있|있)$", pos="^E$") and
    re_morphmatch(morphs, morph="^어야지요?$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "suggestory formula"
    datapoint["rule"] = "suggestory periphrastic construction"
    datapoint["rule_order"] = "59"

elif re_morphmatch(morphs, morph="^(고[싶지?않?])$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction/plus"
    datapoint["rule_order"] = "60"

elif re_morphmatch(morphs, morph="^[ㄹ을]건가요?", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction/plus"
    datapoint["rule_order"] = "61"

elif re_morphmatch(morphs, morph="^[ㄹ을]것이$", pos="^X$") and
    re_morphmatch(morphs, morph="^ㄴ가요$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction/plus"
    datapoint["rule_order"] = "62"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^겠$", pos="^E$") and
    re_morphmatch(morphs, morph="^습니다$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "63"

elif re_morphmatch(morphs, morph="^(겠|겟)$", pos="^E$") and
    re_morphmatch(morphs, morph="^습니다$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "64"

elif re_morphmatch(morphs, morph="^(려고요|으려고요|려는것이|고자하|ㄹ려고하|기로하|ㄹ려구요|려구요|사려하)$", pos="^(X|E|V)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "65"

elif re_morphmatch(morphs, morph="^(려고요|으려고요|려는것이|고자하|ㄹ려고하|기로하|ㄹ려구요|려구요|사려하)$", pos="^(X|E|V)$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "66"

elif re_morphmatch(morphs, morph="^(려고요|으려고요|려는것이|고자하|ㄹ려고하|기로하|ㄹ려구요|려구요|사려하)$", pos="^(X|E|V)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "67"

```

```

려고요|으려고요|르려고요|려는것이|고자하|르려고하|기로하|르려구요|려구요|사려하)$", pos="^ (X|E|V) $") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "67"

elif re_morphmatch(morphs, morph="^[르을으]려고요?", pos="^E$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want verbal suffix"
datapoint["rule_order"] = "68"

elif re_morphmatch(morphs, morph="^아가$", pos="^X$") and
re_morphmatch(morphs, morph="^도$", pos="^V$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "permission"
datapoint["rule"] = "permission periphrastic construction"
datapoint["rule_order"] = "69"

elif re_morphmatch(morphs, morph="^(아보|어보|보|가보|해보|써보|타보|찾아보|재보|맛보)$",
pos="^ (X|V) $") and
re_morphmatch(morphs, morph="^(아도되|아도돼|르수있|면안되)$", pos="^X$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "permission"
datapoint["rule"] = "permission periphrastic construction"
datapoint["rule_order"] = "70"

elif re_morphmatch(morphs, morph="^(아도되|어도되|도되)$", pos="^X$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "permission"
datapoint["rule"] = "permission periphrastic construction"
datapoint["rule_order"] = "71"

elif re_morphmatch(morphs, morph="^(면되|으면되)$", pos="^X$") and
re_morphmatch(morphs, morph="^겠$", pos="^E$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "72"

elif re_morphmatch(morphs, morph="^(았|었)$", pos="^E$") and
re_morphmatch(morphs, morph="^어야지요?$", pos="^E$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "73"

elif re_morphmatch(morphs, morph="^(아야하|어야하|여야하|야하)$", pos="^X$") and
datapoint["zp_hono"] == "plus":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "obligation statement"
datapoint["rule"] = "obligation periphrastic construction"
datapoint["rule_order"] = "74"

elif re_morphmatch(morphs, morph="^(아야하|어야하|여야하|야하)$", pos="^X$") and
re_morphmatch(morphs, morph="^르겟같$", pos="^X$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "75"

elif re_morphmatch(morphs, morph="^(아야하|어야하|여야하|야하)$", pos="^X$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "obligation statement"
datapoint["rule"] = "obligation periphrastic construction"
datapoint["rule_order"] = "76"

elif re_morphmatch(morphs, morph="^(고싶|고싶지않)$", pos="^X$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "77"

```

```

elif re_morphmatch(morphs, morph="^을건가요$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "78"

elif re_morphmatch(morphs, morph="^[르을]것이$", pos="^X$") and
    re_morphmatch(morphs, morph="^ㄴ가요$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "79"

elif re_morphmatch(morphs, morph="^(을게요|ㄹ게요|게요|을께요|ㄹ께요|ㄹ게|ㄹ께|ㄹ건데요|ㄹ건가|ㄹ거라|"
    을건데요|는건데요|은건데요)?$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "80"

elif re_morphmatch(morphs, morph="^(을래요|ㄹ래요)$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "81"

elif re_morphmatch(morphs, morph="^(을래요|ㄹ래요|ㄹ겁니까)$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "82"

elif re_morphmatch(morphs, morph="^ㄹ것이$", pos="^X$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want verbal suffix"
    datapoint["rule_order"] = "83"

elif re_morphmatch(morphs, morph="^(고싶|ㄹ것이|고싶지않)$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "84"

elif re_morphmatch(morphs, morph="^(생기|나오|멀|멀쩡하|정교하|기품있|편리하|덜컹거리|싱싱하|저렴하|부드럽|"
    아슬하|저렴하|잠기|지나|사라지|조용하|나뉘|멀쩡하|딱맞|유효하|독하|생기|개운하|고급스러|독특하|좋아보이|뿌옇|"
    특이하|가동하|붐비|저렴|유치하|잘어울리|거대하|부풀어오르|특별하|한산하|훌륭하|열|견고하|터지|동일하|"
    알록달록하|달|어둠|저렴하|대단하|상영하|유명하|저렴하|기품있|경쾌하|튼튼하|화려하|많|밀리|열리|작동하|들어있|"
    닫|간단하|담백하|적당하|남|빼듯하|돌|차|늦|복잡하|기|짧|머|저렴하|아담하|가벼우|어둠|밝|쓸리|부드럽|달|다르|"
    크|작|적|활기차|비리|고소하|저렴하|달|싱싱하|맑|유명하|굳|운치있|흔잡하|독특하|똑같|조용하|한산하|매콤하|"
    새콤하|꼬불꼬불하|다르|징그럽|모자르|시끄럽|심하|초과하|위치하|심해지|높|어울리|고장나|남성스럽|상하|의혀지|"
    멋지|딸리|흔들리|작|싱겁|바꿔|깨지|끝나|이루하|귀엽|험하|무섭|문제없|번거롭|어렵|쉽|아름답|예쁘|답답하|더럽|"
    만만하|신기하|나|재밌|재미있|지루하|벗겨지|무겁|깨|튀|마르|위험하|이색적이|가볍|느끼하|맛있|불편하|시원하|덥|"
    춤|편하|불편하|출발하|편안하|아프|관계없|싸|비싸|가능하|불가능하|걸리|고쳐지|깨끗하|나오|내리|다니|줄어들|되|"
    들어오|마련되|나아지|맞|매우|맵|멀|가깝|보이|비슷하|비치되|시원하|시작하|없|아니|이|이기|이색적이|있|잘되|"
    충분하|\w+되|\w+여지|\w+ठ|$, pos="^(J|V|N)$"):
    datapoint["subject_by_rule"] = "none"
    datapoint["group"] = "exclude"
    datapoint["rule"] = "exclude predicate"
    datapoint["rule_order"] = "85"

elif re_morphmatch(morphs, morph="^(어있|아있|아지|어지|지|ㄹ만하|게되어있)$", pos="^X$"):
    datapoint["subject_by_rule"] = "none"
    datapoint["group"] = "exclude"
    datapoint["rule"] = "exclude predicate"
    datapoint["rule_order"] = "86"

elif re_morphmatch(morphs, morph="^(고싶|고싶지않)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"

```

```

datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "87"
elif re_morphmatch(morphs, morph="^(르것이|르거라)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "want statement"
        datapoint["rule"] = "want periphrastic construction"
        datapoint["rule_order"] = "88"
elif re_morphmatch(morphs, morph="^(\w+다고하|\w+라고하|다고하)$", pos="^X$"):
    datapoint["subject_by_rule"] = "none"
    datapoint["group"] = "exclude"
    datapoint["rule"] = "exclude predicate"
    datapoint["rule_order"] = "89"
elif re_morphmatch(morphs, morph="^고도$", pos="^E$") and
    re_morphmatch(morphs, morph="^심$", pos="^J$") and datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "want statement"
        datapoint["rule"] = "want periphrastic construction"
        datapoint["rule_order"] = "90"
else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(을까요|ㄹ까요|을까|ㄹ까)$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "suggestory formula"
        datapoint["rule"] = "suggestory verbal suffix"
        datapoint["rule_order"] = "91"
elif re_morphmatch(morphs, morph="^(을까요|ㄹ까요|을까|ㄹ까)$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "suggestory formula"
        datapoint["rule"] = "suggestory verbal suffix"
        datapoint["rule_order"] = "92"
elif re_morphmatch(morphs, morph="^(ㄴ적이?있|ㄴ적이?없)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "experience periphrastic construction"
        datapoint["rule_order"] = "93"
elif re_morphmatch(morphs, morph="^(ㄴ적이?있|ㄴ적이?없)$", pos="^X$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "experience periphrastic construction"
        datapoint["rule_order"] = "94"
elif re_morphmatch(morphs, morph="^더니$", pos="^E$") and
    datapoint["tense_zp"] == "past":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "experience periphrastic construction"
        datapoint["rule_order"] = "95"
elif re_morphmatch(morphs, morph="^([아어]드리)$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "politeness"
        datapoint["rule"] = "politeness auxiliary verb"
        datapoint["rule_order"] = "96"
elif re_morphmatch(morphs, morph="^(아버지|어버리|잃어버리)$", pos="^(X|V)$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "97"
elif re_morphmatch(morphs, morph="^(아보|어보|가보|해보|써보|타보|찾아보|재보)$", pos="^(X|V)$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "try"
        datapoint["rule"] = "modality auxiliary verb"
        datapoint["rule_order"] = "98"
else:
    datapoint["subject_by_rule"] = "N/A"

```

```

if datapoint["subject_by_rule"] == "N/A":
    if [x for x in performative_head_list if x in datapoint["verb"]] and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "performative"
        datapoint["rule"] = "performative predicate"
        datapoint["rule_order"] = "99"

    elif re_morphmatch(morphs, morph="^부탁하$", pos="^V$") and
        re_morphmatch(morphs, morph="^(면되|아도되|ㄹ수있)$", pos="^X$") and
        datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "permission"
        datapoint["rule"] = "permission periphrastic construction"
        datapoint["rule_order"] = "100"

    elif re_morphmatch(morphs, morph="^(기?원하|바라|기바라|추천하|필요하|필요없|ㄹ필요는없|빌|권하|지않아도되)$",
        pos="^(V|X|J)$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "performative"
        datapoint["rule"] = "performative predicate"
        datapoint["rule_order"] = "101"

    elif re_morphmatch(morphs, morph="^(는줄알|ㄴ줄알|을줄알|ㄹ줄알|ㄹ줄아|잘알|알|아|알게되|잘아|모르|는줄모르|
        을줄모르|ㄹ줄모르|어야할지모르|는지모르|ㄴ지모르|을지모르|ㄹ지모르|을수있을지모르|알아들|궁금하|고집하)$",
        pos="^(X|V|J)$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "cognition predicate"
        datapoint["rule_order"] = "102"

    elif re_morphmatch(morphs, morph="^(잊|잊어버리|믿|알아듣|외우|햇갈리)$", pos="^VS$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "cognition predicate"
        datapoint["rule_order"] = "103"

    elif [x for x in cognition_head_list if x in datapoint["verb"]] and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "cognition predicate"
        datapoint["rule_order"] = "104"

    elif re_morphmatch(morphs, morph="^(보|못보|살펴보|둘러보|보려오|듣|들어서|들|들어보|맛보|배고프|싱거우|
        간지럽)$", pos="^(V|J)$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "sensation predicate"
        datapoint["rule_order"] = "105"

    elif [x for x in sensation_head_list if x in datapoint["verb"]] and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "sensation predicate"
        datapoint["rule_order"] = "106"

    elif re_morphmatch(morphs, morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼)$",
        pos="^(J|V)$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "sensation predicate"
        datapoint["rule_order"] = "107"

    elif re_morphmatch(morphs, morph="^(괜찮|고맙|불편하|피곤하|지루하|흥미롭|곤란하|불쾌하|거북하|불안하|힘들|
        힘드|기힘들|좋|어쩔수없|별수없)$", pos="^(J|X)$") and
        datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "experience"
        datapoint["rule"] = "feeling predicate/plus"
        datapoint["rule_order"] = "108"

    elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하|고민되|고민하|놀라|
        당황하|놀래|놀랍|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이)$", pos="^(V|N)$") and
        datapoint["zp_hono"] == "plus":

```

```

datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "109"

elif re_morphmatch(morphs, morph="^(후회하|감사하|감사드리|감탄하|횡재하|들뜨|떨리|만족하|만족스럽|망설여지|
망설이|설|싫어하) $", pos="^(V|J)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "110"

elif [x for x in feeling_head_list if x in datapoint["verb"]] and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "111"

elif re_morphmatch(morphs, morph="^(의숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
기쁘|행복하|난감하|반갑|환영하|신기하|당황스럽|막막하|안타깝|부럽) $", pos="^(J|V)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "112"

elif re_morphmatch(morphs, morph="^(드시|안드|계시|말씀하|고례시) $", pos="^(V|X)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "113"

elif re_morphmatch(morphs, morph="^(드시|안드|계시|말씀하|고례시) $", pos="^(V|X)$"):
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "114"

elif re_morphmatch(morphs, morph="^말씀하$", pos="^V$") and
    re_morphmatch(morphs, morph="^아드리$", pos="^X$"):
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "politeness"
        datapoint["rule"] = "politeness predicate"
        datapoint["rule_order"] = "115"

elif re_morphmatch(morphs, morph="^축하드리$", pos="^V$"):
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "116"

elif re_morphmatch(morphs, morph="^(을것이|ㄹ것이|을거이|ㄹ거이) $", pos="^(X|E)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want (plus) periphrastic construction"
    datapoint["rule_order"] = "117"

elif re_morphmatch(morphs, morph="^(을것이|ㄹ것이|을거이|ㄹ거이) $", pos="^(X|E)$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "118"

elif re_morphmatch(morphs, morph="^(을것이|ㄹ것이|을거이|ㄹ거이) $", pos="^(X|E)$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "119"

elif re_morphmatch(morphs, morph="^지않을것이$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "guess"
    datapoint["rule"] = "guess periphrastic construction"
    datapoint["rule_order"] = "120"

elif re_morphmatch(morphs, morph="^진않$", pos="^X$") and
    re_morphmatch(morphs, morph="^을거이$", pos="^X$") and

```

```

datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "guess"
datapoint["rule"] = "guess periphrastic construction"
datapoint["rule_order"] = "121"

elif [x for x in cognition_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "122"

elif [x for x in performative_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "performative"
datapoint["rule"] = "performative predicate"
datapoint["rule_order"] = "123"

elif re_morphmatch(morphs, morph="^(기?원하|바라|기바라|추천하|필요하|필요없|ㄹ필요는없|지않아도되|빌|권하)$",
pos="^(V|X|J)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "performative"
datapoint["rule"] = "performative predicate"
datapoint["rule_order"] = "124"

elif re_morphmatch(morphs, morph="^(는줄알|ㄴ줄알|을줄알|ㄹ줄알|ㄹ줄아|잘알|알|아|알게되|잘아|모르|는줄모르|
ㄴ줄모르|을줄모르|ㄹ줄모르|어야할지모르|는지모르|을지모르|ㄹ지모르|을수있을지모르|ㄹ수있을지모르|알아들|궁금하|
고집하)$", pos="^(X|V)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "125"

elif [x for x in cognition_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "126"

elif re_morphmatch(morphs, morph="^(잊|잊어버리|믿|알아듣|외우|헷갈리)$", pos="^VS$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "127"

elif re_morphmatch(morphs, morph="^(보|못보|살펴보|둘러보|보러오|듣|들어서|들|들어보|맛보|배고프|싱거우|
간지럽)$", pos="^V$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "128"

elif [x for x in sensation_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "129"

elif re_morphmatch(morphs, morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼)$",
pos="^(J|V)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "130"

elif re_morphmatch(morphs, morph="^(고맙|불편하|피곤하|지루하|흥미롭|곤란하|불쾌하|거북하|불안하|힘들|힘드|
기힘들)$", pos="^(J|X)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "131"

elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하|고민되|고만하|놀라|

```

```

당황하|놀래|놀랍|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이)$", pos="^(V|J|N)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "132"

elif re_morphmatch(morphs, morph="^(후회하|감사하|감사드리|감탄하|횡재하|들뜨|떨리|만족하|만족스럽|망설여지|
망설이|싫|싫어하)$", pos="^V$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "133"

elif [x for x in feeling_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "134"

elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
기쁘|행복하|난감하|반갑|환영하|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
datapoint["zp_mood"] == "ques":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "135"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if [x for x in cognition_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "136"

elif re_morphmatch(morphs, morph="^(면되|으면되|기만하면되|아도되|어도되|아가|도되|지않아도되|야야하|어야하|
여야하|야하|어야할것이|야야할것이|어야할것같|야야되|어야되|야되|야야죠|야야죠|야지요|으면안되|면안되)$",
pos="^(X|E)$") and
datapoint["zp_hono"] == "plus":
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "permission"
datapoint["rule"] = "permission periphrastic construction/plus"
datapoint["rule_order"] = "137"

elif re_morphmatch(morphs, morph="^(야야겠|어야겠)$", pos="^X$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "138"

elif re_morphmatch(morphs, morph="^(야야하|어야하|여야하|야하)$", pos="^X$") and
re_morphmatch(morphs, morph="^겠$", pos="^E$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "139"

elif [x for x in performative_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "performative"
datapoint["rule"] = "performative predicate"
datapoint["rule_order"] = "140"

elif re_morphmatch(morphs, morph="^(기?원하|바라|기바라|추천하|필요하|필요없|ㄹ필요는없|빌|권하)$",
pos="^(V|X|J)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "performative"
datapoint["rule"] = "performative predicate"
datapoint["rule_order"] = "141"

elif re_morphmatch(morphs, morph="^(는줄알|ㄴ줄알|을줄알|ㄹ줄알|ㄹ줄아|잘알|알|아|알게되|잘아|모르|는줄모르|
줄모르|ㄹ줄모르|어야할지모르|는지모르|ㄴ지모르|을지모르|ㄹ지모르|을수있을지모르|알아들|궁금하|고집하)$",

```

```

pos="^(X|V|J)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "142"

elif re_morphmatch(morphs, morph="^(잊|잊어버리|믿|알아듣|외우|헷갈리)$", pos="^VS") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "cognition predicate"
datapoint["rule_order"] = "143"

elif re_morphmatch(morphs, morph="^(보|못보|살펴보|둘러보|보려오|듣|들어서|들|들어보|맛보|배고프|싱거우|
간지럽)$", pos="^(V|J)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "144"

elif [x for x in sensation_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "145"

elif re_morphmatch(morphs, morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼)$",
pos="^(J|V)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "sensation predicate"
datapoint["rule_order"] = "146"

elif re_morphmatch(morphs, morph="^(괜찮|고맙|불편하|피곤하|지루하|흥미롭|곤란하|불쾌하|거북하|불안하|힘들|
힘드|기힘들|좋|어쩔수없|별수없)$", pos="^(J|X)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "147"

elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하|고민되|고민하|놀라|
당황하|놀래|놀람|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이)$", pos="^(V|N)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "148"

elif re_morphmatch(morphs, morph="^(후회하|감사하|감사드리|감탄하|횡재하|들뜨|떨리|만족하|만족스럽|망설여지|
망설이|싫|싫어하)$", pos="^(V|J)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "149"

elif [x for x in feeling_head_list if x in datapoint["verb"]] and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "150"

elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
기쁘|행복하|난감하|반갑|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
re_morphmatch(morphs, morph="^겠$", pos="^E$"):
datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "experience"
datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "151"

elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
기쁘|행복하|난감하|반갑|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and
datapoint["zp_mood"] == "decl":
datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "experience"

```

```

datapoint["rule"] = "feeling predicate"
datapoint["rule_order"] = "152"
elif re_morphmatch(morphs, morph="^(드리)$", pos="^V$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "politeness"
        datapoint["rule"] = "politeness predicate/auxiliary verb"
        datapoint["rule_order"] = "153"
elif re_morphmatch(morphs, morph="^드리$", pos="^V$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "politeness"
        datapoint["rule"] = "politeness predicate"
        datapoint["rule_order"] = "154"
elif re_morphmatch(morphs, morph="^드리$", pos="^V$") and
    datapoint["verb_syn"] == "clau":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "politeness"
        datapoint["rule"] = "politeness predicate"
        datapoint["rule_order"] = "155"
elif re_morphmatch(morphs, morph="^(아드리|어드리)$", pos="^X$"):
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness auxiliary verb"
    datapoint["rule_order"] = "156"
elif re_morphmatch(morphs, morph="^(가져다드리|알려드리|보여드리|해드리|연락드리|도와드리|돌려드리|전해드리|전화드리|갖다드리)$", pos="^V$"):
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "politeness"
    datapoint["rule"] = "politeness predicate"
    datapoint["rule_order"] = "157"
elif re_morphmatch(morphs, morph="^(아버지|어버리|잃어버리)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "subjectification"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "158"
elif re_morphmatch(morphs, morph="^(아버지|어버리|잃어버리)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectification"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "159"
elif re_morphmatch(morphs, morph="^(ㄹ수밖에없|는수밖에없)$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "possibility"
        datapoint["rule"] = "possibility periphrastic construction"
        datapoint["rule_order"] = "160"
elif re_morphmatch(morphs, morph="^(ㄹ수밖에없|는수밖에없)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "possibility"
        datapoint["rule"] = "possibility periphrastic construction"
        datapoint["rule_order"] = "161"
elif re_morphmatch(morphs, morph="^(야하|어야하|여야하|여야하|야하)$", pos="^X$") and
    re_morphmatch(morphs, morph="^겠$", pos="^E$") and
        datapoint["verb_syn"] == "clau":
            datapoint["subject_by_rule"] = "1sh+"
            datapoint["group"] = "want statement"
            datapoint["rule"] = "want periphrastic construction"
            datapoint["rule_order"] = "162"
elif re_morphmatch(morphs, morph="^야하$", pos="^X$") and
    re_morphmatch(morphs, morph="^는데요?$", pos="^E$") and
        datapoint["zp_mood"] == "decl":
            datapoint["subject_by_rule"] = "1sh+"
            datapoint["group"] = "obligation statement"
            datapoint["rule"] = "obligation periphrastic construction"
            datapoint["rule_order"] = "163"
elif re_morphmatch(morphs, morph="^여보$", pos="^X$") and
    re_morphmatch(morphs, morph="^려고요$", pos="^E$") and
        datapoint["zp_mood"] == "decl":

```

```

datapoint["subject_by_rule"] = "lsh+"
datapoint["group"] = "try"
datapoint["rule"] = "modality auxiliary verb"
datapoint["rule_order"] = "164"
elif re_morphmatch(morphs, morph="^어보$", pos="^X$") and
    re_morphmatch(morphs, morph="^려고요$", pos="^E$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "try"
    datapoint["rule"] = "modality auxiliary verb"
    datapoint["rule_order"] = "165"
else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(아야겠|어야겠|어야|야하|야)$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "obligation statement"
    datapoint["rule"] = "obligation periphrastic construction/plus"
    datapoint["rule_order"] = "166"
elif re_morphmatch(morphs, morph="^(려고하|으려고하|사려하|려|려하|으려하|으려고|려고|어보려고하|려고그러|
    려고요|으려고요|르려고요|려는것이|고자하|르려고하|기로하|르려구요|려구요|을것이|르것이|을거이|르거이)$",
    pos="^(X|E)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction/plus"
    datapoint["rule_order"] = "167"
elif re_morphmatch(morphs, morph="^(아야하|어야하|여야하)$", pos="^X$") and
    re_morphmatch(morphs, morph="^겠$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "obligation statement"
    datapoint["rule"] = "obligation periphrastic construction/plus"
    datapoint["rule_order"] = "168"
elif re_morphmatch(morphs, morph="^가사$", pos="^V$") and
    re_morphmatch(morphs, morph="^(어야하|면되)$", pos="^X$"):
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "obligation statement"
    datapoint["rule"] = "obligation periphrastic construction/plus"
    datapoint["rule_order"] = "169"
elif re_morphmatch(morphs, morph="^으면하$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction/plus"
    datapoint["rule_order"] = "170"
elif re_morphmatch(morphs, morph="^(으면안되|면안되)$", pos="^X$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "permission"
    datapoint["rule"] = "permission periphrastic construction"
    datapoint["rule_order"] = "171"
elif re_morphmatch(morphs, morph="^(으면안되|면안되)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "prohibition"
    datapoint["rule"] = "prohibition periphrastic construction"
    datapoint["rule_order"] = "172"
elif re_morphmatch(morphs, morph="^아해$", pos="^X$") and
    re_morphmatch(morphs, morph="^(ㅂ니다|아요)$", pos="^E$"):
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "obligation statement"
    datapoint["rule"] = "obligation periphrastic construction"
    datapoint["rule_order"] = "173"
else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(아야겠|어야겠)$", pos="^X$") and

```

```

datapoint["verb_syn"] == "clau":
datapoint["subject_by_rule"] = "1sh+"
datapoint["group"] = "want statement"
datapoint["rule"] = "want periphrastic construction"
datapoint["rule_order"] = "174"

elif re_morphmatch(morphs, morph="^(으면하|ㄹ까하)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "175"

elif re_morphmatch(morphs, morph="^(을뿐이|ㄹ뿐이|는것뿐이|ㄴ것뿐이)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "judgement periphrastic construction"
    datapoint["rule_order"] = "176"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(아보|어보|보|가보|해보|써보|타보|찾아보|재보|맛보)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "try"
    datapoint["rule"] = "modality auxiliary verb"
    datapoint["rule_order"] = "177"

elif re_morphmatch(morphs, morph="^(아보|어보|보|가보|해보|써보|타보|찾아보|재보|맛보)$", pos="^X$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "try"
    datapoint["rule"] = "modality auxiliary verb"
    datapoint["rule_order"] = "178"

elif re_morphmatch(morphs, morph="^(어야|아야|야하)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "179"

elif re_morphmatch(morphs, morph="^(면되|으면되|기만하면되|아도되|어도되|아가|도되|지않아도되|야야하|어야하|여야하|여야하|야하|어야할것이|야야할것이|어야할것같|야야되|어야되|야되|야야조|야야조|야지요|으면안되|면안되)$",
    pos="^(X|E)$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "permission"
    datapoint["rule"] = "permission periphrastic construction"
    datapoint["rule_order"] = "180"

elif re_morphmatch(morphs, morph="^(면되|으면되|기만하면되|아도되|어도되|아가|도되|지않아도되|야야하|어야하|여야하|야하|어야할것이|야야할것이|어야할것같|야야되|어야되|야되|야야조|야야조|야지요|으면안되|면안되)$",
    pos="^X$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "permission"
    datapoint["rule"] = "permission periphrastic construction"
    datapoint["rule_order"] = "181"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(ㄹ수있|을수있|ㄹ수는있|을수는있|ㄹ수도있|을수도있|ㄹ수가있|을수가있|ㄹ수없|ㄹ수도없|을수없|ㄹ수는없|을수는없|ㄹ수밖에없|을수밖에없|ㄴ수밖에없|는수밖에없|ㄹ수가없|을수가없|ㄹ리가없|을리가없|ㄹ필요는없|을필요는없|ㄹ필요가없|ㄹ필요없|지않아도되|을필요없|ㄹ필요가있|을필요가있|ㄹ뻔하|을뻔하)$",
    pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction/plus"
    datapoint["rule_order"] = "182"

elif re_morphmatch(morphs, morph="^(나하|ㄹ뻔하|을뻔하)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction"

```

```

datapoint["rule_order"] = "183"
elif re_morphmatch(morphs, morph="^(ㄹ수있|을수있|ㄹ수는있|을수는있|ㄹ수도있|을수도있|ㄹ수가있|을수가있)$",
pos="^X$") and
    re_morphmatch(morphs, morph="^(겠)$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction"
    datapoint["rule_order"] = "184"

elif re_morphmatch(morphs, morph="^(ㄹ수있|을수있|ㄹ수는있|을수는있|ㄹ수도있|을수도있|ㄹ수가있|을수가있)$",
pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction"
    datapoint["rule_order"] = "185"

elif re_morphmatch(morphs, morph="^(ㄹ수없|ㄹ수도없|을수없|ㄹ수는없|을수는없|ㄹ수밖에없|을수밖에없|ㄴ수밖에없|
는수밖에없|ㄹ수가없|을수가없|ㄹ리가없|을리가없)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "impossibility"
    datapoint["rule"] = "impossibility periphrastic construction"
    datapoint["rule_order"] = "186"

elif re_morphmatch(morphs, morph="^(ㄹ수없|ㄹ수도없|을수없|ㄹ수는없|을수는없|ㄹ수밖에없|을수밖에없|ㄴ수밖에없|
는수밖에없|ㄹ수가없|을수가없|ㄹ리가없|을리가없)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "unnecessary"
    datapoint["rule"] = "unnecessary periphrastic construction"
    datapoint["rule_order"] = "187"

elif re_morphmatch(morphs, morph="^(ㄹ수있을것같|을수있을것같|ㄹ수있다는것이|을수있다는것이|ㄹ수있을것이|
을수있을것이)$", pos="^X$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction"
    datapoint["rule_order"] = "188"

elif re_morphmatch(morphs, morph="^(ㄹ수있을것같|을수있을것같|ㄹ수있다는것이|을수있다는것이|ㄹ수있을것이|
을수있을것이)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction"
    datapoint["rule_order"] = "189"

elif re_morphmatch(morphs, morph="^(ㄹ수있|을수있|ㄹ수는있|을수는있|ㄹ수도있|을수도있|ㄹ수가있|을수가있|ㄹ수없|
ㄹ수도없|을수없|ㄹ수는없|을수는없|ㄹ수밖에없|을수밖에없|ㄴ수밖에없|는수밖에없|ㄹ수가없|을수가없|ㄹ리가없|
을리가없|ㄹ필요는없|을필요는없|ㄹ필요가없|ㄹ필요없|을필요없|지않아도되|ㄹ필요가있|을필요가있|ㄹ뻔하|을뻔하)$",
pos="^X$") and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "possibility"
    datapoint["rule"] = "possibility periphrastic construction/plus"
    datapoint["rule_order"] = "190"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^(아주|어주|주|가져다주|찍어주|안해주|틀어주|도와주|넣어주|거슬러주)$",
pos="^(X|V)$") and datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "benefactive"
    datapoint["rule"] = "modality auxiliary verb/plus"
    datapoint["rule_order"] = "191"

elif re_morphmatch(morphs, morph="^(아놓|어놓|열어놓|정해놓|걸어놓)$", pos="^(X|V)$") and
    datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "subjectivication"
    datapoint["rule"] = "aspect auxiliary verb/plus"
    datapoint["rule_order"] = "192"

elif "다놓" in datapoint["verb"] and
    datapoint["zp_hono"] == "plus":

```

```

datapoint["subject_by_rule"] = "2sh+"
datapoint["group"] = "subjectivication"
datapoint["rule"] = "aspect auxiliary verb/plus"
datapoint["rule_order"] = "193"

elif re_morphmatch(morphs, morph="^(아두|어두|걸어두|세워두|놓아두|맡겨두|넣어두)$", pos="^(X|V)$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb/plus"
        datapoint["rule_order"] = "194"

elif re_morphmatch(morphs, morph="^(아놓|어놓|열어놓|정해놓|걸어놓)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "195"

elif "다놓" in datapoint["verb"] and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "196"

elif re_morphmatch(morphs, morph="^(아두|어두|걸어두|세워두|놓아두|맡겨두|넣어두)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "197"

elif re_morphmatch(morphs, morph="^(아주|어주|주|가져다주|찍어주|안해주|틀어주|도와주|넣어주|거슬러주)$",
    pos="^(X|V)$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "benefactive"
        datapoint["rule"] = "modality auxiliary verb"
        datapoint["rule_order"] = "198"

elif re_morphmatch(morphs, morph="^(아주|어주|주|가져다주|찍어주)$", pos="^(X|V)$") and
    re_morphmatch(morphs, morph="^어야하$", pos="^X$"):
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "obligation statement"
        datapoint["rule"] = "obligation periphrastic construction"
        datapoint["rule_order"] = "199"

elif re_morphmatch(morphs, morph="^(아보|어보|가보|해보|써보|타보|찾아보|재보)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "try"
        datapoint["rule"] = "modality auxiliary verb"
        datapoint["rule_order"] = "200"

else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^겠$", pos="^E$") and
    datapoint["zp_hono"] == "plus":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "guess/future/will"
        datapoint["rule"] = "guess/future/will infix"
        datapoint["rule_order"] = "201"

elif re_morphmatch(morphs, morph="^겠$", pos="^E$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "guess/future/will"
        datapoint["rule"] = "guess/future/will infix"
        datapoint["rule_order"] = "202"

elif "겠습니까" in datapoint["verb"] and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "will"
        datapoint["rule"] = "will infix"
        datapoint["rule_order"] = "203"

elif re_morphmatch(morphs, morph="^(아놓|어놓|열어놓|정해놓|걸어놓)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"

```

```

datapoint["rule_order"] = "204"
elif re_morphmatch(morphs, morph="^(아두|어두|걸어두|세워두|놓아두|맡겨두|넣어두)$", pos="^(X|V)$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "205"

elif re_morphmatch(morphs, morph="^(아주|어주|주|가져다주|찍어주|안해주|틀어주|도와주|넣어주|거슬러주)$",
    pos="^(X|V)$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "benefactive"
        datapoint["rule"] = "modality auxiliary verb"
        datapoint["rule_order"] = "206"

elif re_morphmatch(morphs, morph="^어다$", pos="^E$") and
    re_morphmatch(morphs, morph="^놓$", pos="^V$"):
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "subjectivication"
        datapoint["rule"] = "aspect auxiliary verb"
        datapoint["rule_order"] = "207"

elif re_morphmatch(morphs, morph="^(을것같|ㄹ것같|는것같|ㄴ것같|은것같|은듯하|ㄴ거같|을거같|ㄹ거같|는거같|
    던것같)$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "guess"
        datapoint["rule"] = "guess periphrastic construction"
        datapoint["rule_order"] = "208"
else:
    datapoint["subject_by_rule"] = "N/A"

if datapoint["subject_by_rule"] == "N/A":
if re_morphmatch(morphs, morph="^고있$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "progress"
        datapoint["rule"] = "progress aspect"
        datapoint["rule_order"] = "209"

elif re_morphmatch(morphs, morph="^고있$", pos="^X$") and
    datapoint["zp_mood"] == "ques":
        datapoint["subject_by_rule"] = "2sh+"
        datapoint["group"] = "progress"
        datapoint["rule"] = "progress aspect"
        datapoint["rule_order"] = "210"

elif re_morphmatch(morphs, morph="^게도$", pos="^X$") and
    datapoint["zp_mood"] == "decl":
        datapoint["subject_by_rule"] = "1sh+"
        datapoint["group"] = "passive"
        datapoint["rule"] = "passive aspect"
        datapoint["rule_order"] = "211"

elif datapoint["zp_hono"] == "plus":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "honorification"
    datapoint["rule"] = "hono total"
    datapoint["rule_order"] = "212"
else:
    datapoint["subject_by_rule"] = "N/A"

```

Appendix 2: Code für die gesamte Regel für den ersten Schritt des Basissystems

```

indirect_str_list = ["주시겠", "주실수", "주시면안", "시면안", "주겠", "줄수", "주면안"]
cognition_head_list = ["예정", "생각", "기억", "확인", "깜박", "예상", "착각", "이해", "확실", "집중", "계획", "숙지",
                       "오해", "파악", "동의", "사과", "깜빡"]
sensation_head_list = ["시음", "민감", "구경", "시원"]
feeling_head_list = ["실망", "공포", "영광", "장담", "한시름", "미안", "실례", "의아", "질색", "놀랍", "환영"]

if datapoint["zp_mood"] == "impe":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "mood derivable"
    datapoint["rule"] = "impe"
    datapoint["rule_order"] = "1"

elif datapoint["zp_mood"] == "sugg":
    datapoint["subject_by_rule"] = "1ph-"
    datapoint["group"] = "mood derivable"
    datapoint["rule"] = "sugg"
    datapoint["rule_order"] = "2"

elif re_morphmatch(morphs, morph="^(으?세요|으?세여|으?십시오|으?십시오) $", pos="^E$") and
    datapoint["zp_mood"] == "decl":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "mood derivable"
    datapoint["rule"] = "impe"
    datapoint["rule_order"] = "3"

elif [x for x in indirect_str_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "ques":
    datapoint["subject_by_rule"] = "2sh+"
    datapoint["group"] = "indirect speech act"
    datapoint["rule"] = "request"
    datapoint["rule_order"] = "4"

elif re_morphmatch(morphs, morph="^(ㄹ까하|기로하|ㄹ까보|을까보) $", pos="^X$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"
    datapoint["rule"] = "want periphrastic construction"
    datapoint["rule_order"] = "5"

elif re_morphmatch(morphs, morph="^(ㄹ까심|을까심|ㄴ가하|은가하|으?려니하|으?려니심) $", pos="^X$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "guess statement"
    datapoint["rule"] = "guess periphrastic construction"
    datapoint["rule_order"] = "6"

elif re_morphmatch(morphs, morph="^(ㄴ가보|은가보) $", pos="^X$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "judgement statement"
    datapoint["rule"] = "judgements periphrastic construction"
    datapoint["rule_order"] = "7"

elif re_morphmatch(morphs, morph="^(나심|은가심|ㄴ가심|나심|라심|지심) $", pos="^X$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "doubt statement"
    datapoint["rule"] = "doubt periphrastic construction"
    datapoint["rule_order"] = "8"

elif re_morphmatch(morphs, morph="^(아?야하|어야하) $", pos="^X$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["tense_zp"] == "past" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "regret statement"
    datapoint["rule"] = "regret periphrastic construction"
    datapoint["rule_order"] = "9"

elif re_morphmatch(morphs, morph="^(을게요|ㄹ게요|게요|을께요|ㄹ께요) ?$", pos="^E$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "1sh+"
    datapoint["group"] = "want statement"

```

```

datapoint["rule"] = "want verbal suffix"
datapoint["rule_order"] = "10"

elif re_morphmatch(morphs, morph="^(는줄알|ㄴ줄알|을줄알|ㄹ줄알|ㄹ줄아|잘알|알|아|알게되|잘아|모르|는줄모르|
    을줄모르|ㄹ줄모르|어야할지모르|는지모르|ㄴ지모르|을지모르|ㄹ지모르|을수있을지모르|알아들|궁금하|고집하)$",
    pos="^(X|V|J)") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "cognition predicate"
    datapoint["rule_order"] = "11"

elif re_morphmatch(morphs, morph="^(잊|잊어버리|믿|알아듣|외우|헷갈리)$", pos="^V$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "cognition predicate"
    datapoint["rule_order"] = "12"

elif re_morphmatch(morphs, morph="^(보|못보|살펴보|둘러보|보라오|듣|들어서|들|들어보|맛보|배고프|싱거우|
    간지럽)$", pos="^(V|J)$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "sensation predicate"
    datapoint["rule_order"] = "13"

elif [x for x in sensation_head_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "sensation predicate"
    datapoint["rule_order"] = "14"

elif re_morphmatch(morphs, morph="^(춥|따뜻하|덥|후텁지근하|찝찝하|아프|쓰라리|만지|만져보|느끼)$",
    pos="^(J|V)$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "sensation predicate"
    datapoint["rule_order"] = "15"

elif re_morphmatch(morphs, morph="^(괜찮|고맙|불편하|피곤하|지루하|흥미롭|곤란하|불쾌하|거북하|불안하|힘들|
    힘드|기힘들|좋|어쩔수없|별수없)$", pos="^(J|X)$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "16"

elif re_morphmatch(morphs, morph="^(좋아하|부담되|기대되|마음에들|기대하|걱정되|걱정하|고민되|고민하|놀라|
    당황하|놀래|놀랍|행복하|미안하|죄송하|죄송|답답하|부담스럽|의아하|질색이)$", pos="^(V|N)$") and
    datapoint["zp_mood"] == "decl" and datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "17"

elif re_morphmatch(morphs, morph="^(후회하|감사하|감사드리|감탄하|횡재하|들뜨|떨리|만족하|만족스럽|망설여지|
    망설이|싫|싫어하)$", pos="^(V|J)$") and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "18"

elif [x for x in feeling_head_list if x in datapoint["verb"]] and
    datapoint["zp_mood"] == "decl" and
    datapoint["zp_hono"] != "plus":
    datapoint["subject_by_rule"] = "lsh+"
    datapoint["group"] = "experience"
    datapoint["rule"] = "feeling predicate"
    datapoint["rule_order"] = "19"

elif re_morphmatch(morphs, morph="^(익숙하|자부하|망설여지|망설이|아쉽|상관없|즐겁|재미있|재미없|즐거우|무섭|
    무섭)"

```

```
기쁘|행복하|난감하|반갑|신기하|당황스럽|막막하|안타깝|부럽)$", pos="^(J|V)$") and  
datapoint["zp_mood"] == "decl" and  
datapoint["zp_hono"] != "plus":  
datapoint["subject_by_rule"] = "lsh+"  
datapoint["group"] = "experience"  
datapoint["rule"] = "feeling predicate"  
datapoint["rule_order"] = "20"  
elif re_morphmatch(morphs, morph="^(더니)$", pos="^E$") and  
datapoint["verb_syn"] == "clau" and  
datapoint["tense_zp"] == "past" and  
datapoint["zp_hono"] != "plus":  
datapoint["subject_by_rule"] = "lsh+"  
datapoint["group"] = "clau"  
datapoint["rule"] = "clau_class_C"  
datapoint["rule_order"] = "21"  
else:  
datapoint["subject_by_rule"] = "N/A"
```

Lebenslauf

Persönliche Daten

Geburtsdatum: 31.12.1987
Geburtsort: Seoul, Südkorea

Ausbildung

2016 – 2024	Promotion der Linguistik Fakultät für Philologie, Ruhr-Universität Bochum
2010 – 2013	M.A.-Studium der Linguistik Germanistisches Institut, Sungkyunkwan Universität, Südkorea
2006 – 2010	B.A.-Studium der Linguistik Germanistisches Institut, Sungkyunkwan Universität, Südkorea
2003 – 2006	Sudo Girls' High School, Südkorea

Publikation

- Arum Park & Munpyo Hong. 2014. Hybrid approach to zero subject resolution for multilingual MT: Spanish-to-Korean cases. In *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation*, 254–261.
- Arum Park & Munpyo Hong. 2015a. Subjektellipse im Koreanischen und deren Behandlung für die maschinelle Übersetzung ins Deutsche. *Koreanische Gesellschaft für Germanistik* 56(1). 197–223.
- Arum Park & Munpyo Hong. 2015b. Objektellipse im Koreanischen und deren Behandlung für die maschinelle Übersetzung ins Deutsche. *Koreanische Gesellschaft für deutsche Sprache und Literatur* 70. 1–21.
- Arum Park, Seunghee Lim, & Munpyo Hong. 2015. Zero object resolution in Korean. In *Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation*, 439–448.