

Dänische L2-Aussprache von Lernern mit Deutsch als Ausgangssprache. Fehler und Fehlerschwere

Dissertation
zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Lisa Tulaja
Kiel

Februar 2019

Erstgutachter:

Zweitgutachterin:

Tag der mündliche Prüfung:

Durch den Prodekan für Studium und Lehre,
Prof. Dr. Ulrich Müller, zum Druck genehmigt:

Prof. Dr. Steffen Höder

Prof. Dr. Inger Petersen

13. 5. 2019

21. 2. 2020

Danksagung

Die vorliegende Dissertation ist im Rahmen des Graduiertenprogramms *Lehramt in Bewegung* der Christian-Albrechts-Universität zu Kiel entstanden (2015–2018). Den Projektverantwortlichen – insbesondere Vizepräsidentin Prof. Dr. Ilka Parchmann – sowie meinem Betreuer Prof. Dr. Steffen Höder möchte ich an dieser Stelle für die Möglichkeit zur Bearbeitung dieses Projekts herzlich danken.

Von ganzem Herzen möchte ich Prof. Dr. Steffen Höder außerdem für seine unermüdliche Hilfsbereitschaft und die inspirierende Zusammenarbeit danken, mit der er mich von Beginn des Betreuungsverhältnisses an unterstützt und begleitet hat.

Prof. Dr. Inger Petersen danke ich für die Bereitschaft, sich meiner Arbeit als Zweitgutachterin anzunehmen.

Ohne die Erlaubnis des Ministeriums für Schule und Berufsbildung hätten die der Arbeit zu Grunde liegenden Datenerhebung nicht stattfinden können. Für das mir dort entgegengebrachte Vertrauen bin ich deshalb sehr dankbar. Besonders möchte ich allen teilnehmenden Lehrkräften und Schülerinnen und Schülern danken, deren freiwillige Partizipation grundlegend für die Arbeit war. Auch Renate Jacob vom Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH) möchte ich in diesem Sinne für die stets gute und enge Zusammenarbeit danken.

In Hinblick auf die im Ausland durchgeführten Erhebungen bin ich den dänischen Universitäten in Kopenhagen und Aarhus zu Dank verpflichtet. Insbesondere möchte ich Prof. Dr. Ocke-Schwen Bohn herzlich danken, dass er mich nicht nur mit offenen Armen in Aarhus empfangen hat, sondern mich tatkräftig unterstützte und durch konstruktive Anmerkungen inspiriert hat. Für die Möglichkeit zur Datenerhebung in Kopenhagen sei dem Center for Internationalisering og Parallelsproglighed (CIP) und dessen Leitung gedankt sowie insbesondere Martin Carlsholt Unger, der mir während meines Aufenthalts in kleinen und großen Fragen zur Seite stand.

Prof. Dr. Alena Witzlack-Makarevich und Nils Machts danke ich für die bewundernswerte Geduld bei der Beantwortung meiner statistischen Fragen und die Unterstützung bei der Datenaufbereitung, der ich anfangs ohne technische Kompetenz gegenüberstand.

Den studentischen Hilfskräften Johanna Nickel, Lisa-Marie Stein und Merve Rucks danke ich für ihre tatkräftige Unterstützung während und nach der Datenerhebung. Auch Milena Jacobsen und Anna Kaapke danke ich, dass sie mich mit ihrer freiwilligen Mitarbeit bei der Erstellung der Kontrollaufnahmen unterstützt haben.

Für das Durchsehen und die Korrektur der Arbeit sowie fachlich konstruktive und anregende Gespräche danke ich meinen Kollegen und Freunden, die mir in dieser Hinsicht unermüdlich zur Seite standen.

Nicht zuletzt gilt mein Dank meiner Familie und meinen Liebsten, die mich auf dem gesamten Weg unterstützt und motiviert haben.

Meinem Vater gewidmet

Inhalt

Tabellen 11

Abbildungen 13

- 1 Einleitung 15
 - 1.1 Gegenstand und Ziele der Arbeit 15
 - 1.2 Gliederung 20
- 2 Explorative Vorstudie 21
 - 2.1 Befragung 21
 - 2.2 Ergebnisse 24
 - 2.3 Folgerungen 27
- 3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner 29
 - 3.1 Fehler und Lernaltersprache 29
 - 3.2 Fehleranalyse 32
 - 3.2.1 Fehlerklassifikation 34
 - 3.2.2 Fehlerursachen und Fehlertypen 35
 - 3.2.2.1 Contrastive Analysis Hypothesis 37
 - 3.2.2.2 Speech Learning Model 39
 - 3.2.2.3 Perceptual Assimilation Model 41
 - 3.2.2.4 Zusammenfassung und Erklärungsmodell der vorliegenden Studie 43
 - 3.2.3 Vorgehensweise 48
 - 3.3 Kontrastive phonetische Analyse Deutsch–Dänisch 50
 - 3.3.1 Auswahl des Transkriptionssystems 51
 - 3.3.2 Segmentale Merkmale 53
 - 3.3.2.1 Konsonanten 53
 - 3.3.2.2 Monophthonge 57
 - 3.3.2.3 Polyphthonge 60
 - 3.3.3 Weitere Merkmale 62
 - 3.3.3.1 Vokalische Quantität 62
 - 3.3.3.2 Phonologische Prozesse 63
 - 3.3.4 Zusammenfassung 65
 - 3.4 Empirischer Abgleich 67
 - 3.4.1 Vorgehen bei der Datenerhebung und -aufbereitung 67
 - 3.4.1.1 Auswahl der Zielwörter 68
 - 3.4.1.2 Schulen und Probanden 77
 - 3.4.1.3 Datenerhebung 77
 - 3.4.1.4 Datenaufbereitung und Fehleridentifikation 78
 - 3.4.2 Fehleranalyse deutscher L2-Dänischlerner 79

3.4.2.1	Konsonanten	79
3.4.2.2	Monophthonge	85
3.4.2.3	Polyphthonge	91
3.4.2.4	Vokalische Quantität	97
3.4.2.5	Phonologische Prozesse	101
3.4.2.6	Zusammenfassung	105
3.4.3	Auswertung der sprecherbiographischen Daten	106
3.4.3.1	Hypothesen zu Zusammenhängen zwischen sprecherbiographischen Metadaten und Fehlern	107
3.4.3.2	Häufigkeit der Fehler nach sprachlichem Hintergrund	110
3.4.3.3	Häufigkeit der Fehler nach Anwendungshäufigkeit der Zielsprache	112
3.4.3.4	Häufigkeit der Fehler nach Geschlecht	114
3.4.3.5	Häufigkeit der Fehler nach Lehrkraft	114
3.4.3.6	Zusammenfassung	119
3.4.4	Diskussion der Ergebnisse	121
4	Studie 2: Bewertung der L2-Fehler deutschsprachiger Lerner durch dänische L1-Sprecher	123
4.1	L2-Aussprachebewertung durch L1-Sprecher	123
4.1.1	Forschung zur Fehlerschwere bis Mitte der 1990er Jahre	124
4.1.2	Differenzierte Termini zur Bewertung ab 1995	127
4.1.2.1	Akzent	128
4.1.2.2	Verständlichkeit	129
4.1.3	Bewertungsdimension Akzeptabilität	130
4.1.4	Zusammenfassung und Implikationen	135
4.2	Empirische Studie	139
4.2.1	Fragestellung und Hypothesen	140
4.2.2	Design und Methode	145
4.2.2.1	Entscheidungen zu Sprachaufnahmen und Stimuli	145
4.2.2.2	Aufgabenformate und Itemkonstruktion	149
4.2.2.3	Operationalisierung und Itemformulierung	154
4.2.2.4	Layout	160
4.2.2.5	Probanden	162
4.2.2.6	Durchführung	164
4.2.3	Ergebnisse und Analysen	165
4.2.3.1	Bestimmung der Itemreliabilität	165
4.2.3.2	Unterschiedliche Bewertung der Stimulustypen	166
4.2.3.3	Mittelwertunterschiede der einzelnen Fehler	178
4.2.4	Diskussion und Fazit	208

5	Zusammenfassung	211
5.1	Übergreifende Zusammenfassung der Hauptstudien	211
5.2	Erkenntnisse für die Ausspracheförderung im schulischen Dänischunterricht	214
5.3	Ausblick	215
	Literatur	217
	Anhang A: Studie 1	227
A.1	Fragebogen zur Erhebung sprecherbiographischer Daten	227
A.2	Analysebogen zur effizienten quantitativen Auswertung von Lernerfehlern	228
A.3	R-Output Chi-Quadrat-Tests	229
A.3.1	Ergebnisse zu sprachspezifischen Unterschieden (Niederdeutsch)	229
A.3.2	Ergebnisse zu den anwendungshäufigkeitsbedingten Unterschieden	233
A.3.3	Ergebnisse zu geschlechtsspezifischen Unterschieden	237
A.4	Mittelwerte (nicht näher untersuchte Fehler der verschiedenen Klassen)	237
	Anhang B: Studie 2	239
B.1	Onlinefragebogen	239
B.2	Ergebnisse der Regressionsanalysen zu den Unterschiedlichkeiten zwischen den Stimulustypen	240
B.2.1	<i>Accentedness</i>	240
B.2.2	<i>Comprehensibility</i>	242
B.2.3	<i>Acceptability</i>	244
B.3	Mittelwerte und Konfidenzintervalle	246
	Akademischer Werdegang	253

Tabellen

- Tab. 2.1: Übersicht über die Fragen des Fragebogens der explorativen Vorstudie 23
- Tab. 3.1: Zusammenhänge zwischen sprachlichen Unterschieden und kognitiven und motorischen Prozessen von Lernern 46
- Tab. 3.2: Schema zur Vorgehensweise bei der Fehleranalyse 49
- Tab. 3.3: Konsonanten des Dänischen, Deutschen und Norddeutschen (kontrastiv) 54
- Tab. 3.4: Monophthonge des Dänischen, Deutschen und Norddeutschen (kontrastiv) 58
- Tab. 3.5: Kontexte und Effekte der Schwa-Tilgung 64
- Tab. 3.6: Übersicht über die Zielwörter und Prüfsätze 73
- Tab. 3.7: Ergebnisse der Fehleranalyse zum stimmlosen [s] 79
- Tab. 3.8: Ergebnisse der Fehleranalyse zum alveolopalatalen Frikativ [ç] 81
- Tab. 3.9: Ergebnisse der Fehleranalyse zum weichen d [ð] 83
- Tab. 3.10: Ergebnisse der Fehleranalyse zum affrizierten t [tʰ] 85
- Tab. 3.11: Ergebnisse der Fehleranalyse zu den vorderen ungerundeten Vokalen 86
- Tab. 3.12: Ergebnisse der Fehleranalyse zum zentralen [ä] 88
- Tab. 3.13: Ergebnisse der Fehleranalyse zum hinteren ungerundeten Vokal [ɔ] 89
- Tab. 3.14: Ergebnisse der Fehleranalyse zu den hinteren gerundeten Vokalen 90
- Tab. 3.15: Ergebnisse der Fehleranalyse zu den bekannten Polyphthongen 93
- Tab. 3.16: Ergebnisse der Fehleranalyse zu den Diphthongen auf [u] 95
- Tab. 3.17: Ergebnisse der Fehleranalyse zu den Triphthongen 96
- Tab. 3.18: Ergebnisse der Fehleranalyse zur vokalischen Qualität 99
- Tab. 3.19: Ergebnisse der Fehleranalyse zu überlangen Vokalen und Konsonanten in Folge von [ə]-Tilgungsprozessen 102
- Tab. 3.20: Ergebnisse der Fehleranalyse zur Lenisierung nachtoniger Plosive 104
- Tab. 3.21: Relative Fehlerhäufigkeiten der Sprecher mit und ohne niederdeutschen Sprachhintergrund in Prozent 111
- Tab. 3.22: Relative Fehlerhäufigkeiten bei Dänischgebrauch außerhalb der Schule 113
- Tab. 3.23: Absolute Fehlerhäufigkeiten nach Geschlecht 114
- Tab. 3.24: Relative und absolute Fehlerhäufigkeiten der unterschiedlichen Klassen 115
- Tab. 3.25: Mittelwerte Fehlerhäufigkeiten aller Klassen (segmentale Ebene) 116
- Tab. 3.26: Mittelwerte Fehlerhäufigkeiten aller Klassen (Vokalquantität und Prozesse) 118
- Tab. 4.1: Übersicht über die Bewertungsdimensionen 135
- Tab. 4.2: Kombinationen der Bewertungsdimensionen 138
- Tab. 4.3: Übersicht über die untersuchten Merkmale 144
- Tab. 4.4: Items zu den Bewertungsdimensionen 159
- Tab. 4.5: Darstellung über die Bedeutung von Extremwerten der Antwortskala 160
- Tab. 4.6: Übersicht über die Altersverteilung der L1-Rater 163
- Tab. 4.7: Übersicht über die Geschlechterverteilung der L1-Rater 164
- Tab. 4.8: Darstellung der Dummykodierung 167
- Tab. 4.9: Beispiel zur Ergebnisdarstellung und -interpretation 168

Tabellen

- Tab. 4.10: Mittelwertunterschiede aV *accentedness* mit L1-Kontrollsprecher als Bezugskategorie 170
- Tab. 4.11: Mittelwertunterschiede aV *accentedness* mit L2-Kontrollsprecher als Bezugskategorie 171
- Tab. 4.12: Mittelwertunterschiede aV *comprehensibility* mit L1-Kontrollsprecher als Bezugskategorie 172
- Tab. 4.13: Mittelwertunterschiede aV *comprehensibility* mit L2-Kontrollsprecher als Bezugskategorie 173
- Tab. 4.14: Mittelwertunterschiede aV *acceptability* mit L1-Kontrollsprecher als Bezugskategorie 174
- Tab. 4.15: Mittelwertunterschiede aV *acceptability* mit L2-Kontrollsprecher als Bezugskategorie 175
- Tab. 4.16: Übersicht zur inhaltlichen Interpretation 186
- Tab. 4.17: Verteilung der Fehler in den Dimensionen *comprehensibility* und *acceptability* 204

Abbildungen

- Abb. 2.1: Ergebnisse zu den allgemeinen Fragen 25
- Abb. 2.2: Ergebnisse zur Selbsteinschätzung 26
- Abb. 2.3: Ergebnisse zu Materialien und Methoden 26
- Abb. 3.1: Beispiel für eine *Single Category Assimilation* 42
- Abb. 3.2: Beispiel für eine *Two Category Assimilation* 42
- Abb. 3.3: Zusammenhang zwischen sprachlichen Unterschieden, internen Prozessen und Lautrealisierung 44
- Abb. 3.4: Analysebogen zur Fehleridentifikation 78
- Abb. 3.5: Fehlerhäufigkeiten (aufsteigend sortiert) 105
- Abb. 4.1: Modell zur Fehlerevaluation 125
- Abb. 4.2: Exemplarische Seite des Fragebogens 161
- Abb. 4.3: *Acceptability* Item 1: Ausschnitt zur Darstellung der Untersuchung von signifikanten Mittelwertunterschieden anhand von Konfidenzintervallen 180
- Abb. 4.4: *Comprehensibility* Item 1: Ausschnitt zur Veranschaulichung der Gruppenbildung 182
- Abb. 4.5: *Acceptability* Item 2: Ausschnitt zur Veranschaulichung der Gruppenbildung 183
- Abb. 4.6: *Accentedness* Item 1: *Han/hun taler med accent* ‚Er/sie spricht mit Akzent‘ 187
- Abb. 4.7: *Accentedness* Item 2: *Han/hun kommer fra et andet land end Danmark* ‚Er/sie kommt aus einem anderen Land als Dänemark‘ 189
- Abb. 4.8: *Comprehensibility* Item 1: *Det er svært at forstå, hvad han/hun siger* ‚Es ist schwer zu verstehen, was er/sie sagt‘ 192
- Abb. 4.9: *Comprehensibility* Item 2: *Jeg skal koncentrere mig for at forstå, hvad han/hun siger* ‚Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt‘ 196
- Abb. 4.10: *Acceptability* Item 1: *Jeg ville tale et andet sprog end dansk med ham/hende* ‚Ich würde eine andere Sprache als Dänisch mit ihm/ihr sprechen‘ 198
- Abb. 4.11: *Acceptability* Item 2: *På dansk ville jeg tale langsomt med ham/hende* ‚Auf Dänisch würde ich langsam mit ihm/ihr sprechen‘ 201
- Abb. 5.1: Zusammenfassung der Ergebnisse aus den Studien 1 und 2: Fehlerschwere und Fehlerhäufigkeiten 213

1 Einleitung

Hvorfor skal vi overhovedet lære udtalen af de fremmede sprog? Det må dog vel i første række være af hensyn til den mulighed at vi senere kan træffe indfødte. (Jespersen 1901: 136)

Nu er vor udtale efter den gamle skole yderst tarvelig [...]. Den har bl.a. de to ulemper, at man ikke forstår de indfødte, og at man ikke forståes af dem [...] (Jespersen 1901: 137)

1.1 Gegenstand und Ziele der Arbeit

Diese Arbeit ist im Rahmen des Graduiertenprogramms *Lehramt in Bewegung* der Christian-Albrechts-Universität zu Kiel entstanden (CAU-LiB). Die Initiative zielt darauf ab, fachspezifisch relevante Aspekte im Zuge der Qualitätsverbesserung der Lehrkräftebildung zu realisieren. Dänisch hat sprachpolitisch als Sprache des Nachbarlandes für das Bundesland Schleswig-Holstein aus regional-strategischen aber auch aus historischen und minderheitenpolitischen Gründen einen besonderen Stellenwert. Dieser besteht neben unterschiedlichen grenzüberschreitenden Projekten der Zusammenarbeit darin, dass Dänisch an öffentlichen Schulen in Schleswig-Holstein unterrichtet wird. Im Schuljahr 2017/18 wird an 60 allgemeinbildenden und 7 berufsbildenden Schulen sowie an 46 Privatschulen Dänisch angeboten; das Fach wird von 396 Lehrkräften für insgesamt 10.483 Schüler unterrichtet (Ministerin für Bildung, Wissenschaft und Kultur 2018: 2).¹ Damit ist Dänisch die einzige nordische Fremdsprache, die im größeren Umfang an deutschen Schulen angeboten wird. Dies ist auch deshalb ein Alleinstellungsmerkmal, da Grönland das einzige andere nicht-nordischsprachige Land ist, in dem Dänisch als Fremdsprache unterrichtet wird.

Als vorrangiges Ziel schulischen Fremdsprachenunterrichts ist in den Curricula neben den Bereichen *methodischer* und *interkultureller* die *funktionale kommunikative Kompetenz* festgelegt (*Fachanforderungen Dänisch* 2016: 19). Diese besteht aus den Teilkompetenzen Hör- und Hörsehverstehen, Sprechen, Leseverstehen, Schreiben und Sprachmittlung in Verbindung mit den sogenannten sprachlichen Mitteln Wortschatz, Grammatik, Aussprache und Intonation sowie Orthographie. Die funktionale kommunikative Kompetenz soll die Schüler „zu situationsangemessener und sachgerechter Kommunikation in der Fremdsprache“ befähigen (*Fachanforderungen Dänisch* 2016: 19). Die fremdsprachliche Handlungsfähigkeit steht im Vordergrund, die sprachlichen Mittel und mit ihnen die Aussprache werden dabei als stützende Elemente für Kommunikation verstanden (*Fachanforderungen Dänisch* 2016: 16; 28; 53–54; 64). In den Fachanforderungen

¹ Zur besseren Lesbarkeit wird in dieser Arbeit auf Darstellungsformen wie Gendersternchen oder *Gender Gap* verzichtet. Das generische Maskulinum wird zur geschlechtsneutralen Bezeichnung eingesetzt, womit immer Personen jedes Geschlechts eingeschlossen werden. Dieses Verfahren wird im gesamten Text angewandt. Sofern neutrale Begriffe verfügbar sind (*Lehrkräfte, Studierende* etc.) werden diese eingesetzt.

für Englisch wird explizit auf den Vorrang der Kommunikation vor den sprachlichen Mitteln verwiesen (*Fachanforderungen Englisch* 2014: 47): „Die sprachlichen Mittel haben grundsätzlich dienende Funktion, wobei die Kommunikation im Vordergrund steht.“ In den Fachanforderungen für Französisch wird für die Aussprache eine Annäherung an die Norm angestrebt, die „im Allgemeinen keine Missverständnisse entstehen“ lässt (*Fachanforderungen Französisch* 2015: 27). Auch hier zeigt sich, dass die grundlegende Verständlichkeit als vorrangig vor Normkonformität eingeordnet ist. Eine Einteilung der sprachlichen Mittel auf diese Art und Weise entspricht der Ausrichtung der Fremdsprachendidaktik der letzten Jahre. Diese hat sich von einem Anspruch möglichst normsprachlicher Korrektheit zu einem Mindestmaß an zielsprachlicher Norm entwickelt, das Verständlichkeit und Kommunikationsfähigkeit gewährt (Celce-Murcia et al. 2012: 9; Levis 2016: 153). Diese Entwicklung kann grundsätzlich als sinnvoll eingestuft werden, da sie Zeit als begrenzte Ressource schulischen Fremdsprachenunterrichts miteinbezieht. Darüber hinaus zeigt die Forschung, dass das Ziel einer standardsprachlichen Aussprache in der L2 nur von wenigen Lernern erreicht wird (Flege et al. 1995) und für die meisten L2-sprachigen Kommunikationskontexte keine muttersprachliche Kompetenz notwendig ist (Derwing & Munro 2009). Verständlichkeit als Ziel anzustreben, stellt sich somit für die meisten Fremdsprachen als grundsätzlich sinnvoll dar. Für das Dänische ist die Verschiebung des Fokus von sprachlicher Korrektheit auf reine Verständlichkeit allerdings aus dreierlei Gründen problematisch:

- a. *sprachliche Alternativen*: Zur Kommunikation mit dänischen Muttersprachlern (im Folgenden L1-Sprecher) stehen mindestens zwei leicht zugängliche Alternativen zur Verfügung, da dänische L1-Sprecher über hohe sprachliche Kompetenzen im Englischen und teilweise auch im Deutschen verfügen.
- b. *geringe Variationstoleranz*: Anders als zum Beispiel im Englischen besteht im monozentrischen Dänischen gegenüber Standarddivergenzen in der Aussprache wenig Toleranz (vgl. Kristiansen 2003, 2009).
- c. *komplexe Phonetik*: Die dänische Aussprache gestaltet sich als äußerst schwierig beim Erlernen der Fremdsprache (Grønnum 2009).

Die Punkte *sprachliche Alternativen* und *geringe Variationstoleranz* führen dazu, dass dänische L1-Sprecher bereits bei geringen Abweichungen von der Norm geneigt sind, einen Wechsel in eine andere für die Kommunikation auf beiden Seiten gleichwertig funktionierende Sprache (Deutsch oder Englisch) vorzunehmen.² Anders als für andere Fremdsprachen ist also das Ziel der reinen Verständlichkeit nicht hinreichend, um im Dänischen funktional fremdsprachig zu sein, vielmehr muss angestrebt werden, möglichst

² Die Intentionen für einen Sprachwechsel sind dabei individuell. Sie können auf Hilfsbereitschaft seitens der dänischen L1-Sprecher beruhen, deren hohe Sprachkompetenz in leichter zugänglichen Varianten sie dazu befähigt, ihren Kommunikationspartnern eine leichtere Alternative anzubieten oder daher rühren, dass es für dänische L1-Sprecher relativ normal ist, dass ein nicht-muttersprachlicher Kommunikationspartner nicht Dänisch spricht und sie es gewohnt sind auf eine Lingua franca zurückzugreifen.

nicht als normabweichender Sprecher erkannt zu werden. Besonders schnell werden Abweichungen in der Aussprache – noch vor Abweichungen in Lexik oder Grammatik – von L1-Sprechern wahrgenommen (Munsberg 2001: 343; Saito, Trofimovich & Isaacs 2017: 439). Die Kommunikation in der Nachbarsprache kann also nur beibehalten werden, wenn die Kompetenzen der Lerner im Bereich Aussprache ausreichend hoch sind, um einen Sprachwechsel der dänischen Kommunikationspartner zu verhindern. Dabei kommt als besondere Erschwernis des Dänischen zum Ausspracheerwerb hinzu, dass die dänische Phonetik relativ komplex und daher verhältnismäßig schwierig zu erlernen ist. Während es im Deutschen 14 distinktive Vokale gibt und Deutsch somit über dem Durchschnitt der meisten anderen Sprachen liegt, deren Vokalinventar überwiegend knapp 6 bedeutungsunterscheidende Vokalen hat (Maddieson 2012), finden sich je nach Auslegung im Dänischen mindesten 17 unterschiedliche monophthongische Qualitäten. Hansen (1990) beschreibt 30 unterschiedliche Vokale, beruhend auf zwei Quantitätsstufen, während Höder (2016) sogar 42 Unterschiede feststellt, die sich aus drei unterschiedlichen Quantitätsstufen ergeben. Insgesamt lässt sich zusammenfassen, dass für das Erreichen einer funktionalen kommunikativen Kompetenz im Dänischen ein relativ hohes Maß an Aussprachekompetenz notwendig ist. Doch ebenso wie für die anderen Fremdsprachen stellt sich das Ziel muttersprachliche Aussprachekompetenz zu erlangen auch für das Dänische für bis auf einige wenige hochbegabte und hoch motivierte Lerner als unrealistisch heraus.

In den Zielformulierungen der anderen Fremdsprachen – beispielsweise Englisch als Fremdsprache – wird ein Grenzlevel formuliert, das es im Unterricht zu erreichen gilt und das ein Kompetenzlevel darstellt, auf dem die Aussprache nicht länger von der Kommunikation ablenkt (Celce-Murcia, Brinton & Goodwin 2012: 9). Für das Dänische wäre ein vergleichbares Level aus den oben dargestellten Gründen im Verhältnis zu anderen Fremdsprachen wahrscheinlich relativ hoch. Diese Besonderheit sollte in den schulischen Curricula berücksichtigt werden. Die bildungspolitischen Vorgaben haben sich in den letzten Jahren diesen Anforderungen angepasst. So erfuhr der Bereich der Aussprache in der Überarbeitung und Neuherausgabe des Curriculums für das Fach Dänisch in Form der Fachanforderungen im Jahr 2016 einen neuen Stellenwert. Anders als bei den oben genannten großen Fremdsprachen wird bereits in der Beschreibung der sprachlichen Mittel darauf hingewiesen, dass ein hohes Maß an sprachlicher Korrektheit zum Erreichen von funktionaler kommunikativer Kompetenz notwendig ist. So heißt es in den Fachanforderungen Dänisch (2016: 54):

Es ist von Beginn an anzustreben, dass die Schülerinnen und Schüler durch die Annäherung an sprachliche Korrektheit als dänischsprachige Gesprächspartnerinnen und -partner akzeptiert werden, so dass im Nachbarsprachenkontext ein Ausweichen auf andere Sprachen überflüssig wird.

Das Ziel schulischen Unterrichts für das Fach Dänisch orientiert sich somit an einem Maß an sprachlicher Korrektheit, das zur Kommunikation mit L1-Sprechern notwendig ist. Darüber hinaus werden die Zielvorgaben für den Bereich Aussprache bereits relativ präzise, auf linguistisch fachsprachlichem Niveau ausgeführt (*Fachanforderungen Dänisch* 2016: 30). So werden die sprachlichen Besonderheiten anhand von Beispielen veranschaulicht, wie etwa die Darstellung der auditiv zu differenzierenden vokalischen monophthongischen Qualitäten demonstriert. So wird auf die unterschiedlichen Öffnungsgrade anhand der Beispiele *mile* ‚Meile (PL)‘, *mele* ‚mit Mehl bedecken‘, *mæle* ‚Sprachfähigkeit‘, *male* ‚malen‘ verwiesen und die Komplexität spezifischer dänischer Qualitäten beispielhaft anhand von *måle* ‚messen‘ angedeutet (*Fachanforderungen Dänisch* 2016: 30). Angaben zu weiteren Qualitäten werden dabei zwar mitgedacht, allerdings leider nicht weiter ausführlich aufgeführt. Grundsätzlich ist die Hervorhebung der Aussprachekompetenz in den Zielformulierungen richtig und begrüßenswert, auch wenn die Darstellung der sprachlichen Besonderheiten nicht in erschöpfender Weise erfolgt und eine lautschriftliche Umsetzung der Orthographie fehlt. Weiterhin problematisch ist, dass die Zielvorgaben in Hinblick auf ein anzustrebendes Mindestmaß bzw. Grenzlevel, welche sprachlichen Mindestvoraussetzungen Lerner zur gelingenden Kommunikation mit L1-Sprechern erwerben müssten, unpräzise bleiben. Diese Unvollständigkeit mündet in einer fehlenden Orientierung für die Lehrkräfte und ist insofern problematisch, als die Lehrkräfte in der begrenzten Unterrichtszeit nicht wahllos alle phonetischen Phänomene behandeln können.

Dieser Aspekt hängt eng mit der Hintansetzung von Aussprache im Unterricht zusammen, die im Kontrast zur oben geschilderten Relevanz des Kompetenzbereichs im Fremdsprachenunterricht vorherrscht. Dies ist nicht nur für den Dänischunterricht zu verzeichnen. Einschlägige fremdsprachendidaktische Studien zeigen, dass die Förderung von Aussprache von Lehrkräften gegenüber Bereichen wie Grammatik oder Lexik generell vernachlässigt wird (Dieling & Hirschfeld 2000; MacDonald 2002; Celce-Murcia, Brinton & Goodwin 2012; Couper 2012). Als Gründe werden die Unsicherheit beim Unterrichten und der Mangel an Ausbildungs- und Professionalisierungsangeboten genannt (Couper 2012: 30; Darcy, Ewert & Lidster 2012). Auch im Fall schulischen Dänischunterrichts kann fehlendes Engagement beteiligter Fachvertreter oder Lehrkräfte für die mangelnde Ausspracheförderung im schulischen Dänischunterricht ausgeschlossen werden und vielmehr die mangelnde Orientierung in Zielvorgaben und Unterrichtsmaterialien sowie Lehreraus- und -weiterbildungsangeboten herangezogen werden. So nennen Fachvertreter und Lehrkräfte eindeutige Mängel:³

- a. *fehlendes Fachwissen*: Der grundlegende Mangel innerhalb der Forschung zu lerngruppenspezifischer Phonetik findet sich auch in den Inhalten der universitären

3 Vergleiche hierzu auch ausführlicher Kapitel 2.

Lehrerbildung sowie im Mangel von Fort- und Weiterbildungsmaßnahmen für Lehrkräfte wieder.

- b. *fehlendes fachdidaktisches Wissen*: Auch Strategien zur kognitiven Aussprachevermittlung sind bislang nicht in der Lehrerbildung verankert.
- c. *fehlende Materialien*: Es mangelt zwar nicht an Werken über dänische Aussprache im Allgemeinen (z. B. Kirk 2008; Snefstrup & Schmidt 2012), eine Aufbereitung für die spezifische Lernergruppe fehlt allerdings, weshalb die vorhandenen Lehrwerke für Lehrkräfte schwierig einzusetzen sind. Das Erfahrungswissen von Lehrkräften ist bislang noch nicht auf sprachwissenschaftliche bzw. fachdidaktische Weise grundlegend aufgearbeitet. Ausgangs- und zielsprachenspezifische Materialien liegen nur in begrenzter oder teilweise veralteter Form vor. Um den Fachanforderungen gerecht zu werden, müssen die Lehrkräfte Unterrichtsmaterialien selbst produzieren.

Aus den Punkten (a), (b) und (c) geht hervor, dass die momentan gegebenen Bedingungen für eine zufriedenstellende Umsetzung der Ausspracheförderung im Unterricht nicht ausreichend sind. Dieser Umstand resultiert aus der fehlenden grundlegenden wissenschaftlichen Aufarbeitung des fachspezifischen Professionswissens. So ist einschlägige Forschung notwendig, um Einblicke darin zu erhalten, welche phonetischen Elemente sich auf die Verständlichkeit bzw. den Akzent auswirken (Derwing & Munro 2005; Darcy, Ewert & Lidster 2012). Der Entwicklung von Curricula sollten evidenzbasierte Erkenntnisse aus der Forschung vorausgehen. Für das Fach Dänisch bedeutet dies, dass eine Aufarbeitung des von Lehrkräften benötigten fachdidaktischen und fachwissenschaftlichen Wissens durchgeführt werden muss, die auf evidenzbasierten Erkenntnissen der Forschung beruht. Bei der Aufarbeitung sollten zwei Aspekte im Fokus stehen:

- a. Die festgelegten Inhalte müssen im Rahmen schulischen Unterrichts realistisch erreichbar sein, und
- b. sie sollten das Mindestmaß sprachlicher Korrektheit beschreiben, das notwendig ist, um mit Muttersprachlern in der Zielsprache zu kommunizieren.

Es gilt also, ähnlich dem von Celce-Murcia, Brinton & Goodwin (2012) dargestellten notwendigen Grenzlevel, eine realistische Zielnorm festzulegen, in deren Beschreibung genau dargestellt wird, welche sprachlichen Phänomene für deutsche Dänischlerner schwierig und somit im Unterricht zu behandeln sind, und die zugleich berücksichtigt, welche lautlichen Merkmale relevant für die Kommunikation mit Muttersprachlern sind. Gleichzeitig sollte diese ein Pensum darstellen, das im schulischen Unterricht erreichbar ist. Der Wortstamm Norm des Kompositums verdeutlicht, dass für das Dänische tatsächlich ein hohes, nahezu muttersprachliches Aussprachelevel am Ende der schulischen Laufbahn erreicht werden muss. Der Zusatz realistisch stellt dar, dass dies Norm allerdings eine reduzierte Version der Zielsprache darstellt, die sprachliche Merkmale enthält, die minimal notwendig und maximal innerhalb des gegebenen Pensums zu erreichen sind. Die Festlegung der realistischen Zielnorm erfordert Wissen darüber, welche Aussprachephänomene deutschen Schülern Probleme bereiten, d. h. ein Spektrum an

Schwierigkeiten, die im Unterricht behandelt werden sollten. Aus diesem Spektrum müssen diejenigen Fehler ausfindig gemacht werden, die sich besonders negativ auf die zielsprachliche Kommunikation auswirken. Dies ergibt sich aus der Untersuchung, wie sich Fehler in Kommunikationsverhalten dänischer L1-Sprecher auswirken. Die vorliegende Arbeit verfolgt das Ziel, die für eine realistische Zielnorm relevanten Grundlagen festzulegen, wodurch sich drei Teilziele ergeben:

- a. *kontrastiver phonetischer Vergleich*: Anhand einer kontrastiven phonetischen Analyse der Ausgangssprachen Dänisch und Deutsch sollen Gemeinsamkeiten und Unterschiede und daraus resultierende potenzielle Schwierigkeiten ermittelt werden;
- b. *empirischer Abgleich*: Die sich aus dem kontrastiven Abgleich ergebenden potenziellen Ausspracheschwierigkeiten deutscher Dänischlerner im Aussprachebereich sollen anhand von echter Lernersprache abgeglichen werden. Daran soll eine Klassifizierung typischer Fehler in einer Fehleranalyse anschließen;
- c. *Ermittlung der Fehlerschwere*: Anhand der Bewertung der typischen Fehler durch Muttersprachler soll die Auswirkung der Fehler auf kommunikativ relevante Dimensionen ermittelt werden.

Die Teilziele stellen gleichzeitig die Arbeitsschritte dar, die zur Ermittlung einer realistischen Zielnorm für den schulischen Dänischunterricht im Aussprachebereich notwendig sind. Diese werden in der Arbeit nacheinander chronologisch durchgeführt. Im Folgenden wird eine Übersicht über die sich ergebende Kapitelstruktur gegeben.

1.2 Gliederung

Zunächst wird der Chronologie des Dissertationsprojekts folgend in Kapitel 2 die explorative Vorstudie dargestellt, die den Rahmen der Ausspracheförderung im schulischen Dänischunterricht und die Voraussetzung der vorliegenden Arbeit beschreibt. Anschließend werden die Studien, die zur Ermittlung einer realistischen Zielnorm notwendig sind, in einzelnen Kapiteln behandelt. So ermittelt Studie 1 in Kapitel 3 typische Aussprachefehler deutschsprachiger Dänischlerner. Dazu werden im ersten Schritt anhand einer kontrastiven phonetischen Analyse die Lautsysteme untereinander verglichen und Gemeinsamkeiten und Unterschiede herausgearbeitet (Kapitel 3.3). Diese werden anschließend empirisch mit Lernersprache abgeglichen (Kapitel 3.4). Anschließend findet in Kapitel 4 die Bewertung der ermittelten typischen Fehler deutschsprachiger L2-Dänischlerner durch L1-Sprecher statt. Dazu werden zunächst einschlägige Konzepte und Studien näher erläutert (Kapitel 4.1) und anschließend die Studie dieser Arbeit dargestellt (Kapitel 4.2). In der Zusammenfassung in Kapitel 5 werden die Ergebnisse beider Studien zusammengebracht (Kapitel 5.1) und Implikationen für den schulischen Dänischunterricht diskutiert (Kapitel 5.2). Die Arbeit schließt mit einem Ausblick auf Möglichkeiten für weiterführende Forschung (Kapitel 5.3).

2 Explorative Vorstudie

Jacob (2004) formuliert bereits 12 Jahre vor Beginn dieser Arbeit Bedarfe für den schulischen Dänischunterricht. Diese fokussieren sich auf die Aussprache und den Mangel an spezifisch abgestimmte Materialien. So äußert Jacob (2004: 4) Bedarf an abgestimmten Materialien zur Ausspracheschulung, die sowohl „imitativ als auch kognitiv nutzbar“ sein müssten. Daneben fordert sie mit IPA versehene Wörterbücher – ebenfalls auf die spezifischen Schülerbedürfnisse zugeschnitten – sowie Ausbildungsmodule im Bereich Aussprache für sowohl angehende als auch erfahrene Lehrkräfte. Von welchen aktuellen Voraussetzungen zu Beginn dieser Arbeit ausgegangen werden musste, sollte anhand einer explorativen Vorstudie ermittelt werden. In dieser wurden Dänischlehrkräfte an allgemeinbildender Schulen in Schleswig-Holstein zur Ausspracheförderung im Schulfach Dänisch befragt. Die Erhebung liefert Erkenntnisse über den aktuellen Stand der Ausspracheförderung im Schulfach Dänisch und bereitet somit die Grundlage für die anschließende Studie.

2.1 Befragung

Bei der explorativen Befragung zu unterschiedlichen Bereiche der Ausspracheförderung wurde ein Onlinefragebogen (Software des Onlineanbieters *LamaPoll*) mit insgesamt 10 Items eingesetzt. Der Fragebogen umfasste allgemeine Fragen zur Ausspracheförderung, zur subjektiven Selbsteinschätzung sowie Fragen zur didaktischen und methodischen Umsetzung. Der Fragebogen konnte von den Lehrkräften in 10 Minuten Bearbeitungszeit ausgefüllt werden. Der Aufruf zur Teilnahme wurde an 37 allgemeinbildende Schulen (Gesamtschulen und Gymnasien) sowie an 60 Dänischlehrkräfte an allgemeinbildenden Schulen persönlich verschickt. Es nahmen insgesamt 31 Lehrkräfte teil.

Sogar für die größeren Fremdsprachen gilt, dass sich Fremdsprachenlehrer bislang beim Unterrichten von Aussprache weitestgehend auf ihre eigene Intuition verlassen müssen (Derwing & Munro 2015: 7). Auch in der internationalen Forschung zur Ausspracheförderung im Fremdsprachenunterricht zeigt sich, dass sich die Mehrheit der Lehrkräfte nicht genügend auf die Förderung von Aussprache in ihrem Unterricht vorbereitet fühlt und deshalb die explizite Förderung meidet, obwohl sie sie als wichtig ansieht und ihr mehr Raum der Unterrichtszeit zusprechen würde (Derwing & Munro 2015: 78; Couper 2016). Die Vorstudie soll ermitteln, ob sich dieser Zustand auch für die Dänischlehrkräfte in Schleswig-Holstein bestätigen lässt.⁴ Zur Beantwortung der Fragen mussten die Probanden auf einer 4-stufigen Skala (*trifft zu*, *trifft eher zu*, *trifft eher nicht zu*, *trifft nicht zu*) zustimmen bzw. ablehnen. Tabelle 2.1 zeigt die Fragen in der Übersicht.

⁴ Nicht alle der im Fragebogen eingesetzten Items sind für die vorliegende Studie von Relevanz. Näher dargestellt werden diejenigen Items, die für die letztendliche Forschungsarbeit maßgeblich waren.

2 Explorative Vorstudie

Itemnummer und Frage		Antwortformat
Einstellungen zur Ausspracheförderung		
1	Die Förderung der Aussprache im Dänisch- unterricht finde ich wichtig.	geschlossen (trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu)
2	Die Förderung der Aussprache im Dänisch- unterricht finde ich schwierig.	
3	Die Bewertung der Aussprache im Dänisch- unterricht finde ich schwierig.	
optional	Nennen Sie Gründe, warum oder inwiefern sich die Förderung der Aussprache als schwierig darstellt.	offen (Kommentarfunktion)

Itemnummer und Frage		Antwortformat
Subjektive Selbsteinschätzung		
4	Ich fühle mich in der Vermittlung der dänischen Aussprache gut aus- und fortgebildet.	geschlossen (trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu)
5	Ich fühle mich in der dänischen Aussprache sehr sicher. Meine Aussprache ist nahezu akzentfrei bis muttersprachlich.	
6	Ich bin im Unterricht ein kompetenter Ansprech- partner zu Fragen in dänischer Aussprache. Ich kenne Regeln und kann auf Probleme der Aus- sprache eingehen.	

Itemnummer und Frage		Antwortformat
Stand vorhandener Materialien zur Vermittlung der dänischen Aussprache und im Unterricht verwendete Methoden		
7	Ich fördere in meinem Unterricht die Aussprache explizit, indem ich Phänomene direkt bewusst mache und Regelmäßigkeiten verdeutliche.	geschlossen (trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu)
8	Ich fördere in meinem Unterricht die Aussprache, indem ich imitieren und nachsprechen lasse.	
9	Ich nutze Material zur expliziten Förderung der Aussprachekompetenz.	
10	Es sind ausreichend Materialien zur expliziten Förderung der Aussprachekompetenz vorhanden.	
11	In Lehrwerken wird die Förderung der Aussprachekompetenz ausreichend berücksichtigt.	
12	Welches Material nutzen Sie zur Förderung der Aussprachekompetenz in Ihrem Unterricht?	offen (Kommentarfunktion)

Itemnummer und Frage		Antwortformat
Wünsche der Lehrkräfte		
13	Ich interessiere mich für fachliche und didaktische Themen zur dänischen Aussprache.	geschlossen (trifft zu, trifft eher zu, trifft eher nicht zu, trifft nicht zu)
14	Ich würde mir Fortbildungsmöglichkeiten zum Thema Förderung der Aussprachekompetenz im Unterricht wünschen und würde diese wahrnehmen.	

Tab. 2.1: Übersicht über die Fragen des Fragebogens der explorativen Vorstudie

Bei den Fragen zu Einstellungen zur Aussprache (Items 1–3) mussten die Probanden angeben, ob sie die Förderung von Aussprache schwierig finden. Bei zustimmenden Antworten wurde in den Fragebogen die Aktivierung einer weiteren Frage installiert. So wurden diejenigen Lehrkräfte, deren Antwort eher dazu tendierte, die Förderung von Aussprache als schwierig zu empfinden, um eine nähere Auskunft gebeten, wie sich diese Schwierigkeit darstellt. Die Antworten auf diese Frage wurden in einem offenen Frageformat gesammelt.

Neben den Fragen zu Einstellungen zur Ausspracheförderung wurden die Lehrkräfte um eine Selbsteinschätzung gebeten, indem erhoben wurde, ob bzw. inwieweit sich die

Lehrkräfte bei der dänischen Ausspracheförderung als kompetent einschätzen. Dies wurde anhand der Items 4–6 ermittelt, die mit derselben Skala wie oben beantwortet werden mussten. Darüber wurde mit einem offenen Antwortfeld ermittelt, ob neben Dänisch noch eine weitere Fremdsprache unterrichtet wird.

Um den Stand vorhandener Materialien zur Vermittlung der dänischen Aussprache zu ermitteln sowie die von den Lehrkräften angewandten Methoden zur Aussprachevermittlung näher zu beleuchten, wurden sechs Fragen generiert, die aus Zustimmungsbzw. Ablehnungsaufgaben mit derselben Antwortskalierung wie oben bestanden (Item 7–11) sowie offene Antwortformate (Item 12) enthielten. Als letzte Fragekategorie wurden die Wünsche der Lehrkräfte ermittelt (Item 13–14).

Die Ergebnisse geben relevante Aufschlüsse darüber, welche Voraussetzung für diese Studie bzw. den Bereich des dänischen Fremdsprachenunterrichts bestehen. Die quantitative Ausrichtung bzw. die relativ stark gelenkten Antwortmöglichkeiten der Erhebung haben den Vorteil, dass bestehende Vorerwartungen einfach und ökonomisch überprüft werden können. Ein Nachteil ist allerdings, dass individuelle Antworten und Erkenntnisse, die nicht vorab antizipiert wurden, nicht in allen Fragen (besonders nicht in denen ohne optionales Kommentarfeld) erhoben werden können. Um in einer weiterführenden Studie mit diesem Schwerpunkt differenziertere Einblicke in den aktuellen Wissens- und Bedarfsstand der Dänischlehrkräfte in Schleswig-Holstein zu erlangen, sollte eine qualitativ ausgerichtete Studie durchgeführt werden.

2.2 Ergebnisse

Die Antworten zu den generellen Fragen zur Ausspracheförderung im Unterricht ergeben, dass die Mehrheit der befragten Lehrkräfte die Ausspracheförderung zwar als wichtig, aber ebenso als schwierig empfindet. Das Diagramm in Abb. 2.1 zeigt die prozentualen Häufigkeiten der Antworten zu den Items 1–3.

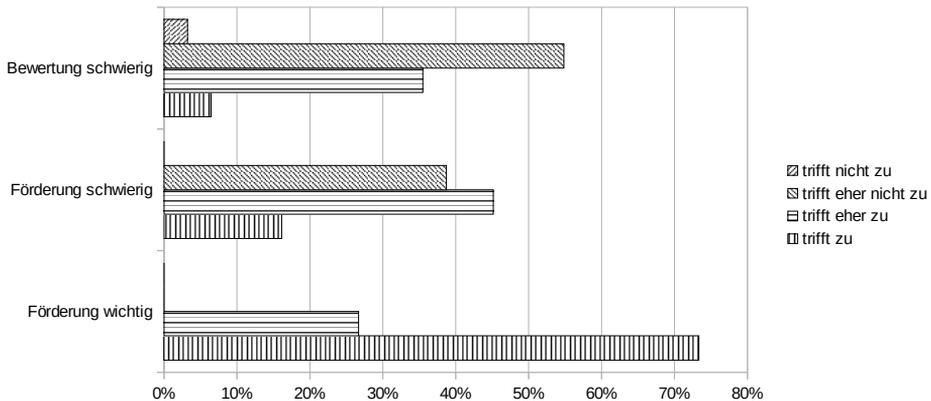


Abb. 2.1: Ergebnisse zu den allgemeinen Fragen

So bewertet die Gesamtheit der befragten Lehrkräfte die Förderung der Aussprache im Dänischunterricht als wichtig (73 % stimmen voll zu, 27 % stimmen eher zu). Die Mehrheit der Lehrkräfte (61 %) empfindet die Förderung der Aussprache im Dänischunterricht als schwierig. 39 % der Lehrkräfte geben an, das Unterrichten von Aussprache sei eher nicht schwierig. Anders verhält es sich mit der Bewertung von Schüleraussprache. Weniger als die Hälfte der befragten Lehrkräfte (41 %) empfindet die Bewertung der Aussprachekompetenz im Dänischunterricht als eher schwierig, während 54 % die Bewertung als eher nicht schwierig und 3 % als überhaupt nicht schwierig empfinden. Aus den offenen Antworten zum Grund der Schwierigkeit ergaben sich folgende Schwierigkeitsbereiche:

- a. Mangel an spezifischen Unterrichtsmaterialien;
- b. anspruchsvolle Phonetik/Phonologie des Dänischen;
- c. problematische Artikulation;
- d. Interferenzen aus dem Deutschen;
- e. Interferenzen aus der Orthographie;
- f. Schüler behalten Regeln nicht und fallen immer wieder in alte Muster.

Eigene Schwierigkeiten oder Mängel in der Ausbildung geben die Lehrkräfte nicht als mögliches Problem an. Diese Ergebnisse entsprechen auch der Auswertung der Fragen zur Selbsteinschätzung. Im Diagramm in Abb. 2.2 lassen sich die Antworten zu den Items 4–6 nachvollziehen.

Es zeigt sich, dass sich die Mehrheit der befragten Lehrkräfte (80 %) als gut ausgebildet und kompetent in der Vermittlung der dänischen Aussprache einschätzt. 32 % bewerten ihre eigene Aussprache als nahezu akzentfrei bis muttersprachlich, 39 % stufen ihre Aussprache eher gut ein, während 29 % sich eher nicht positiv einschätzen würden. 90 % der Gesamtheit sehen sich als kompetente Ansprechpartner zu Regeln und Fragen zur dänischen Aussprache.

2 Explorative Vorstudie

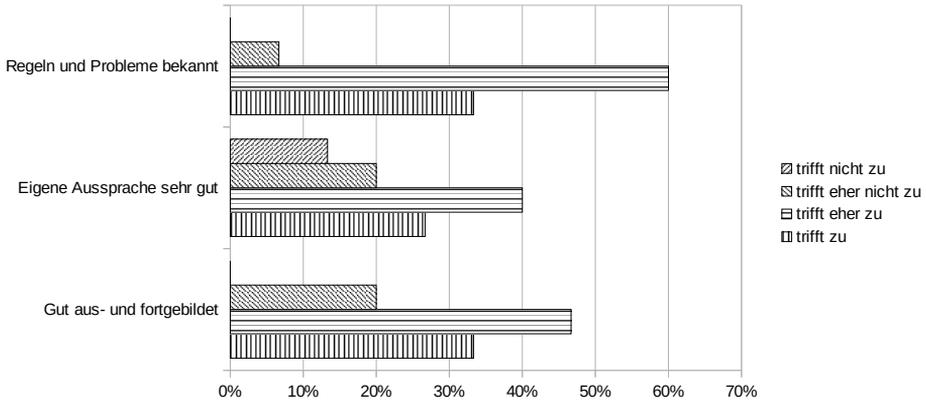


Abb. 2.2: Ergebnisse zur Selbsteinschätzung

Die überwiegende Mehrheit (78 %) der Lehrkräfte, die die Ausspracheförderung als schwierig einstufen, fühlt sich bezüglich ihrer eigenen Aussprache sicher. Lediglich 4 Lehrkräfte dieser Gruppe fühlen sich nicht sicher. Außerdem sehen sich die meisten Lehrkräfte (17 von 19) dieser Gruppe trotzdem als kompetente Ansprechpartner bei Regeln und Fragen zur dänischen Aussprache. 16 der 19 (89 %) der Teilnehmer finden, dass nicht ausreichend Materialien zur expliziten Förderung der Aussprachekompetenz vorhanden sind. Dies wurde anhand der Fragen zum Stand vorhandener Materialien zur Vermittlung der dänischen Aussprache ermittelt. Im Diagramm der Abb. 2.3 lassen sich die Antworten zu den Items 7–11 nachvollziehen.

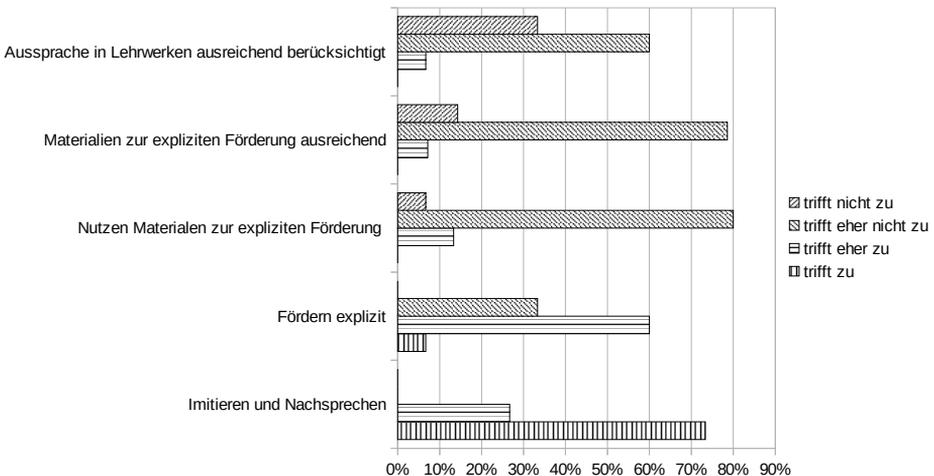


Abb. 2.3: Ergebnisse zu Materialien und Methoden

Zur bevorzugten Vermittlungsstrategie zählt das Imitieren und Nachsprechen. So geben 94 % der Lehrkräfte an, die Aussprache auf diese Art und Weise zu fördern. Die Förderung durch explizite Bewusstmachung von Phänomenen und Regeln wird von 74 % der Lehrer eingesetzt. 81 % der Lehrkräfte geben an, keine konkreten Materialien zur expliziten Förderung einzusetzen. Dieses Ergebnis ist besonders in Hinblick darauf erstaunlich, dass so viele Lehrer angeben, Phänomene und Regelhaftigkeiten konkret zu unterrichten. Die Antworten auf die Aussage zu den Materialien geben näheren Aufschluss zu diesem Sachverhalt. So bestätigt die überwiegende Mehrheit (90 %) der Lehrkräfte, dass nicht genügend Materialien zur expliziten Förderung vorhanden seien und 94 % geben an, dass in Lehrwerken die Förderung der Aussprache nicht ausreichend berücksichtigt werde. Die Ergebnisse lassen darauf schließen, dass die Lehrkräfte die explizite Bewusstmachung von Ausspracheregeln als relevant ansehen, den aktuellen Bestand an Materialien zur expliziten Förderung jedoch als unzureichend bewerten. Auf die offene Frage, welche Materialien die Lehrkräfte zur Förderung kennen bzw. nutzen, nennen die Lehrkräfte Lehrwerke zur Förderung dänischer Aussprache, die in Dänemark von Muttersprachlern entwickelt wurden (Kirk 2008 [*Dansk udtale. En undervisningsvejledning*]), kommunikativ ausgerichtete Lehrwerke, in denen Aussprache in minimaler Form bzw. als Randerscheinung mitbehandelt wird (Pude 2007 [*Vi snakkes ved*]), online im Lehrportal des Instituts für Qualitätsentwicklung an Schulen Schleswig-Holstein (IQSH) zur Verfügung gestellte Materialien (Jacob 2004: [*Et bogstav om dagen*]) sowie dänische Onlineresourcen (DUDA). Darüber hinaus geben die Lehrkräfte an, das jeweils vorhandene Lehrbuch mit eigens erstellten oder authentischen Materialien wie Gedichten und Liedern zu erweitern. Sowohl die Homepages als auch die Übungswerke für Dänisch als Fremdsprache stellen sich als nicht sprachenpaarspezifisch ausgerichtet dar. Das bedeutet, dass auf die sprachspezifischen Probleme deutscher Dänischlerner nicht eingegangen wird. Dementsprechend müssen die Lehrkräfte noch immer viel selbstständige Aufbereitungsarbeit leisten. Aus den Antworten zur letzten Frage nach eventuellen Bedarfen geht hervor, dass die Lehrkräfte Interesse an Fort- und Weiterbildungsangeboten zur Ausspracheförderung im Dänischunterricht haben. So wünschen sich 80 % der befragten Lehrkräfte ein entsprechendes Angebot.

2.3 Folgerungen

Aus der explorativen Vorstudie können Rückschlüsse zum aktuellen Stand der Ausspracheförderung im schulischen Fremdsprachenunterricht Dänisch gezogen werden. Es zeigt sich, dass die Lehrkräfte die Ausspracheförderung als sehr wichtig empfinden, allerdings ein Mangel an sprachenpaarspezifischen Materialien zur Förderung besteht. Die Bewertung der Aussprache ihrer Schüler stufen die Lehrkräfte als unproblematisch ein. Zu den Aspekten, die die Ausspracheförderung als schwierig gestalten, wird von den Lehrkräften die komplexe Phonetik und Phonologie des Dänischen genannt, aus der artikulatorische Probleme resultieren. Darüber hinaus nennen die Lehrkräfte das Phäno-

men, dass die Schüler lautliche Regeln nicht behalten bzw., dass sie bei der Aussprachevermittlung sehr viel wiederholen und von neuem beginnen müssen. Vorstellbar ist, dass der Ansatz der Fehlerkorrektur (z. B. durch Imitation) nicht ausreichend ist, da die Ursache der falschen Realisierung dadurch nicht behoben wird. Eine Fehlerklassifikation für das Dänische besteht bislang leider nicht, weshalb es nicht möglich ist, Schülerfehler systematisch und strukturiert zu analysieren und somit spezifisch zu korrigieren. Dieses Fehlen stellt nicht nur ein praktisches Problem für erfahrene Lehrkräfte dar, sondern bildet auch eine Lücke im Curriculum professionellen Fachwissen angehender Dänischlehrkräfte. Diese sollten im Laufe ihrer Ausbildung mit typischen und wahrscheinlichen Aussprachefehler deutscher Dänischlerner konfrontiert werden. Nur anhand dieses professionellen Wissens ist es möglich, problemspezifisch und angemessen in der Ausspracheförderung und Fehlerkorrektur zu handeln. Die notwendige sprachstrukturelle Aufbereitung hat bislang nicht stattgefunden. Aus dieser Lücke resultiert vermutlich auch die fehlende Aufbereitung entsprechender Materialien für den Unterricht. Die Lehrkräften nehmen diesen Mangel in der Ausbildung offenbar nicht selbst wahr, zumindest formulieren sie ihn nicht. Aus der Selbsteinschätzung ihrer eigenen Aussprachekompetenz und Kompetenz zur Vermittlung von Aussprache geht hervor, dass sich die Lehrkräfte überwiegend sicher fühlen. Im Kontrast dazu steht der Wunsch nach Fort- und Weiterbildungsangeboten. Zu bedenken ist, dass es sich bei der Befragung nach der Sprachkompetenz der Lehrkräfte um eine subjektive Einschätzung handelt, die nicht objektiv überprüft wurde.

Für die Projektarbeit ergeben sich aus der explorativen Vorstudie mehrere Arbeitsziele: Zunächst müssen die sprachenpaarspezifischen Grundlagen zur Aussprachevermittlung bzw. zu deren Inhalten erfolgen. Diese Aufarbeitung bietet die Grundlage für die Ausformulierung professionellen Wissens von Dänischlehrkräften sowie für die Erstellung spezifischer Materialien. Die gewonnenen Erkenntnisse sollen sowohl in der ersten Phase der Lehrerbildung als auch in Fort- und Weiterbildungen implementiert werden sowie als theoretische Grundlage zur Erstellung spezifischer Materialien zur expliziten Vermittlung der Aussprache für Dänischlerner in Norddeutschland dienen. Zusammenfassend lässt sich als Arbeitsziel aus der explorativen Vorstudie ermitteln, dass eine auf den schulischen Dänischunterricht in Schleswig-Holstein ausgerichtete didaktisierte Phonetik zu entwickeln ist, die auf die spezifischen Voraussetzungen der Lerngruppe eingeht und somit als Grundlage für Fachwissen von (angehenden) Lehrkräften sowie zur Erstellung neuer Materialien dienen kann.

3 Studie 1: Fehler deutschsprachiger L2-Dänischler

Die erste Studie untersucht Fehler deutschsprachiger L2-Dänischler genauer. Dabei werden sowohl erwerbsbedingte Hintergründe (Kapitel 3.1) als auch methodische Vorgehensweisen der Analyse (Kapitel 3.2) ausführlich dargestellt. In Kapitel 3.3 erfolgt eine kontrastive phonetische Gegenüberstellung der deutschen und dänischen Lautsysteme. Diese dient dazu, einen Pool hypothetischer Schwierigkeiten für deutsche L2-Lerner zu ermitteln, der anschließend in Kapitel 3.4 anhand von Sprachaufnahmen von realen Lernern abgeglichen wird. In der Analyse (Kapitel 3.4.2) werden sowohl Häufigkeiten als auch Erklärungsansätze dargestellt.

3.1 Fehler und Lernaltersprache

Beim Erwerb einer Fremdsprache (L2) können aus den unterschiedlichsten Gründen Probleme auftreten. Fremdsprachenlehrer begegnen diesen als normabweichende Lernaltersprache, deren Strukturen üblicherweise als Fehler bezeichnet werden. Dieser Begriff ist allerdings nicht unproblematisch, da er in der fremdsprachendidaktischen Forschung mit theoretischen Konzeptionen in Verbindung gebracht wird, die unterschiedlichen Ausrichtungen folgen und im Laufe der Zeit einige Entwicklungen durchlaufen haben. Jeweils aktuelle Spracherwerbstheorien oder didaktisch-methodische Überzeugungen führen zu unterschiedlichen Auslegungen des Konzepts *Fehler* und der damit einhergehenden Vorstellung von Lernaltersprache (Kleppin 2013: 224). So ist einhergehend mit der behavioristischen Fremdsprachendidaktik die vorherrschende Ansicht der 1960er Jahre, dass Fehler Unzulänglichkeiten darstellen, die zu vermeiden und beseitigen sind (Kleppin 1997: 50, 2001: 986). Lernaltersprache wird zu dieser Zeit als nicht vollständige bzw. fehlerhafte Version der Zielsprache verstanden, die sich als unstrukturiertes, durch die Muttersprache (L1) bedingtes Fehlergemenge darstellt. Aus Sicht der damals vorherrschenden Spracherwerbstheorie der *Contrastive Analysis Hypothesis* (im Folgenden CAH) lassen sich Fehler als negative Einwirkungen (Interferenzen) durch die Muttersprache erklären (für eine detaillierte Darstellung vgl. Kapitel 3.2.2.1).

Ende der 1960er bzw. Anfang der 1970er Jahre entwickelt sich, maßgeblich durch die Arbeiten Corders (1967) und Selinkers (erstmalig 1969, veröffentlicht 1972), ein neues Verständnis von Lernaltersprache, das die Einstellung zu Fehlern grundlegend verändert (vgl. Harden 2006: 74). Im Gegensatz zur CAH lehnt Corder (1967) die Annahme ab, dass Interferenzen durch Sprachgewohnheit innerhalb der L1 auftreten. Stattdessen versteht er L1-bedingte Fehler als sichtbar gewordene kognitive L2-Spracherwerbsprozesse. Diese sind übereinstimmend mit denen des L1-Erwerbs: Um eine neue Sprache zu lernen, stellen die Lerner Hypothesen zum System der neuen Sprache auf und gleichen diese mit der Wirklichkeit ab. Diese Auffassung von universalen Spracherwerbsprozessen, die sowohl für den L1-Erwerb wie auch für den L2-Erwerb in gleicher Weise gelten, wird auch als *Identitätshypothese* bezeichnet. Dass Fehler entstehen, ist notwendiger Umstand des Abgleichungsprozesses (Corder 1967: 168). Corder sieht den einzigen Unterschied zwischen L1- und L2-Erwerb in der Grundlage, auf der die Lerner ihre Hypothesen auf-

stellen. Der L2-Erwerb sei leichter, da seine Hypothesenbildung auf dem Fundament des L1-Systems erfolgen könne, während die Hypothesenbildung im L1-Erwerb ohne Grundlage erfolge. Außerdem nimmt Corder an, dass Normabweichungen im L2-Erwerb anders als beim L1-Erwerb als Fehler sanktioniert werden. Corder schließt aus seinen Annahmen, dass es zwingend notwendig sei, Fehler differenzierter zu betrachten als bis dahin geschehen. Aus diesem Grund entwickelt er die *Error Analysis*, in der Fehler als *errors* und *mistakes* unterteilt werden (Corder 1967: 166–167). *Errors* bezeichnen systematische Fehler, d. h. solche, die nicht zufällig entstanden sind, und Aufschlüsse über das vom Lerner genutzte Sprachsystem geben. Durch die *errors* wird die sprachliche Kompetenz von Sprechern sichtbar. *Mistakes* hingegen sind unsystematische situative Fehler oder Versprecher, die von Corder (1967: 166) als *Performanzfehler* bezeichnet werden. Die von Corder vorgenommene Unterscheidung ist für Forschung und Praxis von großer Relevanz, da sie es ermöglicht, Aussagen über die Strukturen von Spracherwerb und die dabei ablaufenden kognitiven Prozesse zu tätigen.

Selinker (1972) definiert Lernersprache mit dem Terminus *Interlanguage*. Ebenso wie Corder (1967) versteht er Lernersprache als eigenes sprachliches System, das Lerner beim Aneignen einer Zielsprache im Verlauf ihres Sprachlernprozesses aufbauen. Dabei hebt er hervor, dass es sich nicht um eine unvollständige Version der Zielsprache, sondern um ein dynamisches, autonomes System stetig wachsender Fertigkeiten handelt, das sowohl Merkmale von Erst- und Zielsprache als auch von weiteren Fremdsprachen enthalten kann und sich ständig verändert und entwickelt. Selinker (1972) teilt Corders Ansicht von Lernersprache als Kontinuum, dessen prozesshafter Charakter durch die normabweichenden L2-Realisierungen sichtbar wird (Ellis 2006: 30). Diese Abweichungen von der Zielsprache lassen wiederum auf die kognitiven Prozesse und Strategien der Lerner Rückschlüsse ziehen. Von Selinker (1972: 216–217) werden fünf solcher die Lernersprache bedingenden Prozesse dargestellt:

- a. *Sprachtransfer*: Die Lerner übertragen L1-Regeln auf die zu erlernende Fremdsprache. Dieser Prozess kann sowohl richtige als auch fehlerhafte Äußerungen zur Folge haben.
- b. *Übergeneralisierung von zielsprachlichen Regeln*: Bei diesem Prozess werden L2-Regeln auf Kontexte übertragen, in denen sie nicht gültig sind. Die daraus entstehenden fehlerhaften Äußerungen (auch Interferenzen genannt) können interlingual (zweisprachlich) und intralingual (innersprachlich) sein. Interlinguale Fehler entstehen aus der Übertragung von Regeln aus der L1 auf die L2, während intralinguale Fehler dadurch entstehen, dass eine neu erworbene Regel in der L2 auf weitere Kontexte in der L2 übertragen werden, in denen diese allerdings nicht richtig ist.
- c. *Kommunikationsstrategien*: Diese Strategien helfen Lernern bei Unsicherheiten oder Unwissen. Typische Strategien sind der Verzicht auf syntaktisch schwieriger Konstruktionen oder die Vermeidung von sprachlich komplexen Themen. Prozesse dieser Art resultieren nicht zwangsläufig in fehlerhaften Äußerungen.

- d. *Spracherwerbsprozesse und -strategien*: Diese Strategien stellen kognitive Prozesse dar, durch die der Lerner fremdsprachliche Kompetenz aufbaut und weiterentwickelt. Dabei werden anhand von sprachlichen Inputs Hypothesen systematisch abgeglichen.
- e. *Transfer durch Training*: Dieser Prozess beschreibt den Einsatz fehlerhafter Merkmale, der durch ungeeignete Vermittlungsmethoden provoziert wurde.

Die Neuerungen Corders und Selinkers ermöglichen einen weitestgehend wertfreien Gebrauch des Fehlerbegriffs, da Abweichungen von der Norm nunmehr als natürlich bzw. als Voraussetzung für ein sich entwickelndes Lerner-Sprachsystem und nicht länger als defizitäre Realisierungen einer Zielsprache verstanden werden. Diese Entwicklung ist positiv einzustufen, da sie die Rolle der Lerner und ihrer Sprache aufwertet und einen konstruktiven und fördernden Umgang zwischen Lernenden und Lehrkraft im Unterricht ermöglicht. Tatsächlich hat sie allerdings auch Konsequenzen für die Ausrichtung fremdsprachlichen Unterrichts, die nicht unhinterfragt bleiben sollten. So ist die Relativierung des Konzepts Fehler vermutlich mit ursächlich für die *kommunikative Wende*. Diese bezeichnet die Veränderung fremdsprachlichen Unterrichts in Theorie und Praxis in Form einer Implementation eines kommunikativen Ansatzes als didaktische und methodische Leitlinie (Kurtz 2013: 80–81). Diese Veränderung zu Gunsten der Förderung kommunikativer Fähigkeiten und interaktiver Handlungsfähigkeit zieht eine Hintanstellung normsprachlicher Korrektheit nach sich (Bohnsteffen 2011: 16). Trotz der scheinbaren Überwindung des wertenden Gebrauchs des Fehlerbegriffs nach Corder und Selinker führt die kommunikative Wende zu einer erneuten Infragestellung des Konzepts. So gehen mit der Ablehnung sprachlicher Korrektheit als Ziel fremdsprachlichen Unterrichts kritische Haltungen gegenüber dem Fehlerkonzept und der einhergehenden Korrektur einher. Diese werden mitunter als schädlich für den Lernprozess beschrieben (Truscott 1996, 2007). Truscott (2007) vertritt die Ansicht, dass der Umgang mit Fehlern und Fehlerkorrektur im Unterricht bislang zu wenig hinterfragt und die positive Wirkung von Fehlerkorrektur unreflektiert als gegeben angenommen wurde. Er belegt seine Ausführungen mit Forschung, die zeigt, dass grammatische Fehlerkorrektur keinen positiven bis sogar negativen Einfluss auf den L2-Spracherwerb hat (Truscott 2007: 270). Diese eher radikale Position, die grundlegend mit der Ablehnung von Fehlerkorrektur einhergeht und eine Polemisierung des Fehlerbegriff einschließt, wird überwiegend als praxisuntauglich eingestuft (Ferris 2004: 58-59; Derwing & Munro 2015: 56). Kleppin & Königs (1997: 272, 292) hinterfragen sogar eine negative Einstellung zum Fehlerkonzept in der Praxis und stellen fest, dass eine aktive Auseinandersetzung mit Lernerfehlern im Unterricht in der Praxis sowohl von Lehrern als auch von Schülern gewünscht wird.

Die vorliegende Arbeit vertritt die Ansicht, dass eine kritische Auseinandersetzung mit dem Fehlerbegriff zwar notwendig sind, diese allerdings nicht zu einer Tabuisierung des Begriffes sowie dessen Verwendung führen sollten. Fehler- und Lernaltersprachenkonzepte bzw. die Debatte um diese müssen außerdem sprachenspezifisch betrachtet werden. So geht die Bedeutung von Englisch als Lingua franca mit einer entsprechenden

Zielsetzung des Unterrichts einher, in der sprachliche Korrektheit untergeordnet werden kann und die Entwicklung fehlertoleranter (wie beispielsweise bei Bohnensteffen 2011) oder kommunikativer Unterrichtsleitbilder (wie beispielsweise bei Kurtz 2013) durchaus ihre Berechtigung haben. In Hinblick auf die eingangs dargestellte Situation für Dänisch (vgl. Kapitel 1.1) ist allerdings zu hinterfragen, ob kommunikative Kompetenz im Dänischen ohne sprachliche Korrektheit auskommt. Wahrscheinlicher ist anzunehmen, dass die sprachliche Korrektheit auf Grund der Variationsintoleranz der Dänen einen notwendigen Bestandteil der Kommunikationsfähigkeit im Dänischen darstellt. Die Frage, ob die besonderen phonetisch-phonologischen Merkmale des Dänischen zu Gunsten von Kommunikation vernachlässigt werden dürfen, stellt sich somit nicht. Innerhalb der Arbeit wird der Terminus *Fehler* deshalb wertfrei entsprechend der gängigen Fehlerdefinitionen der Spracherwerbsforschung als Abweichungen vom Sprachsystem bzw. als Abweichungen von der geltenden sprachlichen Norm einer Sprachgemeinschaft verstanden (Kleppin 1997: 19). Bezogen auf den Bereich der Aussprache fallen darunter somit Laute, die ein L1-Sprecher nicht produzieren oder als Kommunikationspartner nicht verstehen würde (Kleppin 1997: 19). Als Arbeitsdefinition ergibt sich der folgende deskriptive, wertfreie Fehlerbegriff: *Fehler stellen lautliche Normabweichungen dar.*

3.2 Fehleranalyse

In der Fehleranalyse werden Fehler klassifiziert, d. h. nach ihrer Erscheinungsform beschrieben und eingeteilt sowie auf ihre möglichen Ursachen systematisch untersucht. Fehleranalysen geben wichtige Erkenntnisse für den Spracherwerb, da sie Lernprobleme erklärbar machen (Kleppin 1997: 41). Dies kann auf unterschiedlichen sprachlichen Ebenen stattfinden, z. B. innerhalb von Syntax, Lexik, Pragmatik oder Phonetik (Kleppin 1997: 42). In der vorliegenden Arbeit werden Fehler auf der phonetischen Ebene fokussiert.

Als gängige Schritte einer linguistischen Fehleranalyse gelten die Identifizierung, Klassifizierung und Ursachenerklärung von Fehlern (Marx & Mehlhorn 2016: 301). Diese Schritte können erweitert werden. So führt beispielsweise Hirschfeld (2001: 86-88) als weitere Analyseaspekte die Bedeutung der Fehler für die Kommunikation sowie die Häufigkeit der Fehler ein. Zunächst müssen die zu untersuchenden Daten erhoben werden. Bei der Dokumentation des Prozesses sollten die folgenden Fragen beantwortet werden:

- a. Handelt es sich um mündliche oder schriftliche Daten?
- b. Liegen die Daten frei oder gebunden (in Form von vorgelesenen Texten) vor?
- c. Erfolgte die Datenerhebung einmalig oder über mehrere Zeitpunkte?
- d. Wurden die Daten einzeln oder in Dialogform aufgenommen?

In der anschließenden Datenaufbereitung werden die erhobenen Primär- und Metadaten systematisch archiviert und aufbereitet. Anschließend erfolgt die Identifizierung der

Fehler. Dabei dient die L2-Norm als Bewertungskriterium, wobei Abweichungen von dieser Norm als Fehler festgehalten werden (Marx & Mehlhorn 2016: 303).

Nach erfolgreicher Identifikation der Fehler erfolgen als weitere Schritte der Fehleranalyse die Klassifikation und Ursachenerklärung. Während es sich bei der Klassifikation um einen deskriptiven Analyseschritt handelt, bei dem der Unterschied zwischen der tatsächlichen L2-Realisierung und dem angestrebten Zielmerkmal dargestellt wird (Derwing & Munro 2015: 58), ist die Ursachenerklärung ein explanativer Vorgang, in dem lediglich Hypothesen aufgestellt werden können. Diese beiden Schritte sollten unbedingt auseinandergelassen werden, da es bei einer fehlenden Abgrenzung zu Vermischungen von deskriptiven und explanativen Aspekten der Analyse kommen kann. Eine solche Vermischung findet beispielsweise bei Hirschfeld (2001: 87) statt. Hirschfeld (2001) nutzt die folgenden Kategorien zur Klassifizierung, wobei sie sich auf Weinreich (1953) bezieht:

- a. *Phonemverschmelzung*: diese Kategorie wird von Hirschfeld (2001) auch als *underdifferentiation* bezeichnet, womit gemeint ist, dass distinktive Phone nicht auseinandergelassen werden;
- b. *Phonemzerlegung* (auch *overdifferentiation* genannt): nicht distinktive Laute werden fälschlich verschiedenen Kategorien zugeordnet;
- c. *Überdifferenzierung* oder *Differenzierungsdefizit*, auch als *reinterpretation of distinctions* bezeichnet;
- d. *Laut-* bzw. *Phonemersatz*, auch *actual phone substitution*: Laute werden durch andere ersetzt;
- e. *Phonemausfall*: ein Laut wird nicht realisiert.
- f. *Phonemeinschub*: es wird ein zusätzlicher Laut gebildet;
- g. *Phonemvertauschung*: ein normkonformer Laut, wird durch einen nicht-normkonformen Laut ersetzt.

Während die ersten drei Punkte (a–c) explanative Ursachenbeschreibungen für Fehler darstellen, handelt es sich bei den Punkten d–g um konkrete deskriptive Aspekte zur Beschreibung von Erscheinungsformen von Fehlern. Eine ähnliche Vermischung findet bei Böttger (2008) statt. Böttger (2008) gibt in ihrem Handbuch für Lehrende von Deutsch als Fremdsprache einen Überblick über die häufigsten Fehler russischer Deutschlerner. Dazu vergleicht sie lautliche Realisierung russischer und deutscher Orthographie.⁵ Böttger (2008: 24–26) nutzt zur Fehleranalyse insgesamt sieben Kategorien von Fehlertypen. Diese setzen sich aus einer deskriptiven Beschreibung der Erscheinungsform und einer explanativen Ursachenzuschreibung zusammen. Die Bezeichnung der Fehlertypenkategorien haben dabei teilweise deskriptiven (*Ersetzung*), teilweise explanativen Charakter (*Unwissen*). Die unterschiedliche Beschreibungsform von Ursachen (explanativ) und Erscheinungsform (deskriptiv) ist ebenfalls nicht durchgängig konsistent durchgeführt.

5 In ihrer Fehleranalyse stützt sich Böttger (2008) auf Odlin (1996) und James (1971).

Beispielsweise findet sich beim Fehlertyp *Hyperkorrektur* unter der Beschreibung der Erscheinungsform, die eigentlich rein deskriptiv sein sollte, explanative Ansätze zu den Fehlerursachen. Böttger (2008: 26) stellt als Erscheinungsform dar, dass Lernende „die Funktion einer bestimmten Form mit ihrem Gegenteil“ vertauschen. Dies ist zwar eine korrekte Beschreibung der Abläufe, die zu einem Fehler führen, stellen aber nicht das beschreibbare Erscheinungsbild des Fehlers dar. Eine Vermischung von beschreibenden und interpretatorischen Aspekten sollte jedoch vermieden werden, da sie zu Problemen und Unklarheiten während der Fehleranalyse sowie zu Ungenauigkeiten bei der Auswertung der Analyseergebnisse führen kann. Diese Arbeit versucht aus diesem Grund noch strenger der Prämisse zu folgen, die von Marx & Mehlhorn (2016) dargestellten Schritte der Fehleranalyse getrennt voneinander zu behandeln. Dadurch werden nicht nur Probleme bei der Analyse vermieden, sondern außerdem der Grad der Objektivität erhöht. Im Folgenden werden die Schritte der Fehleranalyse, die Klassifikation und die Ursachenbeschreibung, getrennt voneinander detailliert dargestellt. Anschließend werden in Kapitel 3.2.3 die Schlussfolgerungen für die vorliegende Arbeit aus den Darstellungen gezogen und die grundlegenden Arbeitsbegriffe vorgestellt.

3.2.1 Fehlerklassifikation

In der Fehlerklassifikation wird die Erscheinungsform von Fehlern anhand des Unterschieds zwischen der tatsächlichen L2-Realisierung und dem angestrebten Zielmerkmal beschrieben (Derwing & Munro 2015: 58). Wie oben bereits angedeutet, können Ansätze zur Beschreibung bei verschiedenen Forschern sehr unterschiedlich ausfallen. Im Folgenden werden zwei Beschreibungsansätze aus dem englischsprachigen (Derwing & Munro 2015) und deutschsprachigen Raum (Schaden 2006) vorgestellt. Derwing & Munro (2015: 58) unterscheiden zunächst zwischen segmentalen und suprasegmentalen Fehlern und beschreiben die Kategorien zur Klassifizierung segmentaler Fehler anschließend als die Folgenden:

- a. *Insertion* ‚Einfügung‘: ein Segment, das nicht in der Zielform vorhanden ist, wird eingefügt;
- b. *Deletion* ‚Entfernung‘: ein Segment, das in der Zielform vorhanden ist, wird nicht umgesetzt;
- c. *Substitution* ‚Ersatz‘: ein Segment wird durch ein Segment einer anderen phonemischen Kategorie ersetzt;
- d. *Distortion* ‚Verzerrung‘: ein Segment wird nicht zielsprachlich realisiert, ohne die phonemische Kategorie zu wechseln.

Die Einteilung der Fehlertypen aus dem deutschsprachigen Raum ist zu dieser recht ähnlich. Schaden (2006) ermittelt die für seine Arbeit grundlegenden Fehlertypen aus den oben dargestellten sieben Typen des Lautersatzes von Hirschfeld (2001: 87), indem er diese abstrahiert. So beschränkt Schaden (2006: 111) die sieben Typen auf drei. Unter diese lassen sich deskriptiv alle auftretenden Abweichungen einteilen, ohne dass die Lau-

te näher bestimmt werden müssen. Schaden (2006: 111) nutzt die folgende Kurzdarstellung, um die von ihm ermittelten Typen übersichtlich zu beschreiben:

- a. *Einfügung*: $x \rightarrow xy$;
- b. *Tilgung*: $x \rightarrow \emptyset$;
- c. *Substitution*: $x \rightarrow y$.

Die Fehlertypen sind überwiegend mit denen von Derwing & Munro (2015) zu vergleichen. Lediglich der Typ der *distortion* wird von Schaden (2006) nicht als individuelle Kategorie aufgenommen. Beide vorgestellten Schemata ermöglichen eine rein deskriptive Einteilung zur Klassifizierung, die sich von explanativen Ursachenzuschreibungen abgrenzen lässt. Die Termini, die zur Beschreibung eingesetzt werden, dürfen nicht als tatsächlich vom L2-Lerner ausgeführte Handlungen missinterpretiert werden (Derwing & Munro 2015: 58). So darf die *substitution* nicht dahingehend interpretiert werden, dass das Segment zunächst gesprochen und dann getauscht wurde. Die hier ablaufenden Prozesse sind komplex und können teilweise mit den in Kapitel 3.2.2 dargestellten Modellen erklärt werden. Für die Analyse in dieser Arbeit werden die folgenden Termini aus den oben dargestellten Ansätzen eingesetzt:

- a. *Lauteinfügung*: neben dem Ziellaut wird ein weiterer Laut eingefügt;
- b. *Lauttilgung*: der Ziellaut wird nicht realisiert und nicht ersetzt;
- c. *Lautersatz*: der von Derwing & Munro (2015) und Schaden (2006) als *substitution*/Substitution benannte Typ; darunter wird der Ersatz von L2-Segmenten durch Segmente einer anderen phonemischen Kategorie verstanden;
- d. *Lautabwandlung*: wie bei dem von Derwing & Munro (2015: 58) als *distortion* bezeichneten Typ; gemeint ist der Ersatz von Segmenten, der nicht in einem Wechsel der phonemischen Kategorie resultiert.

3.2.2 Fehlerursachen und Fehlertypen

Der nächste Analyseschritt in Marx' & Mehlhorns (2016) Fehleranalyse stellt die Bestimmung der Fehlerursachen dar. In der Forschung existiert eine Vielzahl unterschiedlicher Theorien zur Erklärung von Aussprache Fehlern. Bereits der in Kapitel 3.1 dargestellte Hintergrund zu Corders (1967) Einteilung in *errors* und *mistakes* verdeutlicht, dass die Ursachen für Fehler vielfältig sein können. So können Fehler situativ auftreten (*mistakes*) und somit unterschiedlichen Kontexten oder Situationen zu Grunde liegen. Müdigkeit, Motivation, Lehrpersonen oder das Unterrichtsklima sind nur einige Faktoren, die hier einwirken können. Fehler können auch systematisch verankert sein (*errors*) und im Gegensatz zu *mistakes* nicht zufällig und somit vorhersagbar sein. Eine zentrale Fehlerquelle stellt neben Fehlerursachen wie Übergeneralisierung, Regularisierung und Simplifizierung, Einflüsse durch Elemente des Unterrichts, Fremdsprachenkommunikationsstrategien sowie persönliche und soziokulturelle Faktoren (Kleppin 2001: 989) der Einfluss der L1 auf die Zielsprache dar. Zur Darstellung der Art und Weise wie sich die L1 auswirken kann, wird erneut Schaden (2006) herangezogen, der eine ausführliche Auflistung von

L1-bedingten Ursachen liefert. Ähnlich wie Böttger (2008) definiert er Fehlertypen, die sich als Kombinationen aus Erscheinungsform der Fehler (Lautersatz, Elision, Einfügungen; vgl. Kapitel 3.2.1) und der möglichen Ursache darstellen. Für die Lautsubstitution ergeben sich demnach je nach Ursache die folgenden Fehlertypen (Schaden 2006: 111–112):

- a. *phonetische Äquivalenz*: Lautersatz kann auf Grund von Äquivalenzklassifizierungen bei perzeptorisch und artikulatorisch sehr ähnliche Lauten auftreten;
- b. *Interferenz*: Falls dem L2-Laut im L1-Lautsystem kein äquivalenter Laut zur Verfügung steht, kann der artikulatorisch ähnlichste L1-Laut eingefügt werden. Während die phonetische Äquivalenzsetzung bei ähnlichen Lauten als Fehlerursache herangezogen werden kann, sieht Schaden (2006) Interferenz als Ursache für Fehler bei stärker abweichenden Substitutionen.
- c. *phonetische Approximation*: Lautabweichungen, die dieser Ursache zugeordnet werden können, stellen eine Annäherung an den Ziellaut dar. Erkennen lassen sie sich häufig dadurch, dass sie eine Form haben, die weder dem L1- noch dem L2-Inventar zugeordnet werden kann. Schaden (2006: 111–112) bezeichnet diese Form als Mischform zwischen perzeptorischer und artikulatorischer Schwierigkeit, d. h. Fehler treten auf, obwohl ein Laut zwar bereits richtig perzeptorisch wahrgenommen wird. Er kann allerdings noch nicht artikulatorisch realisiert werden.
- d. *orthographisch motivierte Substitution*: Die letzte aufgeführte Ursache von Lautersatz ist der Einfluss der Schriftsprache. Bei dieser Ursache wendet der Sprecher „Schriftaussprache-Regeln seiner L1 auf zielsprachliches Material an“ (Schaden 2006: 112). Einschlägige Forschung zeigt, dass L2-Lerner bei der lautlichen Umsetzung von Orthographie eher die Regeln ihrer L1, als die Regeln der L2 heranziehen (Nimz 2015). Ebenfalls zur Ursachenkategorie der Orthographie sind die unterschiedlichen Graphem-Inventare zuzuordnen. Schaden (2006: 133) führt darunter Grapheme, die in der L1 des Sprechers „keine genaue Entsprechung haben, d. h. nicht zum L1-Inventar gehören“ und solche, die „im L1-Inventar zwar auftreten, jedoch dort einen anderen Lautwert repräsentieren als in der Zielsprache“.

Neben dem Lautersatz beschreibt Schaden (2006: 116–117) Elisionen als Fehlerkategorie. Als Ursache für Elisionen können Interferenzen, d. h. Übertragungen aus der L1 gezählt werden. Diese sind eher orthographisch als rein lautlich bedingt, d. h. Schaden sieht weniger den Unterschied in den Lautsystemen als Ursache für Auslassung von Lauten, sondern vielmehr die lautliche Umsetzung der Schrift durch L1-Regeln. Eher selten und ebenfalls nur mit einer Fehlerursache erklärt, sind Einfügungen. Schaden (2006: 117–118) erklärt diese anhand von abweichenden Silbenstrukturen. So führt der Transfer von L1-Regeln zu Einfügungen von Lauten, die „letztendlich für den Sprecher die Funktion erfüllen, die Silbenstruktur von L2-Material seiner L1 anzugleichen“ (Schaden 2006: 117). Eine weitere von Schaden (2006: 119) vorgestellte Ursache für Abweichungen behandelt die Nichteinhaltung phonologischer Regeln. So können Fehler dadurch entstehen, dass

Lerner phonologische Regeln aus ihrer L₁ auf die L₂ übertragen und somit phonologische Regeln der L₂ nicht beachten. Wenn die phonologischen Regeln der jeweiligen Sprachen nicht übereinstimmen, kommt es zu Fehlern. Schaden merkt allerdings an, dass dieser Fehler eher bei nicht bedeutungsunterscheidende oder sogar redundanten Merkmalen auftritt.

Da der Fokus der vorliegenden Arbeit auf den durch die L₁ bedingten Fehlern liegt, wird im Folgenden eine Auswahl von Theorien vorgestellt, die die Prozesse erklären, die bei der fehlerhaften Realisierung auftreten (Kapitel 3.2.2.1 bis 3.2.2.3). Die sich teilweise ergänzenden, teilweise widersprechenden Hauptaspekte der Theorien werden im Anschluss (Kapitel 3.2.2.4) zusammengefasst und zum Erklärungsmodell der vorliegenden Studie zusammengeführt.

3.2.2.1 Contrastive Analysis Hypothesis

Einer der meist diskutierten und einflussreichsten Ansätze zur Erklärung von Effekten der L₁ auf den L₂-Erwerbsprozess ist die *Contrastive Analysis Hypothesis* (CAH) nach Fries (1945) und Lado (1957). Diese wurde zwar in einigen Annahmen revidiert, ist aber dennoch als prägende und noch immer bedeutsame Theorie in der Forschung zum Zweitspracherwerb einzuschätzen (Schaden 2006: 13). Die entscheidenden Aussagen der Theorie sind, dass strukturelle Unterschiede in den Sprachsystemen einer L₁ und L₂ die Hauptprobleme für Lerner darstellen und dass die Übertragung von Elementen aus der L₁ auf die L₂ die zentrale Fehlerquelle für L₂-Sprache ist. Die CAH ist grundlegend behavioristisch ausgerichtet. So betrachtet sie Sprache als ein System aus Gewohnheiten. Der L₂-Erwerb wird als maßgeblich vom gewohnten Gebrauch von L₁-Strukturen beeinflusst dargestellt. Lado (1957) postuliert, dass Fehler aus Mustern des Sprachsystems der L₁ und deren Transfer ins Sprachsystem der L₂ resultieren. Dadurch ist es möglich, Probleme und Fehler anhand einer kontrastiven Analyse der zwei Sprachsysteme – Ausgangs- und Zielsprache – vorherzusagen. In Fällen, in denen Segmente des L₁- und L₂-Systems übereinstimmen, wird davon ausgegangen, dass sich der Transfer positiv bzw. als Lernerleichterung darstellt. Im Gegensatz dazu wird bei Unterschieden davon ausgegangen, dass sie zu negativem Transfer bzw. Lernproblemen und Interferenzen führen (Lado 1957: 2). Es existieren zwei Ansätze der CAH. Die sogenannte starke Hypothese geht davon aus, dass sich Fehler in der Lerner Sprache auf Grundlage des kontrastiven Sprachvergleichs vorhersagen lassen – ganz so, wie Lado (1957) es beschreibt. Die schwache Hypothese hingegen relativiert diesen Ansatz und zieht den kontrastiven Sprachvergleich lediglich zur Begründung von Fehlern heran. Die Relativierung ergibt sich durch den Umstand, dass die CAH in ihrer starken Version durch empirische Studien nicht uneingeschränkt bestätigt werden kann. So zeigt beispielsweise Juhász (1970: 22), dass nicht nur Unterschiede, sondern auch Gemeinsamkeiten zu Problemen führen können. Richards (1972: 169) bestätigt, dass der Einfluss der Muttersprache allein als Erklärung von Fehlern zu gering sei und noch viele weitere Faktoren auf die Lerner Sprache einwir-

ken. Die ausbleibende Validierung der CAH als Prognoseinstrument für Fehler sowie die mangelnde praktische Anwendbarkeit führen dazu, dass die CAH in heftige Kritik gerät (Järventausta 2008: 14), bevor sie in relativierter Form als schwache Hypothese rehabilitiert wird. Der genannten Kritik hält James (1971: 54–55) entgegen, dass es nicht der Anspruch der CAH sei, Einflüsse aus der L1 als einzige Fehlerquelle festzulegen. Die L1 als eine mögliche Fehlerquelle zu betrachten, schließe nicht aus, noch weitere Faktoren als auf die Lerner Sprache einflussreich anzunehmen (James 1971: 54). Die CAH sei solange sinnvoll, wie sich einige Fehler mit Hilfe der Theorie erklären ließen. Die CAH kann sich in den 1970er Jahren durch Einführung der Interlanguagehypothese von Selinker (1972) rehabilitieren (Tekin 2012: 47). Dieser Umstand ist insofern erstaunlich, als die Interlanguagehypothese eng mit Corders (1967) Identitätshypothese verbunden ist, die der CAH grundlegend gegenübersteht. Selinker nimmt jedoch im Gegensatz zu Corder einen relativierenden bzw. zwischen Kontrastiv- und Identitätshypothese vermittelnden Standpunkt ein. So bezieht seine Interlanguagehypothese Transferprozesse aus der CAH mit ein und ermöglicht somit eine Neubewertung der CAH (Tekin 2012: 52). Järventausta (2008: 14) empfiehlt für einen reflektierten Standpunkt, die Ansatzpunkte der Kritik an der CAH näher zu betrachten, die sich hauptsächlich auf die unmittelbare praktische Anwendbarkeit der Hypothese bezieht. Die Hypothese kann nicht dazu verwendet werden, Fehler zu prognostizieren. Im Vergleich von zwei Sprachsystemen bestehen allerdings grundsätzlich keine theoretischen oder methodischen Schwächen. In aktueller Forschung nimmt die CAH in Form von kontrastiven Analysen daher noch immer einen didaktischen und analytischen Stellenwert ein, da sie es ermöglicht, Verbindungen zwischen L2-Fehlern und linguistischen Phänomenen wie dem Einfluss der L1 herzustellen (Derwing & Munro 2015: 65). Dabei ist immer zu bedenken, dass die CAH weitere Faktoren auf den Spracherwerbsprozess wie zum Beispiel individuelle Lernprobleme nicht miteinbezieht und zugrundeliegende kognitive Prozesse unbeachtet lässt. Welche Prozesse zu Interferenzen führen, bleibt durch eine reine sprachliche Gegenüberstellung ungeklärt. Hinweise zur Fehlerkorrektur ergeben sich also nicht. Im Bereich Aussprache ist darauf zu achten, dass eine kontrastive Analyse nicht zwischen Perzeption und Produktion der Lerner unterscheidet. In diesen Bereichen können jedoch unterschiedliche Kompetenzstadien vorliegen, wodurch es möglich ist, dass ein Lerner bereits L2-Laute voneinander unterscheiden kann, bevor er sie selbst produzieren kann (zu den Limitationen vgl. Derwing & Munro 2015: 65–66). Viele Erklärungsmöglichkeiten von Ursachen oder Korrekturmöglichkeiten von Fehlern bleiben also unbeachtet.

Bohn (2002: 197) bewertet den Einsatz der CAH zur Ermittlung von Fehlern in Form einer kontrastiven Analyse von Lautsystemen ohne empirische Überprüfung kritisch und bezeichnet die Methode als *armchair method*. Dieser Begriff stellt hauptsächlich die Kritik an der mangelnden empirischen Untermauerung dar. Sprachliche Unterschiede oder Gemeinsamkeiten anhand von Lautinventaren und Symbolen zu ermitteln sei laut Bohn (2002: 198) zwar die einfachste Art des Sprachvergleichs, allerdings auch die unzureichendste, da keine Einsichten zu den zugrundeliegenden, kognitiven Prozessen der

Aussprachefehler gewonnen werden. Daneben kritisiert er die folgenden Mängel, die in der Umsetzung der Methode auftreten und somit zu fehlgeleiteten Ergebnissen führen können (Bohn 2002: 198):

- a. Symbole zur Lautbeschreibung werden arbiträr und unterschiedlich eingesetzt;
- b. Laute, die eigentlich dieselbe Qualität haben, werden unterschiedlich dargestellt;
- c. Laute, die eine unterschiedliche Qualität haben, werden gleich transkribiert.

Im Gegensatz zu Bohn (2002) erläutert Nossok (2009), dass die CAH für sprachliche Bereiche der Syntax oder Morphologie zwar problematisch anzusehen, ihr Einsatz im phonetisch-phonologischen Bereich unter Berücksichtigung der genannten Limitationen und entsprechenden Anpassungen allerdings als gewinnbringend einzustufen sei. Auch Spillner (2009: 551) konstatiert, dass kontrastive Sprachvergleiche trotz aller Kritik „nach wie vor ein wichtiges methodisches Hilfsmittel für die Fremdsprachendidaktik“ sind, solange nicht davon ausgegangen wird, dass alle Fehler durch den negativen Transfer verursacht werden und eine empirische Fehleranalyse an die theoretischen Vergleiche anschließt. So nutzen auch aktuelle Studien kontrastive Analysen, um das Zusammenwirken von Erstsprache, Lernaltersprache und Zielsprache zu untersuchen (Marx & Mehlhorn 2016: 299). Auch Bohn (2002: 198) relativiert seine Kritik insofern, als die Methode des kontrastiven Vergleichs der Lautinventare berechtigt sei, wenn es sich um einen „first approach“ handle und dieser von weiteren Methoden ergänzt und beim Einsatz auf eine „very narrow phonetic transcription“ geachtet werde. Diese ist zwingend notwendig. So ermöglichen Sprachvergleiche auf Grundlage von traditioneller oder sprachspezifischer Transkription unter Umständen keine aussagekräftigen Rückschlüsse, da sich die zu vergleichenden Tradierungen möglicherweise bereits in den der Transkription zu Grunde liegenden Formalitäten unterscheiden (vgl. Kapitel 3.3.1).

3.2.2.2 Speech Learning Model

Das von Flege (1995) entwickelte *Speech Learning Model* (im Folgenden SLM) stellt eine weitere Spracherwerbtheorie dar, die zur Erklärung zur Entstehung von Aussprachefehlern herangezogen werden kann. Im SLM sind insgesamt vier Postulate und sieben Hypothesen zur fremdsprachlichen Lautperzeption und -produktion auf segmentaler Ebene formuliert. Das Modell bildet ein theoretisches Fundament zur Untersuchung von Spracherwerb und Lernaltersprache, indem es die für Fehler ursächlichen kognitiven Prozesse aufarbeitet. Zwei grundlegende Annahmen sind für Fleges SLM Modell ausschlaggebend:

- a. Perzeption wird als obligatorischer Prozess betrachtet, der zu Problemen bei der Produktion führen kann. Im Gegensatz zur in Kapitel 3.2.2.1 dargestellten CAH bezieht Fleges SLM somit Wahrnehmungsprozesse mit ein.
- b. Fleges Theorie beruht auf der Annahme, dass Laute, ebenso wie alle anderen Wissensrepräsentationen, in Kategorien gespeichert und verarbeitet werden und neue L2-Laute nach bereits bestehenden Kategorien des L1-Systems wahrgenommen

werden (Flege 1991: 702). Grundlegend dabei ist die Unterscheidung von *identischen*, *neuen* und *ähnlichen* Lauten (Flege 1987: 10–11). Identische Laute stellen für Lerner kein Problem dar, da sie den Lauten des L1-Systems exakt gleichen und ohne weiteres übernommen werden können. Ähnliche Laute hingegen sind problematisch, da sie, auch wenn sie L1-Lauten gleichen, verschieden von diesen sind. Die Aneignung dieser Laute erfordern die Modifikation bereits etablierter segmentaler Kategorien (Flege 1987: 10). Dies stellt eine größere Schwierigkeit dar, als die Neubildung von Kategorien, wie es neue Laute erfordern. Neue Laute sind auditiv vom L1-System so unterschiedlich, dass es keine Überschneidungen in den Kategorien gibt.

Flege formuliert in seinem SLM vier Postulate. Das erste Postulat besagt, dass Mechanismen des Erstspracherwerbs ein Leben lang erhalten bleiben und auch für den L2-Erwerb genutzt werden. Das zweite Postulat stellt dar, dass Laute in phonetischen Kategorien im Langzeitgedächtnis spezifiziert sind. Das dritte Postulat besagt, dass in der Kindheit angelegte phonetische Kategorien sich ein Leben lang entwickeln können. Das vierte Postulat bedeutet, dass Sprecher mehrerer Sprachen L1- und L2-Laute bzw. deren Kategorien auseinanderhalten können.

Auf den vier Postulaten bauen sieben Hypothesen auf. Sie stellen Ergebnisse Fleges eigener Studien der 1980er 1990er Jahre dar. Die erste Hypothese (H₁) nimmt Bezug auf die perzeptuelle Wahrnehmung von L1- und L2-Lauten. Flege (1995: 239–240) belegt anhand einer Studie von Sheldon & Strange (1982), dass die perzeptuelle Wahrnehmung von Lauten positionssensibel ist. In der Studie fällt japanischen Englischlernern die Wahrnehmung und Produktion eines Kontrasts zwischen /r/ und /l/ in wortfinalen Kontexten leichter als in wortinitialen. Fleges zweite Hypothese (H₂) besagt, dass Lerner in der Lage sind, neue L2-Kategorien zu bilden, sofern sie den Laut als phonetisch verschieden vom nächsten L1-Laut wahrnehmen können. Je größer der wahrgenommene Unterschied zwischen L2- und nächstem L1-Laut, desto wahrscheinlicher ist, dass die Laute auseinandergehalten werden. In dieser dritten Hypothese (H₃) ist impliziert, dass L2-Laute, die L1-Lauten ähnlich sind, am problematischsten zu unterscheiden und neu zu erwerben sind. Dieser Punkt ist einer der entscheidenden Unterschiede zu Lados CAH. Für Lado sind gerade die ähnlichsten Laute am einfachsten zu lernen. Die vierte Hypothese (H₄) bezieht sich auf das Lernalter. Flege, Munro & MacKay (1995) zeigen einen linearen Zusammenhang zwischen L2-Erwerbseintrittsalter und Akzent bei italienischen Englischlernern. Die Ergebnisse lassen darauf schließen, dass die Neuformation von Lautkategorien mit zunehmendem Alter erschwert ist und neue Laute bereits bestehenden Kategorien zugeordnet werden. Dieses Phänomen, bei dem die Neuuzuordnung eines L2-Lauts blockiert wird, bezeichnen Flege, Munro & MacKay (1995: 3133) als *äquivalente Klassifikation*. Dabei werden einer einzelnen phonetischen Kategorie unterschiedliche L1- und L2-Laute zugeordnet und perzeptuell zu sogenannten Diaphonen verbunden. Die fünfte Hypothese (H₅) folgt aus der vierten und besagt, dass sich die Laute der Diaphone in der Produktion ähneln (Flege 1995: 241). Die sechste Hypothese (H₆) gibt an,

dass die gleichen lautlichen Kategorien qualitativ unterschiedlich von denen bilingualer Sprecher sind. Flege erklärt dies dadurch, dass bei der Neubildung einer lautlichen Kategorie im Fremdspracherwerb die neuen L2-Laute von der bereits bestehenden L1-Kategorie *abgespalten* werden (Flege 1995: 242). Bei monolingualen Sprechern geschieht eine solche Abspaltung nicht, wodurch sich der Unterschied ergibt. Die siebte Hypothese (H₇) postuliert schließlich, dass die Eigenschaften der phonetischen Kategorien durch die Lautproduktion repräsentiert werden (Flege 1995: 239). Als entscheidender Unterschied zur CAH sollte festgehalten werden, dass das SLM davon ausgeht, dass besonders ähnliche L2-Laute Lernern vermutlich größere Probleme bereiten. Dies hängt mit dem Umstand zusammen, dass ihnen die Distinktion der Laute vermutlich leichter fällt, wenn sich L2- und L1-Laute stark unterscheiden.

3.2.2.3 Perceptual Assimilation Model

Ebenso wie beim CAH und SLM spielt die Unterscheidung in Ähnlichkeit und Unterschiedlichkeit zwischen L1- und L2-Lauten die zentrale Rolle im *Perceptual Assimilation Model* (im Folgenden PAM) von Best (1995). In Bests Modell stellt die Perzeption als Voraussetzungen für erfolgreiche Produktion den zentralen Punkt dar. Es wird in der fremdsprachendidaktischen Forschung zu einem der wichtigsten Modelle gezählt, da es für viele Studien zur Perzeption von L2-Lernern grundlegend ist (Bohn 2002: 194). Während der 1980er Jahre untersucht Best die Perzeption von erwachsenen Lernern und fasst ihre Ergebnisse im PAM zusammen. Im Jahre 2007 wird das Modell für den L2-Erwerb erweitert (Best & Tyler 2007). Grundlegend für das Modell ist, dass im Erstspracherwerb perzeptuelle Kategorien entwickelt werden, die der Norm einer L1 entsprechen. Im späteren L2-Erwerb versucht der Lerner, unbekannte Laute dem bereits bestehenden System zuzuordnen bzw. Laute zu Kategorien zu assimilieren, die bereits durch die L1 bestehen. Dieser Filterprozess wird gemeinhin mit der Metapher des phonologischen Siebs gleichgesetzt (z. B. bei Derwing & Munro 2015: 67; Nimz 2015: 7; ursprünglich Trubetzkoy 1971 [1937]). Der Grund für Wahrnehmungsschwierigkeiten zwischen ähnlichen Lauten ist demnach, dass beide Laute zu ein und derselben perzeptuellen Kategorie assimiliert werden und die Lerner somit für zwei eigentlich unterschiedliche Laute ein und denselben Laut wahrnehmen. Best & Tyler (2007: 23) bezeichnen dieses Phänomen als *Single Category Assimilation*. Ein für diese Arbeit sprachenpaarspezifisches Beispiel stellt 3.1 dar.

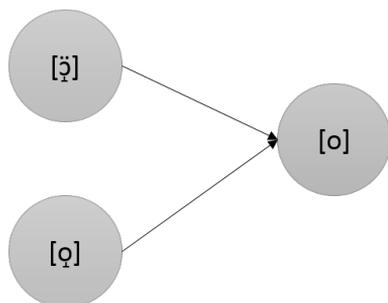


Abb. 3.1: Beispiel für eine *Single Category Assimilation* (angepasste Darstellung nach Derwing & Munro 2015: 67)

Im Beispiel werden die zwei dänischen Laute [ɔ̃] und [ɔ] vom deutschen L2-Lerner mit derselben Kategorie des Lauts [o] assimiliert. Dadurch würde sich für die zwei unterschiedlichen Wörter in (1) keine Bedeutungsunterscheidung geben.

- (1) *låne* ‚leihen‘ [ˈlɔ̃:nə] * [ˈlɔ:nə]
Lone ‚Lone (weiblicher Vorname)‘ [ˈlɔ:nə]

Ein weiteres im Modell behandeltes Phänomen beschreibt die Zuordnung von fremdsprachlichen Lauten zu Kategorien des muttersprachlichen Lautsystems, auch wenn diese sich nicht eins zu eins entsprechen. Dieses Phänomen bezeichnen Best & Tyler (2007: 23) als *Two Category Assimilation*. Der Terminus verdeutlicht im Gegensatz zur *Single Category Assimilation*, dass jeder einzelne neue oder ähnliche Laut einer muttersprachlichen Kategorie zugeordnet wird (verdeutlicht in Abb. 3.2).

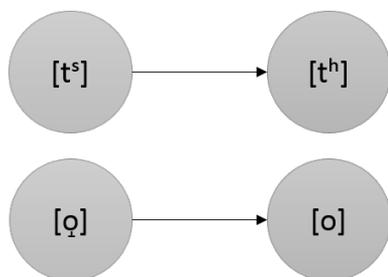


Abb. 3.2: Beispiel für eine *Two Category Assimilation* (angepasste Darstellung nach Derwing & Munro 2015: 67)

Abb. 3.2 demonstriert, dass der dänische Laut [tˢ] von L2-Lernern als deutsches [tʰ] eingeordnet wird, obwohl *t* im Dänischen affriziert wird. Dänisches [ɔ] wird mit deutschem [o] assimiliert, obwohl [o] im Dänischen etwas geschlossener ausgesprochen wird als im Deutschen. Neben diesen beiden Phänomenen beschreibt das Modell außerdem, dass es Lernern möglich ist, Laute voneinander zu unterscheiden, die nicht mit dem L1-System

in Verbindung gebracht werden können. Best, McRoberts & Sithole (1988: 351) zeigen dies anhand von Schnalzlauten im Zulu, die von L1-Englischsprechern problemlos unterschieden werden können, obwohl oder gerade weil sie keiner englischen Lautkategorie zugeordnet werden können. Best & Tyler (2007: 23) bezeichnen solche Laute als *non-assimilable*.

Ähnlich wie das SLM von Flege sagt das PAM von Best also für L2-Laute, die L1-Kategorien ähneln, eine schlechte perzeptuelle Diskriminierung voraus. Die Diskriminierung der Laute fällt den Lernern deshalb so schwer, da diese phonologisch an eine L1-Kategorie assimiliert sind. Die Lerner nehmen also statt des L2-Lauts den L1-Laut wahr. Im SLM wird dieses Phänomen als *äquivalente Klassifikation* bezeichnet, während das PAM das Phänomen als *Single Category Assimilation* bezeichnet. Dieses Problem kann nur überwunden werden, indem der Lerner eine neue phonetische Kategorie entwickelt. Dies gelingt umso problematischer, je besser die Kategorien zusammenpassen.

3.2.2.4 Zusammenfassung und Erklärungsmodell der vorliegenden Studie

Der Hauptunterschied zwischen den drei vorgestellten Modellen besteht darin, dass das CAH das endgültige Sprachprodukt und dessen Fehler fokussiert, während das PAM und das SLM Lernprozesse auf Grundlage von Perzeptions- und Produktionsmechanismen untersuchen (Derwing & Munro 2015: 69). Studien, die auf dem PAM und SLM aufbauen, können miteinbeziehen, dass sich die untersuchte L2-Sprache mit der Zeit verändert und entwickelt. Auf diese Weise werden auch Erwerbsprozessunterschiede individueller Lerner mitbedacht. Außerdem beziehen die Modelle kognitive Mechanismen wie beispielsweise Assimilationsprozesse mit ein, die für Probleme verantwortlich sein können, während die CAH auf eher statische Sprachstrukturen ausgerichtet ist und Schlussfolgerungen aus bestehenden Sachverhalten zieht. Die CAH wird aus diesen Gründen als Prognose- und Erklärungsinstrument in der Forschung häufig kritisiert (z. B. Bohn 2002: 197; Derwing & Munro 2015: 65–66).

Der Einsatz der CAH als Mittel zur Generierung von Ursachenhypothesen und Erklärung von Fehlern erfolgt in der vorliegenden Arbeit nicht ohne kritische Anwendung oder als alleiniges Mittel. So werden zur Hypothesenbildung der Fehlerursachen auch Erkenntnisse der anderen beiden vorgestellten Spracherwerbsmodelle hinzugezogen. Während sich anhand der CAH einige Fehler durch Unterschiede im Lautsystem von L1 und L2 erklären lassen, beruhen andere Fehler möglicherweise auf Ähnlichkeiten oder Gemeinsamkeiten. Auf diese Weise wird beispielsweise die Äquivalenzklassifizierung aus

dem SLM berücksichtigt, aus der hervorgeht, dass neue bzw. unbekannte L2-Laute leichter zu erwerben sind als ähnliche.⁶

Neben der Ähnlichkeit zwischen einzelnen Lauten besteht im Sprachenpaar Deutsch-Dänisch auch häufig lexikalische Ähnlichkeit bis Gleichheit zwischen Wörtern (vgl. (2)).

- (2) dän. *rose* [ˈʁoːsə]
 dt. *Rose* [ˈʁoːzə]

Die lexikalische Ähnlichkeit kann Ursache für lautliche Interferenzerscheinungen sein und wird deshalb zu den möglichen Fehlerursachen ergänzend miteinbezogen. Neben den oben dargestellten grundlegenden theoretischen Modellen orientiert sich die vorliegende Arbeit außerdem an der Fehlertypeneinteilung von Schaden (2006; vgl. Kapitel 3.2.2). Somit ergibt sich ein Pool an möglichen Ursachen (vgl. Tab. 3.1), die zur Hypothesenbildung der Analyse in Kapitel 3.4.2 herangezogen werden. In der Darstellung der Ursachen in Tab. 3.1 werden sowohl sprachlich-strukturellen Unterschiede als auch kognitive und motorische Prozesse aufgeführt. Auf diese Weise können Fehler hinsichtlich drei Fragestellungen analysiert werden:

- Welche sprachlich strukturellen Umstände liegen dem Fehler zu Grunde?
- Welche kognitiven und motorischen Prozesse führen zur fehlerhaften Realisierung?
- Wie stellt sich die Realisierung des Fehlers sprachlich dar?

Diese genaue Unterscheidung ist besonders dann von Interesse, wenn didaktische Implikationen aus der Analyse folgen sollen. Wenn auf Aussprachefehler spezifisch eingegangen werden soll, müssen kognitive und motorische Schwierigkeiten im sprachlich-strukturellen Kontext betrachtet werden. Sprachliche Unterschiede münden nicht zwangsläufig in einer fehlerhaften Realisierung, wie in Abb. 3.3 nachzuvollziehen ist.

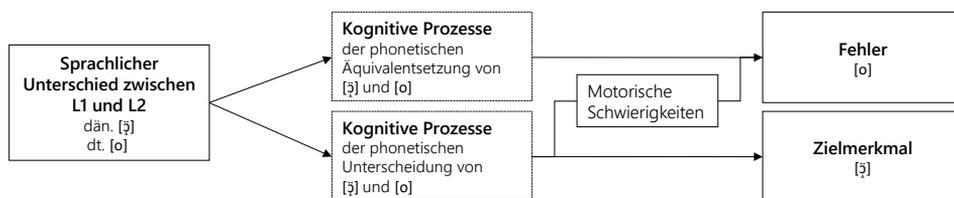


Abb. 3.3: Zusammenhang zwischen sprachlichen Unterschieden, internen Prozessen und Lautrealisierung

⁶ Dass sich die Haupterkenntnisse beider Studien teilweise gegenüberstehen, ist unproblematisch, da bei der Generierung von Ursachenhypothesen zum besseren Verständnis der Fehler alle Möglichkeiten einbezogen werden sollten. Die Fusion der Erkenntnisse aller Modelle stellt somit einen erweiterten Analysehorizont dar.

Dänisches [ɔ̃] und deutsches [o] unterscheiden sich akustisch und artikulatorisch voneinander. Aus diesem Umstand ergibt sich allerdings nicht zwangsläufig eine fehlerhafte Realisierung durch die Lerner. Vielmehr können die kognitiven Prozesse der Lerner unterschiedlich ausfallen. Beispielsweise ist es möglich, dass die Lerner die Unterschiede zwischen den Lauten nicht wahrnehmen und beide gleichsetzen.

In diesem Fall wird es zu einer fehlerhaften Realisierung kommen. Es ist allerdings auch möglich, dass die Lerner die Laute phonetisch sowohl perzeptorisch als auch produktiv voneinander unterscheiden und den Ziellaut – im besten Fall – normkonform realisieren, d. h. trotz sprachlicher Unterschiede kommt es nicht zu einem Aussprachefehler. Da Aussprache neben der kognitiven Komponente allerdings auch von einer motorischen Komponente bestimmt wird, mündet eine bewusste phonetische Unterscheidung nicht zwangsläufig in einer zielsprachlichen Realisierung. Es können an dieser Stelle des Realisierungsprozesses motorische Schwierigkeiten auftreten, die sich beispielsweise auf die Weise darstellen, dass den Lernern die Aussprache noch nicht ganz richtig gelingt, oder sie versuchen den schwierigen Laut zu vermeiden. In diesem Fall kommt es ebenfalls zu einer fehlerhaften Realisierung.

Bei der Analyse von Fehlern ist somit zu bedenken, dass sprachliche Unterschiede lediglich einen Teil der Fehlerursachen erklären, es für die didaktische Implikation allerdings ebenso relevant ist zu untersuchen, welche internen Prozesse beim Lerner plausiblerweise aus diesen Unterschieden hervorgehen. Die Darstellung in Abb. 3.3 ist als vereinfacht zu verstehen, da hier mögliche Zusammenhänge zwischen sprachlichen Unterschieden und kognitiven und motorischen Prozessen von Lernern nur schematisch gezeigt werden. In Tab. 3.1 wird eine ausführlichere Darstellung vorgenommen. Diese ist grundlegend für die Fehleranalyse in Kapitel 3.4.2, in der Hypothesen zu möglichen Fehlerursachen aufgestellt werden.

sprachlicher Unterschied	kognitiver und/oder motorischer Prozess
Ziellaut im L1-Lautinventar nicht vorhanden	Interferenz: Lautersatz des Ziellauts durch L1-Laut
	Vermeidung: Elision des Ziellauts
	Übergeneralisierung: Anwendung einer lautlichen L2-Regel aus einem anderen Kontext
	phonetische Approximation: lautliche Annäherung
Laute der L1 und L2 ähneln sich	Interferenz: Lautersatz des Ziellauts durch L1-Laut
orthographische Regeln unterscheiden sich	orthographische Interferenz: Realisation des Zielworts nach L1-Schriftausspracheregeln
phonologische Regeln unterscheiden sich	phonologische Interferenz: phonologische Regel wird auf Zielwort übertragen
L1- und L2-Wörter ähneln sich lexikalisch	lexikalische Interferenz: Zielwort wird wie L1-Wort ausgesprochen

Tab. 3.1: Zusammenhänge zwischen sprachlichen Unterschieden und kognitiven und motorischen Prozessen von Lernern

Der erste in Tab. 3.1 genannte sprachliche Unterschied beschreibt den Fall, dass der Ziellaut im L1-Lautinventar nicht vorhanden ist bzw. es sich im Sinne von Fleges (1987) SLM um einen neuen Laut handelt. Die Perzeption und Artikulation kann sich als ungewohnt oder schwierig erweisen. Es lassen sich unterschiedliche kognitive und motorische Prozesse vorstellen, die zu Fehlern führen können:

- a. *Interferenzfehler*: Die Lerner nehmen eine phonetische Äquivalentsetzung von unterschiedlichen L1- und L2-Lauten vor, indem sie den neuen Laut einem Laut ihres Lautinventars zuordnen. Bei neuen Lauten besteht – anders als bei ähnlichen Lauten – keine akustische Ähnlichkeit. Trotzdem tritt die Zuordnung auf, da noch keine Kategorie für den neuen Laut besteht. In der Folge ersetzen die Lerner den L2-Laut durch einen artikulatorisch abweichenden, für sie jedoch bekannten L1-Laut (vgl. (3)).

(3) *hedder* ‚heißen (PRÄS)‘ *['hedv] statt ['hɛð²v]

- b. *Elision*: Die Lerner vermeiden den neuen Laut, wodurch es zu einer Lauttilgung kommt (vgl. (4)).

(4) *ud* ‚raus‘ [u] statt [uð²]

- c. *Übergeneralisierung*: Die Lerner wenden eine Regel aus einem anderen lautlichen Kontext zur Realisierung des fremden Lautes an. Ein Beispiel für Dänisch als

Zielsprache wäre die Übertragung von Assimilierungsprozessen auf Wörter, in denen diese normkonform nicht stattfinden (vgl. (5)).

(5) *leve* ‚leben‘ *['lɛ:] statt ['lɛ:ɤ] oder ['lɛ:u]

- d. *phonetische Approximation*: Wenn Lerner die neuen Laute bereits gelernt haben und ihnen kognitiv die Wahrnehmung und Unterscheidung bereits gelingt, ihnen die Produktion in zielsprachlicher Qualität allerdings noch schwer fällt, wird von einer Annäherung an den Ziellaut gesprochen (vgl. Bsp. (6)).

(6) *kåbe* ‚Mantel‘ *['k^ho:ɔ] statt ['k^hö:ɔ]

Die internen Prozesse (a) bis (d) sind ohne genaue Bestimmungsmethoden äußerlich nicht voneinander zu unterscheiden. Um Aussagen darüber treffen zu können, ob der Aussprachefehler auf noch nicht vorhandenem Bewusstsein über Unterschiede bzw. nicht gelingender Perzeption beruht oder aber durch motorische Probleme begründet ist, müssen unterschiedliche Tests durchgeführt werden, die auf die spezifischen Prozesse abzielen.

Die zweite in Tab. 3.1 präsentierte sprachliche Gegebenheit zeigt, dass Ähnlichkeit zwischen dem Ziel- und L1-Laut bestehen kann. Flege (1987) bezeichnet diese Laute – im Gegensatz zu neuen Lauten – als *ähnliche Laute* (*similar sounds*). Diese können auch als eine Unterkategorie der neuen Laute verstanden werden. So ist es möglich, dass sich Laute stark ähneln, aber eben nicht vollständig gleichen. Genau genommen müsste also von unterschiedlichen Lauten gesprochen werden. In dieser Arbeit werden diese Laute allerdings in Anlehnung an Flege (1987) als ähnliche Laute bezeichnet. Diese können ursächlich für die folgenden kognitiven und motorischen Prozesse sein:

- a. *phonetische Approximation*: Erstens ist es – ebenso wie bei den neuen Lauten – möglich, dass die Lerner den Laut bereits wahrnehmen, aber noch nicht korrekt realisieren können. Der kognitive Prozess würde somit eine lautliche Annäherung darstellen.
- b. *Interferenz*: Ebenso wie bei neuen Lauten ist es möglich, dass Interferenzprozesse auftreten, die Lerner allerdings bei ähnlichen Lauten Schwierigkeiten mit der Perzeption haben.

Als weitere mögliche sprachliche Unterschiede sind in Tab. 3.1 die Orthographie, Lexik und Phonologie genannt. Wenn sich die Lerner aus unterschiedlichen Gründen (unbekannter Laut, problematische Artikulation, etc.) bei der Realisation des Ziellauts ausschließlich auf die orthographische Umsetzung verlassen, stellen die kognitiven Prozesse orthographisch Umsetzung anhand von L1-Regeln dar (vgl. (7)).

(7) *hvad* ‚was‘ *['hvät] statt ['væð]

Die Phonologie kann sich dann auswirken, wenn den Lernern die L2-Regeln nicht bekannt sind und sie die Ziellaute bzw. -wörter nach L1-Regeln realisieren (vgl. (8)).

(8) *tak* ‚danke‘ [t^häk^h] statt [t^säg]

Die lexikalische Ähnlichkeit von deutschen und dänischen Wörtern kann zu Fehlern lexikalischer Interferenz führen (vgl. (9)).

(9) *klasse* ‚Klasse‘ [ˈkʰlāsə] statt [ˈkʰlæsə]

Nach dieser detaillierten Darstellung der möglichen sprachlichen Unterschiede und damit zusammenhängender kognitiver Prozesse wird im Folgenden dargestellt, wie die Fehleranalyse in der vorliegenden Arbeit durchgeführt wird.

3.2.3 Vorgehensweise

Die vorangegangenen Kapitel zeigen, dass Marx' & Mehlhorns (2016) Einteilung der Schritte Identifikation, Klassifizierung und Ursachenerklärung unbedingt getrennt voneinander behandelt werden sollten. Die Trennung zwischen deskriptiver und explanativer Beschreibung stellt einen wichtigen Zusatz zur Objektivierung dar. Die Vorgehensweise der Fehleranalyse der vorliegenden Arbeit orientiert sich an den von Marx & Mehlhorn (2016) definierten Schritten einer Fehleranalyse und bezieht die Überlegungen aus Kapitel 3.2.1 und 3.2.2 ein. Es wird sowohl möglich, Fehler deskriptiv zu beschreiben und Hypothesen zu möglichen Ursachen aufzustellen, als auch eine Einordnung in abstraktere Fehlertypenkategorien vorzunehmen. Die Vorgehensweise wird in Tab. 3.2 dargestellt, wobei die Spalten den chronologischen Ablauf der Fehleranalyse demonstrieren.

	untersuchtes Zielmerkmal	Klassifikation (Symptom)	Fehlerursache	Fehlertyp
	hypothetisch	deskriptiv	explanativ	
1a	? statt [ð]	Lautersatz [l] statt [ð]	weiches <i>d</i> [ð] ist in L1 nicht vorhanden (perzeptorische oder artikulatorische Schwierigkeiten führen zu Interferenz- oder Approximationsprozessen)	Lautersatz durch phonetische Approximation
1b				Lautersatz durch Interferenz
2a		Lautersatz [d] statt [ð]	weiches <i>d</i> [ð] ist in L1 nicht vorhanden (perzeptorische oder artikulatorische Schwierigkeiten führen zu Interferenz- oder Approximationsprozessen)	Lautersatz durch phonetische Approximation
2b				Lautersatz durch Interferenz
2c			Orthographie als Realisierungsgrundlage	orthographische Interferenz

Tab. 3.2: Schema zur Vorgehensweise bei der Fehleranalyse

Zunächst wird das lautliche Phänomen ermittelt, von dem anhand des Sprachvergleichs angenommen wird, dass es deutschen Lernern Probleme bereitet (Spalte 1). Während einige Fehler auf Grund von Unterschieden im Lautsystem von L1 und L2 zu erwarten sind, werden auch Hypothesen zu wahrscheinlichen Fehlern auf Grund von Ähnlichkeiten oder Gemeinsamkeiten aufgestellt (vgl. SLM und PAM). In Kapitel 3.3 werden die relevanten Merkmale anhand einer kontrastiven phonetischen Analyse beider Sprachsysteme dargestellt.

Im zweiten Schritt (Spalte 2) erfolgt die Darstellung der tatsächlichen Erscheinungsform der Fehler und deren abstrakte Beschreibung in Form der Klassifizierungskategorie (z. B. Lautersatz). Die Klassifizierung stellt einen deskriptiven Analyseschritt dar. Die hierzu verwendeten Termini sind in Kapitel 3.2.1 aufgeführt.

In einem dritten Schritt (Spalte 3) werden Hypothesen zu den Ursachen der konkreten Fehler aufgestellt. Dieser Analyseschritt ist explanativ. Die zur Hypothesenbildung in Betracht gezogenen Ursachen ergeben sich aus dem Zusammenhang sprachlicher Unterschiede und daraus resultierender kognitiver und motorischer Prozesse. Dieser Zusammenhang wird in Kapitel 3.2.2.4 dargestellt.

Die letzte Spalte definiert den Fehlertyp. Dieser ergibt sich aus der Zusammenfassung von Klassifizierung und Ursachenbestimmung des Fehlers. Dieser Analyseschritt stellt ebenfalls einen interpretatorischen bzw. explanativen Analyseschritt dar.

Als beispielhaftes schwieriges Merkmal wird in Tab. 3.2 das weiche *d* [ð] herangezogen. Für diesen Laut ist anzunehmen, dass er deutschsprachigen L2-Dänischlernern Pro-

bleme bereiten wird, da er im deutschen Lautinventar nicht vorkommt. Es ist zu vermuten, dass als Symptom bzw. konkrete Realisierung ein anderer Laut als [ø] von den Lernern gebildet wird. Wie sich dieser Laut darstellt, lässt sich hypothetisch zwar mutmaßen, zeigt sich empirisch allerdings erst in der Untersuchung realer Lerner Sprache (vgl. Kapitel 3.4). Die Fehleridentifizierung ergibt zwei unterschiedlich konkrete fehlerhafte Lerneräußerungen (Beispielfall 1 und Beispielfall 2), die allerdings beide einen Lautersatz darstellen (Spalte 2). In Beispielfall 1 realisieren die Lerner [l] statt [ø], in Beispielfall 2 realisieren sie [d] statt [ø]. An die Klassifizierung schließt sich die explanative Beschreibung der Ursachen an. Das weiche *d* [ø] ist im deutschen Lautinventar nicht vorhanden und bereitet den Lernern vermutlich artikulatorische Probleme. Da die Zuschreibung von Ursachen immer Interpretation bleibt, solange keine möglichst introspektiven Methoden zur Erkenntnisgewinnung eingesetzt werden, ergeben sich für ein auftretendes Symptom häufig mehrere mögliche Fehlertypen. Beim Lautersatz [l] statt [ø] im Beispielfall 1a wird der Lautersatz dahingehend interpretiert, dass der Lerner versucht den Ziel-laut zu bilden, ohne dass ihm dies normkonform gelingt. Als Fehlertyp ergibt sich eine phonetische Approximation, d. h. lautliche Annäherung an den Ziellaut (Beispielfall 1a). Ebenso gut ist es allerdings möglich, dass es sich bei dem Fehler um eine Interferenz handelt. Das würde bedeuten, dass der Fehler dahingehend zu interpretieren ist, dass der Lerner auf Grund der Fremdheit des Lautes einen ihn bekannten Laut aus seinem Lautinventar einsetzt und es sich somit nicht um eine Annäherung handelt (Beispielfall 1b). Auch im Fall von Beispielfall 2 können mehrere wahrscheinliche Ursachen vorliegen. Erstens könnte ebenso wie in Beispielfall 1 der Fehler aus artikulatorischen oder perzeptorischen Problemen resultieren (Beispielfälle 2a und 2b). Es ist allerdings auch möglich, dass der Lerner die orthographischen Regeln seiner L1 als Realisierungsgrundlage nutzt, und sich somit eine orthographische Interferenz als Fehlertyp ergibt (Beispielfall 2c). In Kapitel 3.4.2 werden die hier dargestellten Schritte einer Fehleranalyse anhand mehrerer lautlicher Phänomene durchgeführt.

3.3 Kontrastive phonetische Analyse Deutsch–Dänisch

In diesem Kapitel erfolgt eine detaillierte kontrastive Darstellung von Abweichungen und Gemeinsamkeiten bzw. Ähnlichkeiten der Lautsysteme Deutsch und Dänisch. Aus dieser Darstellung werden hypothetische Probleme für deutschsprachige L2-Dänischler abgeleitet. Dabei wird zwischen rein lautlich und rein orthographisch begründeten Problemen unterschieden.⁷ Von besonderem Interesse für diese Arbeit sind Lautbildungsprobleme, weshalb bei der Fehleranalyse zwar auf orthographische Realisierungs-

⁷ Bereits bei Jespersen (1901) findet sich diese Unterscheidung. Jespersen trennt zwischen *lyddannelsesfejl* ‚Lautbildungsfehlern‘ und *lydanvendelsesfejl* ‚Lautanwendungsfehlern‘. Erstere entstehen aus Problemen bei der Artikulation, letztere sind Nichtübereinstimmungen innerhalb von lautlicher Umsetzung und Orthographie geschuldet und mit Hilfe von Lautschrift zu vermeiden (Jespersen 1901: 157).

probleme eingegangen wird, diese allerdings nicht im Fokus betrachtet werden. So ist davon auszugehen, dass deutschsprachige L2-Dänischlerner eigentlich kein Problem bei der Artikulation des Diphthongs [ü] haben, die orthographische Umsetzung allerdings Probleme bereiten kann:

(10) *hav* ‚Meer‘ [häv]

In der kontrastiven Gegenüberstellung werden somit lediglich Schwierigkeiten herangezogen, die sich aus lautlichen Kontexten ergeben. In der empirischen Fehleranalyse können Fehler trotzdem auch durch orthographische Interferenz erklärt werden. Im Folgenden wird zunächst die Wahl des dieser Arbeit zugrunde liegenden Transkriptionssystems begründet (Kapitel 3.3.1), um anschließend segmentale, suprasegmentale und prozesshafte Unterschiede des Deutschen und Dänischen zu ermitteln (Kapitel 3.3.2 und 3.3.3). Die Ausführungen resultieren in der übersichtlichen Darstellung der mit der Praxis abzugleichenden Merkmale in Kapitel 3.3.4.

3.3.1 Auswahl des Transkriptionssystems

Auf Grund uneinheitlicher Beschreibungsformen, die besonders im Dänischen vorhanden sind, wird im Folgenden näher begründet, für welche Form der Transkription sich in dieser Arbeit entschieden wurde. Grundlegende Beschreibungen der dänischen Phonetik und Phonologie finden sich bei Grønnum (2009) und Basbøll (2005). Im kontrastiven Bereich existieren Arbeiten von Wagner & Basbøll (1985), Höder (2016) sowie in Teilen Fredsted (1989, 2011). Diese Werke greifen auf verschiedene, teils stark voneinander abweichende Transkriptionssysteme zurück, die zwar auf dem Internationalen Phonetischen Alphabet (IPA; *Handbook* 1999) basieren, aber die Notation der Laute auf unterschiedliche Art und Weise vornehmen. Zudem existiert für Dänisch ein eigenes, traditionelles Transkriptionssystem (*Dania* nach Jespersen 1890–1892; Kristensen 1924), das teilweise zur Beschreibung herangezogen wird und für weitere Varianten sorgt. Das Referenzwörterbuch *Den Danske Ordbog* (DDO), Grønnum (2009), Basbøll (2005) und Fredsted (2011) nutzen das sprachübergreifend einsetzbare IPA vereinfachten Varianten. Die dort eingesetzte Transkription berücksichtigt lediglich bedeutungsdifferenzierenden Merkmale, sie wird als weite Transkription bezeichnet (bei Pompino-Marschall 2009: 268 auch als „breite Transkription“ geführt). Eine solche Form der Lautbeschreibung unterscheidet sich von einer detaillierten Beschreibung, die als enge Transkription bezeichnet wird. Die Arbeiten, die eine vereinfachte Form nutzen, begründen diese damit, dass im untersuchten Sprachsystem kein weiterer zu verwechselnder Laut vorliege und so aus ökonomischen Gründen auf die vereinfachte Form zurückgegriffen werden könne. Wie jedoch Höder (2016) zutreffend darstellt, ist eine weite Transkription nicht ausreichend, um Deutsch und Dänisch kontrastiv bzw. im Kontext des Fremdsprachenerwerbs zu betrachten. Besonders für die kontrastive Analyse bewertet Höder (2016: 9) die weite Transkription als ungünstig, da „phonetisch (fast) Gleiches so häufig unterschiedlich und

phonetisch Verschiedenes gleich notiert wird“. Anhand des prominenten weichen *d* als Beispiel wird dies deutlich. Das DDO transkribiert weiches *d* mit [ð], dem Symbol für einen stimmhaften dentalen Frikativ, obwohl es sich beim weichen *d* artikulatorisch um einen Approximanten handelt (vgl. Höder 2016: 9). Sollte die Lautschrift eine didaktische Stütze zur Lautbildung sein, wäre ein Lerner also schlecht beraten. Höder (2016) ergänzt beim Frikativ deshalb das Diakritikum der Zungensenkung und erhält die detaillierte Umschreibung [ð̞]. Auf diese Weise wird dargestellt, dass kein Verschlusspunkt vorhanden ist und der Laut approximantisch gebildet wird.

Auch anhand der Beschreibung der Diphthonge verdeutlicht sich, weshalb eine einheitliche, möglichst genaue Transkription von Vorteil ist. So notiert die dänische Transkriptionstradition Diphthonge anhand einer vokalischen silbischen und einer approximantischen unsilbischen Komponente (auch als *Halbvokal* bezeichnet, z. B. bei Grønnum 2009: 77). Höder (2016) hingegen argumentiert, dass Approximanten und Vokale in der unsilbischen Komponente artikulatorisch nicht voneinander zu unterscheiden sind, also dass zwischen unsilbischem vokalischem [ɥ] und konsonantischem [w] kein Unterschied wahrzunehmen ist. In der einzelsprachlichen Betrachtung macht es also zunächst keinen Unterschied, ob mit vokalischer oder approximantischer Schreibung gearbeitet wird. In einer kontrastiven Analyse sollte allerdings am besten für beide Sprachen ein einheitliches Transkriptionssystem gewählt werden. Vergleicht man dänisches [ɥ̞] (wie in [hɥ̞] *hav* ‚Meer‘) mit deutschem [ɥ̞] (wie in *hau* (*ihn*)!), ergibt sich lautlich kein Unterschied. Die dänische Schreibung mit einer konsonantischen Komponente könnte an dieser Stelle insofern für Verwirrungen sorgen, als ein Unterschied in der Lautung suggeriert würde, obwohl dieser faktisch nicht vorliegt. Höder (2016: 9) schlägt vor, aus ebensolchen Gründen die unsilbische Komponente als Vokal zu transkribieren. Auch didaktisch würden sich hierdurch Hilfestellungen ergeben, da deutsche Lerner Diphthonge aus ihrer L1-Schriftsprache mit zwei vokalischen Komponenten kennen. Diese Hilfestellung würde durch eine approximantische Notation verloren gehen. Eine lautschriftliche Transkription, die Lernern als Hilfestellung dient, sollte diesen Umstand miteinbeziehen.

Durch diese einzelnen Beispiele wird bereits deutlich, weshalb in der Lautbeschreibung für diese Arbeit eine enge Transkription genutzt wird. Besonders didaktische Implementationen erfordern eine möglichst exakte Darstellung der Laute. Vereinfachungen können keine Unterstützung bei der Artikulation geben und im schlimmsten Fall zu Missverständnissen führen. Im Spracherwerbskontext sollte daher mindestens auf eine richtige und möglichst auf eine exakte Transkription zurückgegriffen werden. Aus diesem Grund basieren die folgenden lautlichen Beschreibungen auf Höder (2016), dessen Transkription sich als ausreichend detailliert für eine exakte phonetische Gegenüberstellung des Deutschen und Dänischen erweist. Die in Kapitel 3.2.2.1 dargestellte Forderung von Bohn (2002: 198) nach einer sehr engen Transkription wird dadurch ebenfalls erfüllt. Eine weitere Überlegung, Höder (2016) als Vorlage zur Analyse heranzuziehen, ist, dass er in die kontrastive Gegenüberstellung der Lautsysteme miteinbezieht, dass Dänisch als schulische Fremdsprache ausschließlich in Schleswig-Holstein unterrichtet

wird. Der hier vorherrschende Regiolekt unterscheidet sich von der Standardnorm und wird daher von Höder in die kontrastive Gegenüberstellung mit aufgenommen. Diese Beschreibung ermöglicht es, nicht nur Hypothesen über die Wechselwirkungen mit der Standardsprache, sondern auch über Wechselwirkungen mit dem Regiolekt miteinzubeziehen. Der Einbezug von Dialekten der Ausgangssprache in die kontrastive und didaktische Bearbeitung von Ziel- und Ausgangssprachen wird bereits von Ternes (1976: 98) als wünschenswert beschrieben. Normalerweise bieten sich zu viele dialektale Hintergründe einer Ausgangssprache, als dass sich diese alle einzeln beschreiben ließen. Im Fall schulischen Dänischunterrichts gibt es allerdings eine regionale Begrenzung und somit die Beschränkung auf einen Regiolekt, dessen Einbezug aus diesem Grund möglich ist.

Neben Höder (2016) werden außerdem Aspekte aus Fredsteds (1989) kontrastiver Gegenüberstellung der Sprachsysteme Deutsch und Dänisch herangezogen. Deren Entstehung liegt zwar bereits zeitlich länger zurück und ist somit in Teilen veraltet bzw. bezieht sich überwiegend auf Fehler, die aus der lautlichen Umsetzung deutscher Orthografie resultieren, doch sie stellt eine der wenigen Untersuchungen dar, deren kontrastive Analyse direkt auf typische Fehler deutscher Dänischlerner abzielt und wird deshalb in Teilen berücksichtigt.

3.3.2 Segmentale Merkmale

Zunächst werden die segmentalen Merkmale betrachtet. Die kontrastive Analyse mit Ableitung resultierender Probleme umfasst die Konsonanten, Vokale und Polyphthonge. Die segmentalen Merkmale werden anhand ihrer Unterschiedlichkeit und Ähnlichkeit einzeln dargestellt. Außerdem wird in Hinblick auf die spätere Fehleranalyse (vgl. Kapitel 3.4.2) das sich daraus ergebende Fehlerpotenzial beschrieben. In der Zusammenfassung der einzelnen Merkmale wird außerdem ein Rückbezug auf Fredsteds (1989) kontrastive Gegenüberstellung genommen.

3.3.2.1 Konsonanten

Aus der engen kontrastiven Transkription von Höder (2016) geht hervor, dass es in beiden Lautinventaren spezifische Laute gibt, die jeweils im anderen Lautsystem nicht vorkommen. Die unterschiedlichen Laute lassen sich in einer von Höder (2016: 22) übernommenen Tabelle (Tab. 3.3) nachvollziehen.⁸

⁸ Bei der hier dargestellten Tabelle handelt es sich um eine vereinfachte Form der Originaltabelle in Höder (2016). Verzichtet wurden auf Laute, die in der vorliegenden kontrastiven Betrachtung nicht von Relevanz sind. Dazu gehören beispielsweise die dentalen Frikative und die nur im Standarddeutschen vorhandenen vibrantischen *r*-Varianten.

	bilabial		labiodental		dental		alveolar		postalveolar		alveolopalatal		palatal		velar		uvular		glottal	
	+s	-s	+s	-s	+s	-s	+s	-s	+s	-s	-s	+s	-s	+s	-s	-s	-s			
Plosiv	b	p p ^h			d	t t ^h t ^s								g	k k ^h					ʔ
Frikativ			v	f			z	s	ʒ	ʃ	ç	j	ç		x					h
Nasal	m m̄						n n̄							ŋ ŋ̄						
Vibranten					r̄														R̄	
lateraler Approximant							l l̄ l̥													
Approximant					ø ø̄ ø̥													ʁ		

Tab. 3.3: Konsonanten des Dänischen, Deutschen und Norddeutschen (kontrastiv) (+s = stimmhaft; -s = stimmlos; vereinfacht nach Höder 2016: 22)

Die Darstellung erklärt sich folgendermaßen: grau markierten Laute sind spezifisch dänisch, mit einer durchgezogenen Linie eingerahmte Laute spezifisch deutsch (sowohl Standard- als auch Norddeutsch) und diejenigen Laute, die mit einer gestrichelten Linie eingerahmt sind, sind für den Regiolekt Norddeutsch spezifisch. Es ergeben sich die folgenden wesentlichen Unterschiede:

- Spezifisch deutsch sind die Frikative [z ʒ ʃ ç x]. Dänisch kennt als alveolare und alveolopalatale Frikative lediglich [s] und [ç].
- Im Deutschen nicht vorkommende Laute sind das alveolopalatale [ç] und das weiche d [ø̥]. Im Norddeutschen existiert vollapproximantisches [ʀ̥], das dem weichen d artikulatorisch ähnelt.
- Der t-Laut wird im Deutschen aspiriert, im Dänischen hingegen affriziert ausgesprochen.

Diese wesentlichen Unterschiede werden nachfolgend detailliert erläutert.

Stimmloses s [s]

Der wesentliche Unterschied bei den Frikativen zwischen Deutsch und Dänisch besteht darin, dass das Deutsche mehr alveolare Frikative hat als das Dänische. Spezifisch deutsch sind die Frikative [z ʒ ʃ ç x]. Im Dänischen gibt es unter den alveolaren und alveolopalatalen Frikativen lediglich [s] und [ç]. Darüber hinaus ist im Dänischen ausschließlich der alveolare stimmlose Frikativ [s] vorhanden, während es im Deutschen auch die stimmhafte Variante [z] gibt. Aus diesem Umstand lassen sich unterschiedliche Schlussfolgerungen für die Fehleranalyse ziehen. So sollten die deutschsprachigen L2-Dänischler in der Lage sein, beide Laute zu bilden; es besteht kein artikulatorisches Problem. Trotzdem ist zu erwarten, dass Interferenzen auftreten, besonders in Vorlesekontexten, in denen nach deutschen Schriftausssprache-Regeln die stimmhafte Variante [z] erwartet würde. Dies ist meistens im Wortanlaut vor Vokalen der Fall (z. B. *Sahne, Sonne, suchen*). Nur in Ausnahmefällen, nämlich bei Entlehnungen aus dem Englischen (z. B. *Softie*) kann wortinitial auch stimmloses [s] stehen (vgl. *Duden* 2015: 120). Problematisch ist somit nicht die Unterscheidung der Laute, sondern der Kontext, in dem die Laute vorkommen. Folgt man der Theorie der CAH, dass große Unterschiede auch große Probleme bereiten, müssten Lerner wortinitial die meisten Fehler produzieren, da im Dänischen wortinitial immer stimmlos [s] realisiert wird, wohingegen im Deutschen immer stimmhaft [z] gesprochen wird. Folgt man dagegen dem Fleges (1987) SLM oder dem PAM von Best (1995), müsste man erwarten, dass – im Gegensatz zur wortmedialen Position – der Kontrast wortinitial so groß ist, dass keine Probleme auftreten. Weiterhin sollte bei der Analyse bedacht werden, dass bei Wörtern, die dem Deutschen lexikalisch ähnlich sind, bei denen allerdings wortmedial stimmhaft [z] gesprochen wird (z. B. *læse* ‚lesen‘), möglicherweise Interferenzen auftreten.

Alveolopalataler Frikativ [ç]

Der spezifisch dänische alveolopalatale Frikativ [ç] liegt artikulatorisch zwischen dem Deutschen [ç] und [ʃ]. Da er im Deutschen nicht auftritt, wird vermutet, dass er entweder perzeptorische oder artikulatorische Probleme bereiten wird. Denkbar ist, dass Fehler im Sinne der *Two Category Assimilation* (Best & Tyler 2007: 23; vgl. Kapitel 3.2.2.3) auftreten, d. h. die Lerner assimilieren den L2-Laut [ç] entweder mit [ç] oder [ʃ]. Es ist vorstellbar, dass sie entweder schriftnah einen Doppelkonsonanten bilden (z. B. *[sjäʊ] für *sjov* [çʊʊ^ʔ] ‚spassig‘) oder in Form einer artikulatorischen Annäherung [ʃ] realisieren (z. B. *[ʃoko'le:ð:] für *chokolade* [çʊʊ^ʔle:ð:] ‚Schokolade‘). Bei der schriftnahen Umsetzung von *sj* zu [sj] ist im eigentlichen Sinne nicht von einem Fehler zu sprechen. So kann diese Realisierung in einigen geographischen Lagen Dänemarks durchaus als soziale Variante auftreten (Grønnum 2009: 305).

Weiches *d* [ð]

Neben dem alveolopalatalen [ç] ist das weiche *d* [ð] spezifisch dänisch und kommt im Deutschen nicht vor. Es handelt sich um einen dentalen Approximanten [ð], bei dem die Luft ungehindert aus dem Mund austritt. Im Norddeutschen existiert vollapproximantisches [ɹ], das dem weichen *d* artikulatorisch ähnelt.

Vorstellbar sind sowohl artikulatorische Probleme als auch Interferenzen bei der lautlichen Realisierung der Orthographie. Weiches *d* wird graphematisch durch ⟨d⟩ oder ⟨dd⟩ realisiert, eine falsche bzw. deutsche lautliche Umsetzung der Orthographie führt zu Fehlern, also beispielsweise der Realisierung als [d]. Ein artikulatorisch nahe liegender Laut wäre im Deutschen der Laut [ɹ]. Hier wird im Gegensatz zum weichen *d* ein medianer Verschlusslaut gebildet, sodass die Luft lateral austritt, während im Dänischen ein Approximant vorliegt, also die Luft ungehindert austritt. Artikulatorisch können hier Probleme auftreten. Im Norddeutschen ist vollapproximantisches [ɹ] vorhanden. Dieses ähnelt auditiv und artikulatorisch dem weichen *d*. Höder (2016: 23) beschreibt den artikulatorischen und auditiven Unterschied zwischen [ɹ] und [ð] als „nicht ohne Weiteres zu ermitteln“. Dänisch *rådden* [ˈʁ̥ð̥ːn] ‚verrottet‘ klingt demnach nahezu genauso wie norddeutsch *rollen* [ˈʁɔɪ̯n]. Dieser Umstand ist insofern eventuell für die Fehleranalyse und sicherlich für die didaktische Implementation interessant, als vorstellbar ist, dass es norddeutschen Lernern bei der Artikulation des schwierigen Lautes helfen könnte.

r-Varianten

Im Standarddeutschen gibt es im Gegensatz zum Dänischen vibrantische *r*-Varianten. Für den Regiolekt Norddeutsch sind sie allerdings nicht typisch. Da von der Mehrheit der Dänischlerner davon ausgegangen werden kann, dass sie in Norddeutschland aufgewachsen und sprachlich sozialisiert sind, braucht diese standarddeutsche Besonderheit also nicht berücksichtigt werden.

Affriziertes *t* [tʰ]

Im Dänischen wird *t* mit einer leichten Affrizierung realisiert [tʰ], hingegen im Deutschen typischerweise mit einer Aspiration [tʰ]. Da der Laut ansonsten gleich ist, ist zu erwarten, dass die deutschen Lerner eine deutsche aspirierte Realisation vornehmen, der Laut also nicht affriziert wird, wie es allerdings im Dänischen üblich wäre. Bei geschlossenen Anschlussvokalen (z. B. *i*) ist es möglich, dass *t* von Lernern eher affriziert wird, da hier der Artikulationsraum bereits geschlossener ist (z. B. bei *ting* [ˈtʰɛŋˀ] ‚Sache‘ oder *Tine* [ˈtʰiːnə] ‚Tine (weiblicher Vorname)‘).

Zusammenfassung

Die sich aus der oben durchgeführten kontrastiven Analyse ergebenden Probleme lassen sich für den konsonantischen Bereich auf folgende Phänomene zusammenfassen:

- a. stimmloses [s], besonders bei Kognaten, die im Deutschen mit stimmhaften [z] realisiert würden;
- b. der alveolopalatale Frikativ [ç];
- c. das weiche *d* [ð];
- d. das affrizierte *t* [tʰ].

Diese Phänomene lassen sich teilweise auch in der Darstellung von Fredsted (1989) wiederfinden. So merkt Fredsted (1989: 100–101) an, dass weiches *d* von Lernern wie [l] ausgesprochen werde. Da Fredsted (1989: 100) in ihrer Fehlerdarstellung einen überwiegend orthographischen Ansatz verfolgt, beschreibt sie als hauptsächliche Probleme die lautliche Realisierung der Grapheme ⟨d⟩ und ⟨g⟩. Im Anlaut stehendes *d* werde überwiegend richtig realisiert, ersetze allerdings auch das weiche *d* und werde auch dort gesprochen, wo ⟨d⟩ stumm ist (Fredsted 1989: 100–101). Für die lautliche Umsetzung von *g* konstatiert Fredsted (1989: 101), dass statt [θ] wie z. B. in *alligevel* ([æ'li:əvɛl] ‚trotzdem‘) [ǰ] realisiert werde (*[æ'li:ǰəvɛl]) oder in finaler Position statt [θ] der Frikativ [ç] realisiert werde (z. B. in *hyggelig* ['hyǰli] ‚gemütlich‘: *[hyǰliç]).

3.3.2.2 Monophthonge

Die Verhältnisse bei den monophthongischen Vokalen sind komplexer als bei den Konsonanten (vgl. Höder 2016: 33). Die enge kontrastive Analyse von Höder (2016: 24) macht deutlich, dass für die kontrastive Analyse die gängigen vier phonetischen Termini zur Bezeichnung der Öffnungsgrade nicht ausreichen. Höder (2016) erweitert diese deshalb durch eine Skala von 1 = geschlossen bis 9 = offen. Die Stufen dazwischen (4 = halbgeschlossen, 6 = halboffen) haben jeweils noch weitere Zwischenstufen, sodass das dänische Vokalsystem sehr eng beschrieben werden kann. Höder (2016: 26) stellt die einzelnen Lautabstufungen auch in einer Übersicht mit Einzelwortbeispielen und den Entsprechungen der gängigen IPA-basierten Transkriptionssysteme dar. Das dänische und das deutsche Lautsystem haben die Laute [i e y ø ə v u] gemeinsam. In Tab. 3.4 werden die Unterschiede in der Übersicht dargestellt (entnommen aus Höder 2016: 34).

		vorn		zentral	hinten		
		-gerundet	+gerundet	-gerundet	-gerundet	+gerundet	
geschlossen	1	i	y			u	
	2	<u>i</u>	<u>y</u>			<u>u</u>	
	3						
halbgeschlossen	4	ɪ~ɛ̣~e	ʏ~ø	ə		ʊ~o~o	
	5	<u>ɛ̣</u>				<u>ö</u>	
halboffen	6	ɛ	œ~œ̣~œ̥~œ̦	ɐ		ɔ~ɔ̣	
	7	<u>æ</u>					<u>ä</u>
	8						
offen	9	<u>a</u>		<u>ä</u>	<u>a</u>		

Tab. 3.4: Monophthonge des Dänischen, Deutschen und Norddeutschen (kontrastiv) (vereinfacht nach Höder 2016: 34)

Ebenso wie für die Abbildung der Konsonanten gilt, dass die grau markierten Laute spezifisch dänisch sind, diejenigen durch eine durchgezogene Linie gekennzeichneten Laute spezifisch deutsch (sowohl Standard- als auch Norddeutsch) und mit einer gestrichelten Linie gekennzeichnete Laute für den Regiolekt Norddeutsch spezifisch sind.⁹ Laute, die nur im Standarddeutschen vorkommen, sind mit einer gepunkteten Linie eingerahmt. In der Darstellung können einige Vereinfachungen vorgenommen werden, da die entsprechenden Laute als akustisch so nah beieinander liegen, dass sie für die hier vorliegenden Interessen als gleichwertig angesehen werden können. So können Unterschiede zwischen dänischem [œ œ̣ œ̥] und deutschem [œ] vernachlässigt werden sowie zwischen deutschem [ɔ] und dänischem [ɔ̣]. Ebenfalls als identisch können deutsches [ɔ], norddeutsches [ʊ] sowie dänisches [o] angesehen werden. Auf Grund der feinen Unterschiede wird standarddeutsches [ɪ] und norddeutsches [i̯] mit dänischem [ɛ̣] zusammengefasst, ebenso wie standarddeutsches [ʏ] und norddeutsches [y̯] mit dänischem [ø]. Durch die Zusammenfassung und den Vergleich mit dem Norddeutschen reduziert sich die Zahl der Unterschiede. Die verbleibenden Ähnlichkeiten und Unterschiede stellen sich als die Folgenden dar:

- a. die vorderen ungerundeten Vokale [ɛ̣ æ];
- b. der zentrale a-Laut [ä];
- c. die hinteren gerundeten ([ö]) und ungerundeten ([ä]) Vokale.

⁹ Auch diese Darstellung wurde nicht in ihrer Gesamtheit von Höder (2016) übernommen. Auf die Darstellung einiger Vokale, die nur im Standarddeutschen vorkommen und in der weiteren Analyse keine Rolle spielen, wurde aus Gründen der Übersichtlichkeit verzichtet.

Diese wesentlichen Unterschiede werden nachfolgend detailliert erläutert.

Vordere ungerundete Vokale [ɛ æ]

Die Laute [ɛ] und [æ] gehören zur Gruppe der vorderen ungerundeten Vokale, in der im Dänischen mindestens sechs bedeutungsunterscheidende Laute vorhanden sind, während das Deutsche nur vier kennt. Die Laute [ɛ] und [æ] kommen im Deutschen nicht vor und können artikulatorische Probleme bereiten. Vorstellbar ist, dass die deutschsprachigen L2-Dänischler die L2-Laute mit dem artikulatorisch am nächsten gelegenen Laut aus der L1 ersetzten. In diesem Fall würde [ɛ] beispielsweise durch einen offeneren Laut [ɛ̃] ersetzt werden und das [æ] zu geschlossen (ebenfalls mit [ɛ̃]) realisiert werden. Deutschsprachige L2-Dänischler würden somit für die drei unterschiedlichen Laute des Dänischen nur einen deutschen bilden, dieses Phänomen wird von Flege, Munro & Fox (1994: 3133) als *äquivalente Klassifikation* bezeichnet. Von den Monophthongen [i ɛ̃ ɛ] wird nicht erwartet, dass sie den deutschsprachigen L2-Dänischlern Probleme bereiten, da [i] und [ɛ̃] auch im Deutschen standardmäßig vorkommen. Darüber hinaus ist [ɛ̃] dem norddeutschen [ɪ] qualitativ so ähnlich, dass hier eventuell auftretende Unterschiede vernachlässigt werden können.

Zentrales a [ä]

Dem Norddeutschen fehlt der zentrale Monophthong [ä], der im Standarddeutschen und im Dänischen vorhanden ist. Da der Laut im Standarddeutschen vorhanden ist, sollte er deutschen Lernern keine Probleme bereiten. Interessant ist, ob durch das Fehlen des Lautes im norddeutschen Regiolekt Unterschiede in der Realisierung zu verzeichnen sind. So ist es vorstellbar, dass die Lerner den Laut fehlerhaft bzw. als weiter hinten artikuliertes [ɑ] oder weiter vorne artikuliertes [a] bilden, wie sie es aus ihrer Sprachumgebung gewohnt sind. Darüber hinaus ist es außerdem möglich, dass Lerner eine falsch gelernte impressionistische Regel anwenden (z. B. „Sprich im Dänischen a wie ä aus!“). Im ersten Fall würde es sich um einen Fehler auf Grund von sprachlicher Interferenz, im letzteren Fall um Übergeneralisierung handeln.

Hinterer ungerundete ([ɤ̃]) und gerundete ([ɔ̃] [ɔ]) Vokale

Der hintere ungerundete Vokal [ɤ̃] ist weder im standard- noch im norddeutschen Lautinventar vertreten. Es ist davon auszugehen, dass hier sowohl Perzeptions- als auch Produktionsprobleme bei Lernern auftreten. Vorstellbar ist, dass sie den Laut, ähnlich dem L1-Laut [ɔ] produzieren.

Der hintere gerundete Vokal [ɔ̃] ist nur im dänischen Inventar vorhanden und wird somit vermutlich perzeptorische und artikulatorische Probleme bereiten. Es ist möglich, dass die Lerner den Laut einen ihnen aus dem L1-Lautinventar bekannten bzw. artikula-

torisch nahegelegenen Laut zuordnen und somit der Ziellaut letztendlich zu offen [ɔ] oder zu geschlossen [o] realisiert wird.

[ɔ] ist ebenfalls ein spezifisch dänischer Laut, wobei hier die Unterschiede zum Deutschen artikulatorisch nicht genauso groß sind, wie bei einigen anderen Lauten. So besteht bei diesem Laut eine große Ähnlichkeit zum deutschen [ɔ] (Höder 2016: 33–34). In einigen norddeutschen Varietäten besteht mit [ɒ] sogar eine annähernd entsprechende Qualität zum dänischen [ɔ] (Höder 2016: 33).

Zusammenfassung

Zusammenfassend können für die Vokale folgende Merkmale als problematisch für deutschsprachige L2-Dänischlerner prognostiziert werden:

- a. vorderer ungerundeter Vokal [e];
- b. vorderer ungerundeter Vokal [æ];
- c. zentraler *a*-Laut [ä];
- d. hinterer ungerundeter Vokal [ɤ];
- e. hinterer gerundeter Vokal [ɔ].

Außerdem wird angenommen, dass keine Probleme auftreten bei:

- f. hinterem gerundetem Vokal [ɔ].

Der Blick in die ältere Forschung zeigt, dass sowohl die Umsetzung der richtigen Qualität als auch Quantität Probleme bereitet (Fredsted 1989: 103). Die Quantität der Laute wird in Kapitel 3.3.3.1 und 3.3.3.2 bearbeitet. Fredsteds (1989) kontrastive Analyse der Vokalsysteme beruht auf einer weiten Transkription. Sie folgert aus den sich ergebenden Unterschieden, dass Lerner bei den Vokalen Probleme haben werden, die es im Deutschen nicht gibt. Die vereinfachte kontrastive Analyse führt allerdings dazu, dass einige Unterschiede unberücksichtigt bleiben und insbesondere Laute, die zwar ähnlich aber nicht gleich sind, nicht erwähnt werden. Besonders problematisch sind laut Fredsted (1989: 105) die Qualitäten der *a*-Laute.

3.3.2.3 Polyphthonge

Die kontrastive Darstellung von Höder (2016: 35–46) zeigt, dass im Dänischen viel mehr Polyphthonge existieren als im Deutschen. Außerdem kennt das Dänische im Gegensatz zum Deutschen auch noch Triphthonge. Höder (2016: 46) teilt für seinen Sprachvergleich alle drei Varietäten (Dänisch, Standarddeutsch, Norddeutsch) in drei Klassen ein (Polyphthonge auf [i̯], auf [u̯] und auf [ɛ̯]). Anders als beispielsweise bei Grønnum (2009) transkribiert Höder (2016) dabei beide Diphthongkomponenten vokalisch.

- (11) *liv* ‚Leben‘
[liw^ɥ] (Grønnum 2009: 260)
[liu̯^ɥ] (Höder 2016: 46)

Sowohl Dänisch als auch Deutsch und Norddeutsch haben vergleichsweise wenige Diphthonge auf [ɪ] und viele Diphthonge auf [ɶ]. Das Dänische hat im Vergleich zu den beiden anderen Varietäten die meisten Kombinationen auf [ɶ]. Tatsächlich lassen sich fast alle monophthongischen Qualitäten mit [ɶ] im Dänischen zu Diphthongen verbinden, dies ist im Standarddeutschen nicht möglich. Auch Verbindungen mit [ɶ] als zweiter Komponente können im Dänischen mit fast allen monophthongischen Qualitäten gebildet werden. Im Standarddeutschen finden sich hier Gemeinsamkeiten, wie die Beispiele in (12) zeigen.

- (12) *Moor* [o:ɶ]
nur [u:ɶ]
leer [e:ɶ]
Tür [y:ɶ]

Noch größere Ähnlichkeit zeigt sich zum Norddeutschen. Hier können, im Gegensatz zum Standarddeutschen, auch [ɶ]-Verbindungen mit kurzen Vokalen als erster Komponente eingegangen werden:

- (13) *zerren* [tseɶn]

Da die untersuchten Probanden alle mit dem Regiolekt vertraut bzw. aufgewachsen sind, sind für diesen Aspekt keine Probleme zu erwarten.

Im Dänischen existieren Kombinationen zwischen [ɶ]-Diphthongen und unsilbischem [ɶ] zu Triphthongen. Deutsch bzw. Norddeutsch kennen diese Triphthonge nicht. Für die Triphthonge sind auf Grund der orthographischen Umsetzung und der ungewohnten phonotaktischen Verbindungen von drei aufeinanderfolgenden Vokalen Probleme bei den Lernern zu erwarten.

- (14) *hvirvle* ['vɪɶlɐ] ‚wirbeln‘

Fredsted (1989: 106), die generell ein Problem in der Umsetzung der vom Lautbild stark abweichenden Orthographie sieht, beschreibt in dieser Hinsicht die Diphthonge für deutschsprachige L2-Dänischlerner als besonders problematisch, da sie als Vokal plus Konsonant geschrieben werden. Die Regeln zur lautlichen Umsetzung der Schrift erklärt Fredsted (1989: 106–107) allerdings als zu komplex zum Erlernen und merkt an, dass selbst Lerner, die das Prinzip verstanden hätten, noch immer Fehler machten.

Da die Polyphthonge im Dänischen sehr vielfältig sind, sollen die erwarteten sprachlichen Probleme in folgenden Gruppen untersucht werden:

- a. bei Polyphthongen auf Grund orthographischer Interferenz;
- b. bei Diphthongen auf [ɶ];
- c. bei Triphthongen.

3.3.3 Weitere Merkmale

Nachdem die segmentalen Eigenschaften beschrieben wurden, erfolgt in der Betrachtung weiterer Merkmale, die teilweise suprasegmentale Eigenschaften haben und teilweise aus sprachlichen Prozessen hervorgehen. Suprasegmentalia definieren sich darüber, dass sie in gewisser Weise von Segmenten unabhängig sind (vgl. Hall 2000: 29). Typischerweise werden zu den suprasegmentalen Eigenschaften u. a. Betonung, Länge und Intonation gezählt. In dieser Studie wird dazu die vokalische Quantität untersucht. Daneben werden phonologische Prozesse untersucht. Darunter werden natürlich auftretende, d. h. phonetisch motivierte Prozesse verstanden, die sich derart gestalten, dass die Realisierung eines Segments oder einer Abfolge von Segmenten davon abhängig ist, in welcher Umgebung sie sich befinden (Hall 2000: 89–99). Verändert sich diese Umgebung, passt sich die lautliche Realisierung unter Umständen dieser Veränderung an. Phonologische Prozesse können durch Regeln genau beschrieben werden. Zu ihnen werden Assimilationen und Dissimilationen, Tilgungen und Längungen gezählt (Hall 2000: 89–99).

3.3.3.1 Vokalische Quantität

Dieses Phänomen kann mit Hilfe der ausführlichen kontrastiven Analyse von Höder (2016) näher erläutert werden. Höder (2016: 54–58) differenziert zwischen kurzen, langen und überlangen Vokalen. Die ersten beiden Quantitätsstufen entsprechen einer *primären Quantität*, d. h. sie sind aus dem Protogermanischen ererbt oder aus zusätzlichen Lautwandelerscheinungen entstanden und somit nicht abhängig von synchronen Prozessen (Höder 2016: 54). Demgegenüber stehen ‚sekundäre Quantitäten‘, die als Nebenerscheinungen von Tilgungsprozessen auftreten. Wenn z. B. ein in der Graphie noch erhaltenes *r* getilgt wird, wirkt sich diese Tilgung auf die Quantität des vorangegangenen Vokals aus (Höder 2016: 54).

- (14) stdt. *Wort* [vɔʁt^h]
 norddt. *Wort* [vɔ:t]
 dän. *vort* ‚unser (N.SG)‘ [vɔ:ɖ]

Wie (14) verdeutlicht, kennt das Standarddeutsche keine sekundäre Quantität, das Norddeutsche und das Dänische hingegen schon. Darüber hinaus spielen noch junge Effekte eine Rolle, z. B. die der Schwa-Tilgung mit Längung bereits langer Vokale zu überlangen Vokalen. Die Schwa-Tilgung bezeichnet den Wegfall des in der Graphie als *e* und in formellem und konservativem Dänisch als [ə] noch erhaltenen Schwa-Lautes. Dieser Wegfall von [ə] in gesprochenem Dänisch wirkt sich auf die Umgebung in Form einer „kompensatorischen Längung bzw. Assimilation“ aus (Höder 2016: 55). (15) verdeutlicht diesen Fall von Schwa-Wegfall mit Längung des vorangegangenen Vokals im Dänischen.

- (15) *lige* [ˈli:] ‚gerade‘ (neben [ˈli:ə])

Die tertiäre Quantität gibt es im Standarddeutschen nicht, im Norddeutschen kann sie allerdings nach Schwa-Synkope nach stamfinalen Vokalen auftreten (Höder 2016: 57). Die Quantitätsprobleme, die bei der tertiären Quantität auftreten werden in Kapitel 3.3.3.2 zu den phonologischen Prozessen näher betrachtet. Im Folgenden steht allerdings die primäre Quantität im Fokus der Betrachtung. So besteht auch bei der primären Quantität ein entscheidender Unterschied zwischen Deutsch und Dänisch. Im Deutschen unterscheiden sich lange und kurze Vokale nicht nur in der Quantität, sondern auch in ihrer Qualität, d. h. im Deutschen sind die Quantitätsstufen auch an Qualität gebunden. So ist „Länge grundsätzlich an relative Geschlossenheit, Kürze grundsätzlich an relative Offenheit gekoppelt“ (Höder 2016: 56):

- (16) *biete* [ˈbi:tʰə]
bitte [ˈbitʰə]

Diese Kopplung besteht im Dänischen nicht. So existiert der geschlossene Laut auch in kurzer Form:

- (17) *trin* [tʰʁin] ‚Stufe‘
trine [tʰʁi:nə] ‚schreiten‘

Auch im Norddeutschen ist die primäre Quantität nicht „als distinktives phonologisches Merkmal erhalten“ (Höder 2016: 36). Es existieren geschlossene Vokale sowohl in langer als auch in kurzer Form, die von unterschiedlichen Kontexten abhängig sind (zur detaillierten Beschreibung vgl. Höder 2016: 56–57):

- (18) *brüte* (norddt. [ˈbʁy:tʰə] vs. stdt. [ˈbʁy:tʰə])
brühte (norddt. [ˈbʁy:tə] vs. stdt. [ˈbʁy:tʰə])

Das von Fredsted (1989: 105) dargestellte Problem, dass Lerner „korte, spændte vokaler“ zu offen aussprechen, beschreibt somit ein kontrastives Problem bei der primären Vokalquantität. So übertragen die deutschsprachigen L2-Dänischler das Phänomen der an die Qualität gekoppelten Quantität auf das Dänische. Es wird vermutet, dass die Lerner die kurzen dänischen Laute vermutlich zu offen produzieren (vgl. (19)).

- (19) *hygge* ‚Gemütlichkeit‘ * [ˈhyɡə] statt [ˈhyɡʊ]

3.3.3.2 Phonologische Prozesse

Phonologische Prozesse umfassen „Zusammenhänge in Form von Modifikationen zugrunde liegender Formen hin zu tatsächlich geäußerten Formen während der Sprachproduktion“ (Höder 2016: 61; vgl. auch Hall 2000: 89). Unter den Modifikationen sind überwiegend Tilgungs- und Längungsprozesse zu verstehen. Bei den zugrunde liegenden Formen kann es sich sowohl um festgeschriebene bzw. schriftliche Normen als auch ältere Varianten handeln. Höder (2016: 61) weist darauf hin, dass der Begriff *Prozess* aus

theoretischer Sicht nicht unproblematisch ist. In sprachlichen Äußerungen, die von der Schriftsprache oder einer älteren Variante abweichen, findet aus synchroner Perspektive nämlich kein Prozess im eigentlichen Sinne statt. So wird nicht in jeder Äußerung die ältere Variante modifiziert. Der Sprachwandel hat in der gesprochenen Sprache längst stattgefunden und die jeweiligen modifizierten Äußerungen sind dort bereits manifestiert. Wenn im Folgenden also von Prozessen gesprochen wird, handelt es sich nicht um Prozesse, die während der Produktion ablaufen, sondern um im Sprachgebrauch fest verankerte Varianten oder gar alternativlose Formen. Im Folgenden werden in dieser Gruppe Schwa-Assimilationsprozesse und Prozesse der Lenisierung von nachtonigen Plosiven untersucht. Beide stellen Wortmodifikationen im Sinne der Abweichung von älteren bzw. schriftlichen Varianten dar.

Schwa-Assimilation

Schwa-Assimilationsprozesse stellen die Anpassungen anderer Teile des Worts auf Grund der Tilgung des Schwa-Lauts dar. Die Schwa-Assimilation wurde bereits im Rahmen der tertiären Quantität von Vokalen angesprochen (vgl. Kapitel 3.3.3.1). Die Schwa-Tilgung gibt es sowohl im Dänischen als auch im Norddeutschen. Im Standarddeutschen kommt sie lediglich als Synkope vor und zwar in Verbindungen, in denen auf den Schwa-Laut ein Liquid oder Nasal folgt (vgl. Höder 2016: 64), wie beispielsweise in:

- (20) *laufen* ['lɔʊfn]
- Waffel* ['vɔʏfl]

Höder (2016: 62) beschreibt 7 Klassen unterschiedlicher Auswirkungen, die die Schwa-Apokope im Dänischen haben kann. Die folgende Tabelle ist übernommen aus Höder (2016: 62) und fasst die möglichen Prozesse zusammen.

	ohne Schwa-Tilgung	mit Schwa-Tilgung	Effekt
Klasse 1	V:ə	V::	langer → überlanger Vokal
Klasse 2	ʏə	V	unsilbischer → silbischer Vokal
Klasse 3	{ð ð̥}ə	{ð ð̥}::	kürzerer → längerer Konsonant
Klasse 4	{l N}ə _ω	{l N}:: _ω	kürzerer → längerer Konsonant
Klasse 5	V:K ₁ ⁿ ə _ω	V::K ₁ ⁿ :: _ω	langer → überlanger Vokal
Klasse 6	VK ₁ ⁿ ə _ω	VK ₁ ⁿ ə _ω	kein Effekt
Klasse 7	Kə{l N}	K{l N}	unsilbischer → silbischer Konsonant

Tab. 3.5: Kontexte und Effekte der Schwa-Tilgung (nach Höder 2016: 62)

Klasse 1 umfasst die Längung langer Vokale zu überlangen Vokalen durch Schwa-Tilgungsprozesse (vgl. *nye* ‚neu (DEF)‘ [ˈny:ə] und [ny:]). Kurze unsilbische Vokale werden nach Schwa-Tilgung silbisch (vgl. *leve* ‚leben‘ [ˈlɛ:ʊə] und [ˈlɛ:u]). Klasse 3 und 4 umfassen die Effekte auf den approximantischen Dental und alle weiteren Liquide und Nasale, die jeweils bei der Schwa-Apokope eine Längung erfahren (vgl. beispielsweise *komme* ‚kommen‘ [ˈkʰɔ̃mə] und [kʰɔ̃m:]). Klasse 6 zeigt diejenigen Konsonanten, bei denen kein Effekt auftritt (vgl. *ikke* ‚nicht‘ [ˈɛ̃ɡə] und [ɛ̃ɡ]). Die in Klasse 7 enthaltenen Phänomene stellen Schwa-Tilgungsprozesse dar, die in der Umgangssprache als „vollständig durchgeführt“ beschrieben werden können (Höder 2016: 62). Im Schriftbild tritt [ə] allerdings noch als *e* auf.

Lenisierung nachtoniger Plosive

Ein im Standarddeutschen nicht vorkommender phonologischer Prozess ist die für das Dänisch typische Lenisierung nachtoniger Plosive. So werden in Wörtern, die nach der akzentuierten Silbe keine Vollvokale mehr enthalten, statt stimmloser Plosive nur stimmhafte realisiert (vgl. Höder 2016: 64):

- (21) *pakke* [ˈpʰä̃ɡə] ‚Päckchen‘ (nicht *[ˈpʰäkʰə])
Mette [ˈmɛ̃ðə] ‚Mette (weiblicher Vorname)‘ (nicht *[ˈmɛtʰə])

Relevant ist, dass im norddeutschen Regiolekt ein ähnlicher Prozess fakultativ stattfindet (Höder 2016: 65). Ebenso wie im Dänischen können stimmlose Plosive nach der akzentuierten Silbe stimmhaft werden:

- (22) *alte* [ˈãl̥ðə] (stdt. [ˈältʰə])
klacken [ˈkʰl̥ä̃ŋ] (stdt. [ˈkʰl̥äkŋ])

Die deutschsprachigen Dänischlerner werden vermutlich sowohl bei Prozessen der Schwa-Assimilation als auch bei Prozessen der Lenisierung nachtoniger Plosive Probleme in der Realisierung aufweisen.

3.3.4 Zusammenfassung

Im Vorangegangenen wurden die drei phonetischen Inventare des Dänische, Standard- und Norddeutschen kontrastierend zueinander untersucht. Dabei wurden diejenigen Merkmale besonders fokussiert, die entweder sehr unterschiedlich oder ähnlich sind, da diese als problematisch für Sprachlerner vermutet werden. Die im folgenden Kapitel anschließende empirische Untersuchung gleicht die kontrastierenden Merkmale und erwartete Schwierigkeiten mit der Praxis ab. Im Folgenden werden diejenigen Merkmale zusammengefasst, die auch in der empirischen Analyse Beachtung finden werden.

Bei den Konsonanten werden auf Grund der in Kapitel 3.3.2.1 dargestellten Unterschiede der Konsonantenqualitäten beider Lautsysteme folgende Merkmale untersucht:

- a. affriziertes *t* [tʰ];¹⁰
- b. stimmloses *s* [s] (z. B. in *vase* [ˈvɛːsə] ‚Vase‘ (nicht *[ˈvɛːzə]));
- c. weiches *d* [ð];
- d. der alveolopalatale Frikativ [ç].

Die vermutlich problematischen Vokalqualitäten stellen sich als die folgenden dar:

- a. vordere ungerundete Vokale ([ɛ ɛː æ]);
- b. zentrales *a* ([ä]);
- c. der hintere ungerundete Vokal ([ɤ]);
- d. die hinteren gerundeten Vokale ([ɔ̞ ɔ]).

Die schwierigen Polyphthonge werden nicht in allen möglichen Kombinationen abgeprüft. Vielmehr werden Stellvertreter für die möglichen Probleme ermittelt. Es ergeben sich Diphthonge, die lautlich für deutschsprachige Dänischlerner unproblematisch sein sollten, allerdings auf Grund ihrer Orthographie Probleme bereiten können. Diese sind:

- a. bekannte Polyphthonge, die auf Grund der Orthographie problematisch sein können ([äy] in z. B. *savner* [ˈsäv̥nɐ] ‚vermissen (PRÄS)‘; [üi] in z. B. *jeg* [ˈjüi] ‚ich‘; [ɤi] in z. B. *løj* [ˈlɤi] ‚Zwiebel‘);
- b. Triphthonge;
- c. Diphthonge auf [y] (z. B. in *leve* [ˈlɛv̥ə] ‚leben‘).

Die primäre Vokalquantität, die sich insofern von der Deutschen unterscheidet, als sie nicht an einen Qualitätswechsel gekoppelt ist, wird ebenfalls anhand von Stellvertreterlauten getestet. Die geschlossenen Vokale [i] und [y] kommt im Deutschen nur in langer Quantität vor. Zu überprüfen ist, ob die deutschsprachigen L2-Dänischlerner diese auch in kurzer Quantität realisieren:

- a. Vokalquantität [i] und [y] (z. B. *mit* [mið] ‚mein (N.SG)‘ (nicht *[mitʰ])).

Schließlich werden die Prozesse untersucht. Die Durchführung der Schwa-Assimilationsprozesse ist fakultativ. Äußerungen in gesprochener Sprache, in denen die Durchführung fehlt, klingen allerdings auffällig bzw. können betont wirken. Die Realisierung von [ə] ohne Assimilationsprozesse ist daher kein Fehler im eigentlichen Sinne. Gesprochene Sprache, die allerdings ganz ohne Assimilationsprozesse realisiert wird, ist stark auffällig. Ebenso wie die anderen Phänomene werden die [ə]-Tilgungsprozesse bei den Vokalen nicht in allen, sondern in jeweils exemplarisch ausgewählten Qualitäten untersucht. Dies geschieht sowohl im Kontext betonter als auch unbetonter Silben:

- a. überlange Vokale in Folge von [ə]-Tilgungsprozessen in betonten Silben (z. B. *lige* [ˈliː:] ‚gerade‘ (neben [ˈliːə]));
- b. überlange Vokale in Folge von [ə]-Tilgungsprozessen in unbetonten Silben (z. B. *kortere* [ˈkʰɔːd̥ɐː:] ‚kürzer‘ (nicht *[ˈkʰɔːd̥ɐʁə])).

¹⁰ Das affrizierte *t* wird dabei in zwei Kontexten untersucht: vor geschlossenerem Vokal und vor offenerem Vokal. So ist es vorstellbar, dass die Affrizierung vor geschlossenerem Vokal leichter fällt, da hier der Artikulationsort bereits ähnlich ist.

Auswirkungen von Schwa-Assimilationsprozessen auf die Konsonanten werden bei Liquiden, Nasalen und beim langen weichen *d* untersucht:

- c. lange Konsonanten in Folge von [ə]-Tilgungsprozessen bei [l: m: n: ŋ: ð:] (z. B. in *komme* [ˈkʰɔ̃m:] ‚kommen‘ (neben [ˈkʰɔ̃mə])).

Als zuletzt ermittelte Schwierigkeit wird der Prozess der Lenisierung nachtoniger Plosive untersucht. Dieser wird sowohl intervokalisch als auch pränasal abgeprüft:

- d. Lenisierung nachtoniger Plosive intervokalisch (z. B. *ikke* [ˈẽʝə] ‚nicht‘ (nicht *[ˈẽkʰə]));
- e. Lenisierung nachtoniger Plosive pränasal (z. B. *katten* [ˈkʰæ̃dn] ‚Katze (DEF)‘ (nicht*[ˈkʰæ̃tn])).

3.4 Empirischer Abgleich

Im folgenden Kapitel werden die in Kapitel 3.3 ermittelten potenziellen Probleme für deutschsprachige L2-Dänischler mit der Praxis bzw. mit authentischer Schülersprache abgeglichen. Dazu werden Aufnahmen von Schülern in der Schule gemacht, aufbereitet und ausgewertet.¹¹ Die Vorgehensweise bei der Erhebung, Auswertung und Interpretation der Daten ist Inhalt der folgenden Kapitel. Zunächst wird in Kapitel 3.4.1 das Vorgehen bei der Datenerhebung und -aufbereitung beschrieben. Daran anschließend erfolgt in Kapitel 3.4.2 die Analyse der erhobenen sprachlichen Daten in Form einer Fehleranalyse. Kapitel 3.4.3 geht auf die nichtsprachlichen biographischen Metadaten der Sprecher ein und ermittelt mögliche Zusammenhänge zu den identifizierten Fehlern. Kapitel 3.4.4 fasst die Ergebnisse abschließend zusammen.

3.4.1 Vorgehen bei der Datenerhebung und -aufbereitung

Vor der Datenerhebung musste entschieden werden, ob die Sprachdaten in freier oder gebundener Form (z. B. in Form von vorgelesenen Texten), einzeln oder als Dialog erfolgen sollte. Die vorliegende Arbeit nutzt eine kontrollierte Form zur Erhebung der Sprachdaten. So werden von den Lernern Sätze vorgelesen. Die so gewonnen Sprachdaten können auf ökonomische Art und Weise ausgewertet werden, was eine quantitative Einsicht in Lernaltersprache und -fehler ermöglicht. Die Schüler lesen alle dieselben vorgegebenen Sätze vor, die in der Auswertung anhand eines Analysebogens effizient ohrenphonetisch ausgewertet werden (vgl. Kapitel 3.4.1.4). Neben der quantitativen Auswertung ermöglicht diese Form der Daten ebenso den direkten Vergleich zwischen den Lernern. Darüber hinaus ergibt sich durch diese Erhebungsmethode der Vorteil, dass die Sprecher sich nur auf die Aussprache und nicht auch noch auf Syntax und Lexik konzentrieren müssen. Dies ist insofern von Vorteil, als der Untersuchungsgegenstand nicht der gesamte Sprachstand der Schüler, sondern deren bestmögliche Aussprache ist. Die Schü-

¹¹ Die Erhebung erfolgte mit Genehmigung des Ministeriums für Schule und Berufsbildung.

ler erhalten vor der Erhebung die vorzulesenden Sätze, die sie gemeinsam mit der Lehrkraft üben sollen.

Trotz der genannten Vorteile ist die Vorlesemethode zur Erhebung von Sprachdaten nicht kritiklos einzusetzen. So muss berücksichtigt werden, dass eine schlechte Lesekompetenz Auswirkungen haben kann (vgl. Piske, MacKay & Flege 2001: 194). Es können somit bedingt durch orthographische Interferenz Fehler auftreten, die in freier Sprache gar nicht vorhanden wären. Dieser Umstand muss bei der Fehleranalyse berücksichtigt werden. Da es jedoch für die vorliegende Studie relevant ist, dass die Probanden alle dieselben Sätze produzieren, wurde sich trotz der genannten Einschränkung für die Vorlesemethode entschieden. Neben dem Aspekt der Ökonomisierung des Auswertungsverfahrens bieten die auf diese Weise erhobenen Schüleraufnahmen durch ihre Vergleichbarkeit die Möglichkeit, auch in Studie 2 (vgl. Kapitel 4) eingesetzt zu werden. Da die Sprachdaten in dieser Studie zur Untersuchung der Effekte einzelner Fehler eingesetzt werden, müssen die jeweiligen untersuchten Fehler isoliert auftreten, um reliable Aussagen treffen zu können. Der Ausschluss lexikalischer und syntaktischer Fehler sowie die Auswahlmöglichkeit zwischen vielen gleichen Sätzen erleichtert an dieser Stelle das Verfahren. Um die Effekte der orthographischen Interferenz möglichst gering zu halten, wurden die vorgegebenen Sätze den Schülern schon vorher zum Üben zugestellt. Um welches Zielwort es genau geht, ist ihnen allerdings nicht bekannt.

Die Überlegungen zur Konstruktion der Zielwörter werden im folgenden Kapitel 3.4.1.1 dargestellt. Daran anschließend erfolgt die Darstellung der Probandengruppe bzw. der teilnehmenden Schulen (vgl. Kapitel 3.4.1.2). Die Beschreibung des konkreten Ablaufs der Datenerhebung erfolgt in Kapitel 3.4.1.3. Kapitel 3.4.1.4 stellt die Datenaufbereitung als letzten Schritt vor der eigentlichen Fehleranalyse in Kapitel 3.4.2 dar.

3.4.1.1 Auswahl der Zielwörter

Die vorzulesenden Sätze sind so konstruiert, dass möglichst nur das zu untersuchende Phänomen Probleme bereitet.¹² Für die aus der kontrastiven Analyse erarbeiteten Phänomene in ihren Gruppen werden dabei zunächst ‚problematische‘ Wörter und daran anschließend Sätze entwickelt. Als problematisch sind folgende segmentale und supra-segmentale Eigenschaften zu verstehen:

- a. vokalische Laute, die in ihren Öffnungsgraden nur feine Unterschiede aufweisen;
- b. konsonantische Laute, die im Deutschen nicht vorkommen (z. B. weiches *d*);
- c. vokalische (monophthongische und polyphthongische) Laute, die in im Deutschen nicht vorkommen (z. B. [ɔ̃]);

¹² Dieser Umstand ist auch für die zweite Studie (vgl. Kapitel 4) notwendig, in der die Fehler möglichst isoliert stehen sollten, damit deren Wirkung auf die Kommunikation mit L1-Sprechern gemessen werden kann (zur Erläuterung vgl. 3.4.1).

- d. finale Konsonanten, die im Deutschen auf Grund der deutschen Auslautverhärtung nicht stimmhaft realisiert werden;
- e. Diphthonge, die im Deutschen nicht vorkommen.

Um dabei möglichst isolierte Fehler zu erzeugen, müssen die Wörter folgende Kriterien erfüllen: sie sollen den Fehler hervorrufen und dürfen darüber hinaus möglichst keine weitere problematische Eigenschaft haben. Die Stimuli sind also so aufgebaut, dass zum Beispiel beim Überprüfen von konsonantischen Problemen die Folgevokale keine Probleme bereiten, da sie im deutschen Lautinventar vorhanden sind. Die meisten der oben genannten weiteren Probleme werden als eigenständiges Phänomen untersucht und sollten deshalb in den anderen Stimuli bestenfalls nicht vorkommen. Bei Phänomenen, die in Vokalverbindungen vorkommen, werden nicht immer alle Kontexte überprüft, d. h. sie werden nicht mit jeder einzelnen Vokalqualität kombiniert, sondern nur mit einer exemplarischen.¹³ Die Fehlerursache ist für die Auswahl der Wörter nicht entscheidend. Ob also ein Phänomen in einem Wort auf Grund von Nicht-Vorhandensein im muttersprachlichem Lautinventar artikulatorische Probleme bereitete oder durch Interferenzen aus dem lexikalischen Inventar eine falsche Realisation bewirkt, ist bei der Erzeugung der vorzulesenden Sätze nicht entscheidend.¹⁴ Im Folgenden werden die Überlegungen zur Auswahl der Zielwörter der prognostizierten Probleme ausführlich dargestellt.

¹³ Auf Grund der Vielzahl der sich ansonsten ergebenden zu überprüfenden Fälle wurde diese ökonomisierende Entscheidung getroffen.

¹⁴ Im Hinblick auf Studie 2 (vgl. Kapitel 4) sollen die im Wort enthaltenen Phänomene außerdem in einen Satz eingebettet werden. Damit ist gewährleistet, dass den muttersprachlichen Bewertern aus der nachfolgenden Studie mehr als nur ein Wort präsentiert wird und eine natürliche Sprachumgebung entsteht (in der Realität treten Aussprachefehler auch nicht komplett isoliert auf).

3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner

Konsonanten

Problem	Zielwort	Prüfsatz
1 affriziertes t [tʰ]	to	[tʰoˀ] ,zwei [‘] <i>Jeg har kun to kroner.</i> <i>„Ich habe nur zwei Kronen.“</i>
	Tines	[tʰiːnəs] ,Tine (GEN) [‘] <i>Det er Tines bil.</i> <i>„Das ist Tines Auto.“</i>
3 stimmloses [s] wortinitial	synger	[ˈsøŋˀe] ,singen (PRÄS) [‘] <i>Han synger gerne.</i> <i>„Er singt gerne.“</i>
	sødt	[søð] ,süß [‘] <i>Det er sødt af dig.</i> <i>„Das ist nett von dir.“</i>
5 stimmloses [s] wortmedial	rose	[ˈʁoːsə] ,Rose [‘] <i>Tak for den dejlige rose.</i> <i>„Danke für die schöne Rose.“</i>
	næse	[ˈnɛːsə] ,Nase [‘] <i>Min bror har en stor næse.</i> <i>„Mein Bruder hat eine große Nase.“</i>
7 weiches d [ð] wortmedial	mødes	[møːðˀs] ,sich treffen (PRÄS) [‘] <i>Hvornår mødes vi?</i> <i>„Wann treffen wir uns?“</i>
	nødder	[ˈnøðˀe] ,Nüsse [‘] <i>Jeg elsker nødder.</i> <i>„Ich liebe Nüsse.“</i>
9 weiches d [ð] wortfinal	ud	[uðˀ] ,hinaus [‘] <i>Vi skal gå ud i haven.</i> <i>„Wir sollen in den Garten gehen.“</i>
10 alveolopalataler Frikativ [ç]	chok	[çʰʁ̥ç] ,Schock [‘] <i>Det var et chok at se ham.</i> <i>„Es war ein Schock, ihn zu sehen.“</i>
	sjette	[ˈçɛːdø] ,sechste [‘] <i>Det var sjette gang nu!</i> <i>„Das war jetzt das sechste Mal!“</i>
	sjældent	[çɛːlːnd̥] ,selten [‘] <i>Det sker bare sjældent!</i> <i>„Das passiert nur selten!“</i>

vordere ungerundete Vokale

Problem	Zielwort	Prüfsatz
13 [ɛ]	melde	[ˈmɛlə] ,melden [‘] <i>Vi skal melde os mere i timen.</i> <i>„Wir sollen uns in der Stunde öfter melden.“</i>
	ældste	[ˈɛlˀsðə] ,ältest (DEF) [‘] <i>Han var den ældste.</i> <i>„Er war der älteste.“</i>
15 [ɛː]	maler	[ˈmɛːlə] ,malen [‘] <i>Vi maler vores hus.</i> <i>„Wir streichen unser Haus.“</i>
	stave	[ˈsðɛːuə] ,buchstabieren [‘] <i>Kan du stave dit navn?</i> <i>„Kannst du deinen Namen buchstabieren.“</i>
17 [æ]	kat	[kʰæð] ,Katze [‘] <i>Lise har en kat.</i> <i>„Lise hat eine Katze.“</i>
	mand	[mænˀ] ,Mann [‘] <i>Lines mand er venlig.</i> <i>„Lines Mann ist freundlich.“</i>

zentrale Vokale

Problem		Zielwort			Prüfsatz
19	[ä]	<i>sang</i>	[sǎŋ [?]]	‚Lied‘	<i>Jeg elsker den sang.</i> ‚Ich liebe dieses Lied.‘
20		<i>træt</i>	[t ^s ʁǎð]	‚müde‘	<i>Jeg er træt.</i> ‚Ich bin müde.‘
21		<i>ret</i>	[ʁǎð]	‚recht‘	<i>Du har ret.</i> ‚Du hast Recht.‘
22		<i>lam</i>	[lǎm]	‚lahm‘	<i>Den gamle mands ben er lam.</i> ‚Das Bein des alten Mannes ist lahm.‘

hinterer ungerundeter Vokal

Problem		Zielwort			Prüfsatz
23	[ǎ]	<i>hånd</i>	[hǎn [?]]	‚Hand‘	<i>Giv mig din hånd!</i> ‚Gib mir deine Hand!‘

hintere gerundete Vokale

Problem		Zielwort			Prüfsatz
24	[ö]	<i>låne</i>	[lǔ:nə]	‚leihen‘	<i>Må jeg låne din bil?</i> ‚Darf ich dein Auto leihen?‘
25		<i>gå</i>	[gǔ: [?]]	‚gehen‘	<i>Han skal gå!</i> ‚Er soll gehen!‘
26	[ø]	<i>vores</i>	[vø::s]	‚unser‘	<i>Vores bil er gul.</i> ‚Unser Auto ist gelb.‘
27		<i>kort</i>	[k ^h ø:d]	‚kurz‘	<i>Det var alt for kort.</i> ‚Das war viel zu kurz!‘

Polyphthonge

Problem		Zielwort			Prüfsatz
28	bekannte Polyphthonge	<i>savner</i>	[sǎŋnə]	‚vermissen (PRÄS)‘	<i>Vi savner solen.</i> ‚Wir vermissen die Sonne.‘
29	[ǎy], [ǎi], [ǎi]	<i>regner</i>	[ʁǎŋnə]	‚regnen (PRÄS)‘	<i>Det regner.</i> ‚Es regnet.‘
30		<i>rejse</i>	[ʁǎjsə]	‚reisen‘	<i>Jeg vil rejse til et varmt land.</i> ‚Ich möchte in ein warmes Land reisen.‘
31		<i>løj</i>	[lǎi [?]]	‚lügen (PRÄT)‘	<i>Han løj over for mig.</i> ‚Er hat mich angelogen.‘
32	Triphthonge	<i>hvirvler</i>	[ʁivvulə]	‚wirbeln (PRÄS)‘	<i>Sneen hvirvler rundt.</i> ‚Der Schnee wirbelt herum.‘
33		<i>spurv</i>	[sɔvurv]	‚Spatz‘	<i>Der sidder en spurv på taget.</i> ‚Da sitzt ein Spatz auf dem Dach.‘
34	Diphthonge auf [y]	<i>liv</i>	[liu [?]]	‚Leben‘	<i>Jeg har et dejligt liv.</i> ‚Ich habe ein schönes Leben.‘

3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner

Polyphthonge

Problem	Zielwort			Prüfsatz
35		<i>flyver</i>	[^h flyʷvɛ]	‚fliegen (PRÄS)‘ <i>Vi flyver til Mexico.</i> ‚Wir fliegen nach Mexiko.‘
36		<i>støvler</i>	[^s støʷvlɛ]	‚Stiefel (PL)‘ <i>Har du nye støvler?</i> ‚Hast du neue Stiefel?‘

Vokalquantität

Problem	Zielwort			Prüfsatz
37	[i] [i:]	<i>milde</i>	[^h milə]	‚mild (DEF)‘ <i>Vi gik en tur i det milde vejr.</i> ‚Wir sind im milden Wetter spazieren gegangen.‘
38		<i>milen</i>	[^h mi:lɛn]	‚Wanderdüne (DEF)‘ <i>De besøgte milen ved Skagen.</i> ‚Sie haben die Wanderdüne in Skagen besucht.‘
39	[y]	<i>hyggelig</i>	[^h hyɡli]	‚gemütlich‘ <i>De har en hyggelig stue.</i> ‚Sie haben ein gemütliches Wohnzimmer.‘

überlange Vokale in Folge von [ə]-Tilgungsprozessen

Problem	Zielwort			Prüfsatz
40	betonte Silben	<i>lige</i>	[^h li:]	‚eben‘ <i>Du skal lige se her.</i> ‚Du musst dir das eben ansehen.‘
41		<i>uge</i>	[^h u:]	‚Woche‘ <i>Vi ses om en uge.</i> ‚Wir sehen uns in einer Woche.‘
42	unbetonte Silben	<i>kortere</i>	[^h kɔ:ɾɔɛ:]	‚kürzer‘ <i>Den her vej er kortere.</i> ‚Dieser Weg ist kürzer.‘
43		<i>lærere</i>	[^h lɛ:rɛ:]	‚Lehrer‘ <i>Vi har mange lærere i skolen.</i> ‚Wir haben viele Lehrer in der Schule.‘

lange Konsonanten in Folge von [ə]-Tilgungsprozessen

Problem	Zielwort			Prüfsatz
44	[l:], [m:], [n:], [ŋ:]	<i>falde</i>	[^h fæl:]	‚fallen‘ <i>Pas på! Det er nemt at falde her.</i> ‚Pass auf! Es ist leicht hier zu fallen.‘
45		<i>komme</i>	[^h kɔ:m:]	‚kommen‘ <i>De siger vi skal komme i aften.</i> ‚Sie sagen, wir sollen heute Abend kommen.‘
46		<i>finde</i>	[^h fɛ:n:]	‚finden‘ <i>Jeg skal finde min bil.</i> ‚Ich muss mein Auto finden.‘
47		<i>konge</i>	[^h kɔ:ŋ:]	‚König‘ <i>Han var en god konge.</i> ‚Er war ein guter König.‘
48	langes weiches d in betonten	<i>ballade</i>	[^h bæ:lɛ:ð:]	‚Ärger‘ <i>Der var meget ballade.</i> ‚Es gab viel Ärger.‘

lange Konsonanten in Folge von [ə]-Tilgungsprozessen				
Problem		Zielwort		Prüfsatz
49	Silben	<i>betyde</i>	[ˈbət̪ˀyð̥ˀ]	‚bedeuten‘ <i>Hvad skal det betyde?</i> ‚Was soll das bedeuten?‘
Lenisierung nachtoniger Plosive				
Problem		Zielwort		Prüfsatz
50	intervokalisches	<i>klipper</i>	[ˈkʰlɛ̃p̥p̥]	‚schneiden (PRÄS)‘ <i>Han klipper med en saks.</i> ‚Er schneidet mit einer Schere.‘
51		<i>Mette</i>	[ˈmɛ̃d̥ə]	‚Mette‘ <i>Mette er min vens kone.</i> ‚Mette ist die Frau meines Freundes.‘
52		<i>pakke</i>	[ˈpʰä̃g̥ə]	‚Paket‘ <i>Der er en pakke.</i> ‚Da ist ein Paket.‘
53	pränasal	<i>mappen</i>	[ˈmä̃p̥m]	‚Mappe (DEF)‘ <i>Må jeg se mappen?</i> ‚Darf ich die Mappe sehen?‘
54		<i>nitten</i>	[ˈnɛ̃d̥n]	‚neunzehn‘ <i>Min bror er nitten.</i> ‚Mein Bruder ist neunzehn.‘
55		<i>jakken</i>	[ˈjä̃g̥ŋ]	‚Jacke (DEF)‘ <i>Må jeg få jakken?</i> ‚Darf ich die Jacke haben?‘

Tab. 3.6: Übersicht über die Zielwörter und Prüfsätze

Die in Tab. 3.6 dargestellten Sätze und Zielwörter werden im Folgenden anhand ihrer Satznummern näher erläutert.

- a. Satz 1 und 2 (affriziertes *t* [tʰ]): Im Dänischen wird *t* mit einer leichten Affrizierung realisiert, im Deutschen typischerweise mit einer Aspiration. In der kontrastiven Analyse wird deshalb davon ausgegangen, dass die deutschen Lerner den Laut aspirieren, anstatt ihn zu affrizieren. Außerdem wird erwartet, dass vor geschlosseneren Vokalen eher eine Affrizierung realisiert wird, da der Öffnungsgrad sich ähnelt. Die Zielwörter sind zum Vergleich einmal mit einem offenerem Folgevokal [ɔ] (*to* ‚zwei‘) und einmal mit geschlossenerem Folgevokal [i] (*Tines* ‚Tine (GEN)‘) gewählt.
- b. Satz 3–6 (stimmloses *s* [s]): Im Dänischen ist ausschließlich der alveolare stimmlose Frikativ [s] vorhanden, während es im Deutschen auch die stimmhafte Variante gibt. In der kontrastiven Analyse wird vermutet, dass Interferenzen in solchen Kontexten auftreten, in denen im Deutschen die Umsetzung der orthographischen Regeln stimmhaftes [z] vorgibt. Das Phänomen wurde sowohl in wortinitialen als auch in wortmedialen Kontexten untersucht. Dazu wurden Zielwörter gewählt, die wortinitial (*synger* vs. *singen*; *sødt* vs. *süß*) und -medial (*rose* vs. *Rose*; *næse* vs. *Nase*) lexikalisch deutschen Wörtern ähneln.
- c. Satz 7–9 (weiches *d* [ð]): Spezifisch dänisch ist der dentale Approximant [ð], der im deutschen Lautinventar nicht vorhanden ist. In der kontrastiven Analyse wird vermutet, dass sowohl artikulatorische Probleme als auch Interferenzen bei der

lautlichen Realisierung der Orthographie auftreten. In den Zielwörtern wird diese Problem sowohl medial (*mødes* ‚sich treffen (PRÄS)‘, *nødder* ‚Nüsse‘) als auch final (*ud* ‚hinaus‘) überprüft.

- d. Satz 10–12 (alveolopalataler Frikativ [ç]): Der spezifisch dänische alveolopalatale Frikativ [ç] wird in zwei Kontexten mit unterschiedlicher Schreibung abgeprüft: mit initialem ⟨ch⟩ und ⟨sj⟩. Denkbar ist, dass die Schüler sich von der Orthographie leiten lassen und bei ⟨sj⟩ den Doppelkonsonanten bilden.
- e. Satz 13–18 (vordere ungerundete Vokale): Das Dänische kennt zahlreiche monophthongische Vokalqualitäten, die sich nur geringfügig unterscheiden, wobei „nur die wenigsten Qualitäten [...] in allen drei Varietäten (annähernd) gleich vorhanden [...]“ sind (Höder 2016: 33). Als Zielwörter wurden diejenigen Vokale aufbereitet, die spezifisch für eine Varietät und/oder artikulatorisch problematisch auseinanderzuhalten sind. In der kontrastiven Analyse wurde herausgearbeitet, dass wahrscheinlich solche Vokallaute problematisch sind, die spezifisch dänisch sind und im deutschen/norddeutschen Lautinventar nicht vorkommen sowie diejenigen, die annähernd gleich im deutschen und dänischen Lautinventar vorkommen. Die Vokallaute [ɛ æ] gehören zur Gruppe der vorderen ungerundeten Vokale, in der im Dänischen mindestens sechs bedeutungsunterscheidende Laute vorhanden sind, während das Deutsche nur vier kennt. Die Stimuli sind zu den zwei fehlenden Lauten konstruiert. Es kann davon ausgegangen werden, dass die deutschen L2-Dänischlerner anstelle von [ɛ] eine zu offene ([ɛ̃]) und anstatt [æ] eine zu geschlossene (ebenfalls [ɛ̃]) Variante produzieren, weil es sich dabei um den artikulatorisch und auditiv nächstliegenden deutschen Laut handelt. Deutschsprachige L2-Dänischlerner würden somit für im Dänischen drei distinktive Laute nur einen bilden.
- f. Satz 19–22 (zentrale Vokale): Bei zentralem *a* können mehrere Aspekte problematisch sein. Dem Norddeutschen fehlt der zentrale Monophthong [ä], der im Standarddeutschen und im Dänischen vorhanden ist. Für Lerner ohne Regiolekt ist er also artikulatorisch unproblematisch. Da die Lerner allerdings alle vom norddeutschen Regiolekt beeinflusst sind, ist es möglich, dass artikulatorische Probleme auftreten. So ist es vorstellbar, dass die Lerner eher einen hinteren Laut ([ɑ]) produzieren, wie sie es aus ihrer Sprachumgebung gewohnt sind. Ebenso gut ist es möglich, dass den Lernern eine zielsprachliche Realisierung gelingt, denn sie beherrschen neben dem Regiolekt auch Standarddeutsch. Trotzdem können Probleme bei der Realisierung entstehen. So führen falsche Lautbildungsregeln („Im Dänischen wird *a* immer wie *ä* ausgesprochen.“) häufig zu Fehlern. Geprüft wird der Laut anhand von Zielwörtern (*sang* ‚Lied‘, *lam* ‚Lamm‘), deren lautliche Umsetzung der Orthographie nach deutschen Regeln zu einer zielsprachlich korrekten Form führen würde. Dadurch kann ermittelt werden, ob die Lerner eine falsche Regel internalisiert haben. Die Zielwörter *træt* ‚müde‘ und *ret* ‚recht‘ sind in dieser Hinsicht problematisch, da ihre Orthographie einen anderen Laut suggeriert.

- g. Satz 23 (hintere ungerundete Vokale): Der hintere ungerundete Vokal [ɤ] ist weder im standard- noch im norddeutschen Monophthonginventar vertreten. Es ist davon auszugehen, dass hier sowohl Perzeptions- als auch Produktionsprobleme bei den Schülern auftreten. Vorstellbar ist, dass die Schüler den Laut zu geschlossen und gerundet produzieren, ähnlich des in ihrem Lautinventar vertretenen [ɔ].
- h. Satz 24–27 (hintere gerundete Vokale): [ɔ̃] ist nur im dänischen Inventar vorhanden und wird somit vermutlich perzeptorische und artikulatorische Probleme bereiten. In den Hypothesen wird davon ausgegangen, dass ein aus dem L₁-Lautinventar artikulatorisch naheliegender Laut eingesetzt (*Single Category Assimilation*) und somit der Ziellaut zu offen [ɔ] oder zu geschlossen [o] realisiert wird. Abgeprüft wird der Laut mit den Zielwörtern *låne* ‚leihen‘ und *gå* ‚gehen‘. Für den Laut [ɔ̃] werden zwei Typen von Zielwörtern konstruiert. Bei der lautlichen Umsetzung der Orthographie bei *vores* ‚unser‘ würde der Laut nach deutschen Regeln an einer offenen Silbe stehen, während er bei *kort* ‚kurz‘ in geschlossener Silbe steht. Im Deutschen kommt [ɔ̃] nur in geschlossener Silbe vor, in offener Silbe wird [o] realisiert. Bei *kort* wird daher davon ausgegangen, dass die Lerner hier nahezu richtig [ɔ̃] realisieren, während sie bei *vores* vermutlich eher fehlerhaft [o] bilden. Dies würde zumindest der deutschen Phonotaktik entsprechen.
- i. Satz 28–36 (Polyphthonge): Um die Polyphthongbildung zu überprüfen, werden nicht alle Qualitäten überprüft, sondern einzelne Laute exemplarisch ausgewählt. Dabei werden für die erste Komponente nicht solche Monophthonge gewählt, die an sich schon problematisch sein könnten. Es werden also nur diejenigen Laute in Betracht gezogen, die auch im deutschen Lautinventar vorkommen und somit unproblematisch in der Realisierung sein sollten. Damit soll erreicht werden, dass wenn Probleme auftreten, diese auf die ungewohnte Verbindung mit der zweiten Komponente zurückzuführen sind. Um die problematischen Diphthonge auf [ɥ] zu testen, werden die Zielwörter *liv* ‚Leben‘, *flyver* ‚fliegen (PRÄS)‘ und *støvler* ‚Stiefel (PL)‘ gewählt. Außerdem wird in der Kategorie der bekannten Polyphthonge noch *savner* ‚vermissen (PRÄS)‘ geprüft. Alle vier Zielwörter weichen in der lautlichen Umsetzung der Orthographie von der gewohnten deutschen Aussprache ab. Anhand von *savner* kann ersichtlich werden, ob es die lautliche Umsetzung der Orthographie ist, die den Lernern Probleme bereitet, oder ob ein artikulatorisches Problem vorliegt. Der Diphthong [äy] sollte den Lernern artikulatorisch keine Schwierigkeiten bereiten. Liegen beim Zielwort *savner* dennoch Fehler vor, kann hier von orthographischer Interferenz ausgegangen werden.
- j. In *liv*, *flyver* und *støvler* kommt neben der ungewohnten orthographischen Umsetzung die im Deutschen nicht vorhandene Kombinierbarkeit von [i], [y], [ø] und [o] mit [ɥ] als Problem hinzu. Die Zielwörter *savner* ‚vermissen (PRÄS)‘, *regner* ‚regnen (PRÄS)‘ und *rejse* ‚reisen‘ überprüfen Diphthonge, die artikulatorisch unproblematisch sein sollten. Die Diphthonge [äy] und [äi] kommen auch im deutschen Lautinventar vor, werden lediglich orthographisch anders umgesetzt.

Das Zielwort *løj* ‚(lügen (PRÄT))‘ testet ebenfalls einen aus dem Deutschen annähernd bekannten Diphthong (vgl. [ɔ̃]) wie in *Leute*; die Unterschiede zu Dänisch [l̥j] sind marginal und zu vernachlässigen). Problematisch ist, dass die erste Komponente auf Grund des Schriftbildes Probleme bereiten könnte. Die für das Dänische typischen und für deutsche Lerner problematischen Triphthonge wurde anhand der Zielwörter *spurv* ‚Spatz‘ und *hvirvler* ‚wirbeln (PRäs)‘ überprüft.

- k. Satz 37–39 (Vokalquantität): Um die Quantität bei den Vokalen zu überprüfen, wird ein Set von Zielwörtern mit unterschiedlichen Vokalqualitäten gewählt, da nicht die einzelnen Qualitäten, sondern lediglich die richtige Umsetzung der Länge im Fokus steht. Ausgewählt wurde der Vokal *i*. Die Zielwörter *milde* ‚mild (DEF)‘ und *milen* ‚Wanderdüne (DEF)‘ werden im Dänischen mit [i] und [i:] realisiert. Erschwerend für das Wort *milde* ist die lexikalische Interferenz zum kognaten Wort *milde* im Deutschen. Das Schriftbild dürfte deutsche Lerner Kürze vermuten lassen und somit einen Qualitätswechsel hervorrufen. Realisieren die Lerner den Vokal trotz der orthographischen bzw. lexikalischen Interferenz kurz und geschlossen als [i], kann davon ausgegangen werden, dass sie das Zielmerkmal beherrschen. Neben diesem Set wurde das frequente Wort *hyggelig* ‚gemütlich‘ abgeprüft.
- l. Satz 40–43 (überlange Vokale in Folge von [ə]-Tilgungsprozessen): Das Dänische kennt die tertiäre Quantität aus der Folge von [ə]-Tilgung. Auch im Norddeutschen sind diese Längungsprozesse zu finden. Die Zielwörter für überlange Vokale sind *lige* ‚eben‘, *uge* ‚Woche‘, *kortere* ‚kürzer‘ und *lærere* ‚Lehrer (PL)‘. Die Zielwörter weisen die Tilgung und Längung in zwei unterschiedlichen Kategorien auf. So tritt die [ə]-Tilgung in *lige* und *uge* in einer betonten Silbe auf, während sie in *kortere* und *lærere* eine unbetonte Silbe betrifft.
- m. Satz 44–49 (lange Konsonanten in Folge von [ə]-Tilgungsprozessen): Auch die konsonantische Längung ist im Standarddeutschen nicht vorhanden, tritt allerdings im Regiolekt auf. Zielwörter sind *falde* ‚fallen‘, *komme* ‚kommen‘, *finde* ‚finden‘ und *konge* ‚König‘. Die Längung von weichem *d* wird anhand der Zielwörter *ballade* ‚Ärger‘ und *betyde* ‚bedeuten‘ getestet.
- n. Satz 50–55 (Lenisierung nachtoniger Plosive): Der letzte untersuchte Prozess ist die Lenisierung der stimmlosen Plosive. Diese werden im Dänischen nach akzentuierten Silben ohne Vollvokal stimmhaft. So gibt es im Dänischen keine wortinternen stimmlosen Plosive. Im Standarddeutschen gibt es diese Lenisierung nicht, wohingegen sie im Regiolekt relativ ausgeprägt anzutreffen ist. Die Zielwörter wurden so gewählt, dass ihre lautliche Umsetzung der Orthographie durch standarddeutsche Regeln nicht in lenisierten Konsonanten resultieren würde (z. B. *pakke* ‚Paket‘), bei einer Realisierung nach norddeutschen Regeln allerdings durchaus eine stimmhafte Variante auftreten könnte ([p^hügə]). Außerdem wurden die Zielwörter sowohl intervokalisch als auch pränasal abgeprüft. Die

Zielwörter sind *klipper* ‚schneiden (PRÄS)‘, *Mette* ‚Mette (weiblicher Vorname)‘, *pakke* ‚Paket‘, *mappen*, ‚Mappe (DEF)‘, *nitten* ‚neunzehn‘ und *jakken* ‚Jacke (DEF)‘.

3.4.1.2 Schulen und Probanden

Um eine repräsentative Stichprobe zu erhalten, werden Sprachdaten aus ganz Schleswig-Holstein gesammelt. Insgesamt werden Schüler an vier Standorten aufgenommen, wobei drei Schulen im Landesteil Schleswig liegen (Husum, Niebüll, Flensburg) und eine Schule in Holstein liegt (Barmstedt). Diese Verteilung repräsentiert die Verteilung der Schulen auf die Landesteile Schleswig und Holstein. Insgesamt wurden 91 Schüler aus 8 Klassen von insgesamt 8 unterschiedlichen Lehrern aufgenommen. Die Geschlechterverteilung nach Jungen und Mädchen ist zufällig gemischt. Insgesamt nahmen 29 Jungen und 62 Mädchen an der Erhebung teil. Die Schüler sind in Norddeutschland aufgewachsen und somit mit dem Regiolekt Norddeutsch rezeptiv und produktiv vertraut. Die Schüler lernen Dänisch als dritte Fremdsprache im fortgeschrittenen Stadium im dritten Lernjahr. Die Daten bilden daher den Aussprachestand ab, der zum Ende der schulischen Laufbahn maximal zu erreichen ist. Dadurch wird es möglich abzugleichen, ob das im Curriculum formulierte Ziel (funktionale Dänischsprachigkeit) erreicht wurde. Alle Daten wurden anonymisiert erhoben, indem anstelle von Namen ein Code zur Probandenidentifizierung genutzt wurde, der nur von dem Probanden persönlich identifiziert werden kann.

3.4.1.3 Datenerhebung

Vor der Erhebung bekommen die Lehrer der Klassen einen Brief für die Eltern zur Einverständniserklärung sowie die vorzulesenden Sätze zugeschickt. Außerdem erhalten die Schüler den Selbstauskunftsbogen zu den sprecherbiographischen Daten (vgl. Anhang A.1). Die Datenerhebung erfolgt klassenweise, d. h. pro Erhebungstag wird eine Klasse aufgenommen. Die Aufnahme wird an den Schulen direkt in einem isolierten Raum anhand eines Audiorecorders Olympus LS-14 von einem Versuchsleiter durchgeführt. Die Probanden erhalten die Sätze auf einzelnen Papierbögen im Format A5. Pro Seite gibt es einen vorzulesenden Satz. Die Sätze werden nacheinander vorgelesen. Die Reihenfolge der Sätze ist festgelegt, unterscheidet sich allerdings von der Reihenfolge in Tab. 3.6. Auf diese Weise werden die Sätze randomisiert, und es wird verhindert, dass ein Merkmal mehrmals hintereinander abgefragt wird. Die Aufgabenstellung lautet, dass der Proband den Text vorlesen soll und bei Fehlern oder Stocken nochmal ansetzen bzw. den Satz wiederholen darf. Jeder Schüler liest insgesamt 55 Sätze vor (vgl. Tab. 3.6). Am Ende erhält jeder Schüler ein Geschenk in Form eines Schreibblocks, Kugelschreibers und Schokolade.

3.4.1.4 Datenaufbereitung und Fehleridentifikation

Sowohl sprecherbiographische Daten als auch die Audiodateien werden gesichert und anschließend ausgewertet. Zur Fehleridentifikation liegt ein Analysebogen vor (vgl. Abb. 3.4).¹⁵

dans-Projekt

Auswertungsbogen für Sprecher (deutsche Dänischler)

Code.: _____



Sprecher: _____

Datum der Aufnahme: _____

Auswerter: _____

Datum der Auswertung: _____

Zutreffende Realisierungen werden angekreuzt oder farblich markiert.

Satz		Kodierung						
1	Jeg har kun to kroner.	K_ts_ts	K_ts_th			K_ts_mx	K_ts_?	
24	Jeg er træt .	V_a_a	V_a_ä			V_a_mx	V_a_?	
2	Det er Tines bil.	K_ts_ts	K_ts_th			K_ts_mx	K_ts_?	
28	Må jeg låne din bil?	V_gaa_gaa	V_gaa_o	V_gaa_c	V_gaa_aa	V_gaa_mx	V_gaa_?	
5	Tak for den dejlige rose .	K_sm_s	K_si_z			K_sm_mx	K_sm_?	
31	Det var alt for kort .	V_c_c	V_c_o			V_c_mx	V_c_?	
61	Hvad skal det betyde ?	Pro_QuKbS_wd:	Pro_QuKbS_e			Pro_QuKuS_mx	Pro_QuKuS_?	
8	Jeg elsker nødder .	K_mwd_wd	K_mwd_l	K_mwd_d		K_mwd_mx	K_mwd_?	
9	Vi skal gå ud i haven.	K_fwd_wd	K_fwd_l	K_fwd_d		K_fwd_mx	K_fwd_?	
3	Han synger gerne.	K_si_s	K_si_z			K_si_mx	K_si_?	
12	Det var sjætte gang nul!	K_sj_1K [c]	K_sj_sj	K_sj_sk	K_sj_sch	K_sj_mx	K_sj_?	
38	Der sidder en spurv på taget.	PolyTr_uau_uau	PolyTr_uau_uav	PolyTr_uau_ua		PolyTr_uau_mx	PolyTr_uau_?	
17	Vi maler vores hus.	V_3_3	V_3_e	V_3_ä	V_3_a	V_3_mx	V_3_?	
58	Jeg skal finde min bil.	Pro_QuKuS_n:	Pro_QuKuS_e			Pro_QuKuS_mx	Pro_QuKuS_?	
40	Vi flyver til Mexico.	Poly_yu_yu	Poly_yu_yv	Poly_yu_y		Poly_yu_mx	Poly_yu_?	

Abb. 3.4: Analysebogen zur Fehleridentifikation

In der ersten Spalte des Analysebogens befindet sich die Satznummer in der Reihenfolge, in der die Sätze von den Schülern vorgelesen wurden. Die Aufnahmen werden ohrenphonetisch von zwei phonetisch ausreichend geschulten Prüfern ausgewertet. Die Prüfer hören die Aufnahmen und kreuzen an, ob die Realisierung zielsprachlich (erste Spalte) oder abweichend (folgende Spalten) ist. Anhand der Kodierung konnte bereits bei der Auswertung festgehalten werden, welche Form der Abweichung die Lerner produzierten. Dieser Vorgang erleichterte die quantitative Auswertung. Auf die Transkription der von den Schülern realisierten Sätze konnte auf Grund der Kodierung verzichtet werden. Diese wurden nur durchgeführt, wenn sie zur näheren Erklärung benötigt wurden.

¹⁵ Der vollständige Analysebogen über alle abgeprüften Sätze kann im Anhang A.2 nachvollzogen werden.

3.4.2 Fehleranalyse deutscher L2-Dänischlerner

Die Auswertung der Erhebung findet in zwei Teilen statt. In diesem Kapitel wird die Fehleranalyse der sich aus der Identifikation in Kapitel 3.4.1.4 ergebenden Fehler durchgeführt. Dabei wird nach dem in Kapitel 3.2 dargestellten Verfahren vorgegangen: Zunächst werden die Fehler klassifiziert und daran anschließend Hypothesen zu möglichen Ursachen aufgestellt. Im Kapitel 3.4.3 erfolgt die Auswertung der nichtsprachlichen bzw. sprecherbiographischen Metadaten.

3.4.2.1 Konsonanten

Bei den Konsonanten wurden das stimmlose [s] im wortinitialen und wortmedialen Kontext, der alveolopalatale Frikativ [ç], das weiche *d* [ð] sowie die Affrizierung von *t* [tʰ] untersucht. Die Analyse dieser Merkmale wird im Folgenden dargestellt.

Stimmloses s [s]

In der kontrastiven Analyse wird angenommen, dass [s] weniger Probleme bei der Artikulation, als vielmehr in der richtigen Anwendung im lautlichen Kontext bereitet. So ist anzunehmen, dass die Lerner in manchen Fällen auch auf die stimmhafte Variante [z] zurückgreifen, da diese im Deutschen ebenfalls vertreten ist. Es ist zu erwarten, dass besonders wortinitial ein Unterschied zwischen Deutsch und Dänisch zu verzeichnen ist, da im Standarddeutschen wortinitiales [s] nicht bzw. nur bei Entlehnungen aus dem Englischen vorkommt (vgl. Kapitel 3.3.2.1; Duden 2015: 120).

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
stimmloses [s] wortinitial	normkonforme Realisierung	100 %		
stimmloses [s] wortmedial	normkonforme Realisierung	77 %		
	Lautersatz [z]	23 %	Orthographie als Realisierungsgrundlage L1- und L2-Wörter ähneln sich	orthographische Interferenz lexikalische Interferenz

Tab. 3.7: Ergebnisse der Fehleranalyse zum stimmlosen [s]

Die Klassifikation der Fehler ergibt, dass stimmhaftes [z] in den Aufnahmen nie wortinitial auftritt. Auch wortmedial sind 77 % der Aufnahmen richtig; in 23 % der Aufnahmen realisieren die Lerner stimmhaftes [z] für die Zielwörter *rose* und *næse*.

In den Hypothesen zu möglich auftretenden Fehlern wurde angenommen, dass Lerner wortinitial fehlerhaft [z] statt [s] bilden. Dieser Fehler würde darauf beruhen, dass die Lerner die Wörter aus ihrer L₁ kennen, in der sie allerdings mit stimmhaften [z] ausgesprochen werden. Beim Übertragen der L₁-Aussprache auf die Aussprache des L₂-Wortes, entsteht der Fehler, weshalb von lexikalischer bzw. orthographischer Interferenz gesprochen wird. Durch die Ergebnisse der Erhebung kann diese Vermutung allerdings nicht bestätigt werden. Obwohl die Zielwörter so gewählt sind, dass sie dem Deutschen stark ähneln (*synger* ‚singen (PRÄS)‘; *sødt* ‚süß (N.SG)‘) enthalten keine der Aufnahmen lexikalische bzw. orthographische Interferenzfehler.

Alveolopalataler Frikativ [ç]

Für den alveolopalatalen Frikativ [ç] wird in der kontrastiven Analyse in Kapitel 3.3.2.1 angenommen, dass er artikulatorische und perzeptorische Probleme bereiten wird und Lerner anstelle von [ç] entweder deutsches [ç] oder [ʃ] bilden.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
Alveopalataler Frikativ [ç] Schreibung mit <i>ch</i>	normkonforme Realisierung	2 %		
	Lautersatz [k] und [x]	46 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	Lautersatz [ʃ]	48 %	L1- und L2-Wörter ähneln sich	Lexikalische Interferenz
			L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz Lautersatz auf Grund phonetischer Approximation
	Lautersatz [tʃ]	2 %	L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Lautersatz auf Grund phonetischer Approximation
Alveopalataler Frikativ [ç] Schreibung mit <i>sj</i>	normkonforme Realisierung	2 %		
	Lautersatz [sʃ]	98 %	L1- und L2-Wörter ähneln sich	Lexikalische Interferenz
			L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz
			L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Lautersatz auf Grund phonetischer Approximation

Tab. 3.8: Ergebnisse der Fehleranalyse zum alveopalatalen Frikativ [ç]

Die Auswertung der Daten zeigt, dass für die Zielwörter mit unterschiedlichen Schreibweisen ähnliche Fehlerhäufigkeiten, aber unterschiedliche Fehlerspektren auftreten bzw. unterschiedliche Fehlertypen vorliegen.

Insgesamt bilden nur 2 % der Lerner den Laut richtig (sowohl im Kontext ⟨ch⟩ als auch für ⟨sj⟩). Für [ç] mit der orthographischen Realisierung *ch* sprechen 48 % Lerner [ʃ] aus. 46 % realisieren [k] bzw. [x]. Dies entspricht einer lautlichen Umsetzung der Schrift nach deutschen Lautregeln. So wird *ch* vor hinteren halbgeschlossenen Vokalen wie [k] ausgesprochen (vgl. *Duden* 2015: 89). Die fehlerhafte Realisierung von [x] wäre auch im Deutschen falsch. So kann *ch* zwar auch als [x] realisiert werden, allerdings nur nach

hinteren halbgeschlossenen Vokalen. Bei den Zielwörtern mit *sj* bilden die Lerner in 98 % der Fälle einen doppelten Konsonanten.

Die Ursachen für die Probleme können vielfältig sein. Zum einen können sie durch orthographische Interferenz bedingt bzw. die lautliche Umsetzung der Orthographie den Lernern in diesem Fall nicht bewusst sein. Darüber hinaus ist es möglich, dass die Lerner nicht zwischen [ʃ] und [ç] unterscheiden bzw. die Laute assimilieren. Dadurch nehmen sie den lautlichen Unterschied nicht wahr und unterscheiden auch bei der Lautproduktion nicht zwischen den Lauten. Diese Fehler werden als *phonetische Äquivalenzfehler* bezeichnet. Als weitere Ursache kann in Betracht gezogen werden, dass die Laute zwar als unterschiedlich wahrgenommen werden, allerdings artikulatorisch Probleme bereiten. Diese Fehler werden als phonetische Approximation bezeichnet. Die Lerner nähern sich in der Lautbildung bereits dem richtigen L2-Laut an.

Die Ergebnisse entsprechen den Erwartungen, die aus der kontrastiven Analyse in Kapitel 3.3.2.1 hervorgehen.

Weiches *d* [ð]

Das weiche *d* kommt im deutschen Lautinventar nicht vor, weshalb in der kontrastiven Analyse davon ausgegangen wird, dass der Laut artikulatorische Probleme bereiten wird.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
weiches <i>d</i> [ð] wortmedial	normkonforme Realisierung	33 %		
	Lautersatz [l]	47 %	L2-Laut in L1 nicht vorhanden (art. ungewohnt/schwierig)	Interferenz: Ersatz des Ziellauts durch L1-Laut
				Lautersatz auf Grund phonetischer Approximation
	Lautersatz [d]	14 %	L2-Laut in L1 nicht vorhanden (art. ungewohnt/schwierig)	Interferenz: Ersatz des Ziellauts durch L1-Laut
				Lautersatz auf Grund phonetischer Approximation
				Orthographie als Realisierungsgrundlage

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
	artikulatorisch weit abweichender Lautersatz	6 %		
weiches <i>d</i> [ð] wortfinal	normkonforme Realisierung	47 %		
	Lautersatz [l]	47 %	L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Interferenz: Ersatz des Ziellauts durch L1-Laut Lautersatz auf Grund phonetischer Approximation
	Lautersatz [t]	0 %	phonologische Regeln unbekannt	Lautersatz auf Grund phonologischer Interferenz
	artikulatorisch weit abweichender Lautersatz	6 %		
	Elision [∅]	1 %	L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Vermeidung des Phänomens auf Grund artikulatorischer Probleme Übergeneralisierung: Übertragung von Regeln zur Tilgung von Wortenden aus einem anderen Zusammenhang

Tab. 3.9: Ergebnisse der Fehleranalyse zum weichen *d* [ð]

Die Lerner bilden weiches *d* wortmedial häufiger fehlerhaft als wortfinal. 67 % der Aufnahmen zum medialen weichen *d* sind fehlerhaft, während wortfinal nur 53 % der Aufnahmen Fehler aufweisen. Normkonform wird der Laut wortmedial in 33 % der Aufnahmen und wortfinal in 47 % der Fälle gebildet. Sowohl in wortmedialer als auch in wortfinaler Position ersetzen 47 % der Lerner weiches *d* fehlerhaft durch [l]. In wortmedialer Position wird der Laut auch durch [d] ersetzt, wortfinal tritt diese Form der Substitution

nicht auf. Dafür findet sich wortfinal allerdings auch die Elision des Lautes. Vorstellbar wäre, dass die Schüler hier auf Grund der im Deutschen phonotaktischen Regel der Auslautverhärtung [t] wortfinal realisieren. Dieser Fall tritt in den Sprachaufnahmen allerdings nicht auf.

Der Laut [ð] kommt im Deutschen nicht vor, weshalb davon auszugehen ist, dass die Lautbildung für die deutschsprachigen L2-Dänischlerner problematisch sein könnte. Um den Lautersatz zu erklären wird daher von Interferenzprozessen ausgegangen. Besonders wenn Lerner zum ersten Mal mit dem Laut konfrontiert werden, ist es möglich, dass sie diesen perzeptiv nicht einordnen können und somit auf die auditiv und artikulatorisch nächste Kategorie ihres Lautsystems zurückgreifen. Für die vorliegende Probandengruppe wird allerdings angenommen, dass den Lernern der Laut perzeptorisch bereits bekannt und bewusst gemacht worden ist, da sie sich im dritten Lernjahr befinden. Daher kann davon ausgegangen werden, dass es sich bei dieser Gruppe eher um ein artikulatorisches Problem handelt und als Fehlertyp also als phonetische Approximation genannt werden muss. Den Lernern gelingt es noch nicht, den Laut korrekt zu artikulieren, sie nähern sich jedoch bereits der richtigen Realisation an. Auch bei der Ersetzung von [ð] durch [d] ist es möglich, dass artikulatorische Probleme bestehen und es sich somit um den Fehlertyp der phonetischen Approximation handelt. Es ist allerdings ebenso möglich, dass die Lerner eine deutsche Umsetzung der Orthographie vornehmen. In diesem Fall wäre der Fehlertyp als orthographische Interferenz zu definieren.

In wortfinaler Position resultieren diese Probleme auch darin, dass der Laut komplett weggelassen wird, d. h. eine Elision stattfindet. Als Fehlertyp kann hier die Vermeidung vorgeschlagen werden. Wenn die Lerner [ð] in wortfinaler Position nicht aus Gründen der Vermeidung absichtlich weglassen, sondern weil sie der Ansicht sind, eine richtige Regel zu befolgen (z. B. „Im Dänischen lässt man meistens die Wortenden weg.“) oder einen aus einem anderen Zusammenhang bekannten Prozess anwenden, der an dieser Stelle nicht normkonform ist, muss von Übergeneralisierung als Fehlertyp gesprochen werden.

In beiden Kontexten (medial und final) treten insgesamt 6 % artikulatorisch weit vom Ziellaut entfernte Substitutionen auf. Bei diesen ist es problematisch, eine Ursache festzumachen. So ist es möglich, dass es sich um nicht-systemische, sondern situative Fehler handelt oder um artikulatorische Probleme, die zur Überforderung führen, dass nicht einmal phonetische Approximation stattfindet. Diese Erklärungsansätze bleiben allerdings spekulativ.

Affriziertes t [tʰ]

Für t wird in der kontrastiven Analyse in Kapitel 3.3.2.1 angenommen, dass die Lerner anstelle der dänischen Affrizierung [tʰ] eine deutsche Aspiration [tʰ] vornehmen. Allerdings wird dort auch die Möglichkeit angesprochen, dass die Lerner bei geschlossenerem Folgevokal t affrizieren.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
affriziertes <i>t</i> [tʰ]	normkonforme Realisierung	0 %		
	Lautersatz [tʰ]	100 %	phonologische Regeln unbekannt	Lautersatz auf Grund phonologischer Interferenz
			Orthographie als Realisierungsgrundlage	Orthographische Interferenz

Tab. 3.10: Ergebnisse der Fehleranalyse zum affrizierten *t* [tʰ]

Die Fehlerklassifizierung bestätigt jedoch, dass die Lerner in allen Sprachaufnahmen weder vor offenem Vokal (Zielwort *to* ‚zwei‘) noch vor geschlossenem Vokal (Zielwort *Tines* ‚Tine (GEN)‘) eine Affrizierung vornehmen. Die deutsche lautliche Umsetzung von aspiriertem *t* kann sowohl phonotaktische Ursachen haben als auch durch die orthographische Umsetzung begründet sein.

Die Sprachaufnahmen bestätigen somit die Vermutungen der kontrastiven Analyse. Weder vor offenem noch geschlossenem Vokal realisieren die Lerner *t* affriziert.

3.4.2.2 Monophthonge

Bei den Vokalen wurden die vorderen ungerundeten Vokale, der zentrale *a*-Laut, der hintere ungerundete Vokal [ɤ] sowie die hinteren gerundeten Vokale untersucht. Die Analyse dieser Merkmale wird im Folgenden dargestellt.

Vordere ungerundete Vokale [ɛ æ]

Für den vorderen ungerundeten Vokal [ɛ] wird in der kontrastiven Analyse davon ausgegangen, dass die Lerner eine zu offene Variante realisieren ([ɛ]). Für den zweiten getetzten vorderen ungerundeten Vokal ([æ]) der im deutschen Lautinventar nicht vorkommt, wurde ebenfalls angenommen, dass dieser als [ɛ] bzw. zu geschlossen realisiert werde. Der Laut [ɛ:] wird in der kontrastiven Analyse nicht aufgegriffen, da sich für diesen kein Unterschied zum deutschen Lautinventar ergibt (vgl. Kapitel 3.3.2.2). In der Erhebung wird trotzdem untersucht, ob die Lerner den Laut in orthographisch schwierigen Wörtern (*maler* [ˈmɛ:lɐ] ‚malen (PRÄS)‘; *stave* [ˈsdɛ:ʊə] ‚buchstabieren‘) umsetzen.

3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
vorderer ungerundeter Vokal [ɛ]	normkonforme Realisierung	9 %		
	Lautersatz [ɛ]	91 %	L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz
				Lautersatz auf Grund phonetischer Approximation
			L1- und L2-Wörter ähneln sich	oexikalische Interferenz
vorderer ungerundeter Vokal [ɛ:]	normkonforme Realisierung	49 %		
	Lautersatz [ä]	31 %	L1- und L2-Wörter ähneln sich	oexikalische Interferenz
			Orthographie als Realisierungsgrundlage	orthographische Interferenz
	artikulatorisch weit abweichender Lautersatz	20 %		
vorderer ungerundeter Vokal [æ]	normkonforme Realisierung	25 %		
	Lautersatz [ɛ]	21 %	L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz
	Lautersatz [ä]	53 %		Lautersatz auf Grund phonetischer Approximation
				L1- und L2-Wörter ähneln sich

Tab. 3.11: Ergebnisse der Fehleranalyse zu den vorderen ungerundeten Vokalen

Lediglich in 9 % der Aufnahmen realisieren die Lerner den Laut [ɛ] richtig. 91 % der Lerner realisieren fehlerhaft [ɛ]. Die Annahme aus der kontrastiven Analyse in Kapitel 3.3.2.2 kann somit bestätigt werden. Als Ursache können perzeptorische Probleme als wahrscheinlich angenommen werden, da sich die Laute auditiv sehr ähnlich sind. Es ist

anzunehmen, dass die Lerner die Laute [ɛ] und [ɛ] als äquivalent einordnen. Beim Fehlertyp würde es sich somit um eine Äquivalenzklassifizierung handeln. Gelingt bereits die Wahrnehmung, die vollständig richtige Artikulation allerdings noch nicht, handelt es sich um den Fehlertyp der phonetischen Approximation. Dies ist möglich, da der Laut im L₁-Lautinventar der Lerner nicht vorkommt und die feinen Artikulationsunterschiede schwierig zu differenzieren sind. Als weiterer Fehlertyp darf die Substitution durch lexikalische Interferenz nicht ausgeschlossen werden. So ist es möglich, dass die Lerner die Wörter deutsch artikulieren, da es sich um Kognaten handelt (*melde* ‚melden‘; *ældste* ‚ältest (DEF)‘; vgl. Tab. 3.11).

Für den zweiten getesteten vorderen ungerundete Vokal ([æ]), der im deutschen Lautinventar nicht vorkommt, wurde ebenfalls angenommen, dass dieser als [ɛ] bzw. zu geschlossen realisiert wird. In diesem Fall zeigen die Aufnahmen allerdings ein breiteres Fehlerspektrum. Insgesamt sind 25 % der Aufnahmen richtig artikuliert. 53 % der Lerner substituieren stattdessen zentrales [ü], 21 % wie angenommen [ɛ] und ein Schüler substituiert [ɔ]. Interessant zu betrachten, ist die Fehlerfrequenz der unterschiedlichen Zielwörter. Beim Zielwort *kat* ([^hkʰæd] ‚Katz(e)‘) realisieren die Lerner den Vokal häufiger als zentrales [ü], während sie das Zielwort *mand* ([mæn^ʔ] ‚Mann‘) falsch entweder mit [ü] oder mit [ɛ] aussprechen. Dafür können verschiedene Ursachen verantwortlich sein. So ist davon auszugehen, dass diejenigen Lerner, die *mand* mit [ü] aussprechen, vermutlich einer deutschen Umsetzung der Orthographie folgen, oder eine lexikalische Interferenz vorliegt. Wohingegen Lerner, die [ɛ] realisieren das Zielwort zwar ebenfalls falsch realisieren, aber bereits wissen, dass es anders als im Deutschen nicht mit [ü] ausgesprochen wird. Dass sie den vorderen gerundeten Laut nicht ganz treffen und zu geschlossen realisieren, kann damit zusammenhängen, dass sie entweder den Unterschied noch nicht wahrnehmen (Äquivalenzklassifizierung) oder noch nicht bilden können (phonetische Approximation).

Der dritte vordere ungerundete Vokal [e:] wird mit den Zielwörtern *maler* ([^lmɛ:lɘ] ‚malen (PRÄS)‘) und *stave* ([^ssdɛ:ɰə] ‚buchstabieren‘) getestet. Die lautliche Realisierung sollte für die deutschsprachigen L2-Dänischlerner unproblematisch sein. Tatsächlich realisieren im Vergleich zu den anderen beiden Qualitäten relativ viele Lerner den Laut normkonform (49 %). 31 % der Lerner substituieren den Ziellaut durch [ü]. Bei diesem Fehler kann von einer orthographischen Interferenz bei *stave* und einer lexikalischen Interferenz bei *maler* ausgegangen werden. Diejenigen Aufnahmen, die artikulatorisch und auditiv weit vom Ziellaut abweichen (20 %) können als Ausdruck einer Unsicherheit bezüglich der dänischen Aussprache gedeutet werden. So haben die Lerner vermutlich in ihrer Spracherwerbslaufbahn des Dänischen bereits die Erfahrung gemacht, dass die *a*-Laute in vielen Fällen ‚anders‘ oder ‚schwierig‘ auszusprechen sind. Die sich hier ergebene starke Abweichung bei einem eigentlich ‚einfachen‘ Laut (auch in der L₁ vorhanden) verdeutlicht dies.

Die Vorannahmen aus der kontrastiven Analyse in Kapitel 3.3.2.2 konnten für die vorderen ungerundeten Vokale überwiegend bestätigt werden. Anstelle des vorderen un-

gerundeten Vokals [ɛ] realisieren die Lerner eine zu offene Variante ([ɛ]). Für den vorderen ungerundete Vokal ([æ]) konnte die Annahme (Lautersatz durch den zu geschlossenen Laut [ɛ:]) zwar ebenfalls in 21 % der Aufnahmen nachgewiesen werden, eher substituierten die Lerner allerdings mit dem Laut [ü].

Anhand des Lauts [ɛ:] der zwar sowohl im dänischen als auch im deutschen Lautsystem vorkommt (vgl. Kapitel 3.3.2.2) zeigt sich in orthographisch schwierigen Wörtern eine Unsicherheit bezüglich der lautlichen Umsetzung und das Zurückgreifen auf orthographische Regeln der L1. In der Erhebung wird untersucht, ob die Lerner den Laut in orthographisch schwierigen Wörtern (*maler* ['mæ:lɐ] ‚malen (PRÄS)‘; *stave* ['sdɛ:ʊə] ‚buchstabieren‘) umsetzen.

Zentrales a [ü]

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
zentralen a-Laut [ü] (<i>sang, lam</i>)	normkonforme Realisierung	90 %		
	Lautersatz [ɛ]	10 %		
zentraler a-Laut [ü] (<i>træt, ret</i>)	normkonforme Realisierung	24 %		
	Lautersatz [ɛ]	69 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	Lautersatz [e]	5 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	artikulatorisch weit abweichender Lautersatz	2 %		

Tab. 3.12: Ergebnisse der Fehleranalyse zum zentralen [ü]

Untersucht wurde, ob der zentrale Monophthong [ü] Probleme bereitet. In der kontrastiven Analyse (Kapitel 3.3.2.2) wird dargestellt, dass der Laut eigentlich keine Probleme bereiten sollte, die Lerner aber trotzdem [ü] eventuell anders realisieren, da der Laut im norddeutschen Regiolekt nicht bzw. eine vordere Variante [a] vorkommt.

Insgesamt kann bestätigt werden, dass der Laut unproblematisch ist. Die Zielwörter, deren lautliche Umsetzung der Orthographie mit der deutschen Umsetzung vereinbar ist (vgl. Tab. 3.6, Satz 19 *sang* und Satz 22 *lam*), werden überwiegend richtig gebildet (90 %). Die fehlerhaften Realisierungen stellen sich als Lautersatz durch [ɛ] dar. Diese Fälle kön-

nen als Verunsicherung bezüglich der dänischen Aussprache oder als fehlerhaft erlernte impressionistische Lautbildungsregel („sprich *a* wie *ä*“) bzw. als Fehlertyp Übergeneralisierung gedeutet werden.

Das Fehlerspektrum für die Zielwörter *ret* und *træt* verteilt sich anders. Vermutlich ist hierfür die orthographische Interferenz ursächlich. In den fehlerhaften Äußerungen (69 %) werden *ret* und *træt* mit [ɛ] realisiert. In 5 % der Aufnahmen realisieren die Lerner [e].

Hinterer ungerundeter Vokal [ɤ]

In der kontrastiven Analyse in Kapitel 3.3.2.2 wird davon ausgegangen, dass der hintere ungerundete Vokal [ɤ] sowohl Probleme in der Perzeption als auch in der Produktion bereitet.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
hinterer ungerundeter Vokal [ɤ]	normkonforme Realisierung	70 %		
	Lautersatz [ɛ]	21 %		
	Lautersatz [ü]	6 %	L1-Wort und L2-Wort ähneln sich stark	lexikalische Interferenz
	artikulatorisch weit abweichender Lautersatz	3 %		

Tab. 3.13: Ergebnisse der Fehleranalyse zum hinteren ungerundeten Vokal [ɤ]

Insgesamt wird der Laut relativ häufig normkonform gebildet (70 %). Zum Fehlerspektrum des Zielworts *hånd* gehören [ü] (6 %), [ɛ] (21 %) und artikulatorisch stark abweichende Laute (3 %), die vermutlich auf Versprecher oder Unsicherheit auf Grund von Unkenntnis oder mangelnder Beherrschung bzw. falsch vermittelten Regeln zurückzuführen sind. Die Lautersetzung mit [ü] könnte daneben lexikalische Interferenzen zur Ursache haben (vgl. *hånd* und dt. *Hand*). So erklärt Schaden (2006: 133), dass Lerner bei unbekanntem Graphemen dazu neigen, sich bei der Interpretation von „orthographischen Basisgraphemen und ihren phonetischen Entsprechungen ihrer L1 leiten [zu] lassen“. Da im deutschen Grapheminventar kein ⟨â⟩ vorhanden ist, kann erwartet werden, dass sich die Lerner stattdessen für das Basisgraphem ⟨a⟩ entscheiden. Besonders wenn die Lerner unsicher in der lautlichen Umsetzung des ungewohnten Graphems sind, ist diese Möglichkeit vorstellbar. [ɛ] könnte als Resultat von Interferenzen aus anderen Fremdsprachen entstehen, z. B. (amerikanisch) engl. *hand* [hænd].

Hintere gerundete Vokale [ö ɔ]

Für den hinteren gerundeten Vokal [ö] werden perzeptorische und artikulatorische Probleme angenommen. Für den hinteren gerundeten Vokal [ɔ] wird in den Hypothesen angenommen, dass er den Lernern keine Probleme bereitet.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
hinterer gerundeter Vokal [ö]	normkonforme Realisierung	6 %		
	Lautersatz [ɔ]	80 %	L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz
				Lautersatz auf Grund phonetischer Approximation
	Lautersatz [o]	5 %	L1- und L2-Laute ähneln sich	Lautersatz auf Grund von phonetischer Äquivalenz
				Lautersatz auf Grund phonetischer Approximation
artikulatorisch weit abweichender Lautersatz	9 %			
hinterer gerundeter Vokal [ɔ], <i>vores</i>	normkonforme Realisierung	38 %		
	Lautersatz [o] und [o:]	62 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
			phonologische Regeln der L2 sind unbekannt	phonologische Interferenz
hinterer gerundeter Vokal [ɔ], <i>kort</i>	normkonforme Realisierung	98 %		
	Lautersatz [oɹ] und [o:]	2 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz

Tab. 3.14: Ergebnisse der Fehleranalyse zu den hinteren gerundeten Vokalen

Die Auswertung der Aufnahmen bestätigt eine große Fehlerhäufigkeit für den Vokal [ɔ̃]. Lediglich 6 % der Aufnahmen sind normkonform realisiert. Die Lerner ersetzen den Ziellaut häufig durch [ɔ] (80 %). In wenigen Fällen realisieren die Lerner auch [o] (5%) und artikulatorisch weit abweichende Laute (9 %).

Wie häufig im Dänischen und der angewandten Vorlesemethode geschuldet, muss davon ausgegangen werden, dass ebenso wie beim Zielwort *hånd* der hinteren ungerundeten Vokale die fehlerhaften Realisierungen mit einer orthographisch bedingten Unsicherheit zusammenhängen. Als Zielwörter dienen *lâne* und *gå*. Besonders das im Deutschen fremde Graphem ⟨â⟩ kann hierbei zu Unsicherheiten führen. Dass es sich bei diesem Fehler allerdings wie oben um orthographische Interferenz handelt, ist eher unwahrscheinlich. Wie bereits dargestellt, würde der fremde Buchstabe mit dem aus der L1 bekanntestem Basisgraphem ersetzt werden, in diesem Fall müsste demnach [ü] realisiert werden. Wahrscheinlicher sind die folgenden Möglichkeiten:

- a. Die Lerner wenden eine impressionistische Lautbildungsregel an („Sprich dänisches *å* wie [ɔ] aus!“). In diesem Fall handelt es sich um einen Regelfehler.
- b. Der Fehler beruht auf artikulatorischen oder perzeptorischen Ursachen. Der Laut wird entweder von den Lernern perzeptorisch nicht von einem [ɔ] oder [o] unterschieden (Äquivalenzklassifizierung) oder die Bildung der Laute bereitet Probleme (phonetische Approximation).

Für den hinteren gerundeten Vokal [ɔ̃] wird in der kontrastiven Analyse angenommen, dass er den Lernern keine Probleme bereitet. Die Aufnahmen zeigen diesbezüglich ein unterschiedliches Bild für die zwei Zielwörter. Für das Zielwort *vores* ([vɔ̃::s]) ergeben sich entgegen den Erwartungen 62 % fehlerhafte Realisierungen, für das Zielwort *kort* ([k^hɔ̃:d̥]) lediglich 2 %. Das Fehlerspektrum ist für beide Zielwörter gleich. Sowohl für *vores* als auch für *kort* realisieren die Lerner fehlerhaft [oɤ̃] und [o:̃].

Warum der Fehler häufiger bei *vores*, also in offener Silbe häufiger auftritt, lässt sich mit der Interferenz des Schriftbildes bzw. der deutschen Phonotaktik erklären. So suggeriert die Schreibung Zweisilbigkeit, und [ɔ] tritt im Deutschen nur in geschlossener Silbe auf (vgl. *Duden* 2015: 112). Ursächlich für diesen Fehler sind somit vermutlich eher nicht artikulatorische oder perzeptorische Probleme. Dass beim Zielwort *kort* trotzdem Fehler gemacht werden, kann damit erklärt werden, dass kurzes [o] im Deutschen auch in geschlossenen Silben vorkommen kann. Die lexikalische Interferenz kann somit diesen falschen Laut begünstigen.

3.4.2.3 Polyphthonge

Bei den Polyphthongen wurden bekannte Polyphthonge mit schwieriger orthographischen Umsetzung, Diphthonge auf [ɥ] sowie Triphthonge untersucht. Die Analyse dieser Merkmale wird im Folgenden dargestellt.

Bekannte Polyphthonge

In der kontrastiven Analyse der Polyphthonge wird davon ausgegangen, dass bekannte Lautverbindungen keine Probleme bereiten (vgl. Kapitel 3.3.2.3). Lediglich das Schriftbild kann zu Unsicherheiten führen. Getestet wurden die Zielwörter *savner*, *regner*, *rejse* und *løj*.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
bekannte Polyphthonge [ḁ̈]	normkonforme Realisierung [ḁ̈]	75 %		
	Lautersatz [äv]	6 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	Lautersatz [ɔ]	8 %		
	artikulatorisch weit abweichender Lautersatz	11 %		
bekannte Polyphthonge [ä̇], orthographisch problematisch	normkonforme Realisierung [ä̇]	96 %		
	Lautersatz [eg]	4 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
bekannte Polyphthonge [ä̈], orthographisch leicht	normkonforme Realisierung [ä̈]	98 %		
	artikulatorisch weit abweichender Lautersatz	2 %		
bekannte Polyphthonge [ǟ]	normkonforme Realisierung [ǟ]	66 %		
	Lautersatz [ø̄]	26 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	artikulatorisch weit abweichender Lautersatz	8 %		

Tab. 3.15: Ergebnisse der Fehleranalyse zu den bekannten Polyphthongen

Am wenigsten Fehler produzieren die Schüler beim Zielwort *rejse*, lediglich 2 % der Aufnahmen sind fehlerhaft. Eine ebenfalls niedrige Fehlerquote ergibt sich für *regner* (4 %). 34 % der Aufnahmen zu *løj* und 25 % der Aufnahmen zu *savner* sind fehlerhaft.

Für das Zielwort *savner* substituieren die Schüler fehlerhaft [äv] (7 %) oder [ɔ] (8 %). In 11 % der fehlerhaften Ausprägungen handelte es sich um weit abweichenden Lautersatz. Die Lautsubstitution [äv] lässt sich mit dem orthographischen Einfluss erklären.

Bei der Lautsubstitution [ɔ] ist es komplizierter. So kann davon ausgegangen werden, dass den Schüler bereits bewusst ist, dass die Buchstabenkombination *av* anders als erwartet bzw. nicht nach deutschen Umsetzungsregeln ausgesprochen wird. Da nicht davon ausgegangen wird, dass ihnen der Laut [äy] Probleme bereitet, wird es sich bei der Lautsubstitution nicht um eine phonetische Approximation handeln. Vielmehr ist anzunehmen, dass die Lerner entweder eine Regel oder einen Laut aus einem anderen Kontext einsetzen. In diesem Fall müsste als Fehler von einer Übergeneralisierung bzw. von einer Unsicherheit bei der Aussprache gesprochen werden.

Bei den Zielwörtern *rejse* und *regner* wird der Ziellaut [äi] angestrebt. Die Lautsubstitution [eg] in 4 % der Aufnahmen des Zielwortes *regner* kann mit orthographischer Interferenz erklärt werden. Die 2 % stark abweichenden Fehler bei *rejse* werden als situative Fehler gewertet. Die Fehlerquoten der beiden Zielwörter sind ungefähr gleich, obwohl die lautliche Umsetzung der Orthographie von *regner* den Lernern deutlich schwerer fallen müsste. Trotzdem wird *regner* überwiegend richtig (96 %) ausgesprochen. Anscheinend sind die Lerner mit dem Wort vertraut. Interessant dabei ist, dass Orthographie eine weniger relevante Rolle zu spielen scheint, wenn ein Wort gelernt wurde bzw. es eine hohe Frequenz hat.

Die Probleme beim Zielwort *løj* besteht in der scheinbar problematischen ersten Komponente, bei der das Graphem ⟨ø⟩ unbekannt ist. Tatsächlich sollte die angestrebte Lautkombination [äi] für deutsche Lerner allerdings nicht problematisch sein, da sie dem deutschen [ɔi] stark ähnelt. Die Lautsubstitution [øi] kann sich aus der orthographischen Interferenz ergeben, wenn die Lerner beim Lesen ⟨ø⟩ durch deutsches ⟨ö⟩ ersetzen.

Diphthonge auf [y]

Die kontrastive Analyse zu den Polyphthongen sagt Probleme bei den Diphthongen auf [y] voraus. Hiermit sind solche Diphthonge gemeint, deren erste Komponente sich von denen aus im deutschen möglichen Kombinationen unterscheidet. Getestet wird die Hypothese anhand der Zielwörter *liv*, *flyver* und *støvler*.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
Diphthonge auf [ɥ]	normkonforme Realisierung [iɥ], [yɥ], [øɥ]	35 %		
	Lautersatz zweite Komponente [v] [iv], [yv], [øv]	16 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	Elision zweite Komponente plus Längung [i:], [y:], [ø:]	28 %	L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Übergeneralisierung: Übertragung von L2-Regeln aus einem anderen Kontext
	artikulatorisch weit abweichender Lautersatz	21 %		

Tab. 3.16: Ergebnisse der Fehleranalyse zu den Diphthongen auf [ɥ]

Für diese Polyphthonge sind wesentlich mehr Fehler zu verzeichnen als für die bekannten Polyphthonge. Lediglich 35 % aller Aufnahmen sind normkonform realisiert. Die Lautsubstitution von [v] in der zweiten Komponente (16 %) kann durch orthographische Interferenz erklärt werden. 28 % der Lerner lassen die zweite Komponente weg und längen die erste. Statt einer Diphthongierung wird eine Tilgung plus Längung durchgeführt. Im Fall *liv* sind diese Apokopen kritisch, da semantisch ein neues Wort entsteht (*liv* [ˈliɥ[?]] ‚Leben‘ vs. *lig* [ˈli:] ‚Leiche‘). Als Ursache für die Elision kann angenommen werden, dass die Schüler bereits Tilgungs- und Assimilationsprozesse in der dänischen Aussprache kennengelernt haben. Sie wenden diese Lautbildungsregeln nun in diesem Kontext – hier allerdings falsch – an. Aus dieser Ursache ergibt sich der Fehlertyp der Übergeneralisierung.

Für das Zielwort *flyver* (mit dem Diphthong [yɥ]) sind mehr Fehler zu verzeichnen als für das Zielwort *støvler* ([øɥ]). Beim Zielwort *liv* ([iɥ]) entstehen am wenigsten Fehler. Der häufigste Fehler bei *støvler* sind artikulatorisch weit abweichende Substitutionen, die vermutlich dadurch indiziert sind, dass das Wort den Lernern unbekannt ist und die Orthographie ihnen Probleme bereitet.

Auch wenn die artikulatorisch weit abweichenden Substitutionen vermutlich einige nicht systematische Fehler (*mistakes*) beinhalten, weist doch die hohe Anzahl darauf hin, dass das Phänomen Diphthonge auf [ɥ] den Lernern Probleme bereitet.

Triphthonge

Von den Triphthongen wird angenommen, dass sie deutschen Lernern Probleme bereiten, da es im Deutschen keine Triphthonge gibt. Getestet werden die Triphthonge anhand der Zielwörter *spurv* und *hvirvler*. Lediglich 2 % der Aufnahmen sind normkonform ([u̥ɐ̯] bzw. [i̥ɐ̯]).

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
Triphthonge	normkonforme Realisierung [u̥ɐ̯], [i̥ɐ̯]	2 %		
	[u̥ɐ̯v], [i̥ɐ̯v]	43 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	[i̥ɐ̯], [u̥ɐ̯]	37 %	L2-Laut in L1 nicht vorhanden (art./perz. ungewohnt/schwierig)	Übergeneralisierung: Übertragung von Regeln zur Tilgung von Wortenden aus einem anderen Kontext
	artikulatorisch weit abweichender Lautersatz	18 %		

Tab. 3.17: Ergebnisse der Fehleranalyse zu den Triphthongen

43 % der Lerner realisieren schriftnah [u̥ɐ̯v] bzw. [i̥ɐ̯v] und 37 % führen eine Elision durch und realisieren [u̥ɐ̯] bzw. [i̥ɐ̯]. In 18 % der Sprachaufnahmen werden weit abweichende Substitutionen durchgeführt.

Ein erster Erklärungsansatz ist die problematische Orthographie. Dafür sprechen zumindest die Äußerungen, die sich artikulatorisch nah an der Schriftsprache befinden. Die Elisionen können auf Interferenzen aus der L1 hindeuten. Im Deutschen sind Triphthonge nicht üblich, was eine Verkürzung zu Diphthongen aus L1-phonotaktischen Gründen bedeuten würde. Darüber hinaus ist es möglich, dass die Lerner bereits bekannte Tilgungs- und Assimilationsprozesse kennengelernt haben oder die impressionistische Regel anwenden, dass im Dänischen die Wortenden häufig nicht mitgesprochen werden. In einem solchem Fall würde es sich um den Fehlertyp der Übergeneralisierung handeln.

Da nur 2 % der Schüleraufnahmen normkonform realisiert wurden, kann die Auswertung der Sprachaufnahmen die Annahmen der kontrastiven Analyse bestätigen: Die Triphthonge bereiten den deutschsprachigen L2-Dänischlernern große Probleme.

3.4.2.4 Vokalische Quantität

Das Problem für deutschsprachige L2-Dänischlerner besteht bei der Quantität der Vokale darin, dass sie aus dem Deutschen bei Änderung der Quantität auch eine Änderung der Qualität gewohnt sind. Sie übertragen das deutsche Phänomen des Quantitätswechsels mit einhergehendem Qualitätswechsel auf das Dänische, weshalb in den Prognosen davon ausgegangen wird, dass die Lerner die kurzen geschlossenen Vokale zu offen aussprechen.

3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
[i]	normkonforme Realisierung	18 %		
	Lautersatz [i]	78 %	kurze L2-Qualität in L1 nicht vorhanden	Interferenz: Ersatz der L2-Qualität durch L1-Qualität
			Orthographie als Realisierungsgrundlage	orthographische Interferenz
			L1- und L2-Wörter ähneln sich	lexikalische Interferenz
			phonologische Regeln der L2 sind unbekannt	phonologische Interferenz: phonologische L1-Regel wird auf L2 übertragen
	Quantitätsänderung [i:]	3 %	kurze L2-Qualität in L1 nicht vorhanden	Interferenz: Ersatz der L2-Quantität durch L1-Quantität
				phonetische Approximation
Phonologische Regeln der L2 sind unbekannt			phonologische Interferenz: phonologische L1-Regel wird auf L2 übertragen	
[i:]	normkonforme Realisierung [i:]	82 %		
	Lautersatz [i]	11 %		
	Quantitätsänderung zu [i]	5 %		
	artikulatorisch weit abweichender Lautersatz	1 %		

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
[y]	normkonforme Realisierung	47 %		
	Lautersatz [ʏ]	14 %	kurze L2-Qualität in L1 nicht vorhanden	Interferenz: Ersatz der L2-Qualität durch L1-Qualität
			Orthographie als Realisierungsgrundlage	orthographische Interferenz
			Phonologische Regeln der L2 sind unbekannt	phonologische Interferenz: phonologische L1-Regel wird auf L2 übertragen
	Quantitätsänderung zu [y:]	38 %	kurze L2-Qualität in L1 nicht vorhanden	Interferenz: Ersatz der L2-Quantität durch L1-Quantität
				phonetische Approximation
		Phonologische Regeln der L2 sind unbekannt	phonologische Interferenz: phonologische L1-Regel wird auf L2 übertragen	
artikulatorisch weit abweichender Lautersatz	1 %			

Tab. 3.18: Ergebnisse der Fehleranalyse zur vokalischen Quantität

Die Sprachdaten zeigen, dass wie erwartet langes [i:] den Lernern keine großen Probleme bereitet (82 % normkonforme Äußerungen),¹⁶ jedoch der kurze geschlossene Vokal [i], der im deutschen Lautinventar nicht vorhanden ist, nur in 18 % der Aufnahmen normkonform realisiert wird. 78 % der Lerner substituieren mit dem offenen [ɪ], wie es im Deutschen zu erwarten wäre. Vier Fehlertypen kommen zur Erklärung der Substitution in Frage:

- a. Kurzes [i] ist im deutschen Lautinventar nicht vorhanden. Die Lerner ersetzen im Zuge einer Interferenz diesen Laut mit den für sie an Kürze gekoppelten Laut [ɪ].

¹⁶ Das Fehlerspektrum weist qualitative ([ɪ] 11 %) und quantitative ([i] 5 %) Substitutionen auf.

- b. Die Lerner realisieren das Zielwort *milde* nach deutschen Regeln lautlicher Umsetzung von Orthographie. In diesem Fall würde es sich um den Fehlertyp der orthographisch motivierten Substitution handeln.
- c. Daran angelehnt besteht darüber hinaus die Möglichkeit, dass es sich um eine lexikalische Interferenz aus dem Deutschen handelt. Sowohl im Deutschen als auch im Dänischen hat *milde* dieselbe Bedeutung und so ist es vorstellbar, dass die Lerner die ihnen bekannte Aussprache übertragen. In diesem Fall würde es sich um eine Substitution durch Interferenz handeln.
- d. Die Lerner kennen die phonologische L2-Regel nicht, dass die geschlossenen Vokale auch in kurzer Quantität vorkommen. Sie wenden deshalb ihre eigene phonologische Regel zur Umsetzung an. Dieser Fehlertyp ist als phonologische Interferenz zu bezeichnen.

Im Gegensatz zum kurzen [i], für das die meisten Lerner eine Qualitätsänderung vornehmen, bestehen die meisten Fehler für kurzes [y] in der Quantitätsänderung (38 %). Außerdem sprechen 47 % der Lerner den Laut richtig aus. Nur 14 % substituieren durch offeneres [ʏ]. Die unterschiedlichen Fehlerfrequenzen können damit zusammenhängen, dass das für [y] gewählte Zielwort *hyggelig* den Schüler bereits häufiger begegnet ist als das für [i] gewählte *milde*. Interessant ist allerdings, dass nur 14 % der Lerner auf [y] zurückgreifen, was auf eine orthographische oder lautliche Interferenz aus dem Deutschen hindeuten würde. Am häufigsten tritt der Fall ein, dass den Schülern die richtige Qualität gelingt, sie diese allerdings in der falschen Quantität realisieren. Es ist insofern erstaunlich, dass die Schüler [y:] realisieren, als das Schriftbild nach deutschen Regeln auf Grund des nachfolgenden doppelten Konsonanten einen kurzen Vokal vermuten ließe. Da *hyggelig* allerdings ein relativ häufig vorkommendes Wort ist, kann davon ausgegangen werden, dass die Lerner den Laut qualitativ kennen und dieser auch im eigenen phonetischen Inventar vorkommt. Den Schülern fehlt allerdings Regelwissen an der Schnittstelle von segmentalen und suprasegmentalen Merkmalen, nämlich, dass Qualität im Dänischen nicht an Quantität gebunden ist und somit Laute in langen und kurzen Versionen vorkommen. Sie wenden daher die ihnen einzig bekannte Regel an und sprechen [y] lang aus. Die phonetische Annäherung hat bereits stattgefunden (phonetische Approximation), als Fehlertyp wäre hier die Interferenz phonologischer Regeln anzuführen. Die zweite Erklärungsmöglichkeit ist eng mit der ersten verbunden. So kann der Fehler einerseits als fehlendes phonologisches Regelwissen bezeichnet werden, andererseits als Interferenz aus der L1. Die Lerner übertragen die deutschen strukturellen Merkmale auf die dänischen. Sie übernehmen die Länge, an die der Laut [y] im Deutschen gekoppelt ist. Als Fehlertyp ließe sich also eine Substitution durch Interferenzformen heranziehen.

3.4.2.5 Phonologische Prozesse

Bei den phonologischen Prozessen wurden die [ə]-Tilgungsprozesse sowie die Lenisierung nachtoniger Plosive untersucht. Die Analyse der Merkmale wird im Folgenden vorgenommen.

Überlange Vokale und Konsonanten in Folge von [ə]-Tilgungsprozessen

Die Prognose zu Tilgungs- mit nachfolgenden Assimilationsprozessen gehen davon aus, dass die Lerner diese Phänomene nicht durchführen.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
überlange Vokale in Folge von [ə]-Tilgungsprozessen, betonte Silben	normkonforme Realisierung	3 %		
	Realisierung [ə]	97 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
			phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen
überlange Vokale in Folge von [ə]-Tilgungsprozessen, unbetonte Silben	normkonforme Realisierung	6 %		
	Realisierung [ə]	75 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
	Realisierung [ɐ]	5 %	L1- und L2-Wörter ähneln sich	lexikalische Interferenz
			phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen
	artikulatorisch weit abweichender Lautersatz	14 %		
lange Konsonanten in Folge von [ə]-Tilgungs-	normkonforme Realisierung	2 %		
	Realisierung [ə]	97 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz

3 Studie 1: Fehler deutschsprachiger L2-Dänischlerner

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp	
prozessen [l:], [m:], [n:], [ŋ:]			phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen	
	artikulatorisch weit abweichender Lautersatz	1 %			
lange Konsonanten in Folge von [ə]-Tilgungsprozessen, Längung weiches <i>d</i> ([ð:])	normkonforme Realisierung	6 %	phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen	
	Realisierung [ə]	88 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz	
	Lautelision		1 %	phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen
				Ziellaut ([ð]) in der L1 nicht vorhanden	Vermeidung: Elision des Ziellauts
artikulatorisch weit abweichender Lautersatz	4 %				

Tab. 3.19: Ergebnisse der Fehleranalyse zu überlangen Vokalen und Konsonanten in Folge von [ə]-Tilgungsprozessen

Die Zielwörter für die betonten Silben sind *lige* ([ˈli:] ‚eben‘) und *uge* ([ˈu:] ‚Woche‘). Die Lerner führen überwiegend keine Tilgung durch und realisieren mit Schwa-Laut [ˈli:ə] und [ˈu:ə] (97 %). Entsprechend bilden sie die Vokale auch nicht überlang. In den unbetonten Silben ergibt sich eine ähnliche Fehlerfrequenz. 6 % der Aufnahmen sind normkonform realisiert. Die Zielwörter hier sind *kortere* ([ˈkʰɔ:ɾɐ:] ‚kürzer‘) und *lærere* ([ˈlɛ:ɾɐ:] ‚Lehrer (PL)‘). Das Fehlerspektrum stellt sich anders dar, als für die Zielwörter mit betonten Silben. 75 % der Lerner sprechen die Wörter so nah an der Schrift aus, dass sie für den ersten [ə]-Laut [e] realisieren, d. h. [ˈkʰɔ:ɾɐɐə] und [ˈlɛ:ɾɐɐə] sprechen. 5 % der Lerner führen eine Tilgung ohne anschließende Längung durch, d. h. [ˈkʰɔ:ɾɐ] und [ˈlɛ:ɾɐ]. 14 % artikulieren stark abweichend ohne erkennbaren Zusammenhang. Die Werte für die Längung von Konsonanten nach Tilgungsprozessen entsprechen denen

der Vokale. Die Zielwörter der Liquide und Nasale *falde, komme, finde, konge* werden in 97 % der Fälle mit finalem [ə]-Laut gebildet. Die Lerner nehmen keine Schwa-Tilgung mit einhergehenden Längungsprozessen ([fæ:l], [kʰɔ̃m:], [fɛ:n] und [kʰɔ̃ŋ:]) vor, wie es in der Umgangssprache üblich wäre. Stattdessen realisieren sie [ˈfælə], [ˈkʰɔ̃mə], [ˈfɪnə] und [ˈkʰɔ̃ŋə]. Zur Überprüfung der Längung des weichen *d* nach Tilgungsprozessen wurden die Zielwörter *betyde* und *ballade* abgefragt. 6 % der Lerner führen eine Schwa-Tilgung mit anschließender Assimilation bzw. Längung des Approximanten durch. 88 % der Lerner realisieren den [ə]-Laut. Ein Schüler führt eine Lautelision durch und spricht [bəˈtʰy:]. Dabei wird zwar eine Längung durchgeführt, allerdings des Vokals und nicht, wie normkonform üblich, des Konsonanten. 4 % der Aufnahmen weisen artikulatorisch weit abweichende Formen auf.

Bevor die Ursachen für die Fehler diskutiert werden, muss festgehalten werden, dass es sich bei diesem Phänomen nicht unbedingt um einen Fehler handelt. Die Realisierungen ohne [ə]-Tilgung mit nachfolgenden Vokallängungsprozessen ist, wenn auch in der normalen Alltagssprache unüblich, standardsprachlich möglich. Ursache für die schriftnahe Aussprache können Effekte fremdsprachlichen Unterrichts sein, in dem auf deutliche Aussprache geachtet wird, um somit der Lehrkraft zu ermöglichen, zu überprüfen, ob die Schüler morphematische Strukturen bzw. grammatikalische Endungen richtig bilden. Darüber hinaus ist sicherlich auch die Erhebungsmethode des Vorlesens mit für die von den Schülern nah an der Schrift realisierte Aussprache verantwortlich. Obwohl es sich definitiv also nicht unbedingt um einen Fehler handelt, wird im Folgenden trotzdem eine Fehlerklassifikation durchgeführt.

In diesem Sinne kann bei der Realisierung von [ə] bzw. [ɐ] von einer Einfügung gesprochen werden, die durch die Schriftsprache begünstigt wird. Die meisten hier dargestellten nicht durchgeführten [ə]-Tilgungsprozesse beruhen auf der Umsetzung der Orthographie oder der nicht Umsetzung der phonologischen Regel. Für diejenigen Lerner, die bei den unbetonten Silben zwar eine Tilgung durchführen, jedoch lediglich [ɐ] ohne Längung realisieren, kann angenommen werden, dass sie das Prinzip der Tilgung von Wortenden in bestimmten Kontexten schon verstanden haben, aber noch nicht vollständig durchführen. Im Fall der Elision beim Zielwort *betyde* ist vorstellbar, dass dem Schüler das Regelwissen fehlt, welcher Laut in diesem Tilgungsprozess assimiliert werden muss. Eine weitere Möglichkeit ist, dass er die Bildung bzw. Längung des problematischen Lauts [ø] vermeidet. Die Fehler, die beim Zielwort *lærere* entstehen, können auch durch lexikalische Interferenz, d. h. auf Grund der Ähnlichkeit zum Deutschen entstanden sein.

Lenisierung nachtoniger Plosive

In der kontrastiven Analyse in Kapitel 3.3.3.2 wird davon ausgegangen, dass die Lenisierung nachtoniger Plosive den Lernern Probleme bereiten wird. Dies hängt damit zusammen, dass es diese Prozesse im Standarddeutschen nicht gibt. Es wird allerdings ange-

merkt, dass den Lernern ihr norddeutscher Regiolekt zur normkonformen Realisierung verhelfen könnte.

untersuchtes Zielmerkmal	Klassifikation (Symptom)	relative Häufigkeit	Fehlerursache	Fehlertyp
Lenisierung nachtoniger Plosive intervokalisches	normkonforme Realisierung	16 %		
	Lautersatz stimmlose Variante [p], [t], [k]	84 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
			L1- und L2-Wörter ähneln sich	lexikalische Interferenz
			phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen
Lenisierung nachtoniger Plosive pränasal	normkonforme Realisierung	10 %		
	Lautersatz stimmlose Variante [p], [t], [k]	90 %	Orthographie als Realisierungsgrundlage	orthographische Interferenz
			L1- und L2-Wörter ähneln sich	lexikalische Interferenz
			phonologische Regel der L2 nicht bekannt	phonologische Interferenz: phonologische L1-Regel wird auf Zielwort übertragen

Tab. 3.20: Ergebnisse der Fehleranalyse zur Lenisierung nachtoniger Plosive

Die Fehlerhäufigkeiten intervokalisches und pränasal entsprechen sich überwiegend. Intervokalisches realisieren 16 % der Lerner normkonform die Plosive stimmhaft, pränasal sind es 10 %. Das Fehlerspektrum ist übersichtlich. Wenn die Lerner fehlerhaft realisieren, handelt es sich um die stimmlose Variante [p], [t] oder [k].

Zurückzuführen sind diese Fehler auf Interferenzformen. Bei den Zielwörtern *Mette*, *pakke*, *mappen* und *jakken* handelt es sich um Kognaten, also um mit dem Deutschen verwandte Wörter. Im Deutschen werden die Wörter mit stimmlosen Plosiven gebildet, weshalb bei den Fehlern von lexikalischer Interferenz ausgegangen werden kann. Darüber hinaus ist vorstellbar, dass die Aussprache der Schüler auf Grund der Vorlesemethode durch orthographische Interferenzen beeinflusst ist. Die doppelten Konsonanten im Schriftbild werden nach deutschen Leseregeln stimmlos realisiert. Als weitere mögliche Ursache darf nicht ausgeschlossen werden, dass die Schüler auch ohne Schriftbild

und lexikalischer Verwandtschaft die Konsonanten stimmlos realisieren, weil der Kontext nachtoniger Plosive im Deutschen Stimmlosigkeit voraussetzt. Auch in diesem Fall würde es sich um den Fehlertyp Lautsubstitution durch Interferenzformen handeln, allerdings um eine phonologische. Insgesamt lässt sich festhalten, dass die Lerner trotz norddeutschen Regiolektiv Schwierigkeiten mit der Lenisierung nachtoniger Plosive aufweisen.

3.4.2.6 Zusammenfassung

In der detaillierten Fehleranalysen in den Kapiteln 3.4.2.1–3.4.2.5 wurden die in der kontrastiven Analyse in Kapitel 3.3 vorhergesagten möglichen Probleme näher untersucht. Dazu wurden die Häufigkeit des Auftretens, eine deskriptive Klassifikation zur Erscheinungsform der Fehler sowie Interpretationsmöglichkeiten zu Ursachen und Typen der auftretenden Fehler dargestellt.

Überwiegend alle aus der kontrastiven Analyse angenommenen Probleme wurden in der Lernaltersprache aufgefunden. Lediglich der erwartete Fehler beim wortinitialen stimmlosen [s] tritt in keiner der Sprachaufnahmen auf. Am häufigsten tritt das Phänomen nicht affriziertes t auf. Die Ergebnisse lassen sich in Abb. 3.5 nachvollziehen. In dieser werden alle Fehler anhand ihrer Häufigkeit sortiert dargestellt.

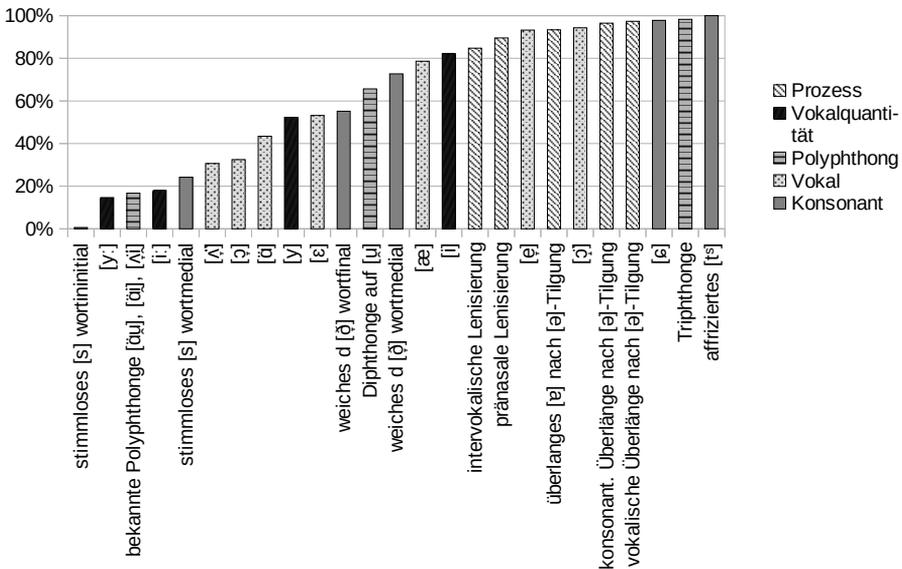


Abb. 3.5: Fehlerhäufigkeiten (aufsteigend sortiert)

Teilt man das Diagramm in eine untere (eher weniger häufig auftretende Fehler) und obere Hälfte (eher häufig auftretende Fehler) bei 50 %, so ist auffällig, dass in der Hälfte der häufigen Fehler alle abgeprüften Prozesse und nur wenige Diphthonge verortet sind. Möglich ist, dass die Prozesse entweder häufig von Lernern falsch durchgeführt werden oder diese auftreten, da die Lerner sich in der Erhebung und im Kontext Unterricht sehr nah an der Schrift orientieren. Vorstellbar ist allerdings auch, dass dieses Phänomen im schulischen Kontext weniger geübt wird als andere Phänomene. So ist auffällig, dass das relativ schwierige Zielmerkmal weiches *d* bereits in vielen Aufnahmen korrekt realisiert wurde und der Fehler zwar eher häufig, aber weniger häufig als andere Fehler auftritt. Die eher weniger häufig auftretenden Fehler lassen sich weiter in sehr selten auftretende Fehlern (0–25 %) unterteilen. Zu diesen gehören Fehler innerhalb der Zielmerkmale der Vokalqualität und wortinitiales [s] sowie der hintere gerundete Vokal [ɔ]. Interessant ist, welche Fehlertypen hinter den hier vertretenen sprachlichen Kategorien stehen. So ist erkennbar, dass es sich bei den häufig auftretenden Fehlern um solche Phänomene handelt, die in der explanativen Ursachenzuschreibung mit Interferenzen erklärt werden können. Bei einigen der mittelmäßig bis relativ häufig auftretenden Fehlern handelt es sich um solche, die vermutlich im Unterricht bislang wenig behandelt werden. So könnten diese durch relativ einfache Strategien vermieden werden, wie beispielsweise Fehler bei affriziertem *t*, den Lenisierungsprozessen und den Triphthongen.

3.4.3 Auswertung der sprecherbiographischen Daten

Während in der vorangegangenen Fehleranalyse in Kapitel 3.4.2 die Häufigkeiten der Fehler dargestellt wurden, werden im Folgenden diese Häufigkeiten mit nichtsprachlichen bzw. sprecherbiographischen Metadaten in Beziehung gesetzt. Die dazu verwendeten Daten entstammen der in Kapitel 3.4.1.3 beschriebenen Datenerhebung. Da die sprecherbiographischen Daten lediglich als Zusatz der eigentlich Sprachdaten erhoben wurden, kann die folgende Analyse lediglich explorative Funktion haben und sollte durch anschließende Forschung gestützt werden.

Da es sich um nominalskalierte Daten handelt, wird die Beziehung der Prädiktoren zu den Fehlern anhand eines Chi-Quadrat-Tests (χ^2) errechnet. Bei diesem Test wird untersucht, ob sich Häufigkeiten bzw. Prozentwerte in verschiedenen Populationen unterscheiden (Bühner & Ziegler 2012: 295–310). Die Nullhypothesen in diesem Verfahren sagen aus, dass keine Häufigkeitsunterschiede in der Grundgesamtheit vorliegen. Die Alternativhypothese ist ungerichtet und lautet, dass Häufigkeitsunterschiede in der Grundgesamtheit vorliegen (Bühner & Ziegler 2012: 296). Die Nullhypothese wird bei einem Signifikanzniveau von $p = 0,05$ abgelehnt.

Die praktische Bedeutsamkeit der Ergebnisse wird durch die Effektstärke geprüft, die in der Regel dem Phi-Korrelationskoeffizienten (Φ) oder Cramérs *V* entspricht (Bühner

& Ziegler 2012: 303).¹⁷ Die Effektstärke ist ein Maß für den Zusammenhang zwischen der abhängigen und unabhängigen Variable. Dabei gelten die folgenden Konventionen (Bühner & Ziegler 2012: 302):

- a. $\leq 0,10$ = kleiner Effekt;
- b. $\leq 0,30$ = mittlerer Effekt;
- c. $\leq 0,50$ = starker Effekt.

Die Berechnung erfolgt mit Hilfe der Software *R* (Version 3.5.1 Feather Spray, 2018), einer quelloffenen Software.¹⁸ Die bei der statistischen Hypothesenüberprüfung entstandenen *R*-Outputs können im Anhang A.3 nachvollzogen werden. Die in *R* eingegebene Formel zur Berechnung des Chi-Quadrat-Tests lautet `chisq.test()`.

Im folgenden Kapitel 3.4.3.1 werden Hypothesen zu sprecherbiographischen Daten aufgestellt, die als Prädiktoren für Fehler verantwortlich sein könnten. Dazu gehören der sprachliche Hintergrund (vgl. Kapitel 3.4.3.2), die Häufigkeit der Verwendung der Zielsprache (vgl. Kapitel 3.4.3.3) und das Geschlecht (vgl. Kapitel 3.4.3.4) der getesteten Lerner. Zuletzt werden die Unterschiede zwischen den Klassen miteinander verglichen (vgl. Kapitel 3.4.3.5).

3.4.3.1 Hypothesen zu Zusammenhängen zwischen sprecherbiographischen Metadaten und Fehlern

In Kapitel 3.1 wurde dargestellt, dass Fehler nicht nur durch den Einfluss der *L1*, sondern noch durch weitere Prädiktoren erklärt werden können. So können weitere Spracherfahrungen sowie nichtsprachliche Faktoren eine Rolle bei der Produktion von (Aus)Sprache spielen. Bei den nicht-sprachlichen Faktoren ist es schwierig bzw. unnötig festzumachen, auf welche spezifischen lautlichen Phänomene diese sich genau auswirken. In den meisten Fällen wird deshalb überwiegend untersucht, ob Lerner, auf die dieser außersprachliche Faktor zutrifft, insgesamt mehr oder weniger Fehler machen. Bei den weiteren sprachlichen unabhängigen Variablen hingegen ist es möglich, Hypothesen zu lautlichen Spezifika aufzustellen. Diese können aus kontrastiven Gegenüberstellungen mit der *L3* oder dem Regiolekt ermittelt werden. Innerhalb der getesteten Probandengruppe ergibt sich für einige Probanden, dass diese anstelle von Standarddeutsch mit Niederdeutsch als Erst- oder Zweitsprache aufgewachsen sind. Bei diesen Lernern ist es vorstellbar, dass ihnen einige problematische Laute des Dänischen leichter fallen, als den Lernern ohne niederdeutschen Sprachhintergrund. So zeigt die kontrastive Analyse in Kapitel 3.3, dass

¹⁷ Wenn die unabhängige Variable mehr als zwei Ausprägungen hat (z. B. wenn die acht Lehrkräfte verglichen werden), erfolgt die Berechnung der Effektstärke über Cramérs *V* (bzw. Cramérs *Phi*). Haben beide nominalskalierten Daten jeweils zwei Ausprägungen, wird der *Phi*-Koeffizient als Maßzahl der Stärke des Zusammenhangs genutzt. Die *Phi*-Werte können zwischen 0 (kein Zusammenhang) und 1 (perfekter Zusammenhang) liegen.

¹⁸ Die in *R* eingesetzten Pakete waren *ggplot2* (Wickham 2016) und *lme4* (Bates et al. 2015).

zwischen dem norddeutschen Regiolekt und dem Dänischen in einigen lautlichen Aspekten mehr Ähnlichkeit besteht als zum Standarddeutschen. In der dialektologischen Forschung wird allerdings davon ausgegangen, dass norddeutsche Merkmale abgeleitete Merkmale aus dem Niederdeutschen darstellen (Höder 2014: 321–322). Während im Norddeutschen Merkmale jedoch eher fakultativ sind, sind sie im Niederdeutschen obligatorisch. Es ist vorstellbar, dass die Schüler mit niederdeutschem Sprachhintergrund Vorteile bei diesen Phänomenen haben (z. B. bei der Lenisierung nachtoniger Plosive), da sie diese bereits aus dem Niederdeutschen kennen. Bei diesen Phänomenen sollte also überprüft werden, ob Lerner mit niederdeutschem Sprachhintergrund signifikant weniger Fehler machen als die Schülerinnen ohne niederdeutschen Sprachhintergrund. Darüber hinaus ist es möglich, dass den Lernern ganz generell durch ein weiteres Sprachspektrum das Erlernen neuer Phänomene leichter fällt. Es ergibt sich die folgende Hypothese:

H_{1,0} Nullhypothese

Lerner mit niederdeutschem Sprachhintergrund unterscheiden sich in der Anzahl der Aussprachefehler nicht von Lernern ohne niederdeutschen Sprachhintergrund.

H_{1,1} Alternativhypothese

Lerner mit niederdeutschem Sprachhintergrund unterscheiden sich in der Anzahl der Aussprachefehler von Lernern ohne niederdeutschen Sprachhintergrund.

Die Alternativhypothese H_{1,1} ist ungerichtet und geht lediglich von einem Unterschied aus, d. h. sie sagt nichts über die Richtung des Unterschieds aus. So gibt die Überprüfung der Hypothese anhand des Chi-Quadrat-Verfahrens keinen Aufschluss darüber, welche der untersuchten Gruppen ‚besser‘ bzw. ‚schlechter‘ ist.

Vorstellbar auf Grundlagen von Ergebnissen aus der kontrastiven Analyse in Kapitel 3.3 ist, dass Lerner mit niederdeutschem Sprachhintergrund Vorteile bei einigen spezifischen sprachlichen Merkmalen haben. Darunter fallen spezifisch die Laute ([æ] und [ɔ]), Merkmale der Vokalquantität sowie die Lenisierung nachtoniger Plosive.

Ein weiterer zu untersuchender sprachlicher Faktor, der nicht mit der L₁ verknüpft ist, ist das Pensum, wie oft und in welchem Rahmen die L₂ verwendet wird. Anzunehmen ist, dass Lerner, die häufiger und in mehr Kontexten mit der dänischen Sprache umgehen, auch weniger Fehler produzieren. Hierzu ergibt sich die folgende Hypothese:

H_{2,0} Nullhypothese

Lerner, die Dänisch auch in Kontexten außerhalb der Schule sprechen, unterscheiden sich in der Anzahl der Aussprachefehler nicht von Lernern, die Dänisch nur in der Schule sprechen.

H_{2,1} Alternativhypothese

Lerner, die Dänisch auch außerhalb der Schule sprechen, unterscheiden sich in

der Anzahl der Aussprachefehler von Lernern, die Dänisch nur in der Schule sprechen.

Die Alternativhypothese $H_{2,1}$ ist ungerichtet und geht lediglich von einem Unterschied aus und sagt nichts über die Richtung des Unterschieds aus, d. h. die Ergebnisse sagen nichts darüber aus, welche der Gruppen ‚besser‘ bzw. ‚schlechter‘ ist.

Neben diesen sprachlichen Faktoren werden auch nichtsprachliche Faktoren für Fehlerhäufigkeiten miteinbezogen. Geschlechtsspezifische Leistungsunterschiede sind ein häufig diskutiertes Thema innerhalb der Schulforschung (vgl. Faulstich-Wieland 2008: 673). Im fremdsprachendidaktischen Bereich gibt es beispielsweise Untersuchungen, ob Mädchen oder Jungen bessere Leistungen erzielen oder wie ihre Präferenzen und Interessen in Bezug auf sprachliche oder naturwissenschaftliche Schulfächer sind (z. B. Helmke et al. 2008; Zaunbauer, Retelsdorf & Möller 2009). Allgemein fallen die Ergebnisse der Mädchen in der Leistungsmessung im Fremdsprachenunterricht meistens besser aus, im mündlichen Bereich (Flüssigkeit und Aussprache) zeigen Zaunbauer, Retelsdorf & Möller (2009: 162) allerdings, dass die Jungen besser sind. Die Erkenntnisse für diesen Faktor erweisen sich als eher gering, weshalb in der vorliegenden Studie angenommen wird, dass sich Jungen und Mädchen nicht signifikant in der Fehlerhäufigkeit unterscheiden. Die Formulierung der Hypothese stellt sich folgendermaßen dar:

$H_{3,0}$ Nullhypothese

Jungen und Mädchen unterscheiden sich nicht signifikant in der Fehlerhäufigkeit der Aussprachefehler im Dänischen.

$H_{3,1}$ Alternativhypothese

Jungen und Mädchen unterscheiden sich signifikant in der Fehlerhäufigkeit der Aussprachefehler im Dänischen.

Eine weitere unabhängige nichtsprachliche Variable ist die Lehrkraft. Dass das Wissen und Können von Lehrkräften einen bedeutsamen Effekt auf die Leistungen von Lernern hat, wurde in unterschiedlichen Zusammenhängen und einer Vielzahl von Studien nachgewiesen (z. B. Kunter et al. 2011; Jansing et al. 2013). In diesen Bereich tiefer vordringende Studien untersuchen, welche Lehrermerkmale genau ausschlaggebend für die Unterschiede in den Schülerleistungen sind. Dazu werden Bereiche des professionellen Wissens und Könnens oder auch emotionale und motivationale Merkmale sowie negative Faktoren der beruflichen Beanspruchung (z. B. Erschöpfung, Burnout und Stress) als Prädiktoren angenommen (z. B. in Klusmann & Richter 2014). Zu den professionellen Bereichen des Wissens und Könnens werden die eigene Aussprachekompetenz der Lehrkräfte sowie ihre Methode der Vermittlung gezählt. Um nähere Aussagen über diese treffen zu können oder Hypothesen dazu aufzustellen, reicht das in dieser Arbeit angewandte Forschungsdesign nicht aus. Untersucht werden kann jedoch, ob die Klassen der Lehrkräfte untereinander divergieren und wenn ja, ob sie sich in spezifischen Phänomenen unterscheiden. So ist vorstellbar, dass Lehrkräfte selbst Probleme mit einigen Phäno-

menen in der eigenen Aussprache haben oder unterschiedliche Phänomene in ihrem Unterricht fokussieren. Ob sich diese Umstände in den Leistungen der Schüler niederschlagen, kann zwar mit dem in dieser Studie angewandten Forschungsdesign nicht zuverlässig nachgewiesen werden, es können allerdings Tendenzen ermittelt werden, die in nachfolgenden Studien näher verfolgt werden können. Zu der nichtsprachlichen weiteren unabhängigen Variable Lehrkraft ergeben sich somit folgende Hypothesen:

H_{4.0} Nullhypothese

Die Klassen der unterschiedlichen Lehrkräfte unterscheiden sich nicht.

H_{4.1} Alternativhypothese

Die Klassen der unterschiedlichen Lehrkräfte unterscheiden sich signifikant.

In den folgenden Kapiteln werden die aufgestellten Hypothesen geprüft. Zunächst werden die weiteren sprachlichen Aspekte überprüft, die nicht an die L₁ gekoppelt sind. Im Anschluss daran folgt die Untersuchung der nichtsprachlichen Aspekte (Geschlecht und Lehrkraft).

3.4.3.2 Häufigkeit der Fehler nach sprachlichem Hintergrund

Die grundlegende Hypothese H_{1.1} lautet, dass Lerner mit niederdeutschem Sprachhintergrund einen Vorteil gegenüber Lernern ohne niederdeutschen Sprachhintergrund haben und weniger Fehler produzieren. Die Nullhypothese H_{1.0} geht nicht davon aus, dass ein Unterschied besteht. Die Analyse kann nur explorativen Charakter haben, da die sprecherbiographischen Daten aus Selbstauskünften der Sprecher stammen und nicht kontrolliert werden konnten. So ist die im Folgenden dargestellte Einteilung der Probandengruppen kritisch zu betrachten. Von den 91 untersuchten Schülern haben 39 Kontakt mit Niederdeutsch. Dieser Kontakt kann entweder sehr häufig und eng bzw. durch Freunde und Familie geprägt oder sehr weit entfernt sein. So geben die Schüler an, Niederdeutsch zu kennen oder zu verstehen, wenn die Nachbarn es sprechen, es selbst aber nicht aktiv zu sprechen. 28 der Schüler haben engeren Kontakt mit Niederdeutsch in ihrem täglichen Leben, z. B. durch Freunde und nahe Verwandte. Eine Schülerin gibt an, eine Niederdeutsch-AG zu besuchen, d. h., dass es diese Möglichkeit an der Schule gibt. In den Berechnungen werden nur die 28 Schüler als Lerner mit niederdeutschem Sprachhintergrund gezählt, die engen Kontakt zum Niederdeutschen haben. Kritisch bleibt zu betrachten, dass von diesen nicht ausnahmslos bekannt ist, ob sie Niederdeutsch auch produktiv selbst gebrauchen. Tab. 3.21 stellt die relativen Fehlerhäufigkeiten der Sprecher mit und ohne niederdeutschem Sprachhintergrund gegenüber.

untersuchtes Merkmal	% fehlerhafter Aufnahmen	
	Lerner mit niederdt. Sprachhintergrund	Lerner ohne niederdt. Sprachhintergrund
über alle Fehler hinweg	59	67
vorderer ungerundeter Vokale [æ]	76	80
hinterer gerundeter Vokal [ɔ]	32	33
Vokalquantität	76	64
Lenisierung nacht. Plosive intervokalisch	83	86
Lenisierung nacht. Plosive pränasal	94	88

Tab. 3.21: Relative Fehlerhäufigkeiten der Sprecher mit und ohne niederdeutschen Sprachhintergrund in Prozent

Die Sprachaufnahmen der Lerner mit niederdeutschem Sprachhintergrund sind in 59 % der Fälle fehlerhaft, während die Aufnahmen der Lerner ohne niederdeutschen Sprachhintergrund in 67 % der Fälle fehlerhaft sind. Der χ^2 -Test zeigt allerdings, dass dieser Unterschied nicht signifikant ist ($\chi^2 = 2,784$, $df = 1$, $p < 0,095$). Die Nullhypothese $H_{1.0}$ kann somit nicht verworfen werden.

Die ausgewählten Phänomene werden im Folgenden näher betrachtet. Die Auswahl resultiert hypothesengeleitet aus der in Kapitel 3.3 durchgeführten kontrastiven Analyse, aus der sich Vermutungen zu sprachlichen Merkmalen ergeben, die niederdeutschen Lernern eventuell leichter als deutschen Lernern fallen. Dazu gehören der vordere ungerundete Vokal [æ], der hintere gerundete Vokal [ɔ], die Vokalquantität sowie die Lenisierung nachtoniger Plosive. Die Ergebnisse des Chi-Quadrat-Tests für diese ausgewählten Merkmale stellen sich folgendermaßen dar:

- a. vorderer ungerundeter Vokal [æ]: Dieser Laut wird in den Aufnahmen der Lerner ohne niederdeutschen Sprachhintergrund in 80 % der Aufnahmen fehlerhaft realisiert, die Aufnahmen der Sprecher mit niederdeutschem Sprachhintergrund sind in 76 % der Fälle falsch. Die Unterschiede erweisen sich im χ^2 -Test als statistisch nicht signifikant ($\chi^2 = 2,555$, $df = 1$, $p < 0,109$).
- b. hinterer gerundeter Vokal [ɔ]: Die Lerner mit niederdeutschem Sprachhintergrund erzielen leicht bessere Mittelwerte (32 %) als diejenigen ohne niederdeutschen Sprachhintergrund (33 %). Diese Unterschiede ergeben in der statistischen Analyse keine Signifikanz ($\chi^2 = 2,555$, $df = 1$, $p < 0,109$).
- c. Vokalquantität: Die Lerner ohne niederdeutschen sprachlichen Hintergrund stellen sich als besser (64 %) heraus als die Lerner mit niederdeutschem Sprachhintergrund (76 %). Dieses Ergebnis ist statistisch allerdings nicht signifikant ($\chi^2 = 2,555$, $df = 1$, $p < 0,109$).

- d. Lenisierung nachtoniger Plosive: Die Aufnahmen für das Merkmal intervokalisch sind in 83 % der Fälle der Sprecher mit niederdeutschem Sprachhintergrund falsch und in 86 % der Fälle der Sprecher ohne niederdeutschen Sprachhintergrund falsch. Pränasal erreichen die Sprecher mit niederdeutschem Sprachhintergrund höhere Fehlerwerte von 94 %, während die Sprecher ohne niederdeutschen Sprachhintergrund nur in 88 % falsch realisieren. Die Unterschiede bei der Lenisierung nachtoniger Plosive stellen sich als statistisch nicht signifikant dar (intervokalisch: $\chi^2 = 0,183$, $df = 1$, $p < 0,669$, pränasal: $\chi^2 = 2,095$, $df = 1$, $p < 0,148$).

Sowohl in der Gesamtfehlerhäufigkeit als auch in der Häufigkeit innerhalb einzelner Merkmale zeigen sich keine signifikanten Zusammenhänge zum sprachlichen Hintergrund Niederdeutsch. Die Hypothese $H_{1,1}$ muss somit für das vorliegende Setting verworfen werden.

3.4.3.3 Häufigkeit der Fehler nach Anwendungshäufigkeit der Zielsprache

Ausgangspunkt für die folgende Analyse ist die Hypothese $H_{2,1}$, dass Lerner, die Dänisch auch außerhalb der Schule sprechen, sich in der Anzahl der Aussprachefehler von Lernern unterscheiden, die Dänisch nur in der Schule sprechen. Sechs Schüler gaben an, auch außerhalb des Unterrichts Dänisch zu sprechen, zum Beispiel mit Freunden, Familie, im Nebenjob oder im Urlaub. Die Aufnahmen der Lerner, die angeben auch außerhalb der Schule Dänisch zu sprechen, waren in 63 % fehlerhaft, die Aufnahmen der übrigen Schüler in 65 % der Fälle. Dass die Unterschiede nicht zufällig entstanden sind, kann nicht belegt werden. Die Signifikanzanalyse zeigt keine nachweisbaren statistischen Zusammenhänge ($\chi^2 = 0,918$, $df = 1$, $p < 0,338$). Die Nullhypothese $H_{2,0}$ kann aus diesem Grund nicht verworfen werden.

In der Gesamtübersicht der Daten verteilen sich die Fehlerhäufigkeiten der beiden Gruppen (diejenigen Schüler, die auch außerhalb der Schule Dänisch sprechen und diejenigen, die es nicht tun) relativ gleichmäßig. Lediglich in einigen Phänomenen werden Unterschiede deutlich, die im Folgenden genauer dargestellt werden.

Untersuchtes Merkmal	% fehlerhafter Aufnahmen	
	auch außerhalb der Schule Dänisch	kein Dänisch außerhalb der Schule
über alle Fehler hinweg	63 %	65 %
weiches <i>d</i> wortmedial	62 %	70%
hinterer ungerundeter Vokal [ḁ̈]	17 %	34 %
bekannte Polyphthonge [ḁ̈y]	11 %	29 %
bekannte Polyphthonge [ḁ̈i]	26 %	36 %
Diphthonge auf [y]	32 %	71 %

Tab. 3.22: Relative Fehlerhäufigkeiten bei Dänischgebrauch außerhalb der Schule

Tab. 3.22 zeigt die relativen Häufigkeiten der näher untersuchten Ergebnisse für die Schüler, die außerhalb der Schule Dänisch sprechen und die es nicht tun. Beim weichen *d* sind keine nennenswerten Unterschiede zu verzeichnen. Während 62 % der Aufnahmen der Schüler, die auch außerhalb der Schule Dänisch sprechen, fehlerhaft sind, sind es bei den anderen Schülern 70 % der Aufnahmen. Der χ^2 -Test ergibt, dass dieser Unterschied nicht statistisch signifikant ist ($\chi^2 = 0,856$, $df = 1$, $p < 0,355$).

Für den hinteren ungerundeten Vokal [ḁ̈] stellen sich die Unterschiede zwischen den Schüler, die Dänisch auch außerhalb der Schule sprechen und denen, die es nicht tun, als größer dar. 17 % der Aufnahmen der Schüler, die auch außerhalb der Schule Dänisch sprechen sind fehlerhaft, während die Aufnahmen der anderen Schüler in 34 % der Fälle fehlerhaft sind. Auch für diese Unterschiede gibt der χ^2 -Test jedoch an, dass sie nicht statistisch signifikant sind ($\chi^2 = 1,976$, $df = 1$, $p < 0,159$).

Die Unterschiede in den arithmetischen Mitteln beim bekannten Polyphthong [ḁ̈y] mit unbekannter Schreibweise zeigen, dass 11 % der Aufnahmen der Schüler, die auch außerhalb der Schule Dänisch sprechen, fehlerhaft sind, während die anderen Schüler in 29 % der Aufnahmen falsch realisierten. Der im χ^2 -Test ermittelte *p*-Wert ist zwar zu hoch, um von einer statistischen Signifikanz ausgehen zu können, lässt aber eine Tendenz vermuten, dass die Unterschiede nicht zufällig sein müssen ($\chi^2 = 2,707$, $df = 1$, $p < 0,099$).

Bei den bekannten Polyphthongen [ḁ̈i] realisieren 26 % der Schüler, die auch außerhalb der Schule Dänisch sprechen, fehlerhaft, bei den anderen sind es 36 %. Diese Unterschiede sind nicht signifikant ($\chi^2 = 0,651$, $df = 1$, $p < 0,419$).

Bei den Diphthongen auf [y] lassen sich die größten Unterschiede verzeichnen. Nur 32 % der Aufnahmen der Schüler, die Dänisch auch außerhalb der Schule sprechen, sind falsch, während es bei den übrigen Schülern 71 % sind. Der χ^2 -Test mit abgeleitetem Phi-

Koeffizienten bestätigt die Unterschiede als statistisch signifikant, allerdings mit schwachem Zusammenhang ($\chi^2 = 8,768$, $df = 1$, $p < 0,003$, $\Phi = 0,184$).

In der Gesamtverteilung zeigen sich keine signifikanten Unterschiede zu denjenigen Lernern, die Dänisch auch außerhalb der Schule sprechen und denjenigen, die dies nicht tun. In der Häufigkeit einzelner Merkmale zeigt sich bei den Diphthongen auf [u] ein signifikanter Unterschied zwischen den untersuchten Gruppen. Dieser ist allerdings mit einem schwachen Effekt zu bewerten. Die Stichprobengrößen sind nicht gleich groß und so ist davon auszugehen, dass die Gegenüberstellung keine repräsentative Aussagekraft haben kann. Aus diesem Grund ist es empfehlenswert, die grundlegende Hypothese mit einem anderen Setting und einer besseren Konstellation der Probandengruppen noch einmal zu untersuchen.

3.4.3.4 Häufigkeit der Fehler nach Geschlecht

Die grundlegende Nullhypothese $H_{3,0}$ der folgenden Analyse besagt, dass es keinen geschlechtsspezifischen Unterschied in der Fehlerverteilung gibt. Unter den Probanden sind insgesamt 62 Mädchen und 29 Jungen. Die relativen Häufigkeiten für die Sprachaufnahmen zeigen, dass 63 % der Aufnahmen der Jungen und 60 % der Aufnahmen der Mädchen fehlerhaft sind. Die absoluten Häufigkeiten stellen sich folgendermaßen dar:

	männlich	weiblich	Zeilensumme
kein Fehler	501	1.168	1.669
Fehler	1.063	2.080	3.143
Spaltensumme	1.564	3.248	4.812

Tab. 3.23: Absolute Fehlerhäufigkeiten nach Geschlecht

Es wird ein χ^2 -Vierfeldertest gerechnet. Dieser bestätigt einen hoch signifikanten Zusammenhang zwischen Geschlecht und den Normabweichungen ($\chi^2 = 7,015$, $df = 1$, $p < 0,008$, $\Phi = 0,039$). Die Effektstärke, ermittelt über den Phi-Koeffizienten, ist allerdings als sehr gering einzustufen. Für die Interpretation der Ergebnisse ist die Effektstärke relevant, die in diesem Fall sehr gering ist. Die Nullhypothese $H_{3,0}$ kann somit zwar verworfen werden, der tatsächliche Effekt, den das Geschlecht auf das Auftreten von Fehlern in der Aussprache hat, ist allerdings als sehr gering einzustufen.

3.4.3.5 Häufigkeit der Fehler nach Lehrkraft

Die für die folgende Analyse grundlegende Nullhypothese $H_{4,0}$ geht davon aus, dass zwischen den verschiedenen Klassen keine Unterschiede in der Fehlerhäufigkeit bestehen. Die Unterschiedlichkeit der Gruppen bzw. Klassen wird keiner statistischen Signifikanz-

untersuchung unterzogen. Worin die Unterschiedlichkeit begründet liegt, kann auf Grund des Untersuchungssettings nicht eindeutig ermittelt werden. Unterschiedlichkeiten lassen sich außerdem nicht eindeutig auf die Lehrkraft zurückführen, da sie ebenso lerngruppenspezifisch sein können oder von weiteren nicht einschätzbaren Parametern abhängen. Dies wurde im vorliegenden Untersuchungssetting nicht kontrolliert. Die Auswertung der sprecherbiographischen Daten zum Faktor Lehrkraft erfolgt daher rein deskriptiv und ermittelt lediglich Vermutungen, die in einer weiterführenden Untersuchung gestützt werden sollten.

	LK1	LK2	LK3	LK4	LK5	LK6	LK7	LK8
relative Fehlerhäufigkeit aller Schüleraufnahmen in %	66 %	62 %	62 %	59 %	59 %	52 %	55 %	72 %

Tab. 3.24: Relative und absolute Fehlerhäufigkeiten der unterschiedlichen Klassen

Die relativen Häufigkeiten über alle Fehler hinweg zeigen große Übereinstimmungen zwischen den Lehrkräften. Allerdings sind, wie in Tab. 3.24 ersichtlich, für Lehrkraft 6 am wenigsten (52 %) und für Lehrkraft 7 am meisten (72 %) fehlerhafte Aufnahmen zu verzeichnen.

Im Folgenden wird die unterschiedliche Fehlerverteilung innerhalb ausgewählter Merkmale deskriptiv beschrieben. Diese weiteren Untersuchungen lassen Vermutungen darauf zu, welchen Einfluss die Lehrkraft durch beispielsweise explizites Unterrichten oder eigene Aussprache auf einzelne Merkmale haben könnte. Im Gegensatz zur hypothesengeleiteten Auswahl der Merkmale bei der Untersuchung zum Einfluss von niederdeutschem Sprachhintergrund (vgl. Kapitel 3.4.3.2) findet die Auswahl für die folgende Analyse anhand der divergierenden Häufigkeiten statt. So werden nur diejenigen Merkmale näher beschrieben, die zwischen den Klassen sehr unterschiedliche Fehlerwerte erzielen. Diejenigen sprachlichen Merkmale, die keine deutlichen Unterschiede in den relativen Fehlerhäufigkeiten zeigen, werden nicht näher dargestellt.¹⁹ Bei diesen handelt es sich um die folgenden:

- a. affriziertes *t* [tʰ];
- b. stimmloses *s* [s] wortinitial;
- c. vorderer ungerundeter Vokal [æ];
- d. hinterer ungerundeter Vokal [ɔ̃];
- e. hinterer gerundeter Vokal [ɔ̞];
- f. hinterer gerundeter Vokal [ɔ̠];
- g. Triphthonge.

¹⁹ Die Mittelwertunterschiede der nicht näher dargestellten segmentalen Merkmale lassen sich im Anhang B.3 nachvollziehen.

Das affrizierte *t* wird einheitlich in allen Lerngruppen falsch realisiert, während das stimmlose *s* wortinitial in allen Lerngruppen richtig realisiert wird. Die Triphthonge werden über alle Gruppen hinweg falsch gebildet. Der vordere ungerundete Vokal [æ] wird von allen Gruppen in mehr als der Hälfte der Aufnahmen fehlerhaft realisiert. Die Häufigkeitswerte liegen allerdings nah beieinander und unterscheiden sich nicht wesentlich. Die Prozessfehler überlange Vokale und Konsonanten in Folge von [ə]-Tilgungsprozessen werden ebenfalls nicht in die Untersuchung mit aufgenommen.

h. überlange Vokale und Konsonanten in Folge von [ə]-Tilgungsprozessen.

Zu den überlangen Vokalen in Folge von [ə]-Tilgungsprozessen werden alle Untersuchungen in betonten und unbetonten Silben gezählt, zu den Konsonanten alle Nasale und Liquide sowie das weiche *d*. Die Mittelwerte waren für alle Merkmale ungefähr gleich hoch.²⁰ In Tab. 3.25 lassen sich die untersuchten segmentalen Merkmale zusammengefasst nachvollziehen:

untersuchtes Merkmal	Lehrkräfte und relative Fehlerhäufigkeiten in %							
	LK1	LK2	LK3	LK4	LK5	LK6	LK7	LK8
stimmloses <i>s</i> [s] wortmedial	54	21	11	23	25	0	5	50
weiches <i>d</i> [ð] wortmedial	77	75	75	50	81	57	54	83
weiches <i>d</i> [ð] wortfinal	64	43	57	25	50	57	73	89
alveolopalataler Frikativ [ç]	100	100	98	100	100	95	97	93
vorderer ungerundeter Vokal [e]	100	93	89	96	100	64	100	94
vorderer ungerundeter Vokal [ɛ:]	62	61	43	38	44	43	45	100
zentrales <i>a</i> [ä], orthographisch leicht	4	11	18	12	6	14	5	17
zentraler <i>a</i> [ä], orthographisch problematisch	100	85	93	100	88	14	23	89
hinterer gerundeter Vokal [ɔ]	69	71	64	58	62	100	55	33
bekannte Polyphthonge [äy] und [äi]	15	12	7	5	8	0	0	37
bekannte Polyphthonge [ɛi]	8	36	50	38	50	57	18	33
Diphthonge auf [y]	90	69	60	58	38	38	58	100

Tab. 3.25: Mittelwerte Fehlerhäufigkeiten aller Klassen (segmentale Ebene)

²⁰ Die Mittelwerte der nicht näher untersuchten Merkmale bei den Prozessen finden sich ebenfalls im Anhang B.3.

Aus den deskriptiven Beschreibungen der Unterschiede zwischen den Fehlermittelwerten zu den einzelnen Merkmalen ergeben sich die folgenden Ergebnisse:

- a. *stimmloses s [s] wortmedial*: Das stimmlose *s* wird medial im Wort häufiger falsch produziert. Während in den Klassen von Lehrkraft 3 und Lehrkraft 7 das Phänomen von einer Überzahl der Schüler richtig realisiert wird, sind die Aufnahmen der Klassen von Lehrkraft 1 und 8 zu 50 % fehlerhaft. Die Schüler von Lehrkraft 6 produzieren den Laut ausnahmslos richtig. Die Klassen der Lehrkräfte 2, 4 und 5 zeigen in ca. einem Viertel der Aufnahmen Normabweichungen.
- b. *weiches d ([ð]) wortmedial*: In jeder Klasse ist mindestens die Hälfte der Aufnahmen fehlerhaft. Die Klasse von Lehrkraft 4 hat mit 50 % fehlerhafter Realisierungen den niedrigsten Schnitt. Die höchste Fehlerquote hat die Klasse von Lehrkraft 8 mit 83 %.
- c. *weiches d ([ð]) wortfinal*: Die arithmetischen Mittel der Fehlerhäufigkeiten bei wortfinalelem weichen *d* zeigen größere Unterschiedlichkeiten über die Klassen hinweg. Während in den Klassen von Lehrkraft 1 (64 %), 3 (57 %), 5 (50 %), 6 (57 %), 7 (73 %) und 8 (89 %) mindestens die Hälfte oder deutlich über die Hälfte der Aufnahmen fehlerhaft sind, liegt die Fehlerhäufigkeit in der Klasse von Lehrkraft 2 bei 43 % und in der Klasse von Lehrkraft 4 sogar nur bei 25 %.
- d. *vorderer ungerundeter Vokal [ɛ]*: Der vordere ungerundete Vokal [ɛ] wird in fast allen Gruppen mehrheitlich normabweichend realisiert, in allen Gruppen sind über 50 % der Aufnahmen fehlerhaft. Die Klasse von Lehrkraft 6 erzielt mit 64 % allerdings den niedrigsten Wert innerhalb der Gruppen. Die anderen Klassen liegen bei arithmetischen Mitteln zwischen 89 % und 100 %.
- e. *vorderer ungerundeter Vokal [ɛ:]*: Der vordere ungerundete Vokal [ɛ:] wird in einigen Klassen (3, 4, 5, 6, 7) in weniger als der Hälfte der Aufnahmen fehlerhaft realisiert, in anderen Klassen (1 und 2) in mehr als der Hälfte und in der Klasse von Lehrkraft 8 sogar in allen Aufnahmen fehlerhaft realisiert.
- f. *zentrales a ([ä]) orthographisch unproblematisch und problematisch*: Der zentrale Vokal [ä] wurde in Sätzen mit unterschiedlicher Schreibweise erhoben. In zwei Sätzen wird der Laut mit einer Schreibung erhoben, deren lautliche Umsetzung der deutschen sehr ähnelt. Lesen die Schüler das Wort also mit ihren Kenntnissen aus der L1 vor, realisieren sie es richtig. Normabweichungen können auftreten, wenn falsche Regeln zur lautlichen Umsetzung bestehen, z. B. wenn die Schüler gelernt haben, dass orthographisches ⟨a⟩ auf Dänisch anders auszusprechen ist. Die arithmetischen Mittelwerte der Fehlerhäufigkeiten liegen allerdings in allen Gruppen unter 20 %. In der Klasse 1, 5 und 7 sogar unter 10 %. Für denselben Laut, diesmal von Zielwörtern abgefragt, deren lautliche orthographische Umsetzung von der deutschen abweicht (*ret* [ʁäð], ‚recht‘, *træt* [tʰʁäð], ‚müde‘), ergeben sich sehr hohe Häufigkeitswerte in den meisten Gruppen bei über 80 %. Lediglich in den Klassen von Lehrkraft 6 (14 %) und 7 (23 %) wird der Fehler nicht häufig gemacht.

- g. *bekannte Polyphthonge* [ḁ̈], [ǟ] und [ä̈]: Für die in dieser Kategorie abgefragten Polyphthonge bestehen lautlich keine Probleme für deutsche Muttersprachler. Die hier abgefragten Lautverbindungen finden sich identisch ([ḁ̈] und [ǟ]) oder nahezu identisch ([ä̈] ~ [ɔ̈]) auch im Deutschen. Problematisch kann die lautliche Umsetzung der von der deutschen stark abweichenden Orthographie sein. Für [ḁ̈], das lautlich nicht problematisch, im abgefragten Zielwort allerdings orthographisch abweicht, stellen sich die Fehlerhäufigkeiten der Klassen ganz unterschiedlich dar. In der Klasse von LK 8 sind 56 % der Aufnahmen fehlerhaft, während in den Klassen von LK 7 und 6 keine Fehler gemacht werden. Die Fehlerhäufigkeiten von LK 1 bis 5 variieren zwischen 15 % und 38 %. Alle Aufnahmen zum Zielwort für [ǟ] mit nicht abweichender Orthographie sind normkonform, bis auf die Klasse von LK 8. Hier ergibt sich ein arithmetisches Mittel von 22 %. Die Mittelwerte bei [ä̈] reichen von 8 % (LK 1) bis 57 % (LK 6).
- h. *Diphthonge auf [y]*: Für die unbekanntes Diphthonge auf [y] ergeben sich unterschiedliche Werte in den arithmetischen Mitteln über die Klassen verteilt. Während in den Klassen von Lehrkraft 1 und 8 fast alle Aufnahmen von der Norm abweichen (90–100 %), liegen die fehlerhaften Aufnahmen in den Klassen 5 und 6 bei 38 %. Die Klassen von LK 2, 3, 4 und 7 erreichen Häufigkeiten zwischen 58 % und 69 %.

Die bei den Vokalqualitäten und Prozessen analysierten Merkmale werden in Tab. 3.26 dargestellt.

untersuchtes Merkmal	Lehrkräfte und relative Fehlerhäufigkeiten in %							
	LK1	LK2	LK3	LK4	LK5	LK6	LK7	LK8
[i]	69	86	79	85	88	100	73	89
[i:]	31	31	7	8	0	0	9	44
[y]	15	43	71	62	38	71	64	57
Lenisierung nachtoniger Plosive intervokalisch	NA	83	77	86	90	76	73	96
Lenisierung nachtoniger Plosive pränasal	69	100	79	97	95	100	94	100

Tab. 3.26: Mittelwerte Fehlerhäufigkeiten aller Klassen (Vokalquantität und Prozesse)

Die Mittelwerte lassen sich folgendermaßen beschreiben:

- a. *Vokalquantität* [i] und [i:]: Die Fehlerhäufigkeiten für den kurzen vorderen geschlossenen ungerundeten Vokal sind hoch, es gibt allerdings auch richtige Realisierungen, z. B. weisen Klasse 1 (69 %) und Klasse 7 (73 %) niedrigere Werte als die anderen auf (Werte zwischen 79 % und 100 %). Der lange Laut [i:] ist im Dä-

nischen und Deutschen gleich. Die Realisierung sollte für die Schüler keine Probleme darstellen. Trotzdem weisen die Fehlerhäufigkeiten Unterschiede in den arithmetischen Mitteln auf. Die Klassen von LK 1, 2 und 8 weisen Fehlerhäufigkeiten zwischen 31 % und 44 % auf, während in den Klassen von LK 3 und 7 die arithmetischen Mittel zwischen 7 % und 9 % liegen und in den Klassen der LK 5 und 6 gar keine Fehler gemacht werden.

- b. [y]: Die arithmetischen Mittel der Fehlerhäufigkeiten variieren zwischen selten (LK 1 15 %) bis relativ häufig (LK 3 und 6 71 %).
- c. *Lenisierung nachtoniger Plosive*: Insgesamt ergeben sich sowohl intervokalisches als auch pränasal erneut relativ hohe arithmetische Mittel zu den Fehlerhäufigkeiten. Intervokalisches liegen die Werte zwischen 73 % und 96 %, pränasal zwischen 69 % und 100 %. Intervokalisches erreicht die Klasse von LK 7 die niedrigsten Fehlerhäufigkeiten, die höchsten Werte werden für LK 8 verzeichnet.

Die Ergebnisse geben Grund zur Annahme, dass in einem Untersuchungssetting, das die weiteren Parameter besser kontrolliert, die eingangs aufgestellte Nullhypothese $H_{4.0}$, vermutlich verworfen werden könnte. So regt die nähere deskriptive Darstellung einzelner begründet ausgewählter Merkmale zu Vermutungen an, dass explizites Fokussieren im Unterricht sich positiv auf die Fehlerhäufigkeit auswirkt. Besonders beim zentralen *a*-Laut [ä] in orthographisch problematischen Wörtern zeigen sich Unterschiede für die unterschiedlichen Klassen. Daneben können bei wortmedialem, stimmlosen [s], den vorderen ungerundeten Vokalen [ɛ] und [ɛ:] sowie der richtigen Umsetzung des schwierigen Schriftbildes bei bekannten Polyphthongen Unterschiede beobachtet werden. Bei den Vokalquantitäten zeigen sich ebenfalls Unterschiede.²¹

3.4.3.6 Zusammenfassung

Unterschiedliche Überlegungen geben Anlass zu weiteren Analysen von Fehlerhäufigkeiten. So wurde im Vorangegangenen getestet, ob die Häufigkeiten mit dem niederdeutschen Sprachhintergrund, dem Gebrauch des Dänischen auch außerhalb der Schule sowie mit geschlechts- oder gruppenspezifischen Merkmalen zusammenhängen.

Aus der Analyse zum niederdeutschen Sprachhintergrund kann nicht hervorgehen, dass die Sprecher statistisch signifikant besser einzustufen sind. Bei der Bewertung dieser Ergebnisse sollte allerdings bedacht werden, dass die Probandengruppe nicht ausreichend kontrolliert werden konnte. So beruhen die Einstufungen zur niederdeutschen Sprachkompetenz auf den Selbsteinschätzungen, die nicht überprüft wurden. Darüber hinaus gaben einige der Probanden zwar an, häufigen Kontakt zum Niederdeutschen zu haben, ob dieser Kontakt allerdings hauptsächlich auf Perzeption und Verstehen beruht oder auch produktive Sprachkompetenz mit einschließt, ist nicht in allen Fällen bekannt.

²¹ Es ist davon auszugehen, dass die Lehrkräfte die Phänomene im Unterricht unterschiedlich fokussiert haben. Diese Annahme müsste allerdings durch ein angepasstes Setting geprüft werden.

Zur genaueren und aussagekräftigeren Analyse der Hypothese sollte eine erneute Untersuchung unter Einbeziehung der genannten Kritikpunkte erfolgen.

Für die Schüler, die auch außerhalb der Schule Dänisch sprechen, lässt sich die Hypothese $H_{2,1}$ ebenfalls nicht für die Gesamtzahl der Fehler bestätigen. Weder sind die Unterschiede im Allgemeinen statistisch signifikant, noch können diese für genauer untersuchte Phänomene bestätigt werden. Lediglich bei den unbekanntem Diphthongen auf [ɥ] scheinen die Schüler, die auch außerhalb der Schule Dänisch sprechen, einen Vorteil zu haben. Dies wird durch eine Tendenz zur Signifikanz für Unterschiede bei bekanntem [œ] bestätigt. Allerdings muss bei der Interpretation der Ergebnisse bedacht werden, dass die Anzahl bzw. die Qualität der außerschulischen Begegnungskontexte, die die Schüler angeben, nicht erhoben ist. Dadurch ist es nicht möglich zu sagen, welcher Art die außerschulischen Begegnungen sind und ob diese sich mit Kontexten vergleichen lassen, die einen tatsächlichen Effekt auf die Aussprache bzw. Fehlerhäufigkeit hätten. Auch um für diese Hypothese aussagekräftigere Ergebnisse zu erhalten, sollte die Untersuchung in einem weiteren Kontext noch einmal aufgegriffen werden. Insgesamt sind die untersuchten Schülerzahlen (6 Schüler, die angeben auch außerhalb der Schule Dänisch zu sprechen) als zu niedrig einzuschätzen, um allgemeingültige Ergebnisse zu postulieren. Es kann allerdings festgehalten werden, dass eine Tendenz feststellbar ist, dass Lerner, die auch außerhalb der Schule Dänisch sprechen, zumindest bei einigen problematischen Segmenten Vorteile gegenüber den anderen Lerner haben.

Geschlechtsspezifische Unterschiede zeigen sich im Chi-Quadrat-Test als statistisch signifikant. Die Ergebnisse zeigen allerdings eine schwache Effektstärke. Der tatsächliche Effekt, den das Geschlecht auf die Fehlerhäufigkeit hat, bestätigt damit zwar Ergebnisse früherer Forschung (Zaubauer, Retelsdorf & Möller 2009), ist hier allerdings aber als gering einzustufen und kann vernachlässigt werden.

Die Unterschiede zwischen den Klassen geben aus der deskriptiven Beschreibung einiger ausgewählte Phänomene Grund zur Annahme, dass die explizite Förderung von Aussprache eine Rolle bei der Fehlerhäufigkeit spielen könnte. So zeigen sich große Unterschiede beim medialen stimmlosen [s], dem vorderen ungerundeten Vokal [ɛ:], dem vorderen ungerundeten Vokal [e], dem zentralen *a*-Laut ([ä]) in orthographisch problematischen Wörtern sowie bei den bekannten Polyphthongen. Auffällig ist, dass einige Klassen häufig hohe bzw. niedrige Fehlerzahlen erzielen. Zum Beispiel werden für die Klasse von LK 8 häufig hohe Fehlerzahlen ermittelt, während für die Klasse von LK 7 häufig niedrige Fehlerzahlen auftreten. Bei den bekannten Polyphthongen [œ] wird deutlich, dass nur Fehler bei LK 1 und LK 8 auftreten. Wieder lässt sich durch dieses Phänomen vermuten, dass die Lehrkraft einen Einfluss auf die Aussprache der Schüler hat, denn um den Polyphthong an dieser Stelle richtig realisieren zu können, ist Wissen bzw. Übung notwendig. Derselbe Umstand zeigt sich bei den Diphthongen auf [ɥ]. Die hohen Fehlerhäufigkeiten in den Klassen 1 und 8 deutet darauf hin, dass diese Klassen sich noch nicht mit den Polyphthongen und ihrer orthographischen Umsetzung auseinandergesetzt haben.

Diese Beobachtungen geben Grund zur Annahme, dass einigen Lerngruppen die Wörter bekannt sein müssen bzw. das Phänomen mit ihnen geübt wurde. Dies würde darauf hindeuten, dass die Lehrkraft einen Effekt auf die Aussprache der Schüler hat. Dies kann anhand des gegebenen Untersuchungssetting nicht belegt werden und so sollten sich weitere Untersuchungen dieser anschließen.

3.4.4 Diskussion der Ergebnisse

Für die Fehleranalyse in Kapitel 3.4.2 lässt sich zusammenfassen, dass einige Fehler deutlich häufiger bzw. seltener vorkommen als andere. Zum Beispiel zeigt sich, dass die Phänomene der Prozesse, wie zum Beispiel die Längung von Konsonanten oder Vokalen nach Schwa-Reduktion deutlich häufiger vorkommen als die falsche Artikulation von weichem *d*. Auch innerhalb der Phänomengruppen lassen sich Unterschiede festhalten. So bereitete stimmloses [s] in wortinitialen Kontexten kein Problem, wohingegen es wortmedial in mehreren Fällen auftritt. Womit die unterschiedliche Häufigkeit zusammenhängt, ist unsicher. Eine Vermutung ist, dass einige Phänomene im Unterricht häufiger behandelt bzw. geübt werden als andere (z. B. weiches *d*) und deswegen nicht so häufig vorkommen wie erwartet bzw. weniger häufig als andere. Entsprechend kann davon ausgegangen werden, dass einige Phänomene eher vernachlässigt werden (z. B. affriziertes [tʰ]), und wieder andere trotz Bewusstmachung und Übung problematisch sind, sei es aus artikulatorischen Gründen (z. B. [ç]) oder aus Gründen der orthographischen Interferenz (z. B. Triphthonge). Einige Phänomene treten sicherlich auch auf, weil unterrichtlicher Kontext besteht und möglichst korrekt bzw. schriftnah gesprochen werden soll, um morphematische Endungen überprüfen zu können (z. B. Schwa-Assimilationsprozesse). Weiterführende Forschung müsste an dieser Stelle ansetzen und untersuchen, ob Bewusstmachung und Übung von Phänomenen zu einer Verbesserung der Phänomene führt.

Die auftretenden Fehler wurden anschließend an die deskriptive Häufigkeitsuntersuchung mit Hilfe eines Chi-Quadrat-Tests auf Gleichverteilung untersucht. Dazu wurden die mit erhobenen außersprachlichen Metadaten herangezogen. Aus den Zusammenhangberechnungen zwischen den sprecherbiographischen Metadaten der Probanden und der Fehler ergibt sich, dass das Geschlecht sich über alle Fehler hinweg als statistisch signifikant auf die Unterschiedlichkeit auswirkt. Obwohl die geschlechtsspezifischen Unterschiede statistische Signifikanz aufweisen, sind diese Effekte jedoch als minimal zu werten.

Die deskriptiv ermittelten Unterschiede zwischen den Klassen wurden in einzelnen ausgewählten Merkmalen näher betrachtet. Die Verknüpfung dieser deskriptiven Beschreibung mit den Ergebnissen der quantitativen Fehleranalyse aus Kapitel 3.4.2 lassen Vermutungen darüber zu, dass der Faktor Lehrkraft einen Effekt auf die Fehlerhäufigkeit hat. So zeigt sich, dass anscheinend alle Lerner ähnliche bzw. dieselben Fehler produzieren (vermutlich ausgelöst durch denselben L1-Hintergrund, orthographische Interferenz,

etc.), einige Lerner allerdings signifikant besser in einigen Phänomenen abschneiden. Es wird vermutet, dass diese explizit geübt bzw. bewusstgemacht wurden. Da die Lehrkraft dafür verantwortlich ist, ob ein Phänomen geübt wurde oder nicht, besteht Grund zur Annahme, dass die Lehrkraft mit der Unterschiedlichkeit in der Fehlerhäufigkeit in Verbindung gebracht werden kann. Das vorliegende Untersuchungssetting lässt allerdings keine generalisierende Aussage zu, da erstens die Anzahl der untersuchten Lehrkräfte als Stichprobe substanzuell zu klein ist und zweitens die weiteren möglichen verantwortlichen Parameter (z. B. Unterschiedlichkeit der Lerngruppe) nicht kontrolliert wurden und somit nicht vergleichbar sind.

Ein weiterer interessanter Analyseaspekt wäre, ob die Aussprachekompetenzen der Lehrkräfte für die richtige bzw. falsche Aussprachekompetenz ihrer Schüler verantwortlich ist. Die erhobenen Daten würden es erlauben, die quantitativen Ergebnisse zur Heterogenität der Lehrkräfte innerhalb der Fehler ihrer Schüler qualitativ abzugleichen. So wurden Lehrerinterviews auf Dänisch geführt, anhand derer die von den jeweiligen Schülern der Lehrkraft gemachten Fehlern mit denen der Lehrkraft abgeglichen werden könnten. Eine solche Untersuchung würde es ermöglichen, den Zusammenhang zwischen auftretenden Schülerfehlern und Lehrkraft auch qualitativ zu stützen. Im Rahmen der vorliegenden Arbeit würde diese Untersuchung zu weit führen, sie sollte allerdings für Folgeuntersuchungen bedacht werden.

Die Untersuchung der weiteren sprecherbiographischen Merkmale – niederdeutscher Sprachhintergrund und Verwendung des Dänischen auch außerhalb der Schule – konnten keinen signifikanten Unterschied verzeichnen. Auch für diese Gruppen lässt sich allerdings festhalten, dass die untersuchte Stichprobe nicht groß genug für generalisierende Aussagen ist. Außerdem wurden die Selbstangaben der Sprecher nicht kontrolliert. In diesem Bereich sollten unbedingt weitere Untersuchungen erfolgen, die die in dieser Studie vorhandenen Restriktionen beachtet.

Die entscheidende Erkenntnis aus den Ergebnissen der Fehleranalyse, dass quantitative Aussagen dazu möglich sind, welche Fehler häufig und welche eher selten vorkommen. Dadurch ist allerdings noch keine Aussage über die Fehlerschwere möglich, d. h. ob es sich bei den jeweilig eher häufig oder eher selten vorkommenden Fehlern um besonders relevante oder eher wenig relevante Fehler handelt. Relevant im Kontext Aussprachefehler bedeutet, inwiefern sich die Fehler auf die funktionale Dänischsprachigkeit, d. h. auf die Kommunikation mit Muttersprachlern auswirkt. Im folgenden Untersuchungsschritt in Kapitel 4 soll dieser Zusammenhang zunächst ausführlich geklärt und anschließend ermittelt werden, auf welche der Fehler dies zutrifft.

4 Studie 2: Bewertung der L2-Fehler deutschsprachiger Lerner durch dänische L1-Sprecher

Nachdem in der vorangegangenen Studie geklärt wurde, welche Probleme bei deutschen L2-Dänischlernern überhaupt bestehen, wird in der nachfolgenden Studie der Frage nachgegangen, welche dieser Fehler für die Kommunikation mit L1-Sprechern relevant sind. Dies wird anhand einer Rating-Studie untersucht, in der Muttersprachler L2-Äußerungen mit isolierten Fehlern bewerten. Im Folgenden wird zunächst der Hintergrund von L1-Ratingstudien (Kapitel 4.1) dargestellt, indem auf den Kontext verwiesen wird, in welchem diese Studien situiert sind sowie die üblichen Methoden und Herangehensweisen erläutert werden. Anschließend erfolgt in Kapitel 4.2 die Darstellung der empirischen Studie.

4.1 L2-Aussprachebewertung durch L1-Sprecher

Fremdsprachenerwerb steht immer innerhalb eines kommunikativen Kontexts und die erfolgreiche Kommunikation in der jeweiligen Sprache stellt das Ziel fremdsprachlichen Unterrichts dar. Erfolgreiche Kommunikation ist dabei von verschiedenen Faktoren abhängig, die mit der Wahrnehmung von Hörern zusammenhängen. Innerhalb der Sprechwirkungsforschung werden die Einflüsse verschiedener Faktoren auf die Wirkung von Äußerungen untersucht (Bose et al. 2016: 1–5). Neben Aspekten wie der der Stimme oder dem Geschlecht spielt besonders die Aussprache eine entscheidende Rolle. In Kommunikationssituationen senden Sprecher eine Äußerung mit einer intendierten Wirkung, die in Form von sprechsprachlichen Zeichen beim Hörer ankommt und nun zunächst innerlich aufgenommen und anschließend identifiziert werden muss. Bei diesem Vorgang werden „die für die Kommunikation unabdingbaren Merkmale erkannt und mit bereits gespeicherten Mustern verglichen“ (Nossok 2009: 35). Anschließend werden Bedeutungen und Funktion der Äußerung gedeutet, es entstehen „hörerinterne Informationen“, die zu einer Reaktion führen (Nossok 2009: 35). Während das sinnliche Wahrnehmen als Perzeption bezeichnet wird, stellt der Begriff der Rezeption die Aufnahme und Verarbeitung der sprechsprachlichen Merkmale dar. Die Perzeption steht vor der Rezeption und beeinflusst diese insofern, als schon hier verschiedene unbewusste Bewertungsprozesse ablaufen. Bereits die Wahrnehmung beeinflusst also die kommunikative Reaktion von Hörern und somit die gesamte Kommunikationssituation, wobei Aussprache zwar lediglich einen von vielen, jedoch wesentlichen Faktor darstellt. Die Wirkung von Aussprache im fremdsprachlichen Kontext steht auch in internationaler fremdsprachendidaktischer Forschung im Fokus des Interesses. Die dort angewandten Konzepte zur Messung der Wirkung sprachlicher Strukturen auf Hörerreaktionen sind Mittelpunkt des vorliegenden Kapitels.

4.1.1 Forschung zur Fehlerschwere bis Mitte der 1990er Jahre

Der wesentliche Interessenschwerpunkt liegt auf den im L2-Erwerb auftretenden Fehlern und deren kommunikativer Wirkung. Die Prämisse dabei ist, dass sich einige Fehler schwerwiegender auswirken als andere (Derwing & Munro 2015: 72). Bereits Mitte der 1970er Jahre entwickeln sich zu dem Forschungsfeld der *Fehlerschwere* (*error gravity*) die ersten Studien (Johansson 1973, 1975, 1978; Dretzke 1985; Khalil 1985; Norell 1991; Koster & Koet 1993), die sich mit unterschiedlichen Aspekten beschäftigen. Während einige Studien die Unterschiedlichkeit sprachlicher Kategorien (grammatische vs. phonetische) untersuchen (Politzer 1978), konzentrieren sich andere ganz auf den Bereich der Phonetik und Phonologie und vergleichen beispielsweise segmentale und suprasegmentale Aussprachefehler miteinander (Anderson-Hsieh, Johnson & Koehler 1992; Johansson 1978). Innerhalb dieser Studien fällt die Beurteilung der Aussprache durch die Bewertung von muttersprachlichen Laien aus, den sogenannten L1-Ratern.²² Von Interesse ist dabei nicht der aktuelle Sprachstand von Lernern zur Bestimmung eines Kompetenzniveaus, sondern die Wirkung einzelner sprachlicher Merkmale auf die Bewertung von Muttersprachlern. Insgesamt gestaltet sich der Forschungsbereich und die angewandten Methoden als heterogen, eine einheitliche Vorgehensweise ist nicht erkennbar. Mitte der 1970er Jahre entwirft Johansson (1975: 25-29) ein Modell zur Messung der Fehlerschwere, das neben der Verständlichkeit auch die *conformity* (‘Übereinstimmung’) misst. *Conformity* stellt für Johansson (1975) ein Maß dafür dar, inwieweit ein Sprecher der Norm einer Sprechergemeinschaft entspricht. Diese Studie ist im Hinblick auf spätere Forschung und besonders mit dem Fokus der vorliegenden Arbeit als innovativ zu bewerten, da sie – wie für aktuelle Studien überwiegend zutreffend – nicht allein Verständlichkeit als wichtigstes Unterrichtsziel voraussetzt, sondern eine Variable, die scheinbar über reine Verständlichkeit hinausgeht und erheblichen Einfluss auf die Bewertung der Muttersprachler hat. So postuliert Johansson bezogen auf das Ziel fremdsprachlichen Unterrichtens, dass Lerner nicht nur von L1-Sprechern verstanden, sondern auch als Kommunikationspartner akzeptiert werden sollten. Das von Johansson entworfene Modell zur Fehlerevaluation zielt darauf ab, Fehler zu hierarchisieren bzw. nach ihrem Schweregrad zu ordnen. Dazu wird anhand einer schrittweisen Abfrage ein *score* ‚Punktwert‘ für jeden überprüften Fehler ermittelt.

²² Diese Form der Untersuchung ist nicht mit der professionellen Bewertung von Aussprache zu verwechseln, bei der standardisierte Sprachtests eingesetzt werden oder L2-Sprache von professionellen Prüfern evaluiert und getestet wird.

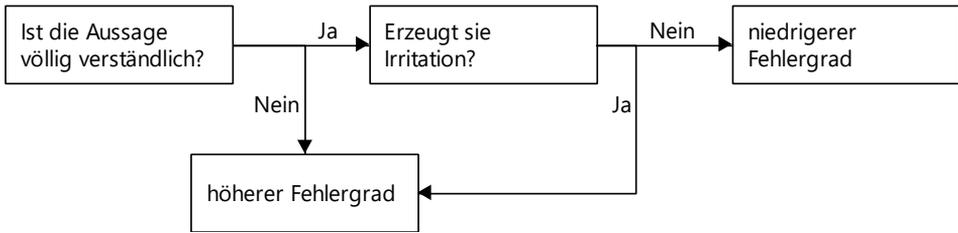


Abb. 4.1: Modell zur Fehlerevaluation
(aus dem Englischen nach Johansson 1978: 5)

Im ersten Schritt der Fehlerevaluation wird untersucht, ob ein Fehler eine Äußerung für L₁-Rater unverständlich macht. Wenn der Fehler die Äußerung unverständlich macht, handelt es sich um einen schwerwiegenden Fehler. Ist eine Äußerung trotz Fehler noch verständlich, wird der nächste für die Kommunikation ausschlaggebende Aspekt überprüft, nämlich ob der Fehler eine Form der *irritation* beim Hörer erzeugt. Falls dies zutrifft, handelt es sich ebenfalls um einen schwerwiegenden Fehler. Falls keine Irritation auftritt und die Aussage voll verständlich ist, kann von einem schwachen Fehler gesprochen werden. Unter Hinzunahme weiterer überprüfter Aspekte, wie beispielsweise der auftretenden Häufigkeit des Fehlers, ermittelt sich nach diesem Schema der endgültige Punktwert. Johansson setzt in seinen Messungen Verständlichkeit als Voraussetzung, setzt allerdings die Akzeptanz gleichwertig mit dieser im Hinblick auf die Einteilung der Fehlerschwere. Dieses Vorgehen ist insofern innovativ, als die zeitgenössischen Messungen der Auswirkung von Fehlern auf die Kommunikation mit L₁-Sprechern überwiegend eindimensional über die Verständlichkeit erfolgt und Akzeptanz als Dimension von Aussprache überwiegend unbeachtet bleibt (Thomson 2018).

Neben dieser relativ modernen Auffassung von Wirkung von Aussprache von Johansson, besteht zu dieser Zeit allerdings überwiegend grundlegend Unklarheit und Uneinigkeit im Forschungsbereich. Je nach Studie werden die Dimensionen zur Bewertung von Aussprache unterschiedlich benannt und definiert. Häufig verwendete Termini sind *comprehensibility* und *intelligibility* (beides wörtlich ‚Verständlichkeit‘, zur näheren Begriffsdefinition vgl. Kapitel 4.1.2.2). Aber auch *accent* wird teilweise synonym gebraucht oder auch häufig als dasselbe Phänomen betrachtet bzw. nicht näher definiert. Dies führt zu terminologischen Verwirrungen und der Gleichsetzung unterschiedlicher Aspekte von Aussprache. Ergebnisse unterschiedlicher Studien lassen sich weder vergleichen noch als allgemeingültig einstufen (Rifkin & Roberts 1992: 522). Insgesamt herrscht bis Mitte der 1990er Jahre Uneinigkeit in der Bezeichnung der Konzepte zur Beschreibung und Messung der Wirkung von L₂-Aussprache auf die Wahrnehmung der Hörer innerhalb der Forschung der damaligen Zeit (Derwing & Munro 2015: 72–73). Forscher wie Johansson (1978: 34–49) und Khalil (1985: 345) zeigen sich als Vordenker, indem sie zwischen subjektiv wahrgenommener und objektiv messbarer Verständlichkeit unterscheiden. So misst Johansson (1978) Verständlichkeit auf zwei verschiedene Arten, erstens als

subjektives Urteil von L1-Ratern von empfundener Verständlichkeit von Sprachproben und zweitens in objektiven Verfahren, in denen Probanden sprachliche Äußerungen im Wortlaut exakt wiederholen müssen. Er führt außerdem kritisch an, dass bei einem Mangel von hinreichender Konzeptdefinition nicht hinreichend bestimmt werden könne, ob die Effekte als Verständlichkeit oder Irritation bzw. Akzeptabilität (*acceptability*) zu bezeichnen seien (Johansson 1978: 47-48). Khalil (1985: 345) problematisiert ebenfalls den vorherrschenden Mangel an Differenziertheit zwischen objektiver und subjektiver Messung von Verständlichkeit. Seine Forschung zeigt, dass die subjektive Empfindung eine Äußerung verstanden zu haben, nicht unbedingt mit der objektiv messbaren Verständlichkeit zusammenhängt. Ähnliche Erkenntnisse sind später grundlegend für die Unterscheidungen der Dimensionen *comprehensibility* und *intelligibility* bei Munro & Derwing (1995a).

Dieser kurze Einblick in die Forschung bis Mitte der 1990er Jahre verdeutlicht, dass obgleich bereits wichtige Grundlagen und -fragen diskutiert werden, bis dahin Unklarheit in der Definition der Bewertungsdimensionen besteht. Dies erschwert bzw. macht den Vergleich von Studien untereinander kaum möglich. Mitte der 1990er Jahre fordern Rifkin & Roberts (1995) aus diesem Grund eine Aufarbeitung der grundlegenden Konzepte zur Messung der Wirkung von L2-Aussprache. Dieser Forderung kommen Munro & Derwing (1995a) für die Konzepte Verständlichkeit und Akzent nach. Sie weisen in einer von ihnen durchgeführten Studie statistisch nach, dass es sich bei den drei Konzepten (*comprehensibility*, *intelligibility* und *accent*) um drei unterschiedliche Dimensionen handelt. Dieser Aufsatz prägt den Forschungsbereich entscheidend. Nachfolgende Forschung beruft sich auf die von Munro & Derwing (1995a) durchgeführte Studie und ihre Konzepte, wodurch es zu einer Vereinheitlichung im Gebrauch von Termini und Messmethoden kommt. Die Aufarbeitung der Termini Irritation bzw. Akzeptabilität erfolgt erst später und wird noch in aktueller Forschung diskutiert (z. B. Beinhoff 2014).

Ein Aspekt, der durch Munros und Derwings Arbeit nicht zur endgültigen Einigkeit gebracht wurde, ist die übergreifende Bezeichnung der Dimensionen zur Bewertung von L2-Aussprache. Diese variiert noch immer in den Studien. In einigen Studien werden *comprehensibility*, *intelligibility* und *accent* unter dem Überbegriff *aspects* (Derwing & Munro 2005: 385) oder *dimensions* (Munro, Derwing & Morton 2006: 112) geführt, in anderen Studien als „factors which affect native speaker perception“ (Matsuura, Chiba & Fujieda 1999: 49). Pilott (2016: 36) führt die Konstrukte unter dem Begriff *global measure*, ein Begriff, der darstellt, dass es sich bei *comprehensibility*, *intelligibility* und *accent/accentedness* um Konstrukte handelt, die anhand einer Äußerung die gesamte Aussprache eines Sprechers messen. In der vorliegenden Arbeit wird als Überbegriff *Bewertungsdimensionen von Aussprache* festgelegt. Dieser Überbegriff erklärt sich dadurch, dass es sich bei den zu untersuchenden Konstrukten um Dimensionen von Aussprache handelt, die durch Evaluation von L1-Ratern gemessen werden können.

Im Folgenden werden die von Munro & Derwing (1995a) aufgearbeiteten Bewertungsdimensionen vorgestellt (Kapitel 4.1.2), da sie grundlegend für die Operationalisie-

rung der vorliegenden Studie sind. Die daran anschließende Weiterentwicklung des Forschungsbereichs um die Bewertungsdimension der Akzeptabilität bis hin zur Darstellung aktueller Studien erfolgt in Kapitel 4.1.3. In der abschließenden Zusammenfassung (Kapitel 4.1.4) werden die Arbeitsdefinitionen für die vorliegende Arbeit festgelegt, bevor sich das daran anschließende Kapitel 4.2 der empirischen Studie widmet.

4.1.2 Differenzierte Termini zur Bewertung ab 1995

Munro & Derwing (1995a) kommen Mitte der 1990er Jahre Rifkin & Roberts (1995) Forderung einer einheitlichen Operationalisierung der Bewertungsdimensionen nach und differenzieren in der Folge zwischen verwandten und trotzdem voneinander unabhängigen Dimensionen von Aussprache. Sie unterscheiden zwischen Akzent und Verständlichkeit und operationalisieren die Dimensionen jeweils in einer objektiv und subjektiv messbaren Dimension. Munro & Derwing (1995a) unterscheiden:

- a. subjektiv wahrgenommenen Akzent (*accentedness*);
- b. wahrgenommene Verständlichkeit (*comprehensibility*);
- c. objektiv messbare Verständlichkeit (*intelligibility*).

Diese Operationalisierung etabliert sich für die fremdsprachliche Ausspracheforschung und hat einen immensen Einfluss auf den gesamten Forschungsbereich (Thomson 2018: 11).²³ Neben der einheitlichen Differenzierung und Definition der Bewertungsdimensionen weisen Munro & Derwing (1995a) außerdem in ihren Studien nach, dass die Dimensionen partiell unabhängig voneinander sind. Ihre Ergebnisse zeigen, dass ein starker Akzent nicht notwendigerweise zur Unverständlichkeit führen muss. Während Fehler und wahrgenommener Akzent stark miteinander korrelierten, korrelierte die wahrgenommene Verständlichkeit (*comprehensibility*) weniger mit dem wahrgenommenen Akzent. Noch weniger Korrelation wurde für objektive Verständlichkeit (*intelligibility*) und Akzent nachgewiesen. In Anlehnung an Munro & Derwing (1995a) werden im Forschungsbereich noch weitere Dimensionen von Aussprache aufgearbeitet. Besonders die Kategorie der *acceptability* ‚Akzeptabilität‘ rückt weiter in den Fokus (Sewell 2012; Pilott 2016). Der vereinheitlichende Effekt, den die Operationalisierung von Munro & Derwing (1995a) auf die anderen Dimensionen hatte, fehlt bei der Akzeptabilität allerdings. So tritt das Konzept unter unterschiedlichen Termini innerhalb der Forschung auf, beispielsweise als soziale Akzeptanz bei Settineri (2011) oder *acceptability* bei Thomson (2018: 25), im Gegensatz zu *pleasantness* bei Lee, Kim & Park (2019).

Für die vorliegende Arbeit sind sowohl die Dimensionen nach Munro & Derwing (1995a) als auch die Akzeptabilität relevant. In den Kapiteln 4.1.2.1 und 4.1.2.2 werden die

²³ In deutschen Studien im Forschungsbereich werden die Begriffe teilweise übersetzt (z. B. bei Settineri 2011). In der vorliegenden Arbeit werden die deutschen Begriffe als Überbegriffe genutzt und die englischen Begriffe zur Bezeichnung des jeweiligen konkreten Begriffs, wie er von Munro & Derwing (1995a) operationalisiert wurde, verwendet. Zur Übersicht vgl. Tab. 4.1.

von Munro und Derwing geprägten Bewertungsdimensionen im Forschungskontext vorgestellt. Kapitel 4.1.3 behandelt die Bewertungsdimension der Akzeptabilität.

4.1.2.1 Akzent

Derwing & Munro (2015: 5) definieren Akzent als „particular pattern of pronunciation that is perceived to distinguish members of different speech communities“. Akzent wird über die Wahrnehmung der Hörer bestimmt bzw. auf welche Art und Weise sie empfinden, dass eine Äußerung von einer lokalen Varietät abweicht, „as how different a pattern of speech sounds compared to the local variety“ (Derwing & Munro 2009: 478). Anders als in phonetisch ausgerichteten Studien operationalisieren Munro und Derwing Akzent also nicht anhand von objektiven Messverfahren und über akustisch gemessene phonetische Abweichungen, sondern in subjektiven Bewertungsverfahren. Dazu nutzen sie 9-stufige Ratingskalen von 1 (kein Akzent) bis 9 (sehr starker Akzent) (Munro & Derwing 1995a: 79, 1995b: 294; 2008: 479; Derwing & Munro 1997: 5). Synonym zu dem von ihnen gewählten Terminus *accent* benennen sie die von ihnen gemessene Dimension auch als *accentedness* (z. B. Derwing & Munro 2015: 5). Der auf diese Weise gemessene Akzent ist von einer objektiv messbaren Größe zu unterscheiden, dem akustisch messbaren Unterschied zwischen Norm und Lernaltersprache. Munro und Derwing nehmen diese Unterscheidung zwar wahr, setzen diese allerdings nicht terminologisch um und untersuchen in ihren Studien auch nur die subjektive Dimension.

In Studien, die sich an die Forschung von Munro und Derwing anlehnen und Akzent untersuchen, können unterschiedliche Ratergruppen eingesetzt werden. So wird zwischen erfahrenen und unerfahrenen (oder naiven) Ratern unterschieden, d. h. zwischen Ratern, die keine Ausbildung oder Erfahrung im Bereich von Fremdsprachenunterricht oder L2-Bewertung haben (naive Rater) und solchen, die in diesem Bereich eine Ausbildung erfahren haben, beispielsweise Sprachlehrer oder -tester (erfahrene Rater). Einige Studien sind besonders an dem Unterschied zwischen diesen beiden Ratergruppen interessiert und vergleichen eventuelle Unterschiede im Bewertungsverhalten der beiden Ratergruppen. Beispielsweise untersuchen Kennedy & Trofimovich (2008), wie sich Hörererfahrung und semantischer Kontext auf die Messung der Dimensionen von Aussprache auswirken. In ihrer Studie bewerteten 24 englische L1-Rater, von denen die eine Hälfte erfahrene, die andere Hälfte naive Hörer sind, L2-Sprachaufnahmen von chinesischen Sprechern. Die Skala zur Messung von *accentedness* reicht von 1 (kein nichtmuttersprachlicher Akzent) bis 9 (starker nichtmuttersprachlicher Akzent). Da es in der vorliegenden Studie allerdings um die Bewertung durch naive L1-Rater geht, werden im Folgenden vornehmlich solche Studien vorgestellt, die dieselbe Ratergruppe untersuchen.

Kraut & Wulff (2013) untersuchen *foreign-accented speech* anhand vorgelesener Sprache von 24 internationalen Studenten, die von 87 englischen Muttersprachlern bewertet werden. Die L1-Rater bewerteten *accent*, *comprehensibility* und *communicative ability* („Kommunikationsfähigkeit“). Als Kontrollvariablen erheben Kraut & Wulff (2013) au-

ßerdem das Geschlecht, das Sprachniveau und die L1 der Sprecher sowie die Bekanntheit der L1-Rater mit Ratingverfahren. Die Stärke des Akzents wird anhand der Frage *How strong was this speaker's accent?* auf einer Skala von 1 bis 7 erhoben (Kraut & Wulff 2013: 7).

Saito et al. (2016) setzen eine Skala von 1 (*heavily accented*) bis 1000 (*not accented at all*) mit einem Schieberegler ein, um *accentedness* zu messen. Die Skala ist somit in der Wahrnehmung der L1-Rater praktisch stufenlos. Bewertet werden die Sprachdaten 40 französischer Muttersprachler von 10 L1-Ratern auf *accentedness* und *comprehensibility*. Die L1-Rater können mit einem Schieberegler zwischen den beschrifteten Enden hin- und herfahren. Die jeweiligen Pole sind mit einem traurigen und einem fröhlichen Emoticon bebildert. Über dem Schieberegler finden sich die Definitionen der Bewertungsdimensionen (Saito et al. 2016: 11).

Der Einblick in die oben dargestellten Studien verdeutlicht, dass die Dimensionen von Munro & Derwing (1995a) bzw. Derwing & Munro (2015) in unterschiedlichsten Kontexten aufgenommen und untersucht werden, wobei die Fragestellungen, die Skalierungen der Bewertung oder auch die Aufgaben und Hinweise für die L1-Rater variieren. Die für die vorliegende Arbeit getroffenen Entscheidungen bezüglich der Definition und Operationalisierung von Akzent findet in Kapitel 4.2.2.3 statt.

4.1.2.2 Verständlichkeit

Munro & Derwing (1995a) unterteilen die Dimension Verständlichkeit in eine objektiv messbare (*intelligibility*) und eine subjektiv empfundene Größe (*comprehensibility*), die sich als tatsächlich wahrgenommener Inhalt und empfundene Schwierigkeit der Dekodierung einer Aussage unterscheiden (Munro & Derwing 2011): „We define COMPREHENSIBILITY as the listener's perception of how easy or difficult it is to understand a given speech sample. This dimension is a judgement of difficulty and not a measure of how much actually gets understood.“ Der deutschen Übersetzung ‚Verständlichkeit‘ entsprechen also im englischen Forschungsraum sprachlich als auch inhaltlich zwei voneinander getrennte Kategorien. Im Folgenden wird die deutsche Übersetzung (‚Verständlichkeit‘) als Überbegriff für die Termini *comprehensibility*, der subjektiv wahrgenommenen Verständlichkeit, und *intelligibility*, der objektiv messbaren Verständlichkeit, verstanden. Munro & Derwing (1995a) präsentieren in ihrer Studie Vorgehensweisen zur objektiven Messung von Verständlichkeit. So lässt sich die tatsächliche Verständlichkeit messen durch

- a. die exakte wörtliche Wiedergabe der Aussage;
- b. die Prozentzahl genannter Schlüsselwörter;
- c. die Fähigkeit zu paraphrasieren;
- d. die Bewertung über Likert-Skalen.

Munro & Derwing (1995a: 77) entscheiden sich selbst für exakte wörtliche Wiedergabe der Aussage. Die subjektiv wahrgenommene Verständlichkeit messen sie anhand einer

Likert-Skala. Die meisten Studien richten sich nach der grundlegenden Definition von Munro & Derwing (1995a), in der Wahl der Skalierung besteht allerdings Uneinheitlichkeit zwischen den Studien. Munro und Derwing nutzen in ihren unterschiedlichen Studien 9-stufige Ratingskalen, wobei die Beschriftungen der Skalen in den unterschiedlichen Studien leicht variieren. Munros & Derwings (1995a: 79) Skala greift von 1 = *extremely easy to understand* bis 9 = *impossible to understand*, während sie in einer späteren Studie (1995b: 293–294) eine Skala von 1 = *not difficult to understand at all* bis 9 = *very difficult to understand* einsetzen.

Auch Studien, die wahrgenommene Verständlichkeit (*comprehensibility*) im Sinne der Dimensionsdefinition von Munro & Derwing (1995a) messen, unterscheiden sich in den angewandten Skalenlängen. So gibt es Studien, die eine 9-stufige Skalierung einsetzen (Kennedy & Trofimovich 2008; Phan & Vo 2012) und andere, die 7-stufige nutzen (Matsuura, Chiba & Fujieda 1999; Kraut & Wulff 2013). Crowther et al. (2015) nutzen Schieberegler, um Verständlichkeit zu messen.

4.1.3 Bewertungsdimension Akzeptabilität

Neben Akzent und Verständlichkeit wird in der vorliegenden Studie Akzeptabilität (*acceptability*) als weitere Dimension von Aussprache untersucht. Die innerhalb der oben dargestellten Forschung immer wieder auftretenden Schwierigkeiten innerhalb der Definition und Operationalisierung von Eigenschaften und Effekten von L2-Aussprache auf Kommunikation mit Muttersprachlern lassen darauf schließen, dass die Dimensionen Verständlichkeit und Akzent dieses Vorhaben nicht vollständig abdecken. Akzeptabilität tritt immer wieder als weitere, vielleicht noch subjektivere Dimension als die bisher vorgestellten, auch in älteren Studien als Schlagwort auf. Thomson (2018: 25) fasst zusammen:

While highly subjective, this dimension of pronunciation is also very real. In a perfect world, comprehensibility and acceptability ought to be the same thing. That is, if speech is intelligible, it should be acceptable. In fact, this is not the case.

In aktuellen Publikationen tritt die Forderung zur Erforschung dieses Phänomens immer weiter in den Vordergrund, denn obwohl Akzeptabilität eine entscheidende Dimension von L2-Aussprache darstellt, ist sie innerhalb der Forschung längst nicht so ausgiebig erforscht wie die anderen Bewertungsdimensionen (Thomson 2018: 14). Darüber hinaus fehlt eine einheitliche Definition und Operationalisierung, wie sie Munro & Derwing (1995a) für die anderen Bewertungsdimensionen vorgenommen haben. Die fehlende Einigkeit führt dazu, dass der Terminus in Studien häufig unterschiedlich benannt und abgegrenzt wird. So wird Akzeptabilität im englischen Sprachraum auch als *pleasantness* (McBride 2015; Lee, Kim & Park 2019) oder als *degree of irritation* (Szpyra-Kozłowska 2015: 76) gemessen. Die fehlende Operationalisierung dieses Gefühls verhin-

dert den Vergleich zwischen den Ergebnissen unterschiedlicher Studien. Darüber hinaus wird der Zeitpunkt schwieriger zu bestimmen, zu dem das Konzept das erste Mal in der Forschung auftritt.

Bereits Mitte der 1960er Jahre messen Quirk & Svartvik (1966) Akzeptabilität. Allerdings prüft ihre Operationalisierung, ob L1-Rater grammatische Fehler erkennen oder nicht erkennen. Sie wenden zwei unterschiedliche Verfahren an, durch die sie Akzeptabilität objektiv und subjektiv messen wollen. Im objektiven Test hören die L1-Rater einen manipulierten, d. h. nachträglich verändert und mit einem fehlerhaften Element versehenen Satz, den sie wiedergeben müssen. Im subjektiven *judgement test* müssen die L1-Rater angeben, ob ein manipulierter Satz grammatisch oder ungrammatisch ist. Quirk & Svartvik (1966) können Korrelationen zwischen den beiden Testmethoden nachweisen und schlussfolgern daraus, dass Ergebnisse objektiver Messmethoden für Akzeptabilität von Sprachproben denen subjektiver Messungen entsprechen. Die von ihnen vorgenommene Operationalisierung von Akzeptabilität weicht stark von aktuellen Vorgehensweisen ab bzw. stimmt eher mit aktuellen Operationalisierungen von Verständlichkeit bzw. *intelligibility* überein. Die subjektive Operationalisierung ähnelt einer Bewertungsmethode aus dem *pronunciation assessment*, d. h. der Messung von korrekter Sprache durch geschulte L1-Rater, um ein Sprachniveau zu ermitteln. Johansson (1975: 26) misst etwas später, ob sprachliche Äußerungen von L1-Ratern als akzeptabel oder nicht wahrgenommen werden. Er misst dies über den Grad der Irritation (*degree of irritation*), welcher durch gelingende bzw. nicht gelingende Kommunikation mit L1-Sprechern sichtbar werde. Beeinflusst ein Fehler die Verständlichkeit oder die sogenannte *conformity*, werde die Kommunikation gestört (Johansson 1973: 102). Unter *conformity* versteht Johansson die Übereinstimmung einer Äußerung mit der Norm. Der Effekt von Fehlern auf letztere bei vollständiger Verständlichkeit ist die von Johansson (1973: 102) definierte Maßeinheit dafür, ob eine Äußerung *acceptable* oder nicht ist. Forschungsmethodisch orientiert sich Johansson an Quirk & Svartvik (1966). So misst er Akzeptabilität objektiv, indem er fehlerhafte Sätze von den Probanden in korrigierter Form wiederholen lässt. Je nachdem, ob und welche Änderungen die Probanden vornehmen, werden *judgement scores* vergeben, die untereinander verglichen werden können (Johansson 1975: 36). Bei der subjektiven Messung von Akzeptabilität lässt Johansson die L1-Rater Sprachproben auf einer 5-stufigen Skala bewerten, die an den Polen mit *the sentence is completely normal and idiomatic* und *the sentence is extremely unnatural and abnormal* beschriftet ist (Johansson 1975: 41). Johansson testet sowohl geschulte als auch ungeschulte L1-Rater. Johanssons objektive und subjektive Messungen erhalten relativ gleiche Ergebnisse. Die geschulten L1-Rater sind allerdings weniger konsistent als die ungeschulten. Johansson (1978: 22) bewertet dieses Ergebnis dahingehend, dass linguistisches Wissen sich eher negativ auf die Messung von Akzeptabilität auswirkt: „These results indicate that linguistic sophistication may be an obstacle rather than an advantage in judgements of acceptability.“ Er schlussfolgert, dass, wenn sich für objektive und subjektive Messungen sowieso gleiche Ergebnisse ergeben, die subjektive Messung von Akzeptabilität vorgezogen werden soll-

te, da sie einfacher durchzuführen sei (Johansson 1978: 21). In diesen frühen Studien zur Akzeptabilität von sowohl Quirk und Svartvik als auch Johansson ergeben sich zwei zentrale Probleme:

- a. Mit der objektiven Messung wird höchstens indirekt das von Johansson als Akzeptabilität benannte und als Irritation innerhalb einer Kommunikationssituation definierte Konstrukt gemessen. Da die L1-Rater den Auftrag erhalten, fehlerhafte Aussagen korrigieren zu müssen, wird allenfalls gemessen, wie gut sie eine Abweichung wahrnehmen. Somit wird eher sprachlich-strukturelles Wissen als eine soziophonetische Einstellung abgefragt. Während Johansson (1975: 163) seine Ergebnisse so interpretiert, dass sich linguistisch geschulte Probanden eher negativ für die Messung von Akzeptabilität darstellen, müsste die Interpretation eigentlich lauten, dass linguistisch geschulte L1-Rater fehlerhafte Sprachproben besser erkennen, als nicht geschulte L1-Rater.
- b. Der Vergleich der subjektiven und der objektiven Messung ist vermutlich nicht gerechtfertigt, da beide Messungen unterschiedliche Phänomene messen. Während in der objektiven Messung die sprachliche Korrektheit erhoben wird, steht im Fokus der subjektiven Messung von Empfindungen der L1-Rater, ob eine Sprachaufnahme von der Norm abweicht. Es ist möglich, dass auf diese Weise indirekt die Irritation bei der Kommunikation gemessen wird, vordergründig wird allerdings zunächst abgefragt, ob die Rater eine Abweichung wahrnehmen.

Diese Probleme verdeutlichen die essentielle Wichtigkeit der Passung zwischen Definition des geprüften Konstrukts (Akzeptabilität) und dessen Operationalisierung in Form von Items. Der Blick in aktuellere Forschung zeigt, dass auch über vierzig Jahre nach Quirk & Svartviks (1966) bzw. Johanssons (1973, 1975, 1978) Arbeiten noch immer definitorische und methodologische Schwierigkeiten bestehen. So bestehen teilweise in den Arbeitsdefinitionen der Studien noch immer Überschneidungen mit den Konzepten *comprehensibility*, *intelligibility* und *accent* oder Akzeptabilität wird gar nicht erst näher definiert und trotzdem als solche gemessen. Beinhoff (2014: 156) erklärt diesen Umstand damit, dass Akzeptabilität ein vermeintlich selbsterklärender Begriff zu sein scheine, dessen zu Grunde liegende Bedeutung auch in anderen alltäglichen Kontexten genutzt werde. Noch weniger als bei den anderen beiden Dimensionen scheint es sich bei der Akzeptabilität um eine sprachlich-strukturelle Dimension als um ein subjektives Gefühl zu handeln. Problematisch an der Uneinheitlichkeit innerhalb der Definition ist außerdem, dass sie zu inkonsistenter Interpretation von Ergebnissen führt.

Im Folgenden werden Studien der neueren Zeit herangezogen, um mehr Einblicke in das Konstrukt der Dimension zu erhalten und daraus Schlüsse zur Definition und Operationalisierung der vorliegenden Studie zu ziehen.

Tomokiyo, Black & Lenzo (2005: 1471) fragen Probanden nach der Akzeptabilität (*acceptability*) von akzentuierter Sprache, indem sie Sprachproben auf einer 5-stufigen Ratingskala von 1 = *bad* bis 5 = *excellent* beurteilen lassen. Dabei wird für die L1-Rater nicht näher definiert, in welchem Kontext die Sprecher einzustufen sind. Es ist somit

nicht eindeutig festzulegen, von welchen Normen die L1-Rater bei ihrer Bewertung ausgehen. Außerdem bleibt ungeklärt, welche sozialen oder kommunikativen Konsequenzen die Einstufung der Sprecher als gut oder schlecht hat.

Settinieri (2011: 67) beschreibt Akzeptabilität als „soziale Implikationen nonnormativer Aussprache“, die „einen zentralen Faktor für einen erfolgreich verlaufenden Ausspracheerwerb dar[stellen]“. Fehlerhafte phonetische Realisierung in der Lerner Sprache kann zu „Sanktionen seitens der Hörer“ führen (Settinieri 2011: 67). Settinieri untersucht die soziale Akzeptabilität unterschiedlicher phonetischer Abweichungen von der Zielnorm in der L2 Deutsch, indem sie die Phänomene von L1-Ratern bewerten lässt und ermittelt, ob es innerhalb der Bewertung signifikante Unterschiede gibt. Als Item erfragt sie ‚Wie bewerten Sie diesen Akzent?‘ und nutzt eine 7-stufige Ratingskala von 1 = *negativ* bis 7 = *positiv*. Durch Settinieris (2011) Studie lassen sich phonetische Abweichungen mit unterschiedlich positiver bzw. negativer Bewertung in Zusammenhang bringen. Settinieri (2011: 70) möchte auf diese Weise die Dimension der sozialen Akzeptanz von Aussprache messen. Sie gibt allerdings kritisch zu bedenken, dass ihr Item nicht eindeutig soziale Akzeptanz umfasst, sondern auch „eine Mischung mehrerer Beurteilungskriterien, z. B. auch Verständlichkeit und/oder Akzentuiertheit“ darstellen könnte (Settinieri 2011: 76). Settinieri definiert soziale Akzeptanz als zentralen Faktor für erfolgreiche Kommunikation. Durch die von ihr durchgeführte Messung bleibt allerdings offen, welche Folgen eine positive oder negative Bewertung für Kommunikation hätte.

In Chang Li-Anns (2008) Studie wird dieser Aspekt berücksichtigt, indem sie Folgen der Bewertung durch einen Kontext impliziert. L1-Rater aus Singapur, Malaysia, Indien, den Philippinen und den USA bewerten die Akzeptabilität unterschiedlicher englischer Akzente mit Hilfe einer 7-stufigen Skala von 1 = *I consider this sort of English unacceptable for international communication* bis 7 = *I consider this sort of English acceptable for international communication*. Aus Chang Li-Anns (2008) Studie ergibt sich, dass alle L1-Ratergruppen, bis auf die indischen, den amerikanischen Akzent favorisieren, gefolgt von ihrem jeweils eigenen. Die Ergebnisse zeigen, dass es zwischen L1-Ratern verschiedener Herkunftssprachen zwar eine favorisierte Norm gibt, Akzente aber auch unterschiedlich bewertet werden.

Sewell (2012) untersucht Akzeptabilität im Hinblick auf den Kontext Lehre im Fremdsprachenunterricht. Sewell prüft, welche sprachlichen Phänomene eines Akzents sich auf Verständlichkeit und Akzeptabilität auswirken und ab wann ein Akzent als nicht mehr akzeptabel in diesem Kontext angesehen wird. L1-Rater bewerten Sprachproben auf 6-stufigen Ratingskalen neben *acceptability* zu *correctness*, *pleasantness* und *familiarity* mit jeweils einem Item. Die Pole sind mit *agree strongly* und *disagree strongly* gekennzeichnet. Das Item zur Ermittlung von Akzeptabilität ist *This speaker's accent is acceptable as a model for pronunciation teaching purposes in HK (Hong Kong)*. Sowohl Chang Li-Ann als auch Sewell beziehen einen konkreten Kontext zur Bewertung von Aussprache mit ein. Eine positive oder negative Bewertung von Aussprache stellt somit implizit

das erfolgreiche kommunikative Bestehen oder Nichtbestehen im jeweiligen Kontext dar.

Die Studie von Pilott (2016) gewährt weiterführende Einblicke in das Konstrukt. So erachtet Pilott (2016: 51) es als erforderlich, bei der Messung von Akzeptabilität die Folgen von Aussprachebewertung mit einzubeziehen, indem man die Ziele von Fremdsprachenunterricht in der Konstruktdefinition berücksichtigt. Positive bzw. negative Bewertung fördert bzw. hindert L2-Lerner im Erreichen ihrer Ziele. Diese Ziele können vielfältig sein. So ist etwa bereits das Kommunizieren in der L2 mit L1-Sprechern ein solches Ziel, genau wie das Erlangen oder die Ausübung einer Arbeit. Die Dimension Akzeptabilität bildet Konsequenzen ab, die L2-Aussprache haben kann. Pilotts Auffassung nach sollten diese Konsequenzen mit in die Messung einfließen. Entsprechend definiert Pilott Akzeptabilität als subjektive Bewertungsdimension, die aussagt, ob Aussprache gut genug für einen festgelegten Zweck ist. Ähnlich wie Chang Li-Ann und Sewell argumentiert Pilott (2016: 55), dass Akzeptabilität aus diesem Grund immer unter dem Gesichtspunkt eines Kontexts jenseits sprachlich-struktureller Eigenschaften beurteilt werden muss, beispielsweise ob Aussprache gut genug ist, um als Lehrer oder Nachrichtensprecher zu arbeiten. Er lässt diese Arbeitsdefinition in die Messung von Akzeptabilität einfließen, indem er die Dimension über das Item *How acceptable is the migrant's pronunciation for a job with your company?* abprüft. Die insgesamt 40 weiblichen und männlichen L1-Rater unterschiedlicher Herkunft²⁴ bewerten auf einer 9-stufigen Likert-Skala von 1 = *not acceptable* bis 9 = *very acceptable*. Pilott weist für die Faktoren Sprachfluss, Prosodie und Segmente signifikante Auswirkungen auf Akzeptabilität nach. In Bezug auf segmentale Fehler zeigt Pilott (2016: 107), dass die Auswirkung auf Akzeptabilität von der L1 abhängt. So werden Sprecher mit einer L1, die phonologisch stark vom Englischen abweicht, signifikant schlechter bewertet.

Neben Pilotts Arbeitsdefinition von Akzeptabilität und der darin enthaltenen Berücksichtigung der kommunikativen Auswirkung von L2-Aussprache in konkreten Gebrauchskontexten ist an der Studie außerdem für die vorliegende Arbeit interessant, dass Pilott die Bewertung von Akzeptabilität als abhängig von den anderen Bewertungsdimensionen darstellt und dies in der Messung berücksichtigt werden sollte. Er misst Korrelationen zwischen den Dimensionen und zeigt, dass Akzeptabilität am ehesten mit den Verständlichkeitsdimensionen (*comprehensibility* und *intelligibility*) korreliert. Zum Akzent (*accentedness*) gibt es lediglich eine schwache Korrelation (Pilott 2016: 192).

Die Darstellung der Forschung zur Akzeptabilität zeigt, dass diese eine weitere wichtige zu untersuchende Dimension von Aussprache ist. Die eher sprachlich-strukturell verankerten Dimensionen Verständlichkeit und Akzent sind eng mit bisherigen Zielen fremdsprachlichen Unterrichts verbunden, die das Erreichen einer Norm und idealisierten Standards darstellen. Die Dimension der Akzeptabilität jedoch bezieht die direkten

²⁴ Die genauen Herkunftsländer werden nicht angegeben. Als geographische Regionen werden Südostasien, Afrika, China, Japan, Westeuropa, Naher Osten, Südamerika und Südasien genannt.

Folgen von Aussprache in Form von L1-Raterreaktionen mit ein und beachtet damit über sprachlich-strukturelle hinaus auch relevante soziophonetische Aspekte, die das untersuchbare Spektrum von Wirkung von L2-Aussprache auf L1-Kommunikation abrunden.

4.1.4 Zusammenfassung und Implikationen

Das vorangegangene Kapitel konnte zeigen, dass Aussprache neben sprachlich-strukturellen Dimensionen auch mindestens in einer weiteren, soziophonetischen Dimension in Studien zur Bewertung von L2-Sprache durch L1-Rater beurteilt wird. Es ergibt sich ein Setting aus drei Bewertungsdimensionen, die sich mit Hilfe der Differenzierung zwischen subjektiven und objektiven Aspekten zu 3×2 Dimensionen ordnen lassen (siehe Tab. 4.1).

übergeordnete Bewertungsdimension	objektive Komponente	subjektive Komponente
Akzent	<i>accent</i> objektiv messbarer Akzent	<i>accentedness</i> subjektiv wahrgenommener Akzent
Verständlichkeit	<i>intelligibility</i> objektiv messbare Verständlichkeit	<i>comprehensibility</i> subjektiv wahrgenommene Verständlichkeit
Akzeptabilität	<i>acceptance</i> objektiv messbare Akzeptanz	<i>acceptability</i> subjektiv wahrgenommene Akzeptanz

Tab. 4.1: Übersicht über die Bewertungsdimensionen

Accent misst die sprachlich-strukturell feststellbaren Abweichungen von L2-Sprache, die von der L1 der Sprecher abhängig sind. *Accentedness* entspricht der von Ratern wahrgenommenen Stärke von Akzent. *Intelligibility* stellt dar, ob L2-Sprache von L1-Ratern objektiv besser bzw. schlechter verstanden wird. Daneben zeigt *comprehensibility*, ob L2-Sprache von L1-Ratern subjektiv als besser bzw. schlechter verständlich bewertet wird. Durch die *acceptance* wird gemessen, ob L2-Sprache ihren kommunikativen Zweck in einem Gespräch mit L1-Ratern erfüllt, wohingegen die subjektive Komponente *acceptability* misst, ob L2-Sprache von L1-Ratern als mehr bzw. weniger akzeptabel in einem bestimmten Kontext bewertet wird.

Zur Begriffsdefinition der vorliegenden Studie werden die oben dargestellten Definitionen und Operationalisierungen vorangegangener Forschung teilweise übernommen, teilweise für das Vorhaben angeglichen. So werden die Definitionen von Verständlichkeit und Akzent von Munro & Derwing (1995a) zwar übernommen und ähnlich operati-

onalisiert (vgl. zur methodischen Operationalisierung und Erstellung der Test-Items Kapitel 4.2.2.3), die objektive und subjektive Komponente von Akzent (*accent* und *accentedness*) werden allerdings in der vorliegenden Arbeit voneinander unterschieden. Munro & Derwing (1995a) nehmen diese Unterscheidung nicht vor, sondern gebrauchen die Termini synonym (vgl. Kapitel 4.1.2.1). Es ergeben sich die folgenden Arbeitsdefinitionen für die vorliegende Studie:

- a. *Akzent* stellt sprachlich-strukturelle Abweichungen von L2-Sprache gegenüber der zielsprachlichen Norm dar, die von der L1 der Sprecher abhängig sind. Akzent als Konzept umfasst sowohl eine objektive anhand von phonetischen Verfahren feststellbare und beschreibbare Komponente (*accent*) als auch eine subjektive, sich durch die Wahrnehmung von L1-Ratern definierende Komponente (*accentedness*).
- b. *Verständlichkeit* ist ein Maß dafür, in welchem Umfang L1-Rater einer Äußerung von L2-Sprechern die intendierte Bedeutung entnehmen können. Verständlichkeit als Konzept umfasst sowohl eine objektiv anhand einschlägiger Verfahren feststellbare und beispielsweise in Prozentzahlen beschreibbare Komponente (*intelligibility*) als auch eine subjektive, durch den von L1-Ratern gefühlten Grad aufzuwendender kognitiver Anstrengung zum Erfassen eines Inhalts einer sprachlichen Äußerung definierte Komponente (*comprehensibility*).

Die deutschen Termini *Akzent* und *Verständlichkeit* werden somit als Überbegriffe geführt, wobei die Strukturierung in jeweils eine objektive und eine subjektive Komponente bestehen bleibt. Diese werden jeweils mit den englischen Termini voneinander unterschieden. Es ergeben sich die vier Bewertungsdimensionen *accent* und *accentedness* sowie *intelligibility* und *comprehensibility*. Wenn im Verlauf der Arbeit in deutscher Sprache deutlich gemacht werden muss, dass in einer Sprachprobe Akzent wahrgenommen wird, wird das Adjektiv *akzentuiert* verwendet. Für die Verständlichkeit wird das Adjektiv *verständlich* verwendet.

Für die Arbeitsdefinition von Akzeptabilität verwendet die vorliegende Studie Erkenntnisse zu angewandten Methoden, Arbeitsbegriffen und Ergebnissen aus den Arbeiten von Beinhoff (2014), Settineri (2011), Chang Li-Ann (2008), Sewell (2012) und Pilott (2016). So wird bei der Definition des Arbeitsbegriffs für Akzeptabilität die von Beinhoff (2014), Sewell (2012) und Chang Li-Ann (2008) geforderte Einbeziehung des Kontexts berücksichtigt und geprüft, ob L1-Sprecher des Dänischen L2-Sprecher für einen geeigneten Kommunikationspartner in der Zielsprache Dänisch halten. Settineris (2011) Definition von Akzeptabilität als soziale Implikation nonnormativer Aussprache ist grundlegend und wird um Pilotts (2016) Kriterium der Einbeziehung der Folgen in konkreten Gebrauchskontexten erweitert. Als Arbeitsdefinition ergibt sich:

- c. *Akzeptabilität* ist die kontextabhängige soziale Implikation nonnormativer Aussprache, die eine Reaktion der L1-Rater hervorruft, die über den unmittelbaren kommunikativen Kontext hinausgeht. Akzeptabilität enthält sowohl eine objektive Komponente, die sich als unmittelbare, von außen beschreibbare Reaktion von

L2-Ratern darstellt (*acceptance*), als auch eine subjektive Komponente, die sich als persönliche Einschätzung von L2-Ratern darstellt, wie ihre Reaktion auf eine Äußerung ausfallen würde (*acceptability*).

Dabei ist die konkrete Reaktion vom jeweiligen Kontext abhängig. Ebenso wie in den anderen beiden Bewertungsdimensionen wird der deutsche Terminus Akzeptabilität als Überbegriff eingesetzt, wohingegen die englischen Termini *acceptance* und *acceptability* die jeweils objektive und subjektive Komponente beschreiben. Wenn im Verlauf der Arbeit in deutscher Sprache dargestellt werden soll, dass L2-Sprache in einem Kontext keine negative Reaktion seitens der L2-Rater auslöst bzw. wenn diese nicht erwarten negativ zu reagieren, wird das Adjektiv *akzeptabel* verwendet.

Während sich die objektive Komponente von Akzent allerdings sprachlich-strukturell beschreiben lässt und auch die objektive Komponente von Verständlichkeit ohne Rückgriff auf Bewertung von L1-Ratern beschreibbar ist, handelt es sich bei objektiver Akzeptabilität um ein soziophonetisches Phänomen. Dieser Unterschied ist gemeint, wenn Akzeptabilität insgesamt in Abgrenzung zu anderen Bewertungsdimensionen als „highly subjective“ beschrieben (Thomson 2018: 25) oder im Gegensatz zu den eher sprachlich-strukturell zu verordnenden Dimensionen als soziale Dimension eingestuft wird (Beinhoff 2014: 156).

Die Messung der objektiven und subjektiven Komponenten der Bewertungsdimensionen gestaltet sich unterschiedlich. Die objektiven Dimensionen werden anhand von wertneutralen Messverfahren ermittelt, während die subjektiven Dimensionen die Einschätzung von L1-Sprechern erfordern und somit durch Ratingverfahren in Form von Bewertungsskalen ermittelt werden. Bezogen auf die einzelnen Komponenten stellen sich die Messmethoden konkret folgendermaßen dar:

- a. Der tatsächliche Akzent einer Äußerung (*accent*) kann objektiv mit Verfahren der akustischen Phonetik oder zumindest annäherungsweise durch die ohrenphonetische Analyse und Transkription durch geschulte Transkribenten sichtbar gemacht werden, während die subjektive Dimension (*accentedness*), die die Wahrnehmung und das Empfinden der Probanden darstellt, in Ratingverfahren mit L1-Sprechern als Probanden von Bewertungsskalen abgefragt wird. Die Skalierung, d. h. die Anzahl und Abstände der von den Probanden auszuwählenden Bewertungspunkte variiert von Studie zu Studie.
- b. Die objektiv messbare Verständlichkeit (*intelligibility*) stellt sich als der tatsächlich verstandene Prozentsatz einer Aussage dar und kann durch Verständnistests gemessen werden. Diese können sich so darstellen, dass L1-Rater die genaue Anzahl der Wörter oder exakt die Wörter, die sie gehört haben, wiedergeben müssen. Die L1-Rater können nicht geschulte Probanden sein. Die subjektiv wahrgenommene Dimension der Verständlichkeit (*comprehensibility*) wird genau wie die *accentedness* durch Ratingverfahren bzw. -skalen erhoben.
- c. Für die objektive Dimension der Akzeptabilität (*acceptance*) stellt sich die objektive Messung in der methodischen Umsetzung komplizierter dar, da sie die tat-

sächliche Reaktion eines L1-Hörers auf eine L2-Äußerung erfassen muss. Soll Akzeptabilität also objektiv gemessen werden, müssen L1-Sprecher-Reaktionen auf eine L2-Äußerung beobachtet werden. Ebenso wie bei den anderen beiden Dimensionen stellt die subjektive Dimension der Akzeptabilität (*acceptability*) die Wahrnehmung der L1-Rater dar und wird als solche über Bewertungsskalen erhoben.

Die in den vorangegangenen vorgestellten Studien zur Messung von Bewertung von L2-Sprache durch L1-Rater zeigen, dass die drei Dimensionen prinzipiell unabhängig voneinander sind, auch wenn Zusammenhänge nachgewiesen werden können (Trofimovich & Isaacs 2012: 913). Bestimmte Zusammenhänge bzw. Kombinationen von Bewertungen sind dabei wahrscheinlicher als andere. In Tab. 4.2 werden vorstellbare Kombinationsmöglichkeiten von Bewertungen dargestellt. Diejenigen, die keine praktische Bedeutung haben, d. h. unwahrscheinlich sind, werden in der Tabelle mit einem Asterisken markiert dargestellt. Die positiven Werte von Bewertungsverfahren werden über Plus-, die eher negativen Werte über Minussymbole dargestellt. Die Darstellung ist somit nicht binär zu lesen, sondern umfasst mehrere Skalierungsmöglichkeiten. L1-Rater können L2-Äußerungen als eher nicht verständlich (1) oder eher verständlich (2) einstufen, als eher akzeptabel oder eher nicht akzeptabel sowie als eher akzentuiert oder nicht akzentuiert.

Verständlichkeit (<i>comprehensibility</i>)	Akzeptabilität (<i>acceptability</i>)	Akzent (<i>accentedness</i>)	Bewertungskombinationen
(1) –verständlich	–akzeptabel	–akzentuiert	* (1a) –verst. –akzept. –akzent.
		+akzentuiert	(1b) –verst. –akzept. +akzent.
	+akzeptabel	–akzentuiert	* (1c) –verst. +akzept. –akzent.
		+akzentuiert	* (1d) –verst. +akzept. +akzent.
(2) +verständlich	–akzeptabel	–akzentuiert	(2a) +verst. –akzept. –akzent.
		+akzentuiert	* (2b) +verst. –akzept. +akzent.
	+akzeptabel	–akzentuiert	(2c) +verst. +akzept. –akzent.
		+akzentuiert	(2d) +verst. +akzept. +akzent.

Tab. 4.2: Kombinationen der Bewertungsdimensionen

Im Folgenden werden die einzelnen Kombinationsmöglichkeiten beschrieben:

- a. Die Kombination, dass eine Äußerung in der Bewertungsdimension wahrgenommener Akzent (*accentedness*) eher niedrige Werte erzielt, gleichzeitig hohe Werte

für die wahrgenommene Verständlichkeit bekommt, aber als nicht akzeptabel eingestuft wird, ist eher unwahrscheinlich.

- b. Eine als eher akzentuiert bewertete Äußerung mit niedrigen Werten für die wahrgenommene Verständlichkeit, also eine von den L₁-Ratern als eher akzentuierte und unverständlich eingestufte Äußerung, erzielt im wahrscheinlichsten Fall ebenfalls keine hohen Werte in der Akzeptabilität.
- c. Dass eine Äußerung als eher nicht verständlich eingestuft wird, obwohl sie ansonsten hohe Werte in den Dimensionen Akzeptabilität und niedrige Werte in der Dimension Akzent erzielt, ist zwar möglich, aber unwahrscheinlich.
- d. Die Kombination, dass eine als akzentuiert eingestufte Äußerung gleichzeitig als nicht verständlich aber akzeptabel eingestuft wird, ist ebenfalls unwahrscheinlich.

Äußerungen können trotz Verständlichkeit inakzeptabel sein:

- a. Im wahrscheinlichsten Fall sind sie dann akzentuiert.
- b. Eher unwahrscheinlich bzw. für den Kontext L₂-Aussprache nicht relevant, ist, dass verständliche Sprache ohne Akzent als nicht akzeptabel eingestuft wird.
- c. Es ist möglich, dass eine Äußerung von L₂-Ratern hohe Werte für wahrgenommene Verständlichkeit und Akzeptabilität bekommt, obwohl sie gleichzeitig hohe Werte in der Dimension Akzent geben. Dies ist beispielsweise bei einem zwar wahrgenommenen, aber sich nicht auswirkenden Akzent der Fall.
- d. Die gleiche Kombination kann auch in muttersprachlicher Aussprache auftreten.

Die hypothetisch durchgeführte Kombination der Bewertungsdimensionen ergibt ein Erwartungsspektrum der im folgenden Kapitel vorgestellten empirischen Studie (Kapitel 4.2).

4.2 Empirische Studie

Bei der folgenden Studie handelt es sich um eine Perzeptionsstudie im oben dargestellten Forschungsbereich zur Bewertung von L₂-Sprache durch L₁-Sprecher. Die Studie erweitert den Forschungsbereich um eine in diesem Zusammenhang wenig beforschte Sprache. Einer Probandengruppe von dänischen L₁-Ratern werden Sprachaufnahmen deutscher L₂-Dänischlerner präsentiert, die sie anhand der Bewertungsdimensionen Akzent, Verständlichkeit und Akzeptabilität bewerten. Die Ergebnisse der Studie ermöglichen es, evidenzbasierte Aussagen über Profile unterschiedlicher Fehler zu tätigen und daraus Schlussfolgerungen für die unterrichtliche Praxis zu ziehen. Im Folgenden werden zunächst die Forschungsfragen (Kapitel 4.2.1) und das Forschungsdesign (Kapitel 4.2.2) vorgestellt und um die Darstellung der konkreten Vorgehensweise bei der Datenerhebung und -analyse erweitert. Das Kapitel wird durch die Präsentation (Kapitel 4.2.3) und übergreifende Diskussion der Ergebnisse (Kapitel 4.2.4) abgeschlossen.

4.2.1 Fragestellung und Hypothesen

Aus Studie 1 (vgl. Kapitel 3) ergeben sich Merkmale, die – wenn auch mit unterschiedlich auftretender Häufigkeit – typisch für die untersuchten Dänischlerner sind. Bislang ungeklärt ist, wie sich diese typischen Fehler auf die Kommunikation mit L1-Sprechern auswirken bzw. wie sie dem Ziel fremdsprachlichen Unterrichts, funktionale Fremdsprachigkeit und Kommunikationsfähigkeit zu erzeugen, im Wege stehen. Zu untersuchen bleibt,

- a. ob die ermittelten Fehler überhaupt einen Einfluss auf die Bewertung durch L1-Sprecher haben;
- b. wenn ja, wie sich dieser bezogen auf konkrete Bewertungsdimensionen darstellt.

Die folgende Studie widmet sich der Klärung dieser Fragen, indem sie untersucht, inwieweit sich die typischen Aussprachefehler deutscher Dänischlerner auf die Bewertung von L1-Sprechern auswirken und ob sich die Merkmale darin signifikant unterscheiden.²⁵

Ebenso wie in den in Kapitel 4.1 vorgestellten Studien zur Bestimmung der Fehler-schwere wird zur Klärung der Fragen ein Forschungsdesign mit L1-Ratingpanel eingesetzt. Dabei bewerten L1-Rater eine Auswahl an Fehlern im Aussprachebereich, die sich aus der vorangegangenen Studie (vgl. Kapitel 3) ergibt und die typischen Fehler der untersuchten Dänischlerner darstellt.

Insgesamt werden 23 phonetische Merkmale getestet. Es werden überwiegend segmentale Phänomene untersucht, d. h. unterschiedliche monophthongische Vokallaute, daneben Konsonanten sowie Polyphthonge. Von besonderem Interesse sind diejenigen Merkmale, die im deutschen Lautinventar nicht bzw. in anderen phonologischen Umgebungen vorkommen. So werden für die Konsonanten wortmediales stimmloses *s* [s], weiches *d* ([ð]) und der alveopalatale Frikativ [ç] untersucht. Bei den Vokalen werden vordere ungerundete ([ɛ], [e:] und [æ]), zentrale ([ü]) sowie hintere gerundete ([ö], [ɔ]) und ungerundete ([ä]) getestet.

Zu den getesteten typischen Fehlern bei den Polyphthongen gehören die fehlerhaften Realisierungen der aus dem eigenen Lautsystem bekannten Diphthonge [äy] und [äi], weiterer Diphthonge auf [y] sowie des Triphthongs [igy].

Bei den suprasegmentalen Fehlern werden die ungewohnten Quantitäten bei [i] und [y] getestet. Diese sind innerhalb der Realisierung allerdings nicht von den segmentalen Fehlern zu unterscheiden, da sie sich als qualitativer Lautersatz (z. B. [ɪ] statt [i]) darstellen. Neben den segmentalen Fehlern werden außerdem nicht realisierte phonologische Prozesse nach Schwa-Tilgung getestet.

In Tab. 4.3 wird die lautliche Kategorie, der zielsprachliche Laut in normkonformer Realisierung (Ziellaut), die als Stimuli eingesetzten Sätze sowie der konkret von den Schülern produzierte und als Stimulus eingesetzte Fehler dargestellt. Das Zielwort, das

²⁵ Die untersuchten phonetischen Merkmale können in normabweichender Form als Fehler oder als normkonform realisierte Zielmerkmale vorliegen.

den isolierten Fehler beinhaltet, wird in Tab. 4.3 dabei fett dargestellt. In den drei äußersten Spalten wird das Zielwort in zielsprachlicher und fehlerhafter Aussprache, wie es den L1-Ratern präsentiert wird, dargestellt sowie die Nummerierung aufgeführt, unter der der jeweilige Stimulus in der Studie geführt wird. Bei der Transkription der fehlerhaften, als Stimuli eingesetzten Merkmale wurde eine Form gewählt, die den Unterschied zwischen der angestrebten normsprachlichen Realisierung und den von den Lernern tatsächlich realisierten Lauten deutlich macht. Auf die Berücksichtigung weiterer phonetischer Merkmale im Sinne einer engeren Transkription wurde dabei verzichtet.

segmentale Ebene: Konsonanten

Zielmerkmal	Stimulussatz mit Fehler	überprüfter Fehler	normgerechte Aussprache	Aussprache im Stimulus	Nr.
stimmloses [s] wortmedial	<i>Tak for den dejlige rose.</i>	Lautersatz [z] für stimmloses [s]	[ˈkɔ:sə]	[ˈkɔ:zə]	1
	<i>Min bror har en stor næse.</i>		[ˈnɛ:sə]	[ˈnɛ:zə]	
weiches d [ð] wortmedial	<i>Hvornår mødes vi?</i>	Lautersatz [d] für weiches d [ð]	[mø:ð̥s]	[ˈmø:dəs]	2
	<i>Jeg elsker nødder.</i>	Lautersatz [l] für weiches d [ð]	[ˈnøð̥v]	[ˈnølv]	3
alveopalataler Frikativ [ç] Frikativ [ç]	<i>Det sker bare sjældent!</i>	Lautersatz [sj] für alveopalatalen Frikativ [ç]	[çɛl:nɔ̃]	[sjɛlnɔ̃]	4
	<i>Det var et chok at se ham.</i>	Lautersatz [ʃ] für alveopalatalen Frikativ [ç]	[çʌŋ]	[ʃɔ̃ŋ]	5

4 Studie 2: Bewertung der L2-Fehler deutschsprachiger Lerner durch dänische L1-Sprecher

segmentale Ebene: Vokale

Zielmerkmal	Stimulussatz mit Fehler	überprüfter Fehler	norm-gerechte Aussprache	Aus-sprache im Stimulus	Nr.
[e]	<i>Vi skal melde os mere i timen.</i>	Lautersatz [e] für [e]	[^h mɛlɔ]	[^h mɛlɔ]	6
	<i>Han var den ældste.</i>		[^h ɛl ^h sɔ]	[^h ɛlsɔ]	
[ɛ:]	<i>Vi maler vores hus.</i>	Lautersatz [ä:] für [ɛ:]	[^h mɛ:lɐ]	[^h mä:lɐ]	7
	<i>Kan du stave dit navn?</i>		[^h sɔ:ɛ:ɥɔ]	[^h sɔä:rɐ]	
[æ]	<i>Lise har en kat.</i>	Lautersatz [ä] für [æ]	[^h kæɔ]	[^h käɔ]	8
	<i>Lines mand er venlig.</i>		[mæn ^h]	[mäɔ]	
[ä]	<i>Jeg elsker den sang.</i>	Lautersatz [ɛ] für [ä]	[sän ^h]	[sɛɔ]	9
	<i>Jeg er træt.</i>		[^h tʁæɔ]	[^h tʁɛɔ]	
[ʌ]	<i>Giv mig din hånd!</i>	Lautersatz [ɛ] für [ʌ]	[hän ^h]	[hen]	10
[ɔ]	<i>Må jeg låne din bil?</i>	Lautersatz [ɔ] für [ɔ]	[^h lɔ:nɔ]	[^h lɔ:nɔ]	11
[ɔ]	<i>Vores bil er gul.</i>	Lautersatz [oɔ] für [ɔ]	[vɔ::s]	[vɔɔs]	12

segmentale Ebene: Polyphthonge

Zielmerkmal	Stimulussatz mit Fehler	überprüfter Fehler	norm-gerechte Aussprache	Aus-sprache im Stimulus	Nr.
bekannte Polyphthonge [ḁ̈], [ḁ̈i], [ḁ̈i]	<i>Han løj over for mig.</i>	Lautersatz [ø̥] für [ḁ̈i]	[lʰḁ̈iʰ]	[lʰø̥i]	13
	<i>Det regner.</i>	Lautersatz [e̥] für [ḁ̈i]	[ʰe̥ḁ̈iɐ]	[ʰe̥:ø̥ɐ]	14
Triphthonge	<i>Sneen hvirvler rundt.</i>	Lautersatz [ɪ̥] für [i̥ɥ̥]	[ʰvi̥ɣ̥ɐ]	[ʰvɪ̥ɐ]	15
unbekannte Polyphthonge auf [ɥ̥]	<i>Vi flyver til Mexico.</i>	Lautersatz [y̥] für [y̥ɥ̥]	[ʰfly̥ɥ̥e]	[ʰfly̥:ve]	16

Quantität

Zielmerkmal	Stimulussatz mit Fehler	überprüfter Fehler	norm-gerechte Aussprache	Aus-sprache im Stimulus	Nr.
Quantität [i]	<i>Vi gik en tur i det milde vejr.</i>	Lautersatz [ɪ] für [i]	[ˈmɪlə]	[ˈmiɫə]	17
Quantität [y]	<i>De har en hyggelig stue.</i>	Lautersatz [ɣ̥] für [y]	[ˈhy̥ɣ̥li:]	[ˈhɣ̥ɣ̥li:]	18

Prozesse					
Zielmerkmal	Stimulussatz mit Fehler	überprüfter Fehler	normgerechte Aussprache	Aussprache im Stimulus	Nr.
überlange Vokale, betonte Silben	<i>Du skal lige se her.</i>	Einfügung von [ə] mit fehlender Realisierung überlanger Vokale in betonten Silben	[li:]	[ˈliə]	19
	<i>Vi ses om en uge.</i>		[u:]	[ˈuə]	
überlange Vokale, unbetonte Silben	<i>Vi har mange lærere i skolen.</i>	Einfügung von [ə] mit fehlender Realisierung überlanger Vokale in unbetonten Silben	[ˈlɛ:ɐ:]	[ˈlɛ:ɐ:ə]	20
lange Konsonanten, unbetonte Silben	<i>Pas på! Det er nemt at falde her.</i>	Einfügung von [ə] mit fehlender Realisierung langer Konsonanten	[fæ:l]	[ˈfælə]	21
	<i>Jeg skal finde min bil.</i>		[fɛ:n]	[ˈfɪnə]	
Lenisierung nachtoniger Plosive intervokalisch	<i>Mette er min vens kone.</i>	nicht realisierte Lenisierung intervokalisch	[ˈmɛɖə]	[ˈmɛtʰə]	22
	<i>Der er en pakke.</i>		[ˈpʰüɖə]	[ˈpʰäkʰə]	
Lenisierung nachtoniger Plosive pränasal	<i>Min bror er nitten.</i>	nicht realisierte Lenisierung pränasal	[ˈnɛɖn]	[ˈnitn]	23
	<i>Må jeg få jakken?</i>		[ˈjüɖŋ]	[ˈjäkŋ]	

Tab. 4.3: Übersicht über die untersuchten Merkmale

Die Untersuchung erfolgt hypothesengeleitet, wobei sich die Arbeitshypothesen aus den oben dargestellten zentralen Fragestellungen ergeben. Zunächst wird anhand von Kontrollstimuli ermittelt, ob die L1-Sprecher Sprachaufnahmen mit Fehlern anders bewerten als solche ohne. Die grundlegende Arbeitshypothese im Hinblick auf jeden einzelnen Fehler lautet:

H_{1,0} Nullhypothese

Die Aussprache von L2-Lernern wird von L1-Ratern nicht anders bewertet als die Aussprache von L1-Sprechern und fortgeschrittenen L2-Sprechern.²⁶

²⁶ Als Kontrollstimuli dienen Aufnahmen von L1-Sprechern sowie auch von fortgeschrittenen L2-Lernern. Zur genauen Definition der Stimulustypen vgl. Kapitel 4.2.3.2.

H_{1.1} Alternativhypothese

Die Aussprache von L2-Lernern wird von L1-Ratern anders bewertet als die Aussprache von L1-Sprechern und fortgeschrittenen L2-Sprechern.

Als nächstes wird untersucht, ob sich Fehler unterschiedlich auf die Kommunikation mit L1-Sprechern auswirken, d. h. ob einige Fehler negativer bewertet werden als andere. Dazu wird die folgende Arbeitshypothese im Hinblick auf jede Bewertungsdimension aufgestellt:

H_{2.0} Nullhypothese

Die unterschiedlichen Fehler unterscheiden sich nicht in ihrer Bewertung durch L1-Sprecher in der jeweiligen Bewertungsdimension.

H_{2.1} Alternativhypothese

Die unterschiedlichen Fehler unterscheiden sich in ihrer Bewertung durch L1-Sprecher in der jeweiligen Bewertungsdimension.

Das zur Überprüfung der Hypothesen eingesetzte Forschungsdesign und die darin angewandten Methoden werden im Folgenden dargestellt.

4.2.2 Design und Methode

Bei der vorliegenden Perzeptionsstudie werden in einem L1-Ratingpanel unterschiedliche Aussprachefehler deutschsprachiger Dänischler von dänischen Muttersprachlern (L1-Rater) bewertet. Dazu werden den L1-Ratern sowohl Schüleraufnahmen des dritten Lernjahres als auch Aufnahmen von zwei Kontrollsprechern vorgespielt. Die L1-Rater bewerten die Aufnahmen in drei Bewertungsdimensionen (*accentedness*, *comprehensibility* und *acceptability*) in einem digitalen Fragebogen. Das zur Erstellung des Fragebogens verwendete Tool ist *SoSci Survey*, ein zu wissenschaftlichen Zwecken kostenlos nutzbares Onlinewerkzeug, mit dem webbasierte Fragebögen auch mit multimedialen Inhalten erstellt und publiziert können. Im Folgenden (Kapitel 4.2.2.1) werden die Stimuli anhand der ihnen zugrundeliegenden Sprachaufnahmen vorgestellt. In Kapitel 4.2.2.2 findet eine detaillierte Begründung der Erstellung der Ratingitems statt. Kapitel 4.2.2.3 stellt anschließend die konkreten Formulierungen vor, die gewählt wurden, um die untersuchten Konstrukte zu operationalisieren. Die Gesamtzusammenstellung des Fragebogens erfolgt in Kapitel 4.2.2.4. Anschließend werden in Kapitel 4.2.2.5 die Teilnehmer der Studie sowie in Kapitel 4.2.2.6 der Ablauf der Erhebung dargestellt.

4.2.2.1 Entscheidungen zu Sprachaufnahmen und Stimuli

In Studien des Forschungsbereichs zur Fehlerevaluation werden in einem experimentellen Setting Stimuli in Form von Lerner Sprache L1-Ratern zur Bewertung präsentiert. Bei der Auswahl der hierzu eingesetzten Sprachaufnahmen stellt die Entscheidung zwischen natürlichen Stimuli, d. h. unveränderter Lerner Sprache oder künstlich manipulierten Sti-

multi einen entscheidender Faktor für die Validität der Studie dar. Validität als Gütekriterium besagt, ob tatsächlich gemessen wurde, was gemessen werden sollte. Je stärker die Erhebungssituation kontrolliert ist, desto höher ist die sogenannte interne Validität, d. h. die eindeutige Interpretierbarkeit der Ergebnisse (Riemer & Settinieri 2010: 770). Die externe Validität, d. h. die Verallgemeinerbarkeit der Ergebnisse, sinkt hingegen mit zunehmender Manipulation. So ermöglicht eine manipulierte Sprachaufnahmen zwar den Einfluss genau eines sprachlichen Feature abzutesten, sie hat allerdings wenig mit realer Lernaltersprache gemein.

Stimuli, die von professionell instruierten Sprechern eingesprochen werden, stellen das Maximum an Kontrolle und Manipulation dar. In diesen spricht derselbe bilinguale Sprecher dieselbe Passage einmal mit und einmal ohne das zu prüfende Merkmal ein. Auf diese Weise werden Effekte, die durch den Sprecher bzw. die Stimme hervorgerufen werden könnten, weitgehend verhindert. Darüber hinaus wird die Isolierung eines Fehlers ermöglicht und neben der zu testenden Variable existiert eine Kontrollversion. Dieser künstliche Ansatz wird als *matched guise technique* bezeichnet, er wird in den 1960er Jahren erstmals von Lambert et al. (1960) eingeführt. Mehr als die Hälfte der Studien bis Mitte der 1990er Jahre bedient sich eines solchen simulierten Ansatzes, um die Fehleranzahl zu kontrollieren (Rifkin & Roberts 1995: 516). Künstlich erzeugte Stimuli haben den Vorteil, dass sie gezielt einen Fehler und dessen spezifischen Einfluss untersuchbar machen. In einem maximal künstlich aufgestellten Setting spricht ein und derselbe L1-Sprecher Sätze ein, die in nur einem Merkmal fehlerhaft sind. Auf diese Weise kann verhindert werden, dass andere, nicht zu untersuchende Fehler oder Sprechereffekte wie beispielsweise die Stimme oder das Geschlecht als Störvariable wirken. Es können in solchen Designs also sowohl soziale Effekte auf die Bewertung der L1-Rater ausgeschlossen, als auch die Aussagekraft bezüglich einzelner Merkmale erhöht werden.

Natürliche Stimuli werden aus realer Lernaltersprache erstellt und ermöglichen somit höchstmögliche Authentizität. Sie bergen allerdings den Nachteil, dass statt eines einzelnen zu testenden Fehlers, eine Vielzahl von Fehlern gleichzeitig auftreten können. Dieser Umstand erschwert es, Reaktionen genau auf ein Merkmal zurückzuführen. Studien, die ihre Stimuli möglichst natürlich gestalten, stoßen deshalb auf Schwierigkeiten in der Aussagekraft ihrer Ergebnisse in Bezug auf den Einfluss einzelner Merkmale. Besonders, wenn auf spontane Sprache zurückgegriffen wird, können multiple Faktoren nicht kontrolliert werden. Außerdem sind neben phonetischen Aussprache Fehlern lexikalische oder grammatische Fehler zu erwarten. Der Effekt auf die Bewertung der L1-Rater einzelner Merkmale kann aus der Vielfalt unterschiedlicher Normabweichungen nicht eindeutig bestimmt werden (interne Validität). Der Verzicht auf maximal authentische Lernaltersprache (in Form spontaner Sprache) kann sich also trotz des Anspruchs auf Natürlichkeit als sinnvoll darstellen, besonders für Studien, deren Forschungsfrage einzelne sprachliche Merkmale fokussieren. Der entscheidende Vorteil künstlicher Stimuli, aufgrund weniger Störvariablen kontrollierbarer als natürliche Sprache zu sein, steht allerdings dem entscheidenden Nachteil gegenüber, Lernaltersprache nicht authentisch abzubil-

den. Sie stellen ein künstliches vom Versuchsleiter erzeugtes Produkt dar. In der Realität würden die einzelnen zu untersuchenden Merkmale nie vereinzelt und isoliert auftreten. Vielmehr befinden sich die identifizierbaren, von der Norm abweichenden Fehler in einem L2-sprachlichen Umfeld, d. h. einer undefinierten Anzahl aus Normabweichungen in z. B. Rhythmus, Intonation oder Suprasegmentalia. Diese, das L2-sprachliche Umfeld ausmachenden Normabweichungen, können für die Sprechergruppe einer gemeinsamen L1 als gleich angenommen werden und müssen daher nicht im einzelnen genauer beschrieben und bestimmt werden. Im Folgenden wird diese Variable mit dem Begriff *L2-Rauschen* bezeichnet. Lautliche, lexikalische oder grammatische Fehler sind fest in das L2-Rauschen eingebunden. Auch die zu untersuchenden Merkmale sollten in einer solchen Verbindung untersucht werden, um die Wirklichkeit möglichst repräsentativ abzubilden (externe Validität). Werden die Stimuli von L1-Sprechern eingesprochen, fällt dieses L2-Rauschen weg, wodurch sich der Mangel an Natürlichkeit ergibt. Dieser bringt folgende Probleme mit sich:

- a. Studien, die diese Form von Stimuli einsetzen, erlauben es nicht, Aussagen über Reaktion von L1-Ratern auf L2-Sprecher zu tätigen, obwohl sie dies intendieren. Der Einsatz der künstlichen Stimuli erlaubt höchstens Aussagen über die Reaktion von L1-Ratern auf einen isolierten segmentalen Fehler in ansonsten perfekter L1-Sprache zu tätigen, da ein solches Setting nicht repräsentativ für die Reaktion auf natürliche L2-Sprache ist.
- b. L1-Rater empfinden die synthetischen Texte als unnatürlich. Einschlägige Studien können zeigen, dass L1-Rater L2-Lerner und L1-Sprecher innerhalb von wenigen Höreindrücken mit großer Sicherheit voneinander unterscheiden können. Die Qualität der Stimuli und somit die Aussagekraft der Ergebnisse ist also stark davon abhängig, wie gut ein Sprecher die Natürlichkeit imitiert. Möglich ist, dass die L1-Rater erkennen, dass die Stimuli von einem L1-Sprecher stammen und die künstlich eingefügten Fehler als unnatürlichen Sprachfehler einstufen.

Um möglichst valide Ergebnisse zur Untersuchung von L1-Reaktionen auf Lernersprache zu erhalten, sollten also Stimuli eingesetzt werden, die so nah wie möglich an authentische Sprache herankommen (vgl. Rifkin & Roberts 1995: 517–518).

In dieser Studie stellt die Auswahl der Sprachaufnahmen für die Stimuli einen Kompromiss zwischen größtmöglicher Kontrolle und größtmöglicher Authentizität dar. Die Stimuli wurden von Schülern eingesprochen und bestehen somit aus authentischer Lernersprache. Auf diese Weise befinden sich die zu untersuchenden Fehler im Umfeld des L2-Rauschen und enthalten somit neben dem abgefragten Merkmal noch unkontrollierte weitere Normabweichungen (z. B. in Wort- und Satzintonation). Aus oben genannten Gründen wird dieser Umstand für die Aussagekraft der Studie als förderlich eingestuft. Trotzdem ist zu bedenken, dass die unkontrollierten sprachlichen Merkmale des L2-Rauschens als Störvariablen auftreten können, d. h. es kann nicht mit ausnahmsloser Sicherheit gesagt werden, dass die L1-Raterreaktion durch das eigentlich zu untersuchende Merkmal erzeugt wurde. Da allerdings alle Stimuli aus authentischer Lernersprache ge-

neriert wurden, diese also alle dieselben Störfaktoren enthalten und sich lediglich in den zu überprüfenden Merkmalen unterscheiden, ist mit großer Sicherheit davon auszugehen, dass das L2-Rauschen als Störvariable ignoriert werden kann.

Die aufgenommenen Sätze wurden von den Schülern nicht frei produziert, sondern abgelesen, um lexikalische oder grammatische Fehler zu vermeiden. Vorgelesene Sprache ermöglicht somit ein gewisses Maß an Kontrolliertheit. Außerdem können die L2-Sprecher die Sätze üben, um somit ihre beste Aussprache in der L2 zu präsentieren.

Die verwendeten Sprachaufnahmen stammen aus Studie 1 (vgl. Kapitel 3), in der zunächst Auslösewörter generiert wurden, die einen typischen Fehler provozieren. Diese Auslösewörter wurden von 92 verschiedenen Schülern vorgelesen, sodass zu jedem Merkmal eine große Auswahlmöglichkeit an Sprachmaterial zur Verfügung steht. Aus den vorgelesenen Sätzen wurden diejenigen Aufnahmen für Studie 2 ausgewählt, in denen das entsprechende Merkmal prominent zu hören ist und keine weiteren ohrenphonetisch auffälligen Fehler enthalten sind. Die Vergleichbarkeit der Sätze wurde ebenfalls durch das Vorlesen ermöglicht. Neben diesem praktischen Vorteil, gilt es allerdings auch Nachteile zu berücksichtigen, die vorgelesene Stimuli mit sich bringen. So ist es möglich, dass die Erkennbarkeit der Stimuli als Vorleseausssprache einen Effekt auf das Ratingverhalten der L1-Rater hat. Da dieser eventuelle Störfaktor allerdings für alle Aufnahmen, einschließlich der Kontrollsprecher, gleich wäre, hat er keine nachteilige Auswirkung auf die Vergleichbarkeit zwischen den Stimuli. Trotzdem sollte bei der Interpretation der Ergebnisse insgesamt berücksichtigt werden, dass es sich um vorgelesene Sprache handelt. Ein daraus resultierender, zu bedenkender Faktor wäre beispielsweise, dass Vorlesen Fehler provozieren kann, die in freier Sprache möglicherweise gar nicht auftreten würden. Die getesteten Normabweichungen der Schüler also vielmehr Schwierigkeiten der orthographischen Umsetzung darstellen als den tatsächlichen Sprachstand. Für die aktuelle Studie ist dieser Umstand allerdings unproblematisch, da nicht der Sprachstand erhoben werden soll, sondern die Bewertung von L1-Ratern.

Es lässt sich zusammenfassen, dass bei der Stimuluserstellung der vorliegenden Studie ein Kompromiss zwischen Natürlichkeit und Künstlichkeit eingegangen wurde. So wird natürliche Lernaltersprache eingesetzt, die allerdings dahingehend reguliert wird, welche Sätze gesprochen werden, um spezifische Aussprachefehler zu provozieren und grammatische und lexikalische Fehler auszuschließen.

Aus allen eingelesenen Sätzen wurden diejenigen von einem phonetisch geschulten Hörer mit ausreichender Sprachkompetenz ausgewählt, in denen der zu untersuchende Fehler prominent und möglichst ohne weitere Fehler zu hören ist.

Neben den fehlerhaften Realisierungen der überprüften Merkmale wurden außerdem Kontrollstimuli, die das jeweilige Merkmal in normkonformer Realisierung enthalten. Die Kontrollstimuli wurden sowohl von einer muttersprachlichen Sprecherin als auch von einer fortgeschrittenen L2-Sprecherin eingesprochen. Es ergeben sich drei Stimulustypen, die im Folgenden näher beschrieben werden:

- a. *L2-Lerner*: Die Stimuli der L2-Lerner enthalten Aufnahmen unterschiedlicher Schüler aus Studie 1 (vgl. Kapitel 3). Es handelt sich gleichermaßen um männliche und weibliche Sprecher. Teilweise ist an der Stimme nicht zu erkennen, um welches Geschlecht es sich handelt. Die Aufnahmen wurden danach ausgewählt, ob die Sprecher das Merkmal fehlerhaft produzieren und die restlichen Wörter des Satz unmarkiert aussprechen. Diese Aufnahmen enthalten somit das zu testende Merkmal in fehlerhafter Realisierung und sind gleichzeitig in L2-Qualität eingesprochen, d. h., das zu testende Merkmal ist als fehlerhaft realisierte Variante in L2-Rauschen eingebettet.
- b. *L1-Kontrollsprecher*: Die Stimuli der L1-Kontrollsprecherin wurden von einer 31-jährigen bilingualen Sprecherin eingesprochen, die aus Kopenhagen stammt und größtenteils in Deutschland aufgewachsen ist. Die zu prüfenden Merkmale sind in diesen Stimuli fehlerlos und in L1-Qualität eingesprochen und können als Standarddänisch und somit unmarkiert gewertet werden.
- c. *L2-Kontrollsprecher*: Die Kontrollstimuli der fortgeschrittenen L2-Lerner wurden von einer Sprecherin eingesprochen, deren Erwerbsbeginn im schulischen Dänischunterricht liegt. Weiterer gesteuerter L2-Erwerb fand durch universitäre Sprachkurse im Rahmen des Skandinavistikstudiums mit Wahlsprache Dänisch statt. Zusätzlich erfolgte ungesteuerter L2-Erwerb durch Auslandsaufenthalte in Dänemark und Praktika (Aarhus). Die Sprecherin befindet sich insgesamt auf C1/C2-Sprachniveau, auch das Ausspracheniveau ist auf diesem Level einzustufen. Die Stimuli enthalten die zu testenden Zielmerkmale in fehlerloser Realisierung, eingebettet in L2-Rauschen.

Von jedem untersuchten Merkmal wurden mindestens vier Stimuli erzeugt. Zwei des Stimulustyps L2-Lerner und jeweils einer für die zwei Kontrollstimulustypen. Die Stimuli sind dabei so aufgebaut, dass derselbe Satz (mit integriertem abzuprüfendem Merkmal) viermal abgespielt wird. Auf diese Weise können sich die L1-Rater einen genügenden Höreindruck verschaffen. Die Sprachaufnahmen wurden dazu im Programm *Audacity* (Version 2.1.2) bearbeitet und zusammengeschnitten. Die zeitlichen Abstände zwischen den Satz wiederholungen sind für alle Stimuli gleich (4 Sekunden). Alle Aufnahmen wurden bezüglich Ton- und Lautstärke angepasst, sodass alle Aufnahmen von gleicher Qualität sind. Insgesamt werden 23 Merkmale (vgl. Tab. 4.3) untersucht. Jedes Merkmal wird durch zwei Stimuli mit dem zu überprüfenden Fehler und zwei Kontrollstimuli abgefragt. Dadurch ergeben sich insgesamt 92 Stimuli. Pro Fragebogenseite wird ein Stimulus durch sechs Items abgefragt, d. h. es ergeben sich insgesamt 92 Fragebogenseiten und 552 zu beantwortende Items.

4.2.2.2 Aufgabenformate und Itemkonstruktion

Im folgenden Kapitel werden die zugrunde liegenden Überlegungen der Itemkonstruktion des vorliegenden Fragebogens dargestellt. Dazu wird zunächst der Terminus *Item* er-

läutert, um anschließend die Möglichkeiten der Einzel- oder Mehrfachitemabfrage eines Konstrukts darzulegen. Der darauffolgende Teil beschäftigt sich mit den Entscheidungen zur Konstruktion der einzelnen Items.

Items sind die in Fragebögen enthaltenen von Probanden zu beantwortenden Einheiten. Sie können in Form von Selbstauskunftsfragen und Aufgaben vorliegen oder, wie es im vorliegenden Fragebogen der Fall ist, Aussagen darstellen, denen die Probanden auf einer Ratingskala zustimmen oder die sie ablehnen können. Items stellen Indikatoren des zu messenden Konstrukts dar. Durch sie wird das Konstrukt operationalisiert, d. h. inhaltlich abgebildet (Bortz & Döring 2006: 267–268).

Items können in Form von Einzelitems oder als Bündel mehrerer Items vorliegen, sogenannte psychometrische Skalen (Bortz & Döring 2006: 267). Der Vorteil der Messung theoretischer Konzepte anhand mehrerer Items besteht darin, dass der Einsatz statistischer Verfahren zur Reliabilitätsmessung ermöglicht wird. Anhand dieser wird deutlich, wie zuverlässig das zu testende Konstrukt durch den Test abgebildet wird (Bortz & Döring 2006: 462). Als weitere Vorteile nennen Bortz & Döring (2006: 268), dass Bündel von Items mehrere inhaltliche Facetten komplexer Merkmale besser abbilden und somit ein zu messendes Konstrukt besser wiedergibt, als Einzelitems. Das ist besonders in sozialwissenschaftlichen und psychologischen Studien von Relevanz. In der vorliegenden Studie wurden pro abzurufendes Merkmal zwei Indikatoren bzw. Items erstellt, obwohl vorangegangene Studien im Forschungsbereich häufig lediglich auf ein Merkmal zur Operationalisierung zurückgreifen. Die Hinzunahme eines zweiten Items soll allerdings die erwähnte genauere Abbildung der Komplexität sowie statistische Auswertungsverfahren zur Testung der Reliabilität des Fragebogens ermöglichen. In der Itemauswertung der vorliegenden Studie werden die Items allerdings als Einzelitems behandelt, d. h. weder Items zusammengefasst, noch deren Mittelwerte aufsummiert. Dies kann für Folgestudien angedacht werden, in denen das Messinstrument statistischen Reliabilitätsprüfungen unterzogen werden müsste. Zwei Merkmale sind nicht für alle Methoden zur Skalenüberprüfung der klassischen Testtheorie ausreichend, allerdings bilden diese beiden Items das Konstrukt genügend ab und verhindern eine Überbelastung der Probanden, die sich irritiert zeigen können, wenn sie zu viele scheinbare gleiche Items beantworten müssen. Bei der Konstruktion der Items wurden grundsätzliche methodische Entscheidungen berücksichtigt, die sich im Wesentlichen auf vier Aspekte beziehen:

- a. Typen von Ratingskalen;
- b. Polarität;
- c. Skalenstufen und Skalenmitte;
- d. Beschriftung der Skala.

Ratingskalen dienen dazu, einen Gegenstand hinsichtlich eines abgefragtes Merkmals zu beurteilen. Es gibt verschiedene Typen von Ratingskalen. Direkte Skalen skalieren das Merkmal in Abstufungen. Ein Beispiel für die Skalierung der Bewertungsdimension wahrgenommener Akzent wurde bereits im Forschungsstand in Kapitel 4.1.2.1 vorgestellt: *Wie stark ist der Akzent des Sprechers?* mit den Antwortkategorien *sehr schwach* bis

sehr stark oder numerisch 1 bis 5. Einen weiteren Typ stellen Tests dar, in denen der Grad der Zustimmung bzw. Ablehnung zu Aussagen (auch Statements) als Indikator für die Ausprägung des Merkmals angesehen wird (Jonkisz, Moosbrugger & Brandt 2012: 50). Dieses Verfahren ermöglicht die Messung latenter, d. h. nicht direkt beobachtbarer oder aus einem theoretischen Konstrukt entstammender Merkmale. Die in den Sozialwissenschaften für dieses Ziel am häufigsten eingesetzte Umsetzung dieses Typs stellen Likert-Skalen dar (Bortz & Döring 2006: 269). In diesen wird dasselbe Merkmal typischerweise durch mehrere Aussagen in 5-stufigen Bewertungsskalen erfasst. Die Werte des gesamten Itembündels werden anschließend in der Auswertung als Summenscore zusammengefasst betrachtet (Bortz & Döring 2006: 269).²⁷ Likert-Skalen ermöglichen es, die gewonnenen Daten in statistischen Methoden einzusetzen, die eigentlich ein höheres Skalenniveau erfordern. Konkret ermöglichen Likert-Skalen die Transformation ordinalskalierte Daten in intervallskalierte Daten. Intervallskalierte Daten bieten deutlich differenziertere Auswertungsmöglichkeiten in statistischen Analysen (Völkl & Korb 2018: 19–20). Messtheoretisch sind Daten, die mit Ratingskalen erhoben werden, jedoch lediglich ordinalskaliert. So kann nicht davon ausgegangen werden, dass die Abstände der konstruierten Antwortkategorien gleich sind (Völkl & Korb 2018: 20). Likert-Skalen stellen allerdings ein etabliertes Verfahren dar, in dem die Abstände von Probanden als gleich groß interpretiert werden, wenn es mindestens fünf Ausprägungen gibt und die Skalenbeschriftung dies zulässt. Aus diesem Grund ist die Interpretation der Daten als Intervalldaten trotz tatsächlicher Ordinalskalierung möglich.

Wie oben bereits erwähnt, wird in Likert-Skalen ein theoretisches Konstrukt typischerweise von mehreren Items gemessen. Die Items werden in sogenannten *Itembatterien* eingesetzt, einer universellen Ratingskala mit einheitlicher Skalenbeschriftung, in denen unterschiedliche Aussagen auf gleiche Weise beantwortet werden.

Im vorliegenden Fragebogen wurden Items in Form des Likert-Typs eingesetzt, obwohl die meisten der im Kapitel 4.1.2 vorgestellten Studien des Forschungsbereichs direkte Skalierungen nutzen.²⁸ Die größte Vorteil von Items des Likert-Typs gegenüber direkten Skalen besteht darin, dass sie die Erhebung latenter Merkmale ermöglichen. Eine direkte Skalierung stellt sich zwar für die Bewertungsdimension *accentedness* als vermutlich reliabel heraus (*Wie stark ist der Akzent des Sprechers?*), schwieriger wird es jedoch bei den anderen beiden Bewertungsdimensionen. Die direkte Frage *Wie verständlich stiften Sie die vorliegende Aussage ein?* ist sowohl für Probanden als auch für Forscher schwierig zu interpretieren, wie auch die Darstellungen in Kapitel 4.1.2.1 zeigen. Noch problematischer verhält es sich für die Bewertungsdimension *acceptability*. Was meinen Probanden, wenn sie auf eine Frage der Art *Wie akzeptabel schätzen Sie die Äußerung*

²⁷ Dazu werden die Werte eines Probanden über alle Items eines Merkmals zusammengezählt und durch die Zahl der von ihm beantworteten Items dividiert (Bortz & Döring 2006: 270).

²⁸ Eine Ausnahme stellt die in Kapitel 4.1.3 vorgestellte Studie von Sewell (2012) dar, der ebenfalls Zustimmung-/Ablehnungsaufgaben zur Messung von *acceptability* einsetzt.

ein? antworten? Wie lassen sich Ergebnisse im Bereich *sehr akzeptabel* oder *überhaupt nicht akzeptabel* interpretieren? Die Auslegung der Zustimmung oder Ablehnung von *Ich würde langsam mit ihm/ihr sprechen* ist hingegen eindeutig.

Darüber hinaus müssten die Bewertungsdimensionen in einer direkten Skalierung durch recht unterschiedliche Antwortformate erhoben werden. So wird zum einen sprachprobenspezifisch gefragt, wie stark ein Akzent ist oder wie schwierig eine Aussage zu verstehen ist, zum anderen wird probandenbezogen gefragt, wie akzeptabel er eine Äußerung wahrnimmt, d. h. wie derjenige auf eine Äußerung reagieren würde. Ein einheitliches Skalenmaß mit einheitlichen Antwortmöglichkeiten ließe sich kaum finden. Dadurch würde sich die ohnehin schon große Anzahl zu bearbeitender Items noch in ihrem Komplexitätsgrad erhöhen (vgl. Kapitel 4.2.2.4). Ratingskalen des Likert-Typs ermöglichen eine kompakte Darstellung der Fragen als Itembatterie, deren identische Antwortmöglichkeiten die Bearbeitung erleichtern. Dadurch wird die Bearbeitungszeit verkürzt.

Der nächste Aspekt, der bei der Erstellung von Ratingskalen berücksichtigt werden muss, bezieht sich auf die Polarität. Die Antwortskalen können unipolar und bipolar gestaltet sein. Unipolare Antwortskalen stellen ein Kontinuum von einer geringen Ausprägung zu einer hohen Ausprägung dar (z. B. *gar nicht akzentuiert* bis *sehr stark akzentuiert*). Sie haben einen ‚Nullpunkt‘ bzw. „einen Bezugspunkt, der das geringste Ausmaß der Zustimmung (bzw. Ablehnung) kennzeichnet“ und einen Endpunkt (positiv oder negativ), „der die stärkste Zustimmung (bzw. Ablehnung) markiert“ (Jonkisz, Moosbrugger & Brandt 2012: 51). Daraus folgt, dass der Grad der Zustimmung (bzw. Ablehnung) nur in eine Richtung verläuft. Anders verhält es sich bei bipolaren Antwortskalen, die sich dadurch auszeichnen, dass sie zwei gegensätzliche Dimensionen abbilden (z. B. *stimme zu* bis *lehne ab*). Items des Likert-Typs stellen ein Beispiel für die Messung der Zustimmung bzw. Ablehnung als Indikator für die Ausprägung eines Merkmals dar. Ihre Antwortskala ist bipolar zwischen Zustimmung (*stimme (voll) zu*) und Ablehnung (*stimme (überhaupt) nicht zu*) gepolt.

Bei den meisten im Forschungsüberblick vorgestellten Studien handelt es sich um direkte Ratingskalen mit unipolaren Antwortskalen. Da es sich bei der vorliegenden Ratingskala jedoch um Items des Likert-Typs handelt, sind die einheitlichen Antwortskalen entsprechend bipolar ausgerichtet (*stimme zu* – *stimme nicht zu*).

Als nächstes wird bei der Erstellung der Items die Anzahl der Antwortkategorien bzw. Skalenstufen bestimmt (Jonkisz, Moosbrugger & Brandt 2012: 50–51). Aus der Darstellung des Forschungsstands (Kapitel 4.1.2 und 4.1.3) wurde bereits deutlich, dass die Zahl der Skalenstufen variieren kann, im Forschungsbereich von Bewertung von L2-Sprache durch L1-Rater allerdings relativ homogen ist. Auch wenn einige Studien Magnitudenskalen einsetzen, d. h. stufenlose Bewertungsskalen, setzen die meisten Studien zwischen 5 und 9 Skalenstufen ein. Bei der Festlegung der Skalenstufen muss auch berücksichtigt, ob es eine Skalenmitte geben soll (bei einer ungeraden Stufenzahl) oder nicht (bei einer geraden Stufenzahl). In bipolaren Bewertungsskalen kann diese unter-

schiedliche Positionen der Probanden darstellen und somit die Interpretation der Skala sowohl für die Probanden als auch für die Forscher erschweren (Krosnick & Fabrigar 2012: 147). Beispielsweise kann eine mittlere Kategorie eine neutrale Position bzw. Indifferenz (*weder – noch*) darstellen oder Ambivalenz (*teils – teils*) zum Ausdruck bringen (Menold & Bogner 2015: 5). In unipolaren Ratingskalen steht die mittlere Kategorie für eine mittlere Position.

In dieser Studie wurden 5 Skalenpunkte gewählt. Diese Entscheidung beruht auf den Erkenntnissen von Krosnick & Fabrigar (2012: 144), die eine optimale Messung in Bezug auf Reliabilität, Validität und Differenzierungsgrad bei einer Kategorienganzheit zwischen 5 und 7 festlegen. Da Likert-Skalen in der Regel fünfstufig sind, wurde diese Anzahl beibehalten. Die Entscheidung für eine mittlere Kategorie ist für die vorliegende Studie unproblematisch, da auf Grund der Operationalisierung der Items nicht zu erwarten ist, dass die L1-Rater Schwierigkeiten haben werden, eine Haltung zu einer Äußerung einzunehmen. Zu erwarten ist, dass sich die Probanden eher für die eine oder die andere Seite entscheiden. Daraus folgt, dass vermutlich eher mit extremen bzw. polnahen Werten gerechnet werden muss, die Sprachdaten aber auch eindeutiger als (nicht) akzentuiert, (in)akzeptabel oder (un)verständlich eingeordnet werden können.

Schließlich muss die Beschriftung der Skalen (auch Skalenlabeling) (d) festgelegt werden, d. h. wie die Beschriftung der Antwortskala ausfallen soll. Diese Entscheidung umfasst zwei Aspekte: Erstens muss entschieden werden, ob alle Skalenpunkte einzeln oder nur die äußersten Skalenpunkte beschriftet werden sollen. Zweitens muss die Form der Bezeichnung der Skalenpunkte festgelegt werden. Dieses kann numerisch, verbal oder optisch in Form von Bildern ausfallen (Jonkisz, Moosbrugger & Brandt 2012: 52). In numerischen Skalen werden die Skalenstufen mit Zahlen bezeichnet. Sie erwecken den Eindruck einer genaueren Messung sowie die Sammlung intervallskalierter Daten. Tatsächlich ist der Abstand zwischen den Skalenstufen allerdings genauso willkürlich wie in anderen Skalen zur Bewertung, d. h. es handelt sich ebenfalls um ordinalskalierte Daten. Der Einsatz numerischer Skalen kann auf Grund dieses Missverständnisses allerdings leicht zu falschen Interpretation und messtheoretischen Problemen führen (Jonkisz, Moosbrugger & Brandt 2012: 55). Im Gegensatz zu numerischen Skalen werden die Skalenstufen verbaler Ratingskalen mit Worten bezeichnet. Dies ist insofern von Vorteil, als „die Interpretation der Skalenpunkte intersubjektiv einheitlicher erfolgt – die Probanden brauchen sich nicht ‚vorzustellen‘, was sich hinter den einzelnen Skalenpunkten verbirgt“ (Jonkisz, Moosbrugger & Brandt 2012: 52). Die dritte Möglichkeit stellen optische Skalen dar, in denen Probanden anhand von Symbolen ihre Bewertung abgeben. Der Vollständigkeit halber sei erwähnt, dass auch die Kombination verschiedener Skalenbeschriftungen beispielsweise verbal-numerisch oder optisch-verbal eine Entscheidungsmöglichkeit darstellt. Die Beschriftung der Items des vorliegenden Fragebogens war teilweise schon durch die Entscheidung zum Einsatz von Likert-Skalen vorgegeben. So besteht im vorliegenden Fragebogen eine verbale Beschriftung, wobei die minimal notwendige Anzahl an Kategorien beschriftet wurde, nämlich lediglich die Endpole (Krosnick & Fabrigar

2012: 149). Diese Minimalbeschriftung wird in der Literatur als vorteilhaft bezeichnet, da die Probanden weniger durch sprachliche Ausführungen verwirrt werden und der Bewertungsprozess somit erleichtert wird. Da Likert-Skalen und das Prinzip der Zustimmung bzw. Ablehnung von Aussagen allgemein häufig beispielsweise in Umfragen vorkommen, kann davon ausgegangen werden, dass die L1-Rater mit diesem Rating-Verfahren vertraut und weitere Beschriftungen überflüssig sind.²⁹ Die Endpole der Antwortskala des vorliegenden Fragebogens sind verbal mit 1 = *uenig* ‚stimme nicht zu‘ und 5 = *enig* ‚stimme zu‘ gekennzeichnet.

4.2.2.3 Operationalisierung und Itemformulierung

Operationalisierung bedeutet die Entwicklung von Indikatoren zur Messung latenter, d. h. nicht direkt beobachtbarer Merkmale (Bortz & Döring 2006: 224), die als Items für den Fragebogen entwickelt werden. Im Folgenden wird dargestellt, welche Überlegungen zur Konstruktion der Items der vorliegenden Studie angestellt wurden. Diese beziehen sich auf die Kategorisierung, die sprachliche Verständlichkeit der Items sowie auf die inhaltliche Eindeutigkeit und folgen den Ausführungen von Jonkisz, Moosbrugger & Brandt (2012: 62–66).

Wie in Kapitel 4.2.2.2 zur Skalenbildung ausgeführt beziehen sich Entscheidungsfragen zur Kategorisierung auf die direkte oder indirekte Formulierung von Items. Direkt formulierte Items haben den Nachteil, dass sie sehr unterschiedlich aufgefasst werden können. So bietet *akzeptabel* in der Frage *Wie akzeptabel finden Sie die Aussprache des Sprechers?* mindestens zwei Auslegungsmöglichkeiten:

- a. ‚Akzeptabel ist eine Äußerung, wenn ich sie verstehen kann.‘
- b. ‚Akzeptabel ist eine Äußerung, wenn sie normsprachlich klingt.‘

Zur leichteren Interpretation empfehlen Jonkisz, Moosbrugger & Brandt (2012: 62–63) auf gut gewählte indirekte Verhaltensindikatoren zurückzugreifen. Für das Verhalten im Fall von akzeptabler bzw. nicht akzeptabler Aussprache lassen sich je nach Kontext unterschiedliche Varianten vorstellen (z. B. Kommunikationsabbruch, Angebot eines Arbeitgebers o. ä.; vgl. Kapitel 4.1.3). In der vorliegenden Studie wurde das Switchingverhalten der Dänen als Verhaltensindikator gewählt. Als eine Vorstufe des tatsächlichen Sprachwechsels wird außerdem das langsame Sprechen hinzugezogen.

Ein weiterer Aspekt der Kategorisierung ist die Entscheidung zwischen personalisiert und depersonalisiert formulierten Items. Der Unterschied zeigt sich in Beispiel (23).

- (23) *personalisiert*: ‚Wie würden Sie die Aussprache einschätzen?‘
depersonalisiert: ‚Wie, glauben Sie, würden andere die Aussprache einschätzen?‘

²⁹ Diese Annahme konnte in der Evaluation des Fragebogens anhand von ausgewählten Probanden bestätigt werden.

Personalisierte Items können leicht unehrlich beantwortet werden, wenn sie persönliche Merkmale abfragen, liefern allerdings zuverlässigere Antworten als depersonalisierte Items (Jonkisz, Moosbrugger & Brandt 2012: 63). Schwierigkeiten können sich bei sehr persönlichen oder intimen Kontexten ergeben. Da die Beurteilung von Aussprache fremder Personen in einem anonymen Rating allerdings keine Verletzung der Privatsphäre erwarten lässt, wurden personalisierte Items gewählt.

Der Aspekt der sprachlichen Verständlichkeit von Items zielt darauf ab, dass Items von Probanden ohne große Mühe und bereits nach dem ersten Durchlesen verstanden werden sollten. Jonkisz, Moosbrugger & Brandt (2012: 64–65) stellen zur Gewährleistung sprachlicher Verständlichkeit von Items die folgenden Punkte auf:

- a. Die Items sollten positiv formuliert sein, da es sonst zu unklaren Antwortmustern kommen kann. So kann auf eine negative Formulierung von Fragen (z. B. *Finden Sie die Aussprache nicht akzeptabel?*) Zustimmung sowohl mit ja (*Ja, ich finde sie nicht akzeptabel*) als auch mit nein (*Nein, ich finde sie nicht akzeptabel*) erfolgen. Auch im Fall bei Itemformaten, die Zustimmung bzw. Ablehnung von Aussagen erfordern, können Missverständnisse auftreten. So führt die Nicht-Zustimmung bei einer negativ formulierten Aussage zu einer doppelten Verneinung (*Ich muss mich nicht konzentrieren, um zu verstehen, was er/sie sagt* führt bei Nicht-Zustimmung zu dem Ergebnis *Ich stimme nicht zu. Ich muss mich nicht nicht konzentrieren, um zu verstehen, was er/sie sagt*).
- b. Komplizierte und lange Sätze sowie zu kurze Sätze mit Abkürzungen sollten vermieden werden. Ebenso sind Fachbegriffe zu vermeiden, die den Probanden ohne Erklärung eventuell nicht geläufig sind.
- c. Intensitätsangaben zur Stärke oder Häufigkeit (*Ich muss mich sehr konzentrieren, um zu verstehen, was er/sie sagt*) sollten ebenfalls nicht gebraucht werden, da sie die Eindeutigkeit verringern und nicht klar zu interpretieren sind. Er ist vorzuziehen, das Item ohne diese Angaben zu formulieren (*Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt*).

Diese von Jonkisz, Moosbrugger & Brandt (2012: 64–65) erarbeiteten Aspekte wurden bei der Formulierung der Items der vorliegenden Arbeit berücksichtigt. Zuletzt werden die Entscheidungskriterien zur Gewährleistung der Eindeutigkeit des Iteminhalts genannt. Eindeutigkeit ist dann erreicht, wenn „alle Probanden den Iteminhalt in gleicher Weise verstehen und die Antworten entsprechend der individuellen Ausprägung des interessierenden Merkmals geben“ (Jonkisz, Moosbrugger & Brandt 2012: 65). Nur durch diese „intersubjektive Verständnisbasis“ wird es möglich, von denselben Bedingungen und einer vergleichbaren Messung auszugehen (Jonkisz, Moosbrugger & Brandt 2012: 65). Die Eindeutigkeit eines Iteminhalts lässt sich durch die Formulierung der Items beeinflussen. Jonkisz, Moosbrugger & Brandt (2012: 65–66) empfehlen die Berücksichtigung verschiedener Regeln. Die für die Konstruktion der Items der vorliegenden Studie relevanten darunter sind:

- a. Universalausdrücke (*immer, nie, überhaupt nicht, total*) sollten vermieden werden.
- b. Pro Item sollte nur eine Aussage enthalten sein. Das Item *Die Aussprache ist schwer zu verstehen und ich muss mich sehr konzentrieren, um zu verstehen was er/sie sagt* sollte lieber auf zwei Items aufgeteilt werden.
- c. Der Iteminhalt sollte nicht auf unterschiedliche Weise interpretierbar sein. *Ich finde die Äußerung verständlich* kann sowohl bedeuten, dass der intendierte Inhalt akustisch und inhaltlich wahrgenommen wurde oder, dass eine Äußerung bzw. Haltung nachvollziehbar erscheint. Eine eindeutiger interpretierbare Formulierung wäre *Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt*.

Die itemspezifischen Überlegungen zur Operationalisierung werden im Folgenden zu jeder abhängigen Variable, d. h. in den einzelnen Bewertungsdimensionen dargestellt.

Accentedness

Die Operationalisierung des wahrgenommenen Akzents folgt inhaltlich der Formulierung früherer Studien. In der Skalierung der Items weicht sie jedoch von diesen ab. So skaliert die Mehrheit der Studien *accentedness* in direkten, meist unipolar skalierten Items von *wenig Akzent* bis *starker Akzent* mit unterschiedlicher Kategorienanzahl (vgl. Kapitel 4.1.2.1). Für die vorliegende Studie wurden hingegen aus in Kapitel 4.2.2.2 genannten Gründen Items des Likert-Typs und somit indirekte bipolare Items erstellt. So ergibt sich als erstes Item *Han/hun taler med accent* ‚Er/sie spricht mit Akzent‘ und als zweites Item *Han/hun kommer fra et andet land end Danmark* ‚Er/sie kommt aus einem anderen Land als Dänemark‘. Die Items bilden auf unterschiedliche Weise den wahrgenommenen Akzent ab. Das erste Item ist konkret formuliert, indem es direkt nach dem Akzent fragt. Akzent ist dabei nicht als schwierig zu interpretierender Fachbegriff einzuordnen, sondern als Alltagsbegriff unter dem die Probanden L2-geprägtes Dänisch verstehen. Das zweite Item testet den wahrgenommenen Akzent nur implizit ab. So kann davon ausgegangen werden, dass Sprecher, die nicht aus Dänemark kommen, im Normalfall kein Dänisch sprechen oder höchstens als Fremdsprache erworben haben. Zustimmung der L1-Rater impliziert, dass bei einem Stimulus Dänisch mit Akzent wahrgenommen wird.

Comprehensibility

Die wahrgenommene Verständlichkeit wird in der vorliegenden Studie über den subjektiv gefühlten aufzuwendenden Grad der kognitiven Anstrengung definiert, der zum Verstehen einer Äußerung notwendig ist (vgl. Kapitel 4.1.4). In der Darstellung des Forschungsstands (vgl. Kapitel 4.1.2.2) zeigt sich, dass die meisten Studien direkte unipolare Skalierungen zur Operationalisierung einsetzen. Die Polbeschriftungen der Studien las-

sen sich von *extrem einfach zu verstehen* bis *unmöglich zu verstehen* zusammenfassen. Ebenso wie *accentedness* wird *comprehensibility* in der vorliegenden Studie anhand von indirekten bipolaren Items gemessen.³⁰ Daraus ergeben sich die Items *Jeg skal koncentrere mig for at forstå, hvad han/hun siger* ‚Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt‘ und *Det er svært at forstå, hvad han/hun siger* ‚Es ist schwierig zu verstehen, was er/sie sagt‘. Das erste Item stellt eine indirekte Erfragung des Merkmals Verständlichkeit dar, das über den Indikator *nötige Konzentration* gemessen wird. Das zweite Item ist relativ direkt formuliert, obwohl Jonkisz, Moosbrugger & Brandt (2012: 62) von einer direkten Formulierung abraten, da nicht immer von einer interindividuellen Übereinstimmung in der Interpretation ausgegangen werden kann (was verstehen Probanden unter *verständlich*?). Probanden können Verständlichkeit entweder auf sprachlicher oder akustischer Ebene bewerten oder auf inhaltlicher Ebene im Sinne von *Ich finde seine/ihre Haltung verständlich*. Das vorliegende Item wird allerdings dadurch interpretierbar, dass es sich auf eine konkrete Äußerung innerhalb eines bestimmten Settings bezieht, das insgesamt so aufgebaut ist, dass den Probanden willkürliche wertfreie Aussagen vorgespielt werden (z. B. *Der er en pakke*. ‚Da ist ein Paket.‘). Es ist davon auszugehen, dass den Probanden bewusst ist, dass sie mit dem Attribut *verständlich* eine Bewertung des insgesamt wahrgenommenen Inhalts bzw. von verständlicher Sprache abgeben und nicht ihre Haltung gegenüber einer Äußerung zum Ausdruck bringen sollen.

Acceptability

Anders als bei den Operationalisierungen von wahrgenommenem Akzent und wahrgenommener Verständlichkeit, bei denen die Überführung bereits vorhandener Operationalisierungen in ein indirektes bipolares Format genügte und die überwiegende Entscheidungsfragen auf die Formulierung der Items bezogen war, gestaltete sich die Operationalisierung von *acceptability* schwieriger. Dies hängt mit dem Umstand zusammen, dass auch innerhalb der Forschung dazu bislang noch kein Konsens besteht. Im Folgenden werden daher zunächst die Überlegungen der vorliegenden Studie zur Operationalisierung der Dimension dargestellt, bevor im Anschluss die Formulierung der Items näher ausgeführt wird.

In Kapitel 4.1.3 wurde bereits dargestellt, dass die Findung einer einheitlichen und angemessenen Definition der Dimension problematisch ist. In Kapitel 4.1.4 findet sich die Arbeitsdefinition der vorliegenden Arbeit, die Akzeptabilität als kontextabhängige soziale Implikation nonnormativer Aussprache beschreibt, die eine Reaktion der Muttersprachler hervorruft, die über den unmittelbaren kommunikativen Kontext hinausgeht und sowohl eine objektive Komponente, als auch eine subjektive Komponente enthält. Die subjektive Komponente (*acceptability*), die im vorliegenden Setting als abhängige

30 Zur Darstellung der Gründe für diese Entscheidung vgl. Kapitel 4.2.2.2.

Variable gemessen wird, stellt sich als persönliche Einschätzung von L2-Ratern dar, wie ihre Reaktion auf Aussprache ausfallen würde.

Zur Operationalisierung dieses theoretischen Konstrukts bieten Chang Li-Anns (2008), Sewells (2012) und Pilotts (2016) Überlegungen einen Ansatz zur Findung valider Indikatoren. Sie beziehen den Zweck von Sprache bei der Messung von Akzeptabilität mit ein bzw. messen *acceptability* anhand von indirekter Verhaltensindikatoren. Diese ergeben sich durch den Kontext des jeweiligen Rahmensettings, das ein gewisses Maß an Englischsprachigkeit voraussetzt. In der vorliegenden Arbeit definiert sich der Zweck von L2-Dänisch als verantwortlicher Aspekt von L2-Sprache für die gelingende Kommunikation mit L1-Sprechern.³¹ Dies wirkt sich so aus, dass die L1-Sprecher Dänisch während der Kommunikation mit L2-Lernern im Dänischen bleiben und nicht die Sprache wechseln. Als indirekte Verhaltensindikatoren für *acceptability* ergeben sich *langsam sprechen* und *in eine Lingua franca wechseln*, wobei *langsam sprechen* eine Vorstufe des Wechsels in eine sprachliche Alternative darstellt. Die Indikatoren liegen inhaltlich also auf unterschiedlichen Niveaus. Aus dieser Operationalisierung ergeben sich zwei Items: *Jeg ville tale et andet sprog end dansk med ham/hende* ‚Ich würde eine andere Sprache als Dänisch mit ihm/ihr sprechen‘ und *På dansk ville jeg tale langsomt med ham/hende* ‚Auf Dänisch würde ich langsam mit ihm/ihr sprechen‘. Die Items sind konkret formuliert und sowohl von den Probanden als auch von Forschern eindeutig interpretierbar.

Überblick über die Items

Die eingeführten Items werden in der folgenden Tab. 4.4 in der Übersicht dargestellt.

³¹ Vgl. Kapitel 1.1. Die funktionale Sprachfähigkeit der L2-Lerner definiert sich als Kommunikationsfähigkeit in der Zielsprache Dänisch, d. h. die L1-Sprecher akzeptieren die L2-Lerner als Kommunikationspartner in der Zielsprache Dänisch.

Bewertungsdimension	Items
<i>accentedness</i> (wahrgenommener Akzent)	(1) <i>Han/hun taler med accent.</i> ,Er/sie spricht mit Akzent.‘
	(2) <i>Han/hun kommer fra et andet land end Danmark.</i> ,Er/sie kommt aus einem anderen Land als Dänemark.‘
<i>comprehensibility</i> (wahrgenommene Verständlichkeit)	(1) <i>Jeg skal koncentrere mig for at forstå, hvad han/hun siger.</i> ,Ich muss mich konzentrieren, um zu verstehen was er/sie sagt.‘
	(2) <i>Det er svært at forstå, hvad han/hun siger.</i> ,Es ist schwer zu verstehen, was er/sie sagt.‘
<i>acceptability</i> (wahrgenommene Akzeptabilität)	(1) <i>Jeg ville tale et andet sprog end dansk med ham/hende.</i> ,Ich würde eine andere Sprache als Dänisch mit ihm/ihr sprechen.‘
	(2) <i>På dansk ville jeg tale langsomt med ham/hende.</i> ,Auf Dänisch würde ich langsam mit ihm/ihr sprechen.‘

Tab. 4.4: Items zu den Bewertungsdimensionen

Die Dimensionen sind inhaltlich unterschiedlich gerichtet. So weist ein hoher Grad an wahrgenommenen Akzent eher auf niedrige Aussprachekompetenz hin, während ein hoher Grad an wahrgenommener Verständlichkeit und Akzeptabilität eine hohe Aussprachekompetenz vermuten lässt.

Die Unterschiede in der inhaltlichen Ausrichtung der Dimensionen werden durch die Formulierung der Items kompensiert. So sind die Items auf diese Weise formuliert, dass in Kombination mit der Antwortskala über alle Dimensionen hinweg niedrige Werte als ‚gut‘ bzw. ‚bessere Bewertung‘ im Sinne von unmarkierter Aussprache zu interpretieren sind (hohes Ausspracheniveau). Höhere Werte weisen auf die Bewertung als ‚schlechter‘ im Sinne von Wahrnehmung von Merkmalen von L2-Aussprache hin (niedriges Ausspracheniveau). Alle Items erhalten dieselbe Antwortskala von 1 = *uenig* ‚stimme nicht zu‘ bis 5 = *enig* ‚stimme zu‘. Es ergibt sich die folgende Tabelle:

Extremwert Antwortskala	Wert in der Bewertungsdimension	Interpretation
1 = <i>stimme nicht zu</i>	niedrige <i>accentedness</i>	hohes Ausspracheniveau
	hohe <i>comprehensibility</i>	
	hohe <i>acceptability</i>	
5 = <i>stimme zu</i>	hohe <i>accentedness</i>	niedriges Ausspracheniveau
	niedrige <i>comprehensibility</i>	
	niedrige <i>acceptability</i>	

Tab. 4.5: Darstellung über die Bedeutung von Extremwerten der Antwortskala

Nachdem im vorliegenden Kapitel die inhaltliche Reliabilität der Items dargestellt wurde, wird diese in Kapitel 4.2.3.1 statistisch anhand der Schätzung des Cronbach- α -Koeffizienten untersucht.

4.2.2.4 Layout

Bei dem in dieser Studie eingesetzten Fragebogen handelt es sich um einen digitalen, online abrufbaren Fragebogen. Das bei der Erstellung verwendete Tool stellt die Software *SoSci Survey* dar, der Fragebogen ist in Seiten gegliedert. Pro Seite wird ein Fehler mit je sechs Items abgefragt. Die Items stellen Aussagen dar, denen die L1-Rater auf einer Likert-Skala zustimmen oder die sie ablehnen können (genauer zur Likert-Skalierung in Kapitel 4.2.2.2). In Tab. 4.3 können die getesteten Fehler und deren Stimulussätze nachvollzogen werden. Jeder Fehler wird mit jeweils zwei Items in den Bewertungsdimensionen Akzent, Verständlichkeit und Akzeptabilität abgefragt.

Der Aufbau des Fragebogens ist in Abb. 4.2 dargestellt.³² Oben auf der Abbildung sind das Logo des Forschungsprogramms und der Universität sowie der Fortschrittsbalken zu erkennen. Dieser ermöglicht den Probanden ihren Bearbeitungsfortschritt auf jeder Seite nachzuvollziehen. Darunter befindet sich die Tonspur, die den Stimulus enthält. Die L1-Rater können selbstständig die Wiedergabe der Audiodatei starten und stoppen sowie die Lautstärke verändern.

³² Im Anhang B.1 können ebenfalls die ersten Seiten des Fragebogens nachvollzogen werden.

cau.kiel.de 7% færdig

0:03 / 0:17

Jeg ville tale et andet sprog end dansk med ham/hende.
helt uenig ○ ○ ○ ○ ○ helt enig

Jeg skal koncentrere mig for at forstå, hvad han/hun siger.
helt uenig ○ ○ ○ ○ ○ helt enig

På dansk ville jeg tale langsomt med ham/hende.
helt uenig ○ ○ ○ ○ ○ helt enig

Det er svært at forstå, hvad han/hun siger.
helt uenig ○ ○ ○ ○ ○ helt enig

Han/hun taler med accent.
helt uenig ○ ○ ○ ○ ○ helt enig

Han/hun kommer fra et andet land end Danmark.
helt uenig ○ ○ ○ ○ ○ helt enig

Næste

dansk-Projekt, CAU Kiel – 2017

Abb. 4.2: Exemplarische Seite des Fragebogens

Unter der Tonspur stehen die sechs Items. Das erste und das dritte Item erfragen die *acceptability*, das zweite und vierte Item erfragen die *comprehensibility* und die letzten beiden Items die *accentedness*. Die ausführliche Darstellung und Begründung zur Erstellung der Items findet sich in Kapitel 4.2.2.3. Am unteren Ende jeder Seite des Onlinefragebogens können die Probanden über den Button *Næste* ‚nächste‘ zur nächsten Seite gelangen. Dies funktioniert nur, wenn alle Items beantwortet wurden. Die Reihenfolge der insgesamt 92 Seiten ist randomisiert, d. h. sie ist bei jedem Probanden unterschiedlich. Auf diese Weise sollen Sequenzeffekte, d. h. Effekte die auf die Reihenfolge zurückzuführen sind, beispielsweise Ermüdungseffekte, vermieden werden (Bortz & Döring 2006: 731). Die Anordnung der Items blieb auf jeder Seite unverändert, sodass die Probanden nicht jedes Mal erneut lesen müssen und Verwirrungen oder Verwechslung verhindert werden. Die Antworten dürfen gewechselt werden, d. h., wenn sich ein Proband auf derselben Seite umentscheiden möchte, ist dies problemlos möglich. Zurückblättern und späteres Umentscheiden ist allerdings nicht möglich. Zu Beginn des Fragebogens werden den Probanden in einem Einleitungstext einige relevante Informationen mitgeteilt. So wird die Länge der Bearbeitungsdauer von einer halben Stunde genannt sowie der Hin-

weis gegeben, kurze Pausen bei Ermüdungserscheinungen zu halten.³³ Außerdem wird darauf hingewiesen, dass der Fragebogen allein auszufüllen und die Teilnahme freiwillig ist und jederzeit abgebrochen werden kann. Darüber hinaus wird die Anonymisierung der Daten versichert. Bevor die erste Fragebogenseite erscheint, werden die Probanden gebeten, Kopfhörer aufzusetzen und den Ton einzuschalten.

Am Ende der Erhebung erhalten die Probanden die Möglichkeit, Kommentare sowie der Wunsch zum Erhalt zu Informationen über Ergebnisse der Studie zu hinterlassen. Außerdem ist die Teilnahme an einem freiwilligen Gewinnspiel möglich. Auf diese Weise soll die Motivation der Probanden gefördert werden bzw. die Probanden zur Teilnahme animiert werden. Die im Gewinnfall zur Kontaktaufnahme benötigte E-Mail-Adresse wird dazu am Ende des Fragebogens aufgenommen von den restlichen Daten getrennt gespeichert, sodass es dem Versuchsleiter nicht möglich ist, die erhobenen Daten mit den Probanden in Verbindung zu bringen. Weitere persönliche Daten wie Namen oder Adressen werden nicht erhoben.

4.2.2.5 Probanden

Die L1-Rater, die vor Ort an der Erhebung teilnehmen, werden ausgehend aus den zwei größten Städten Dänemarks, Kopenhagen und Aarhus, akquiriert. Die Auswahl beruht hauptsächlich auf praktischen Erwägungen. Da die Universitäten als institutionelle Multiplikatoren fungieren, ist es leichter hier Teilnehmer zu akquirieren. In den Innenstädten und Bibliotheken werden Flyer verteilt und so auf die Erhebung aufmerksam gemacht. Als weiteres wichtiges Verbreitungsmedium dient das Internet, indem der Flyer und der daran angehängte Aufruf zur Teilnahme an der Studie in Foren und Gruppen der Universitäten bei *Facebook* (soziales Netzwerk) geteilt werden. Durch die Hinzunahme des Internets als Akquiseinstrument lässt sich die Herkunft der Probanden nicht regional begrenzen. Dieser Umstand bedeutet allerdings keine Einschränkung. Erstens wird die Herkunft der einzelnen Rater im Fragebogen mit erhoben und kann somit in der Analyse berücksichtigt und kontrolliert werden, zweitens ist ein möglichst breites Probandenspektrum für die Stichprobenbeschaffenheit erwünscht.

Insgesamt nahmen 192 Probanden an der Befragung teil.³⁴ 105 der Fragebögen wurden vollständig ausgefüllt. Da die Phänomene in randomisierter Reihenfolge abgefragt, d. h. die Seiten des Fragebogens immer unterschiedlich angeordnet werden, können auch die Fragebögen ausgewertet werden, die nicht vollständig bearbeitet wurden.

Neben den eigentlichen Ratingdaten werden am Anfang der Erhebung folgende (nichtsprachliche) Angaben zu den Ratern erhoben:

³³ Anhand der Evaluation ausgewählter Probanden konnte ermittelt werden, dass die Bearbeitung des Fragebogens im Schnitt eine halbe Stunde beträgt (vgl. Kapitel 4.2.2.6).

³⁴ Diese Anzahl stellt nur die berücksichtigten Fragebögen dar, d. h. diejenigen, die zu mindestens 20 % ausgefüllt waren.

- a. Alter;
- b. Geschlecht;
- c. Schulabschluss;
- d. Ausbildung;
- e. Sprachrepertoire;
- f. Wohn- und Geburtsort;
- g. wie leicht/schwer es den Probanden fällt, neue Sprachen zu erlernen;
- h. die zuerst erworbene Sprache.

Diese raterbiographischen Metadaten sind somit kontrollierbar und können auf ihren Einfluss auf das Ratingverhalten geprüft werden. Beispielsweise ist es möglich zu untersuchen, ob eine Ausbildung im (fremd)sprachlichen Bereich Auswirkungen auf das Ratingverhalten hat oder L1-Rater, denen es selbst leicht fällt neue Sprachen zu lernen, geneigter sind, schneller die Sprache zu wechseln, wenn das Gegenüber Abweichungen in der Aussprache aufweist. Im Folgenden wird ein Überblick über das Probandenprofil bezogen auf Alter und Geschlecht gegeben. Auf die ausführliche Darstellung der weiteren erhobenen nichtsprachlichen Metadaten wird allerdings verzichtet, da sie nicht in der Analyse miteinbezogen werden. In einer Folgestudie ist eine Untersuchung der Auswirkung dieser Aspekte auf das Ratingverhalten möglich.

Die meisten Probanden sind zwischen 21 und 30 Jahren alt. Der jüngste Proband der Studie ist 14 Jahre alt. Der älteste Proband ist 76 Jahre alt (Tab. 4.6).

Alter	Anzahl Probanden
14–20	17
21–25	86
26–30	46
31–40	20
41–60	15
61–76	8
ohne Angabe	0

Tab. 4.6: Übersicht über die Altersverteilung der L1-Rater

Die Mehrheit der Probanden ist weiblich (76 %). Insgesamt nahmen 45 Männer und 145 Frauen an der Erhebung teil. Zwei Probanden gaben kein Geschlecht an (Tab. 4.7).

Geschlecht	Anzahl Probanden
männlich	45
weiblich	145
ohne Angabe	2

Tab. 4.7: Übersicht über die Geschlechterverteilung der L1-Rater

4.2.2.6 Durchführung

Die Befragung wurde online durchgeführt. Ein Viertel der Probanden kam persönlich zur Erhebung, während der Rest den Fragebogen ohne Anleitung online ausfüllte. Um zu gewährleisten, dass eine Mindestanzahl an Probanden befragt wird und auch um das Verfahren anhand einer nachträglichen Befragung der anwesenden Probanden evaluieren zu können, füllen einige Probanden den Fragebogen vor Ort an den Universitäten in Kopenhagen und Aarhus in separaten Räumen mit Tablets und Kopfhörern aus.

Die erste Erhebungsrunde fand am 27. und 28. November 2017 in Kopenhagen an der Universität statt. Sie wurde in einem separaten Raum durchgeführt; den Probanden standen zur Beantworten ein Tablet und Kopfhörer zur Verfügung. Dadurch sollte eventueller Störschall vermindert werden und die Konzentration der Probanden auf den Fragebogen fokussiert werden. Die Termine für die vor Ort anwesenden Probanden waren so getaktet, dass immer maximal zwei Probanden gleichzeitig den Fragebogen beantworteten. Anschließend wurden die Probanden um eine mündliche Evaluation gebeten.

Aus den ersten Evaluationen ergab sich, dass der Fragebogen überwiegend als lang empfunden wurde. In den darauffolgenden Durchführungen wurde deshalb in die einleitende Instruktion des Onlinefragebogens der Zusatz eingefügt, dass die Bearbeitung des Fragebogens circa eine halbe Stunde in Anspruch nimmt und dass es erlaubt sei, bei Ermüdungserscheinungen eine Pause zu machen. Später wurde die Länge des Fragebogens daraufhin nicht mehr beanstandet.

Die Probanden gaben an, dass alle sprachlichen Stimuli akustisch gut zu verstehen gewesen seien. Einige Probanden fragten, aus welchen Ländern die Sprecher kamen und waren sich sicher, asiatische Sprecher gehört zu haben. Einige der vor Ort anwesenden Probanden sprachen bei der Erhebung laut mit und wiederholten die gehörten Fehler. Dies gibt Aufschluss darüber, dass die zu untersuchenden Merkmale tatsächlich prominent wahrgenommen wurden und liefert einen wichtigen Aspekt zur Qualitätssicherung der Studie. Die zweite Erhebung fand am 4. und 5. Dezember 2017 in Aarhus statt. Zeitgleich war die Bearbeitung des Onlinefragebogens möglich.³⁵

³⁵ Alle vor Ort anwesenden Teilnehmer erhielten ein kleines Präsent in Form von Schokolade, eines Schreibblocks und Kugelschreibers und alle Teilnehmer, auch diejenigen, die den Fragebogen nur online ausfüllten,

4.2.3 Ergebnisse und Analysen

Im Folgenden werden die erhobenen Daten und die statistische Analyse vorgestellt. In Kapitel 4.2.3.1 wird zunächst geprüft, ob die jeweiligen Items vergleichbar gut die getesteten Dimensionen abbilden bzw. in welchem Verhältnis die jeweils zwei Items zueinander stehen. Dies wird anhand des Cronbach- α -Koeffizienten ermittelt. Dies ist insofern relevant, als *acceptability* anhand von Items geprüft wird, die jeweils eine unterschiedliche Abstufung der Dimension darstellen. Ob sich diese Unterschiedlichkeit auch in den Ratings wiederfindet, kann anhand der Ergebnisse dieses Kapitels nachvollzogen werden.

In Kapitel 4.2.3.2 wird die erste Hypothese $H_{1.1}$ aus Kapitel 4.2.1 überprüft. Diese besagt, dass die L1-Rater die verschiedenen Stimulustypen (L1-Kontrollsprecher, fortgeschrittene L2-Kontrollsprecher, L2-Lerner) unterschiedlich bewerten und überprüft somit, ob die L1-Rater überhaupt anders auf fehlerhafte Stimuli reagieren. Die Überprüfung findet anhand eines statistischen Vergleichs der unterschiedlichen Bewertung der Stimuli mit Fehler (L2-Lerner) und der Kontrollstimuli (L1-Kontrollsprecher und fortgeschrittener L2-Kontrollsprecher) in Form einer linearen Regressionsanalyse statt. Die Auswahl dieses Verfahrens wird zu Beginn des Kapitels näher erläutert.

Kapitel 4.2.3.3 beschäftigt sich mit der Bearbeitung der zweiten Hypothese $H_{2.1}$, deren Aussage ist, dass sich die unterschiedlichen Fehler in ihrer Bewertung durch L1-Sprecher in der jeweiligen Bewertungsdimension unterscheiden. Um die Hypothese $H_{2.1}$ zu prüfen, werden in der Analyse die Mittelwerte aus dem L1-Rating anhand von Konfidenzintervallen untereinander verglichen. In einer abschließenden Diskussion (Kapitel 4.2.4) werden die Ergebnisse zusammengefasst.

Bei der für die Bearbeitung der Daten eingesetzten Software handelt es sich um *R* (Version 3.5.1 Feather Spray, 2018).³⁶ Diese wird von vielen quantitativ arbeitenden Linguisten eingesetzt, da sie auf nichtproprietären Datenformaten beruht und damit die Arbeit mit den Daten reproduzierbar und nachhaltig macht (Levshina 2015: 21). Die Software ist kommandozeilen- und skriptbasiert und ermöglicht die Interoperabilität mit der zur Datenerhebung verwendeten Software SoSciSurvey. Die bei der statistischen Hypothesenüberprüfung entstandenen *R*-Outputs sind im Anhang B nachzuvollziehen.

4.2.3.1 Bestimmung der Itemreliabilität

Zur Bestimmung der Reliabilität der Items zu den einzelnen Bewertungsdimensionen (interne Konsistenz) wurde der Cronbach- α -Koeffizient für jede Dimension (jeweils zwei Items) geschätzt. Cronbach- α stellt „inhaltlich die durchschnittliche Korrelation zwischen allen Einzelitems“ dar (Bortz & Döring 2006: 271). Die Berechnungen zu

nahmen an einem Gewinnspiel über insgesamt 6×2 Kinokarten plus Getränk und Popcorn teil. Dazu wurde die E-Mail-Adresse der Probanden unabhängig von den restlichen Daten erhoben. Die Gewinner wurden nach der Erhebung per Zufall ausgelost und bekamen die Gutscheine per Post zugeschickt.

³⁶ Die eingesetzten Pakete waren *ggplot2* (Wickham 2016), *dplyr* (Wickham 2018) und *psych* (Revelle 2018).

Cronbach- α werden ebenso wie die Analysen zu den gesammelten Daten in *R* durchgeführt. Die dafür eingesetzte Funktion lautet `alpha()` (Levshina 2015: 364-365). Die interne Konsistenz stellt sich für die Items zur Messung von *acceptability* als hoch (0,81) heraus. Für die Items zur Messung von *comprehensibility* (0,94) und *accentedness* (0,96) ist sie sogar als exzellent einzustufen. Bei der Berechnung von Cronbach- α für die Items der Variable *acceptability* wurde genau wie für die Items der anderen beiden Variablen eine gleiche Itemtrennschärfe (auch tau-Äquivalenz) angenommen, obwohl sie inhaltlich vermutlich zwei unterschiedliche Konstanten eines Konstrukts darstellen.³⁷ Bei solchen Items (essenziell tau-äquivalent) stellt Cronbach- α eine Mindesteinschätzung der Reliabilität dar (Bühner 2010: 133). Dies ist insofern zu erklären, als der Wert lediglich die Reliabilität der Items unter der Prämisse darstellt, dass die wahren Werte für beide Items gleich sind. In dem Fall, dass schon im Vorhinein die Annahme besteht, dass die wahren Werte der Items voneinander verschoben sind, sollte auch der erwartete Wert für Cronbach- α niedriger sein. Der niedrigere Wert für die Variable *acceptability* spiegelt also die inhaltliche Unterschiedlichkeit der Indikatoren (Items) wider, die zwar dasselbe Konstrukt, allerdings in einer unterschiedlichen Ausprägung messen.

4.2.3.2 Unterschiedliche Bewertung der Stimulustypen

Dieses Kapitel diskutiert Hypothese $H_{1,1}$, die postuliert, dass innerhalb der Bewertung der Stimuli mit Fehler (L2-Lerner) und der Bewertung der Kontrollstimuli (L1-Kontrollsprecher und fortgeschrittener L2-Kontrollsprecher) Unterschiede bestehen. Die Mittelwertunterschiede der Stimulustypen werden jeweils innerhalb einer abhängigen Variable – der jeweiligen Bewertungsdimensionen (*accentedness*, *comprehensibility* und *acceptability*) – miteinander verglichen. Dieser Analyseschritt soll Erkenntnisse darüber bringen, inwieweit L1-Rater fehlerhafte Sprache wahrnehmen und inwieweit fehlerhafte Aussprache kommunikativ relevant ist.

Üblicherweise wird zur Prüfung von Mittelwertunterschieden von mehr als zwei unabhängigen Stichproben und einer metrischen, abhängigen Variable eine Varianzanalyse durchgeführt (Bortz & Döring 2006: 709).³⁸ Die Varianzanalyse kann als Spezialfall der linearen Regression verstanden werden. Beide, Varianz- und lineare Regressionsanalysen messen Zusammenhänge zwischen Variablen und liefern praktisch identische Ergebnisse (Levshina 2015: 171). Die Regressionsanalyse wird allerdings als flexibler in Bezug auf die Variablenauswahl angesehen, da sie den Einsatz von sowohl numerischen als auch

³⁷ Die Itemtrennschärfe ist ein Maß dafür, wie „gut die Beantwortung des betreffenden Items das Gesamtergebnis vorhersagt“ (Bortz & Döring 2006: 478). Items mit der gleichen Itemtrennschärfe werden als tau-äquivalent bezeichnet. Man geht davon aus, dass sie denselben wahren Wert erheben. Items, die den wahren Wert jeweils um eine Konstante verschoben erheben, werden als essenziell oder im Wesentlichen tau-äquivalent bezeichnet (Bühner 2010: 125). Zur genaueren Ausführung vgl. Bühner (2010: 124–126).

³⁸ Die Varianzanalyse wird auch genauer unter *one-way analysis of variance* ‚einfaktorielle Varianzanalyse‘ oder kurz ANOVA geführt.

kategorialen Variablen sowie die Kombination beider Variablentypen ermöglicht. Aus diesem Grund wird die statistische Hypothesenüberprüfung der vorliegenden Arbeit anhand einer linearen Regressionsanalyse durchgeführt. Lineare Regressionen modellieren Zusammenhänge zwischen Variablen mit Hilfe einer Regressionsgleichung, die sich durch eine Regressionsgerade darstellen lässt. Wenn „zur Prädiktion nominale Variablen vor[liegen]“, müssen diese allerdings für das Verfahren umkodiert werden (Bühner & Ziegler 2012: 704). Dieser Umkodierungsprozess wird als *Dummykodierung* bezeichnet, die sich daraus ergebenden Variablen als *dummykodiert*. Als unabhängige Variable (Prädiktorvariable) des vorliegenden Settings fungieren die unterschiedlichen Stimulustypen (L2-Lerner, L1-Kontrollsprecher und fortgeschrittener L2-Kontrollsprecher). Somit handelt es sich bei der unabhängigen Variable des vorliegenden Settings um eine nominalskalierte Variable, weshalb das Dummykodierungsverfahren angewandt werden muss. Dieses wird anhand von Tab. 4.8 in Kürze erläutert. Die nominale Variable der Stimulustypen ist in der ersten Spalte dargestellt. Die Stimulustypen teilen sich folgendermaßen auf:

- a. 1 = L2-Lerner;
- b. 2 = L1-Kontrollsprecher;
- c. 3 = L2-Kontrollsprecher.

Bei der Dummykodierung wird die nominale Variable (in diesem Fall Stimulustyp) in mehrere dichotome Variablen umgewandelt. Dazu wurden in der vorliegenden Studie drei neue Variablen erstellt (siehe Tab. L2-Lerner), (L1-Kontrollsprecher) und (L2-Kontrollsprecher). Jeder Datenpunkt bzw. jeder Ratingwert, der zu Stimuli des Typs L2-Lerner abgegeben wurde, wird bei der neuen Variable L2-Lerner der Wert 1 zugewiesen, in den zwei anderen Variablen erhält er den Wert 0. Ebenso verhält es sich mit Daten, die den Stimuli der L2-Kontrollsprecher und L1-Kontrollsprechern zugeordnet werden müssen.

Stimulustyp	L2-Lerner	L1-Kontrollsprecher	L2-Kontrollsprecher
1	1	0	0
2	0	1	0
3	0	0	1

Tab. 4.8: Darstellung der Dummykodierung

Auf diese Weise kann nun in der Datenanalyse eine Variable als Bezugs-kategorie ausgewählt werden, die als Ausgang für den Abgleich mit den Mittelwerten der anderen Gruppen dient. In der Ergebnisausgabe werden die Mittelwertdifferenzen zwischen der Bezugs-kategorie und den jeweils anderen Gruppen ersichtlich. Zur näheren Erläuterung wird im Folgenden eine Ergebnisausgabe dargestellt (vgl. Tab. 4.9), die sich aus der linearen Regressionsanalyserrechnung mit der Hauptfunktion $lm()$ ergibt (Levshina 2015:

144). Der vom Programm erstellte Output beinhaltet unterschiedliche Informationen, die sich aus dem gerechneten Regressionsmodell ergeben.³⁹ Die folgende Tab. 4.9 dient zur Veranschaulichung der Ergebnisdarstellung und -interpretation.

	<i>acceptability 1</i>	<i>acceptability 2</i>
(<i>intercept</i> L1-Kontrollsprecher)	1,04 ***	1,08 ***
	(0,02)	(0,02)
L2-Lerner	0,85 ***	1,62 ***
	(0,02)	(0,03)
L2-Kontrollsprecher	0,27 ***	0,59 ***
	(0,03)	(0,03)
<i>N</i>	8.108	8.108
<i>R</i> ²	0,16	0,29

Tab. 4.9: Beispiel zur Ergebnisdarstellung und -interpretation
 (***) $p < 0.001$; (**) $p < 0.01$; (*) $p < 0.05$)

In den Spalten von Tab. 4.9 werden die Ergebnisse für die zwei Items der jeweils untersuchten abhängigen Variable (im Beispiel *acceptability*) dargestellt. Die erste Zeile stellt die Werte für die Bezugskategorie dar, den sogenannten *intercept*.⁴⁰ Im Beispiel repräsentiert der Stimulustyp L1-Kontrollsprecher die Bezugskategorie. Die zwei nachfolgenden Zeilen entsprechen der Mittelwertdifferenz der jeweiligen Prädiktoren (Stimulustypen L2-Lerner und L2-Kontrollsprecher) zur Bezugskategorie. In Klammern unter dem Ergebnis findet sich die jeweilige Standardabweichung.⁴¹ Bei der Interpretation der Mittelwerte stellen niedrige Werte für die jeweils betrachtete abhängige Variable (aV) ‚positive‘ Bewertung dar (vgl. Kapitel 4.2.2.3). ‚Positiv‘ ist im Sinne von unmarkiert, d. h. ohne wahrgenommene L2-Merkmale zu verstehen. Je niedriger die Werte sind, desto ‚besser‘ sind die Stimuli des jeweiligen Stimulustyps also bewertet. Das Beispiel wäre somit folgendermaßen zu analysieren: Die L2-Lerner werden im Mittel über alle Aufnahmen hinweg im Item 1 der Bewertungsdimension *acceptability* um 0,85 Punkte schlechter bewertet als die L1-Kontrollsprecher. Im Mittel werden sie also mit 1,85 bewertet, während die

³⁹ Die im Fließtext anhand von Tabellen dargestellten Ergebnisse stellen eine vereinfachte Darstellung des tatsächlichen *R*-Outputs dar. Die vollständigen Tabellen lassen sich im Anhang B.2 nachvollziehen.

⁴⁰ In normalen linearen Regressionsmodellen, die nicht dummykodiert sind, stellt der Intercept den erwarteten Wert der unabhängigen Variable dar, wenn alle unabhängigen Variablen den Wert 0 haben. Die in solchen Modellen visualisierbare Regressionsgerade schneidet an diesem Punkt durch die *y*-Achse (Levshina 2015: 147).

⁴¹ Die Standardabweichung stellt die Varianz der Messergebnisse dar, also die Streuung der Messwerte um den Mittelwert. Je geringer der Wert ist, umso näher liegen die Messwerte um den Mittelwert. Zur näheren Ausführung vgl. Bühner & Ziegler (2012: 45–49).

L1-Sprecher mit 1,04 bewertet werden. Die Sternchen geben an, wie signifikant dieser Unterschied einzustufen ist. Dabei wird von einem 5 %igen Signifikanzniveau ausgegangen.⁴² In der Darstellung wird zwischen drei Stufen unterschieden:

- a. $p < 0,001$: hoch signifikant, wird mit drei Sternchen (***) angezeigt;
- b. $p < 0,01$: sehr signifikant, wird mit zwei Sternchen (**) angezeigt;
- c. $p < 0,05$: signifikant, wird mit einem Sternchen (*) angezeigt.

In den letzten beiden Zeilen finden sich als weitere Interpretationshilfen die Anzahl der untersuchten Datenpunkte (N) sowie R^2 als Gütemaß der linearen Regression. R^2 liegt zwischen 0 und 1 und zeigt an, wie gut die unabhängigen Variablen geeignet sind, um die Varianz der abhängigen Variable zu erklären (Bortz & Döring 2006: 835; Levshina 2015: 148). In Tab. 4.9 können die Ausprägungen zu Item 1 zu 16 % im Regressionsmodell durch die unabhängigen Variablen erklärt werden. Gründe, weshalb R^2 nicht 1 ist, sind zum einen Messfehler der einzelnen Variablen, zum anderen der Umstand, dass die Vorhersage der abhängigen Variable neben den gewählten Prädiktoren (unabhängigen Variablen) noch von weiteren, hier nicht berücksichtigten Variablen abhängig ist (Bühner & Ziegler 2012: 658). Dieser Umstand ist als normal zu interpretieren. So kann besonders in Messungen, in denen menschliches Verhalten untersucht wird, davon ausgegangen werden, dass nicht alle verantwortlichen Variablen mit einbezogen wurden. Im vorliegenden Fall ist es wahrscheinlich, dass auch die Varianz im Antwortverhalten der L1-Rater eine Rolle für die Ausprägung innerhalb der abhängigen Variable spielt, die hier in die Rechnung allerdings nicht aufgenommen wurde.

Ob das gerechnete Regressionsmodell generell signifikant ist, zeigen die F - und t -Werte.⁴³ Diese gehören zu den zusätzlich in der Ergebnisdarstellung genannten Gütekriterien, die in der vereinfachten tabellarischen Darstellung nicht aufgeführt sind, im originalen Output (vgl. Anhang B.2) allerdings nachvollzogen werden können. Neben den Angaben zur F - und t -Statistik finden sich darunter Angaben zum korrigierten R^2 (auch angepasstes oder adjustiertes R^2), zur Varianz der Residuen (auch *residual standard error*, in den Analysen abgekürzt mit Std.-Fehler)⁴⁴ und dem p -Wert (auch Irrtumswahrscheinlichkeit). Dieses stellt die Güte des gerechneten Regressionsmodells in Relation zur Anzahl unabhängiger Variablen dar. Pro zusätzlicher unabhängiger Variable wird R^2

⁴² Das Signifikanzniveau gibt den p -Wert an, ab dem die Nullhypothese verworfen wird. Bei einem 5 %igen Signifikanzniveau wird die Nullhypothese bei $p < 0,05$ abgelehnt.

⁴³ Während die F -Werte sich auf die Signifikanzprüfung des Gesamtmodells beziehen, stellen die t -Werte die Signifikanzprüfung für die einzelnen Koeffizienten dar. Der p -Wert zeigt an, ob die jeweiligen F - und t -Werte signifikant sind. Bei der Berechnung des F -Wertes werden die Abweichungen der Mittelwerte von der Regressionsgeraden zur Abweichung der y -Werte von ihrem Gruppenmittelwert in Verhältnis gesetzt. In der zugehörigen Gleichung stellt der Wert im Zähler die Anzahl der Koeffizienten (k) dar, der Wert im Nenner ergibt sich aus $n-1-k$. Die Darstellung der F -Werte in den folgenden Abschnitten enthält somit folgende Parameter: $F(\text{df}_{\text{Zähler}}, \text{df}_{\text{Nenner}}) = F\text{-Wert}$, $p = \text{Signifikanz}$. Zur näheren Erläuterung von F - und t -Werten sowie deren Berechnung siehe Bortz & Döring (2006: 663–664), Levshina (2015: 178) und Meindl (2011: 241).

⁴⁴ Residuen stellen die „Abweichungen zwischen den vom Modell vorhergesagten und den tatsächlichen empirischen Messwerten“ dar (Bortz & Döring 2006: 750).

mit einem Strafwert belegt, woraus sich der Wert für korrigiertes R^2 ergibt. Dieser fällt also kleiner aus als der für einfaches R^2 (Bortz & Döring : 968).

Im Folgenden werden die Ergebnisse der statistischen Hypothesenüberprüfung in den einzelnen Bewertungsdimensionen dargestellt.

Accentedness

Die folgende Analyse bezieht sich auf die Mittelwertunterschiede der unabhängige Variable (uV) (Stimulustypen) innerhalb der aV *accentedness*. In der ersten Analyse stellen die Stimuli der L1-Kontrollsprecherin die Bezugskategorie (*intercept*) dar. Diese Ergebnisse dienen also als Referenzwert (vgl. Tab. 4.10) für den Abgleich der Mittelwerte mit den anderen Gruppen. In der zweiten Analyse stellen die Stimuli der L2-Kontrollsprecherin die Bezugskategorie dar (vgl. Tab. 4.11). Durch den Einsatz unterschiedlicher Kontrollgruppen als Bezugskategorie (*intercept*) ist es möglich, Unterschiede zwischen allen drei Stimulustypen zu analysieren.

	<i>accentedness</i> 1	<i>accentedness</i> 2
<i>(intercept</i> L1-Kontrollsprecher)	1,40 ***	1,39 ***
	(0,02)	(0,02)
L2-Lerner	3,14 ***	3,25 ***
	(0,03)	(0,03)
L2-Kontrollsprecher	1,91 ***	2,01 ***
	(0,03)	(0,03)
<i>N</i>	8.111	8.111
R^2	0,581	0,594

Tab. 4.10: Mittelwertunterschiede aV *accentedness* mit L1-Kontrollsprecher als Bezugskategorie

Das Regressionsmodell zum ersten Item der abhängigen Variable (aV) *accentedness* wurde anhand von 8.111 Datenpunkten durchgeführt. Das Modell korreliert signifikant mit den Daten (korrigiertes R^2 : 0,581, F -Wert (2, 8.108): 5.623, p -Wert: $< 2,2e^{-16}$) und bestätigt signifikante Mittelwertunterschiede in der aV *accentedness* zwischen den Stimulustypen L1-Kontrollsprecher und L2-Lerner (Ergebnisse der Koeffizienten: Stimulustyp L2-Lerner: 3,14, Std.-Fehler: 0,03, t -Wert: 105,97, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 1,91, Std.-Fehler: 0,03, t -Wert: 56,90, p -Wert: $< 2e^{-16***}$).

Das Regressionsmodell zum zweiten Item der aV *accentedness* beruht ebenfalls auf 8.111 Datenpunkten, korreliert signifikant mit den Daten (korrigiertes R^2 : 0,594, F -Wert (2, 8.108): 5.930, p -Wert: $< 2,2e^{-16}$) und bestätigt einen signifikanten Unterschied inner-

halb der Auswirkung auf *accentedness* zwischen der L1-Kontrollsprecherin und den L2-Lernern (Koeffizienten: Stimulustyp L2-Lerner: 3,25 Std.-Fehler: 0,03, t -Wert: 108,88, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 2,301, Std.-Fehler: 0,03, t -Wert: 59,60, p -Wert: $< 2e^{-16***}$).

Die Stimuli der L1-Kontrollsprecherin erhalten für den wahrgenommenen Akzent mittlere Werte von 1,40. Den geringsten Wert, den die Probanden ankreuzen können, stellt der Wert 1 dar. Somit ist dieser Mittelwert als Minimalwert zu interpretieren.

Sowohl die Bewertung der Stimuli der L2-Lerner als auch die Bewertung der Stimuli der fortgeschrittenen L2-Kontrollsprecherin weichen signifikant von derjenigen der L1-Kontrollsprecherin ab. Die Stimuli der L2-Kontrollsprecherin erzielt mittlere Werte von 1,91 für Item 1 und 2,01 für Item 2. Der wahrgenommene Akzent innerhalb der Stimuli der L2-Kontrollsprecherin ist im Mittel also um ca. einen Skaleneinheit höher als derjenige der Stimuli der L1-Kontrollsprecherin. Der wahrgenommene Akzent der Stimuli der L2-Lerner wird im Mittel für Item 1 mit 3,14 und mit 3,25 für Item 2 bewertet. Die Stimuli der L2-Lerner werden also im Mittel um ca. 2 Skaleneinheiten höher bewertet als die Stimuli der L1-Kontrollsprecher. In der folgenden Analyse (Tab. 4.11) dienen die Stimuli der L2-Kontrollsprecher als Referenzgruppe.

	<i>accentedness</i> 1	<i>accentedness</i> 2
(<i>intercept</i> L2-Kontrollsprecher)	3,31 ***	3,40 ***
	(0,02)	(0,02)
L2-Lerner	1,23 ***	1,24 ***
	(0,03)	(0,03)
L1- Kontrollsprecher	-1,91 ***	-2,01 ***
	(0,03)	(0,03)
<i>N</i>	8.111	8.111
<i>R</i> ²	0,58	0,59

Tab. 4.11: Mittelwertunterschiede aV *accentedness* mit L2-Kontrollsprecher als Bezugs-kategorie

Aus Tab. geht hervor, dass die L2-Lerner im Mittel um 1,23 Skaleneinheiten höher als die L2-Kontrollsprecher für Item 1 bewertet werden (p -Wert: $< 2e^{-16***}$). Das gerechnete Modell korreliert signifikant mit den Datenpunkten (8.111 Datenpunkte, korrigiertes R^2 : 0,581, F -Wert (2, 8.108): 5623, p -Wert: $< 2,2e^{-16***}$). Für Item 2 ergeben sich sehr ähnliche Werte (vgl. Anhang B.2). Die L2-Lerner werden im Durchschnitt um 1,24 Skaleneinheiten höher bewertet als die L2-Kontrollsprecher (p -Wert: $< 2e^{-16***}$). Auch dieses Modell ist statistisch signifikant (8.111 Datenpunkte, korrigiertes R^2 : 0,5939, F -Wert (2, 8.108): 5930, p -Wert: $< 2,2e^{-16***}$).

Die Ergebnisse zeigen, dass sowohl die Bewertung der L2-Lerner als auch die Bewertung der L1-Kontrollsprecher signifikant von der Bewertung der L2-Kontrollsprecher abweichen. So werden die L2-Lerner signifikant in 1,23 Skaleneinheiten für Item 1 und 1,24 Skaleneinheiten für Item 2 höher bewertet, während der wahrgenommene Akzent der L1-Kontrollsprecherin in 1,91 Skaleneinheiten für Item 1 und 2,01 Skaleneinheiten für Item 2 niedriger bewertet wird. Der wahrgenommene Akzent der L2-Kontrollsprecher ist also geringer als derjenige der L2-Lerner, sie werden jedoch noch immer deutlich von den L1-Kontrollsprechern unterschieden.

Unterschiede zwischen den beiden Items sind kaum vorhanden. Dies bestätigt die Ergebnisse des Cronbach- α -Koeffizienten in Kapitel 4.2.3.1. Es kann davon ausgegangen werden, dass beide Items die Bewertungsdimension *accentedness* annähernd gleich zuverlässig abbilden.

Comprehensibility

Das folgende Kapitel stellt die Mittelwertunterschiede der abhängigen Variable *comprehensibility* dar. Ebenso wie in der Analyse zur *accentedness* ist in der ersten Analyse die L1-Kontrollsprecherin die Bezugskategorie (*intercept*) (vgl. Tab. 4.12), während in der zweiten Analyse die Mittelwerte der L2-Kontrollsprecherin als Referenzwerte dienen (vgl. Tab. 4.13).

	<i>comprehensibility</i> 1	<i>comprehensibility</i> 2
<i>(intercept</i> L1-Kontrollsprecher)	1,07 ***	1,07 ***
	(0,02)	(0,02)
L2-Lerner	1,31 ***	1,46 ***
	(0,03)	(0,03)
L1- Kontrollsprecher	0,39 ***	0,45 ***
	(0,03)	(0,03)
<i>N</i>	8.108	8.108
<i>R</i> ²	0,24	0,26

Tab. 4.12: Mittelwertunterschiede aV *comprehensibility* mit L1-Kontrollsprecher als Bezugskategorie

Das Regressionsmodell zum ersten Item der aV *comprehensibility* beruht auf 8.108 Datenpunkten, korreliert signifikant mit den Daten (korrigiertes R^2 0,2368, F -Wert (2, 8.105): 1.258, p -Wert: $< 2,2e^{-16}$) und bestätigt einen signifikanten Unterschied innerhalb der Auswirkung auf *comprehensibility* zwischen der L1-Kontrollsprecherin und den L2-Lernern. Ebenso wie bei der aV *accentedness* ist der Unterschied zwischen den Stimuli

der L1- und L2-Kontrollsprecherinnen signifikant, dieser fällt allerdings viel kleiner aus (Ergebnisse der Koeffizienten: Stimulustyp L2-Lerner: 1,31, Std.-Fehler: 0,02, t -Wert: 47,09, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 0,39, Std.-Fehler: 0,03, t -Wert: 12,47, p -Wert: $< 2e^{-16***}$).

Das Regressionsmodell zum zweiten Item der Bewertungsdimension *comprehensibility* beruht ebenfalls auf 8.108 Datenpunkten, korreliert auch signifikant mit den Daten (korrigiertes R^2 : 0,2593, F -Wert (2, 8.105): 1420, p -Wert: $< 2,2e^{-16}$) und bestätigt ebenfalls einen signifikanten Unterschied innerhalb der Auswirkung auf *comprehensibility* zwischen der L1-Kontrollsprecherin und den L2-Lernern sowie der L2-Kontrollsprecherin (Ergebnisse der Koeffizienten: Stimulustyp L2-Lerner: 1,46, Std.-Fehler: 0,03, t -Wert: 50,15, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 0,44, Std.-Fehler: 0,03, t -Wert: 13,61, p -Wert: $< 2e^{-16***}$).

Aus der ersten Analyse zur aV *comprehensibility* geht hervor, dass sich sowohl die L2-Lerner als auch die L2-Kontrollsprecher signifikant in der Bewertung über alle Stimuli hinweg von den L1-Sprechern unterscheiden. Während die L2-Kontrollsprecher nur wenig von den L1-Sprechern abweichen (0,39 bei Item 1 und 0,44 bei Item 2), erzielen die L2-Lerner viel höhere Werte, die sich um einen ganzen Skaleneinheit unterscheiden (1,31 bei Item 1 und 1,46 bei Item 2).

Auch für die aV *comprehensibility* wird überprüft, ob die Bewertung der L2-Kontrollsprecher sich signifikant von derjenigen der L2-Lerner unterscheidet. Dazu dienen die L2-Kontrollsprecher in der folgenden Analyse als Referenzgruppe (vgl. Tab. 4.13).

	<i>comprehensibility</i> 1	<i>comprehensibility</i> 2
(intercept L2-Kontrollsprecher)	1,46 *** (0,02)	1,52 *** (0,02)
L2-Lerner	0,92 *** (0,03)	1,01 *** (0,03)
L1- Kontrollsprecher	-0,39 *** (0,03)	-0,45 *** (0,03)
N	8.108	8.108
R^2	0,24	0,26

Tab. 4.13: Mittelwertunterschiede aV *comprehensibility* mit L2-Kontrollsprecher als Bezugskategorie

In Tab. zeigt sich, dass die bessere Bewertung der L2-Kontrollsprecher gegenüber den L2-Lernern signifikant ist. In Item 1 werden die L2-Kontrollsprecher über alle Fehler hinweg um 0,92 Skaleneinheiten besser bewertet (p -Wert = $< 2e^{-16***}$). Das Modell korreliert statistisch signifikant mit den Daten (8.108 Datenpunkte, korrigiertes R^2 : 0,2368,

F-Wert (2, 8.105): 1.258, p -Wert: $< 2,2e^{-16***}$). Item 2 wird bei den L2-Kontrollsprechern um 1,01 Skaleneinheiten besser bewertet als bei den L2-Lernern (p -Wert $< 2e^{-16***}$). Auch dieses Modell korreliert statistisch signifikant mit den Daten (8.108 Datenpunkte, korrigiertes R^2 : 0,2593, F-Wert (2, 8.105): 1.420, p -Wert: $< 2,2e^{-16***}$).

Für die aV *comprehensibility* kann zusammengefasst werden, dass die L1-Rater die L2-Lerner (2,38) als deutlich unverständlicher einstufen als die L1-Kontrollsprecher (1,07) und die L2-Kontrollsprecher (1,46). Da der Wert 1 der minimale Wert ist, den die L1-Rater angeben können, kann von einer vollständig wahrgenommenen Verständlichkeit der L1-Kontrollsprecher ausgegangen werden. Das L2-Rauschen, das sowohl in der Stimulustyp der L2-Lerner und der L2-Kontrollsprecher vorhanden ist, wirkt sich in der Stimulustyp der L2-Kontrollsprecher leicht negativ aus. Die zusätzlichen Fehler der L2-Lerner haben einen noch größeren Effekt auf die wahrgenommene Verständlichkeit. Zwischen den Items sind keine Unterschiede wahrzunehmen. Es ist davon auszugehen, dass die Items das Konstrukt *comprehensibility* gleich gut abbilden, wie bereits die Ergebnisse des Cronbach- α -Koeffizienten in Kapitel 4.2.3.1 vermuten ließen.

Acceptability

Zuletzt erfolgt die Analyse der Mittelwertunterschiede der abhängigen Variable *acceptability*. Dabei wird wieder in der ersten Analyse die L1-Kontrollsprecherin als Bezugs-kategorie (*intercept*) für den Abgleich der Mittelwerte mit den anderen Gruppen eingesetzt (vgl. Tab. 4.14), um anschließend in der zweiten Analyse die Mittelwerte der L2-Kontrollsprecherin als Referenzwerte zu vergleichen (vgl. Tab. 4.15).

	<i>acceptability 1</i>	<i>acceptability 2</i>
(<i>intercept</i> L1-Kontrollsprecher)	1,04 *** (0,02)	1,08 *** (0,02)
L2-Lerner	0,85 *** (0,02)	1,62 *** (0,03)
L1- Kontrollsprecher	0,27 *** (0,03)	0,59 *** (0,03)
<i>N</i>	8.108	8.108
R^2	0,16	0,29

Tab. 4.14: Mittelwertunterschiede aV *acceptability* mit L1-Kontrollsprecher als Bezugs-kategorie

Das Regressionsmodell zum ersten Item der aV *acceptability* beruht auf 8.108 Datenpunkten, korreliert signifikant mit den Daten (korrigiertes R^2 : 0,1612, F-Statistik (2,

8.105): 780,2, p -Wert: $< 2,2e^{-16***}$) und bestätigt einen signifikanten Unterschied innerhalb der Auswirkung auf *acceptability* zwischen der L1-Kontrollsprecherin und den L2-Lernern. Auch die L2-Kontrollsprecherin wird wieder signifikant höher bewertet (Ergebnisse der Koeffizienten: Stimulustyp L2-Lerner: 0,85, Std.-Fehler: 0,02, t -Wert: 31,23, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 0,27, Std.-Fehler: 0,03, t -Wert: 10,26, p -Wert: $< 2e^{-16***}$).

Das Regressionsmodell zum zweiten Item der aV *acceptability* umfasst ebenfalls 8.108 Datenpunkte, korreliert außerdem signifikant mit den Daten (korrigiertes R^2 : 0,2856, F -Wert (2, 8.105): 1620, p -Wert: $< 2,2e^{-16}$) und bestätigt einen signifikanten Unterschied innerhalb der Auswirkung auf *acceptability* zwischen der L1-Kontrollsprecherin und den L2-Lernern sowie der L2-Kontrollsprecherin (Ergebnisse der Koeffizienten: Stimulustyp L2-Lerner: 1,62, Std.-Fehler: 0,03, t -Wert: 54,57, p -Wert: $< 2e^{-16***}$; Stimulustyp fortgeschrittene L2-Kontrollsprecher: 0,59, Std.-Fehler: 0,03, t -Wert: 17,61, p -Wert: $< 2e^{-16***}$).

Sowohl die L2-Lerner als auch die L2-Kontrollsprecher unterscheiden sich signifikant in der Bewertung von *acceptability* über alle Stimuli hinweg. Während die L2-Kontrollsprecher nur wenig von den L1-Sprechern abweichen (0,27 bei Item 1 und 0,59 bei Item 2), erzielen die L2-Lerner höhere Werte (0,85 bei Item 1 und 0,59 bei Item 2).

In einer nächsten Berechnung wird überprüft, ob die Bewertung der Stimuli der L2-Lerner sich auch signifikant von der Bewertung der Stimuli der L2-Kontrollsprecher unterscheidet.

	<i>acceptability</i> 1	<i>acceptability</i> 2
(<i>intercept</i> L2-Kontrollsprecher)	1,31 *** (0,02)	1,67 *** (0,02)
L2-Lerner	0,59 *** (0,02)	1,03 *** (0,03)
L1- Kontrollsprecher	-0,27 *** (0,03)	-0,59 *** (0,03)
N	8.108	8.108
R^2	0,16	0,29

Tab. 4.15: Mittelwertunterschiede aV *acceptability* mit L2-Kontrollsprecher als Bezugskategorie

Das signifikante Regressionsmodell (8.108 Datenpunkte, korrigiertes R^2 : 0,1612, F -Wert (2, 8.105): 780,2, p -Wert: $< 2,2e^{-16***}$) für Item 1 zeigt, dass sich die L2-Lerner im Durch-

schnitt über alle Fehler in 0,59 Skalenpunkten von den L2-Kontrollsprechern unterscheiden. Diese Unterschiede sind statistisch signifikant (p -Wert $< 2e^{-16***}$).

Das Modell für Item 2 korreliert ebenfalls signifikant mit den Daten (8.108 Datenpunkte, korrigiertes R^2 : 0,2854, F -Wert (2, 8.105): 1.620, p -Wert: $< 2,2e^{-16***}$) und zeigt, dass die L2-Kontrollsprecher im Mittel um 1,03 Skalenpunkte besser bewertet werden als die L2-Lerner (p -Wert $< 2e^{-16***}$).

In allen Stimulustypen sind die Werte bei Item 1 niedriger als bei Item 2. Dies wurde auf Grund der Operationalisierung und der inhaltlichen Ausrichtung von Item 1 als Vorstufe von Item 2 erwartet und kann nun in der Analyse bestätigt werden. Der Unterschied zwischen Item 1 und Item 2 ist größer als bei den anderen Bewertungsdimensionen. Diese Ergebnisse decken sich mit den Ergebnissen des Cronbach- α -Koeffizienten in Kapitel 4.2.3.1.

Zusammenfassung

Der Analyse geht die übergeordnete Fragestellung voraus, ob die L1-Rater fehlerhafte Aussprache generell und die Fehler insbesondere wahrnehmen. Dazu wurde die Hypothese geprüft, ob die Aussprache von L2-Lernern anders bewertet wird, als die von L1-Sprechern und fortgeschrittenen L2-Sprechern (vgl. Hypothese $H_{1.1}$). Die Nullhypothese ($H_{1.0}$) geht entsprechend von keinem nachweisbaren Unterschied aus. In der Analyse der abhängigen Variable konnten die folgenden Beobachtungen gemacht werden:

- a. Die Bewertungen der Stimulustypen unterscheiden sich signifikant voneinander. Wie erwartet werden die Aufnahmen der L1-Kontrollsprecherin über alle Stimuli hinweg in allen drei Dimensionen signifikant besser bewertet als die der L2-Lerner. Außerdem zeigt sich, dass auch die Stimuli der L2-Kontrollsprecher schlechter bewertet werden als die der L1-Kontrollsprecher. Sie werden aber besser als diejenigen der L2-Lerner eingestuft.
- b. Die Betrachtung der einzelnen Bewertungsdimensionen ergibt für die abhängige Variable *accentedness*, dass sich die L2-Kontrollsprecherin in 1,91 (Item 1) und 2,01 (Item 2) Skalenpunkten von der L1-Kontrollsprecherin unterscheidet. Höhere Werte implizieren einen höher wahrgenommenen Akzent. Die L2-Lerner werden um 3,14 (Item 1) und 3,25 (Item 2) Skalenpunkte höher als die L1-Kontrollsprecherin bewertet.
- c. Diese Unterschiede sind für die aV *comprehensibility* nicht ganz so groß, allerdings ebenfalls vorhanden. Die L2-Kontrollsprecherin wird in Item 1 in 0,39 Skalenpunkten schlechter verständlich als die L1-Kontrollsprecherin wahrgenommen und in Item 2 in 0,45 Skalenpunkten. Die L2-Lerner werden in 1,31 Skalenpunkten in Item 1 als schlechter verständlich als die L1-Kontrollsprecherin wahrgenommen bewertet und in Item 2 in 1,46 Skalenpunkten.
- d. Innerhalb der aV *acceptability* stellen niedrigere Werte eine höhere *acceptability* dar. Die Werte der L1-Kontrollsprecherin liegen bei Item 1 bei 1,04 und bei Item 2

bei 1,08. Da der niedrigste von den L1-Ratern ankreuzbare Wert 1 darstellt, sind diese Werte als sehr niedrig einzustufen.

- e. Im Vergleich der jeweils zwei Items pro aV lässt sich für *accentedness* und *comprehensibility* feststellen, dass die Ergebnisse annähernd gleich ausfallen. In der Dimension *acceptability* jedoch fallen die Werte für Item 1 und Item 2 sehr unterschiedlich aus. Dies entspricht der inhaltlichen Abstufung der Items als jeweils zwei unterschiedlich zu gewichtende Indikatoren des Konstrukts (vgl. Kapitel 4.2.3.1).

Es lässt sich zusammenfassen, dass die L1-Rater die L2-Lerner in allen drei aV schlechter bewerten als die L1-Kontrollsprecher und L2-Lerner. Die Nullhypothese $H_{1,0}$ kann auf Grund der signifikanten Ergebnisse verworfen werden. Es besteht somit großer Grund zur Annahme, dass die Alternativhypothese $H_{1,1}$ zutrifft und die L1-Rater die Aussprache der L2-Sprecher signifikant anders bewerten als die der L1- und L2-Kontrollsprecher. Da die uV der L2-Kontrollsprecher und der L2-Lerner beide das L2-Rauschen beinhalten und sich lediglich in den Fehlern unterscheiden, kann dieses Ergebnis dahingehend interpretiert werden, dass die getesteten Fehler einen Unterschied in der Bewertung von L2-Aussprache machen.

Neben der Überprüfung der Arbeitshypothese lassen sich weitere Rückschlüsse aus den Ergebnissen ziehen. So ist es auffällig, dass die L2-Kontrollsprecherin über alle Stimuli hinweg in allen Bewertungsdimensionen schlechter als die L1-Kontrollsprecherin bewertet wird, obwohl ihre Sprachaufnahmen fehlerfrei sind, d. h. die überprüften Merkmale in richtiger Realisation auftreten und auch sonst keine weiteren segmentalen oder suprasegmentalen Fehler vorhanden sind. Außerdem wird sie jedoch über alle Stimuli hinweg in allen Bewertungsdimensionen besser als die L2-Lerner bewertet. Die L2-Kontrollsprecherin hat ein höheres Sprachniveau als die L2-Lerner. Dieser Umstand bietet Grund zu der Annahme, dass eine Verbesserung im Sprachniveau mit einer besseren Bewertung innerhalb der Dimensionen einhergeht.

Die minimalen Werte für *acceptability* der L1-Kontrollsprecherin liegen bei 1,04 bei Item 1 und 1,08 bei Item 2. Die L2-Kontrollsprecherin erreicht Werte von 1,31 bei Item 1 und 1,67 bei Item 2. Ob die L2-Kontrollsprecherin ein Sprachniveau erreicht hat, auf dem die L1-Rater sie als akzeptable Kommunikationspartnerin einstufen, ist nicht eindeutig zu bestimmen. Die Werte der L2-Lerner liegen im Mittel bei 1,90 für Item 1 und 2,70 für Item 2. Diese Ergebnisse lassen sich dahingehend zusammenfassen, dass die L2-Kontrollsprecherin ein Sprachniveau erreicht hat, dass akzeptabler als das der L2-Sprecher ist, wobei sie noch immer als nichtmuttersprachlich erkannt wird. So unterscheidet sie sich in den Ratings zur *accentedness* sogar in 1,91 Skalenpunkten für Item 1 und 2,01 Skalenpunkten für Item 2. Am geringsten ist der Unterschied in der aV *comprehensibility* zu verzeichnen. Die L2-Kontrollsprecherin unterscheidet sich hier nur in 0,39 Skalenpunkten für Item 1 und 0,45 Skalenpunkten für Item 2 von der L1-Kontrollsprecherin. Die L2-Lerner unterscheiden sich in mehr als einem Skalenpunkt (Item 1 = 1,31 und Item 2 = 1,46) von der L1-Kontrollsprecherin.

Die durchgeführten Analysen beziehen sich auf die Unterschiede innerhalb der unterschiedlichen Stimulustypen und die übergreifende Untersuchung der Mittelwerte über alle untersuchten Merkmale hinweg. Dabei bleibt unbeantwortet, wie sich die Unterschiede bezogen auf die einzelnen Fehler darstellen. Diese Fragestellung ist Schwerpunkt der folgenden Ergebnisauswertung und -analyse. So werden in Kapitel 4.2.3.3 die Mittelwertunterschiede der einzelnen Fehler untersucht und die Hypothese $H_{2,1}$ geprüft, die aussagt, dass sich die unterschiedlichen Fehler von den L1-Ratern verschieden bewertet werden.

4.2.3.3 Mittelwertunterschiede der einzelnen Fehler

Aus den vorangegangenen Analysen ergibt sich, dass sich die Bewertung der Stimuli mit Fehlern (L2-Lerner) signifikant von den Bewertungen der Kontrollstimuli (L1-Kontrollsprecher und L2-Kontrollsprecher) unterscheidet. Im Folgenden wird untersucht, wie sich die Bewertung bezogen auf die einzelnen Fehler darstellt. Damit wird die Hypothese $H_{2,1}$ überprüft, die davon ausgeht, dass sich die Fehler in den Mittelwerten signifikant unterscheiden. Innerhalb der Untersuchung werden die Mittelwerte der Bewertungen aller Fehler miteinander verglichen. Dies wird für jede Bewertungsdimension durchgeführt.

Zum Vergleich von Mittelwertunterschieden bieten sich verschiedene Strategien an. Eine Möglichkeit stellt die Anwendung von Post-hoc-Tests dar, die ähnlich wie *t*-Tests und Varianzanalysen den statistischen Vergleich von Mittelwerten ermöglichen (Bühner & Ziegler 2012: 549). Allerdings berücksichtigt diese Form von Tests beim Vergleich, dass die Standardfehler der einzelnen Vergleichsgruppen unter Umständen nicht unabhängig voneinander betrachtet werden können und vermeiden somit die Kumulierung von Messfehlern, auch α -Fehler-Inflation genannt (Bühner & Ziegler 2012: 550). Als α -Fehler wird die statistische Fehlentscheidung beschrieben, von einem Effekt auszugehen, obwohl keiner vorhanden ist (Bortz & Döring 2006: 106).⁴⁵ Post-hoc-Tests wie beispielsweise die Bonferroni-Korrektur erlauben es, „beliebig viele Einzelvergleiche durchzuführen, ohne eine α -Fehler-Inflation zu riskieren“, da ein korrigiertes alternatives α -Niveau (auch adjustiertes α) zum Vergleich eingesetzt wird (Bühner & Ziegler 2012: 554). Das adjustierte α wird für alle Einzelvergleiche insgesamt errechnet und bei steigender Vergleichszahl korrigiert (= kleiner). Dadurch wird bei einer hohen Anzahl von zu untersuchenden Mittelwerten der Einsatz dieser Vorgehensweise leider problematisch. So kann die korrigierte α -Fehlerwahrscheinlichkeit unter Umständen sehr klein werden, was

⁴⁵ α (auch Signifikanzniveau oder Alpha-Fehler-Niveau) ist das Maß für den Fehler und Entscheidungskriterium zur Annahme einer Hypothese. Es zeigt an, wie wahrscheinlich es ist, dass ein statistischer Irrtum vorliegt, wenn die Nullhypothese verworfen wird (Fehler 1. Art oder auch α -Fehler). α wird als Entscheidungskriterium konventionell auf ein Signifikanzniveau von 5 % festgesetzt (Bortz & Döring 2006 665).

dazu führt, dass auch das Ergebnis sehr geringe Werte aufweisen muss, um als statistisch signifikant gewertet werden zu können.⁴⁶

Eine Lösung für dieses Problem bietet das Bonferroni-Holm-Verfahren, bei dem das adjustierte α sich nicht als Gesamtwert aus der Berechnung aller Einzelvergleiche ergibt, sondern für die nach absteigender Größe sortierten Mittelwerte paarweise errechnet wird (Bühner & Ziegler 2012: 554).⁴⁷ Settineri (2011: 71) nutzt eine ähnliche Methode, indem sie für ihren Mittelwertvergleich Bonferroni-korrigierte t -Tests rechnet. Sie vergleicht die unterschiedliche Bewertung von L1-Ratern von neun unterschiedlichen Fehlertypen, stellt diese allerdings nicht in allen möglichen Konstellationen gegenüber, sondern lediglich in spezifisch ausgewählten Fehlertypenpaaren mit stark unterschiedlichen Mittelwerten. Diese Vorgehensweise ermöglicht erstens die Rechnung von t -Tests und zweitens einen übersichtlicheren Vergleich der relevanten Fehlerpaare. Es lässt sich jedoch hinterfragen, ob es relevant ist, die Signifikanz jedes einzelnen Fehlervergleichs zu messen. So hat die signifikante Unterschiedlichkeit zwischen individuellen Fehlern unter Umständen keine praktische Bedeutung. Relevant wäre vielmehr eine Untersuchung dahingehend, welche Art von Fehlern niedriger bzw. höher bewertet wird, um Aussagen darüber treffen zu können, ob es sich beispielsweise um ‚leichte‘ oder eher ‚schwere‘ Fehler handelt. Anstatt vieler empirisch geleiteter Einzelvergleiche sollte also vielmehr eine inhaltliche Analyse von Fehlergruppen gleicher Fehlerschwere erfolgen. Diese Art der Untersuchung wird für die vorliegende Arbeit gewählt. So werden die Fehler innerhalb der jeweiligen Dimensionen anhand ihrer Fehlerschwere in Gruppen eingeteilt. Die Fehlerschwere wird auf Grundlage der Mittelwerte der einzelnen Fehler der Bewertung der L1-Rater bestimmt. Zum Nachweis der signifikanten Mittelwertunterschiede zwischen den Fehlern insgesamt werden Konfidenzintervalle (auch Vertrauensintervalle) genutzt. Konfidenzintervalle geben eine Spannweite an, von der man annimmt, dass sie neben dem gemessenen Mittelwert einer Stichprobe den Mittelwert der Grundgesamtheit mit einer gewissen Wahrscheinlichkeit mit einschließt (Bühner & Ziegler 2012: 163). Sie erlauben es, ohne Angaben abstrakter statistischer Wahrscheinlichkeiten „konkret die Genauigkeit der Parameterschätzungen durch die angegebenen Grenzen“ darzustellen (Bühner & Ziegler 2012: 192). So werden anhand der in die Darstellung integrierten Konfidenzintervalle signifikante Unterschiede zwischen den Fehlern bzw. sich ergebene Fehlergruppen trotz der Vielzahl der untersuchten Fehler leicht ersichtlich. Zur Berechnung von Konfidenzintervallen werden die Eigenschaften der Normalverteilung genutzt. Dabei wird konventionell davon ausgegangen, dass im Bereich $\pm 1,96$ Standardabweichungen um den Mittelwert 95 % aller Werte liegen (Bühner & Ziegler 2012: 163). Der

⁴⁶ Bühner & Ziegler (2012: 554) zeigen, dass bereits bei fünf Gruppen insgesamt 10 Einzelvergleiche nötig sind, bei denen $p = 0,005$ nicht überschritten werden dürfte, damit die Effekte als statistisch signifikant beurteilt werden könnten. Im Fall der vorliegenden Arbeit handelt es sich um 23 zu vergleichende Gruppen (Fehler), das adjustierte α würde daher sehr klein ausfallen. Dieses Verfahren ist deshalb für die vorliegende Arbeit nicht praktikabel.

⁴⁷ Für eine ausführlichere Darstellung dieses Verfahrens vgl. Bühner & Ziegler (2012: 554–555).

Standardfehler der zu untersuchenden Mittelwerte wird mit 1,96 multipliziert. Das Ergebnis definiert oberhalb und unterhalb des Mittelwerts die Grenzen (auch Intervall), innerhalb derer der wahre Wert liegt. Wenn sich zwei Konfidenzintervalle nicht überschneiden, kann mit einer 95 %igen Wahrscheinlichkeit davon ausgegangen werden, dass ein signifikanter Unterschied zwischen zwei Mittelwerten vorliegt und sich die Fehlerbewertung somit bedeutsam unterscheidet (Bühner & Ziegler 2012: 267). Die grundlegende Vorgehensweise wird anhand eines Ausschnitts der Mittelwerte für Item 1 *acceptability* erläutert (vgl. Abb. 4.3).

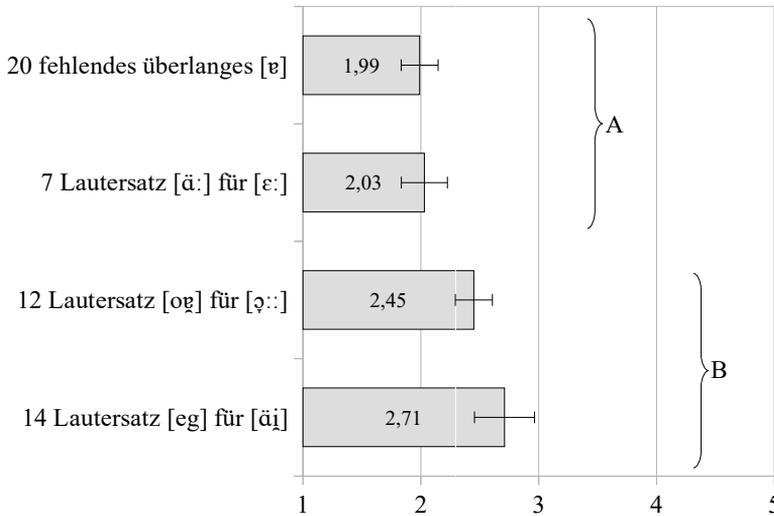


Abb. 4.3: *Acceptability* Item 1: Ausschnitt zur Darstellung der Untersuchung von signifikanten Mittelwertunterschieden anhand von Konfidenzintervallen

Auf der *y*-Achse befinden sich die Fehlernummern und -bezeichnungen, auf der *x*-Achse kann die durchschnittliche Fehlerbewertung auf der 5-stufigen Likert-Skala nachvollzogen werden. Dargestellt sind die Mittelwerte der Stimuli mit den zu überprüfenden Fehlern 20, 7, 12 und 14. In der Abbildung finden sich die Bezeichnung der Fehler sowie der genaue Mittelwert als Zahlenwert (z. B. Fehler 20 wird im Mittel mit 1,99 von den L1-Ratern für Item 1 *acceptability* bewertet). Die mittlere Bewertung des Items über alle Fehler hinweg wird anhand der dünnen weißen Linie angezeigt. Außerdem wird das Konfidenzintervall graphisch angezeigt.⁴⁸ Die Konfidenzintervalle von Fehler 20 und Fehler 7 überschneiden sich. Die Mittelwerte (und die Bewertung) können nicht als signifikant unterschiedlich eingestuft werden, d. h. die Fehler können auf Grund ihrer gemeinsamen Fehlerschwere in dieselbe Gruppe (A) eingeordnet werden. Ebenso verhält es sich

⁴⁸ Die genauen Werte der Konfidenzintervalle lassen sich im Anhang B.3 nachvollziehen.

mit Fehler 12 und Fehler 14 (= Gruppe B). Die Mittelwerte der beiden Gruppen sind signifikant unterschiedlich voneinander (keine Überschneidung der Konfidenzintervalle). Außerdem können die Fehler aus Gruppe A im Verhältnis zu den Fehlern aus Gruppe B als von den L1-Ratern niedriger bewertet und als im Verhältnis leichtere Fehler beschrieben werden.

Die erhobenen Mittelwerte aller Fehler lassen sich weniger leicht voneinander differenzieren und in Gruppen einteilen, als das Beispiel in Abb. 4.3 vorzugeben scheint, weil insgesamt sehr viel stärkere Überschneidungen zwischen den Konfidenzintervallen der einzelnen Fehler bestehen. Dadurch entstehen weniger klar voneinander abgrenzbare Gruppen; vielmehr bildet sich ein Kontinuum. Gruppenbildung ist deshalb zwar methodisch wie oben diskutiert möglich, muss aber auf zusätzlichen Entscheidungskriterien aufbauen.

Eine Einteilung könnte anhand der Skalenpunkte (1–5) erfolgen, die durch den Fragebogen vorgegeben sind: Absolut ähnlich bewertete Fehler würden so gleich eingruppiert. So befinden sich die Mittelwerte zur *acceptability* bei Item 1 beispielsweise überwiegend im Bereich der unteren zwei Skalenpunkte (zwischen 1,16 und 2,71), d. h. Rater bewerten die Fehler insgesamt tendenziell als recht akzeptabel. Eine entsprechende Gruppenbildung würde diesen Umstand abbilden.

Allerdings zielt die Untersuchung weniger auf die absolute Bewertung von Fehlern ab als auf ein Ranking der Fehler untereinander, also auf die Frage, ob ein Fehler als schwerer bewertet wird als ein anderer. Die im Folgenden tatsächlich umgesetzte Gruppenbildung erfolgt deshalb mit Blick auf signifikante Unterschiede zwischen den Daten in dem Bereich, in dem sie jeweils tatsächlich liegen.

Dabei werden in einem ersten Schritt zunächst die Randgruppen (relativ am leichtesten und am schwersten bewertete Fehler) festgelegt, indem derjenige Wert als Referenzwert festgelegt wird, bei dem der am äußersten Rand stehende Fehler noch in die Grenzen des Konfidenzintervalls fällt. Im Beispiel in Abb. 4.4 ist dies für die Gruppe mit den niedrigsten Mittelwerten (A) der Fehler 1. Alle Fehler, die innerhalb der Grenzen des Konfidenzintervalls von Fehler 1 liegen, werden nicht signifikant anders bewertet und werden somit zur selben Gruppe gezählt. Die Bewertung der Fehler in Gruppe A fällt im Verhältnis zu den anderen Gruppen eher leicht aus. Ab Fehler 9 findet eine signifikant höhere Bewertung statt.

4 Studie 2: Bewertung der L2-Fehler deutschsprachiger Lerner durch dänische L1-Sprecher

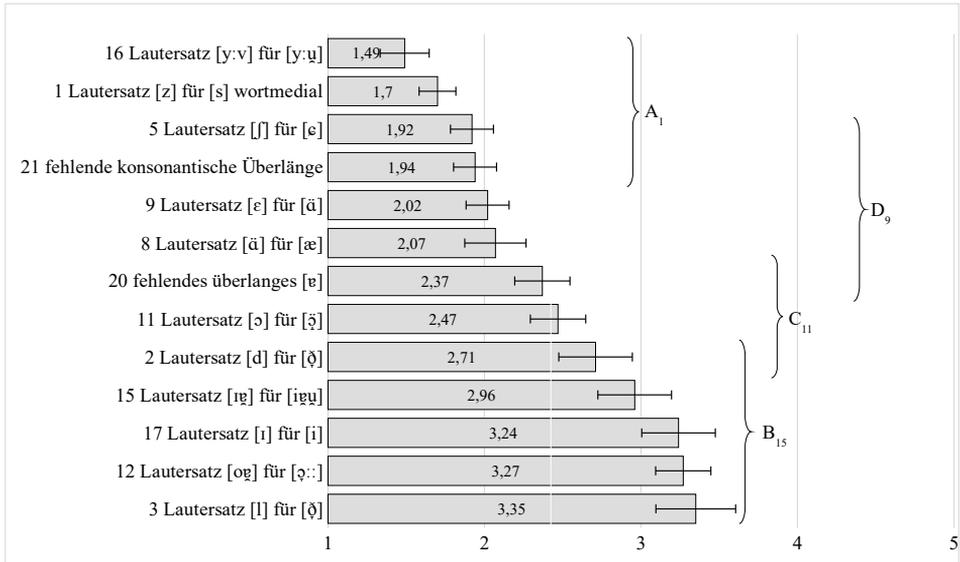


Abb. 4.4: *Comprehensibility* Item 1: Ausschnitt zur Veranschaulichung der Gruppenbildung

Als Referenzwert für die am höchsten bewertete Gruppe (B) wird der Fehler 15 genommen. Dieser stellt den am niedrigsten bewerteten Fehler dar, dessen Konfidenzintervall den am höchsten bewerteten Fehler (3) noch mit einschließt. Im Verhältnis zur Bewertung der Fehler der anderen Gruppen werden die Fehler aus Gruppe B als schwer eingeordnet. Nachdem die äußeren Gruppen (A und B) festgelegt wurden, wird im nächsten Schritt die Gruppe der mittelschweren Fehler bestimmt. Diese orientiert sich am mittleren Wert der Gesamtbewertung (in diesem Fall $M = 2,42$; in der Abbildung an der weißen Linie zu erkennen). So wird Gruppe C um den Fehler 11 ($M = 2,47$) gebildet. Der übrig gebliebene Fehler 9 bildet den Referenzwert für die letzte Gruppe (D). Die Markierungen zur Gruppenbezeichnung zeigen deutlich, dass sich die Fehler der unterschiedlichen Gruppen teilweise überschneiden. Dieser Umstand scheint die Interpretation zu erschweren, bildet allerdings die Wirklichkeit der Daten ab, die sich nicht als klar voneinander abgrenzbare Gruppen, sondern fließende Übergänge gestalten. Teilweise ermöglicht die Datenverteilung lediglich die Einteilung von zwei Gruppen. Eine solche, dichotome Beschreibung bildet allerdings in einigen Fällen nicht das gesamte Spektrum an analysierbaren Erkenntnissen ab, die die Daten liefern. Die in einem solchen Fall getroffenen inhaltlichen Entscheidungen werden in Abb. 4.5 dargestellt.

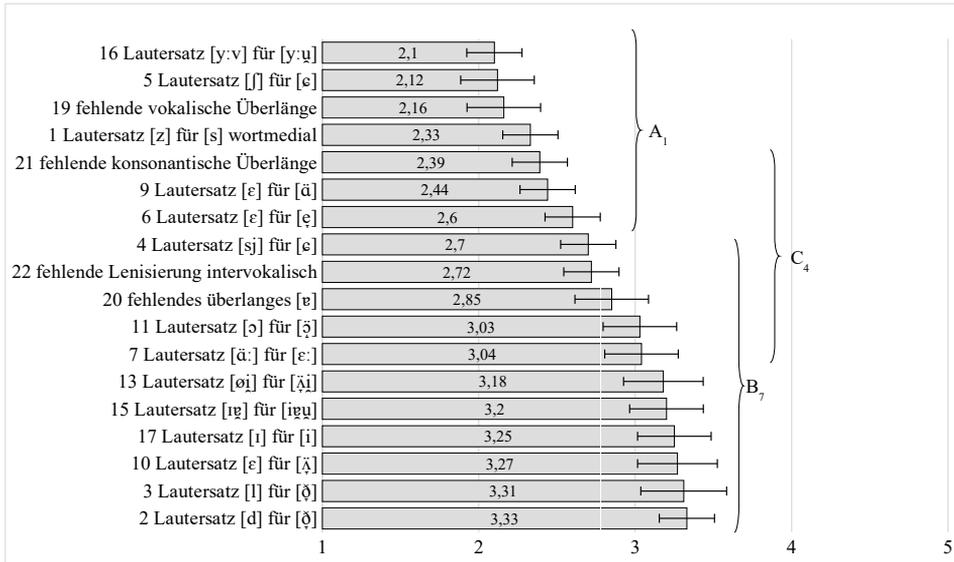


Abb. 4.5: *Acceptability* Item 2: Ausschnitt zur Veranschaulichung der Gruppenbildung

Anhand der Datenverteilung lassen sich auf praktische Weise zwei Gruppen (A und B) erstellen. Damit wäre die Bandbreite der Mittelwerte abgedeckt und die Ergebnisse ließen sich auf die Weise interpretieren, dass es zwei Fehlergruppen gibt, eine akzeptablere und eine eher wenige akzeptable. Mit dieser Interpretation wären die sich aus den Daten ergebenden Erkenntnisse allerdings noch nicht ausgeschöpft. So lässt sich anhand des Konfidenzintervalls von Fehler 4 eine mittlere Fehlergruppe (C) erstellen, die eine Vereinigungsmenge aus Fehlern aus Gruppe A und Fehlern aus Gruppe B darstellt. Diese Fehler können als mittelschwer im Vergleich zu den anderen beiden Gruppen gewertet werden.

Nachdem die Fehler auf Grundlage der Fehlerschwere in Gruppen eingeteilt wurden, erfolgt die inhaltliche Interpretation der sich in den jeweiligen Gruppen befindenden Fehler. Diese erfolgt unter Berücksichtigung der folgenden Aspekte:

- lautliche Kategorie (Vokale, Konsonanten, Polyphthonge, Prozesse);
- Ursache für das Auftreten der Fehler;
- Höreindruck der L1-Rater.

Zunächst werden die Gruppen daraufhin untersucht, ob Tendenzen in der phonetischen Strukturierung bemerkbar sind. Dazu wird differenzierter analysiert, ob es sich um Konsonanten, Vokale oder Polyphthonge bzw. nicht durchgeführte lautliche Prozesse handelt. Im zweiten Interpretationsschritt (b) werden die Fehlerursachen genauer untersucht. Diese können in drei Hauptursachen aufgeteilt werden. Erstens können Fehler auf rein artikulatorischen Schwierigkeit beruhen. Davon ist insbesondere auszugehen, wenn

der geprüfte Ziellaut in der L1 nicht vorkommt. Zu Fehlern dieser Art zählt zum Beispiel Fehler 6:

- (24) Fehler 6: Lautersatz [ɛ] für [ɸ]
melde ‚melden‘ [ˈmɛlə] statt [ˈmɸlə]

Zweitens können Fehler auf Grund von Nichtanwendung einer lautlichen Regel entstehen. Diese Ursache ist wahrscheinlich, wenn der Ziellaut in der L1 vorkommt, vom L2-Lerner allerdings nicht realisiert wird, wie es beispielsweise bei Fehler 1 der Fall ist:

- (25) Fehler 1: Lautersatz [z] für [s]
rose ‚Rose‘ [ˈʁo:zə] statt [ˈʁo:sə]

Die L2-Lerner müssten hier für die richtige Realisierung die Regel kennen, dass stimmhaftes [z] im Dänischen nicht vorkommt und in den Zielwörtern stimmloses s [s] realisiert werden müsste. Zuletzt können Fehler auf rein orthographische Ursachen zurückgeführt werden, nämlich dann, wenn der Ziellaut keine artikulatorischen Schwierigkeiten bereiten dürfte, dem Lerner allerdings auch keine Regel zur Umsetzung der Orthographie zur Verfügung stehen kann (vgl. Beispiel (26)).

- (26) Fehler 14: Lautersatz [ɛɣ] für [äi]
regner ‚regnen (PRÄS)‘ [ˈʁɛ:ɣnɐ] statt [ˈʁäi:nɐ]

Als letzter inhaltlicher Analyseschritt (c) werden Möglichkeiten zur Interpretation des Höreindrucks der L1-Rater gegeben. Dabei können zwei Aspekte von Relevanz sein. Erstens können die Fehler der erweiterten L1-Norm entsprechen. Darunter fallen Realisierungen, die im umgangssprachlichen Gebrauch wenig frequent oder im umgangssprachlichen Rahmen markiert sind, da sie entweder eine regionale, soziale oder altersabhängige Variante darstellen. Diese Formen ähneln überwiegend ihrer orthographischen Umsetzung, deren lautliche Interpretation der deutschsprachigen L2-Lerner vermutlich die Ursache für diese Fehler darstellt. Ein Beispiel für eine im standardsprachlichen Gebrauch unübliche, aber dennoch der erweiterten L1-Norm entsprechenden Form, stellen die Realisierungen in Fehler 16, 4 und Fehler 19 dar:

- (27) Fehler 16: Lautersatz [y:v] für [yʏ]
flyver ‚fliegen (PRÄS)‘ [ˈfly:vɐ] statt [ˈflyʏʔɐ]

Fehler 4: Lautersatz [sj] für alveolopalatalen Frikativ [ç]
sjældent ‚selten‘ [sjel:nɔ] statt [çɛl:nɔ]

Fehler 19: Einfügung von [ə] mit fehlender Realisierung überlanger Vokale in betonten Silben
uge ‚Woche‘ [ˈu:ə] statt [ˈu:]

Die in Fehler 16 als fehlerhafte Form dargestellte Umsetzung existiert als veraltete aber normgerechte Realisierung des Diphthongs bei älteren Sprechern (Grønnum 2009: 295).

Die Aussprache von [ç] als [sj] in Fehler 4 stellt weniger eine veraltete als eine Variante in Form eines markierten Soziolekts dar (Grønnum 2009: 306). Fehler 19 repräsentiert eine Form, die in gesprochener Sprache eher nicht vorkommt bzw. zur Hervorhebung eingesetzt wird (Grønnum 2009: 187). Unter den zweiten Aspekt dieses Analyseschritts (c) fallen diejenigen fehlerhaften Formen, die zwar nicht der normgerechten Umsetzung des Zielworts entsprechen, aber der normgerechten Realisierung eines anderen tatsächlich existenten oder zumindest eines phonotaktisch möglichen Wortes ähneln bzw. gleichen. Dieser Umstand kann leicht zu einer inhaltlichen Verwirrung der L1-Rater führen (vgl. Beispiel (28)).

- (28) Fehler 11: Lautersatz [ɔ] für [ö]
låne ‚leihen‘ [ˈlɔ:nə] statt [ˈl̥ö:nə]

Die fehlerhafte Umsetzung [lɔ:nə] stellt ein segmental und phonotaktisch mögliches, aber nicht existentes dänisches Wort dar. Die Fehler lassen sich nach diesen Kriterien in folgender Tabelle einordnen:

4 Studie 2: Bewertung der L2-Fehler deutschsprachiger Lerner durch dänische L1-Sprecher

Fehler	lautliche Kategorie	Fehlerursache	Interpretation
1	Konsonanten	Nichtanwendung lautliche Regel	
2		artikulatorische Schwierigkeit	
3		artikulatorische Schwierigkeit	phonotaktisch möglich
4		artikulatorische Schwierigkeit	normkonform
5		artikulatorische Schwierigkeit	
6	Vokale	artikulatorische Schwierigkeit	
7		Nichtanwendung lautliche Regel	
8		artikulatorische Schwierigkeit	
9		Nichtanwendung lautliche Regel	
10		artikulatorische Schwierigkeit	phonotaktisch möglich
11		artikulatorische Schwierigkeit	phonotaktisch möglich
12		Nichtanwendung lautliche Regel	
13	Polyphthonge	orthographische Umsetzung	
14		orthographische Umsetzung	phonotaktisch möglich
15		artikulatorische Schwierigkeit	
16		artikulatorische Schwierigkeit	normkonform
17	Vokalische Quantität	Nichtanwendung lautliche Regel	
18		Nichtanwendung lautliche Regel	
19	Prozesse	Nichtanwendung lautliche Regel	normkonform
20		Nichtanwendung lautliche Regel	
21		Nichtanwendung lautliche Regel	normkonform
22		Nichtanwendung lautliche Regel	
23		Nichtanwendung lautliche Regel	

Tab. 4.16: Übersicht zur inhaltlichen Interpretation

Die Ergebnisse der Vergleiche der statistisch signifikanten Mittelwertunterschiede durch die Konfidenzintervalle in den einzelnen Bewertungsdimensionen werden in den anschließenden Kapiteln auf die oben beschriebene Art und Weise dargestellt.

Accentedness

Die Darstellung der signifikanten Mittelwertunterschiede im ersten Analyseschritt innerhalb der aV *accentedness* erfolgt in zwei Abbildungen. Abb. 4.6 zeigt die Mittelwertunterschiede aller Fehler in Item 1, während Abb. 4.7 die Mittelwertunterschiede aller Fehler in Item 2 darstellt. Alle nachfolgenden Tabellen sind so aufgebaut, dass auf der y-Achse die Fehler mit Nummerierung und Bezeichnung dargestellt sind, während die x-Achse die durchschnittliche Bewertung auf der Likert-Skala von 1 (*helt uenig*) bis 5 (*helt enig*) angibt.

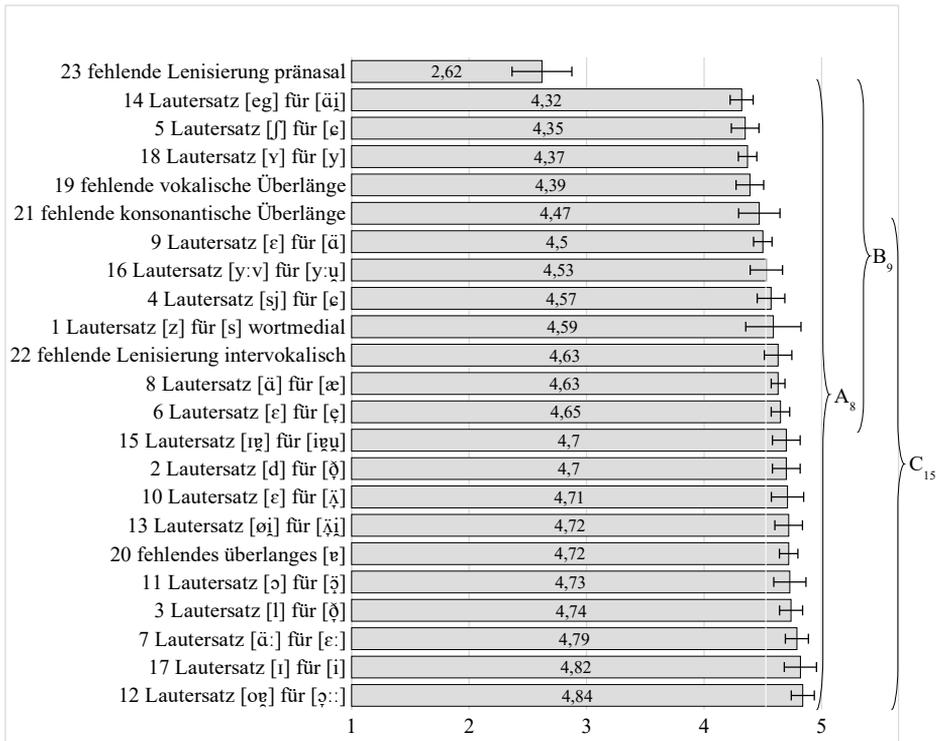


Abb. 4.6: *Accentedness* Item 1: *Han/hun taler med accent* ‚Er/sie spricht mit Akzent‘ (1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Die Gruppeneinteilung für Item 1 erklärt sich folgendermaßen: Der Mittelwert um Gruppe A stellt den Wert (Fehler 8; $M = 4,63$) dar, von dem sich keiner der anderen Feh-

ler signifikant unterscheidet.⁴⁹ Anhand von Gruppe A lässt sich also bestimmen, dass die Fehler im Mittel sehr ähnlich bewertet werden. Innerhalb dieser Gruppe lassen sich allerdings anhand von Gruppe B und C Tendenzen erkennen. So werden einige Fehler in Gruppe B, die sich um den Referenzwert von Fehler 9 ($M = 4,5$) bilden, signifikant niedriger bewertet, als einige Fehler in Gruppe C (Referenzwert Fehler 15; $M = 4,7$). Dieser Unterschied zwischen den Mittelwerten ist allerdings so gering, dass er vielleicht statistisch relevant ist, inhaltlich allerdings keine Aussagekraft hat, zumindest nicht über die eventuelle Einschätzung von Tendenzen bestimmter Fehler. Der Fehler 23 wird aus der Gruppeneinteilung (in beiden Items) ausgeschlossen, da er signifikant niedriger als die anderen Fehler bewertet wird. Bei Fehler 23 handelt es sich sprachlich um einen phonetischen Prozess, der sich als nicht realisierte Lenisierung pränasaler Plosive darstellt. Die L2-Lerner produzieren statt normsprachlich [$^h n \text{ɛ} \text{d} \text{n}$] den nicht lenisierten Laut [$^h n \text{ɪ} \text{t} \text{n}$]. Der Fehler kann auf die Nichtanwendung einer lautlichen Regel zurückgeführt werden.⁵⁰ Die verhältnismäßig niedrige Bewertung zur *accentedness* hängt möglicherweise damit zusammen, dass die L1-Rater diesen Fehler kaum wahrnehmen.

⁴⁹ Alle Mittelwerte und unteren und oberen Grenzen der Konfidenzintervalle lassen sich im Anhang B.3 nachvollziehen.

⁵⁰ Die Regel lautet, dass im Dänischen stimmlose Plosive nur vor dem akzentuierten Vokal vorkommen können (Höder 2016: 64).

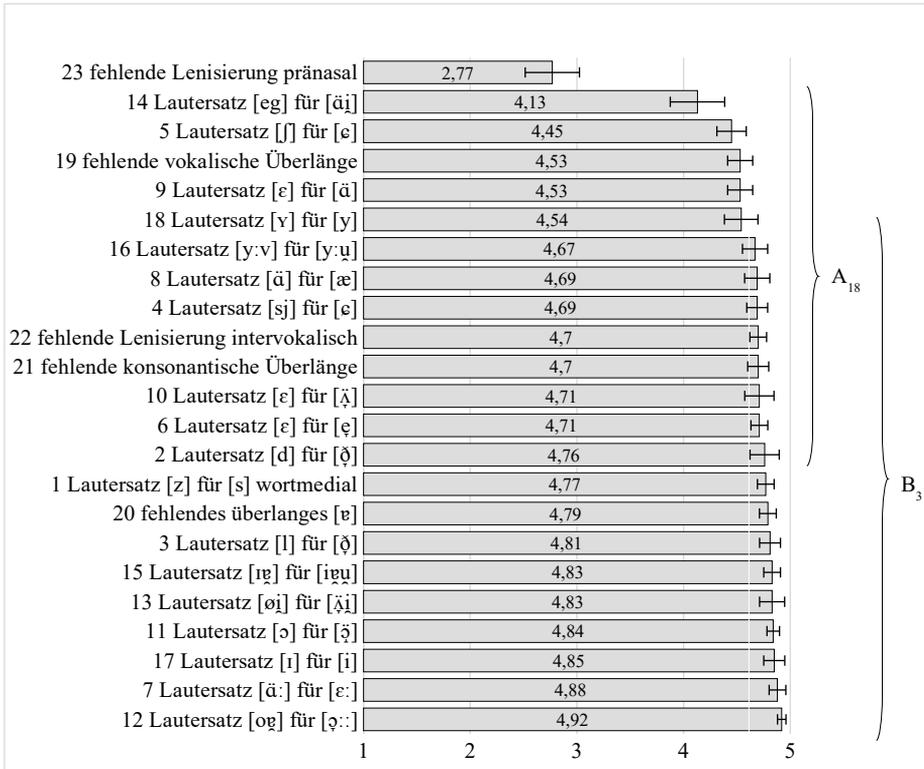


Abb. 4.7: *Accentedness* Item 2: *Han/hun kommer fra et andet land end Danmark* ‚Er/sie kommt aus einem anderen Land als Dänemark‘

(1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Die Gruppeneinteilung in Item 2 verhält sich ähnlich wie beim ersten Item, allerdings ergeben sich bei diesem Item zwei Gruppen. Gruppe A umfasst alle Fehler innerhalb der Grenzen des Konfidenzintervalls von Fehler 18 ($M = 4,54$). Die Fehler der Gruppe B orientieren sich am Referenzwert von Fehler 3 ($M = 4,81$). Auch zwischen diesen beiden Gruppen sind die Unterschiede gering und die Abstufungen lassen sich inhaltlich eher als Tendenzen, denn als aussagekräftige Unterschiede deuten. Von denjenigen Fehlern, die sich innerhalb Gruppe A aber nicht in der Schnittmenge von Gruppe B befinden, lässt sich aussagen, dass sie (Fehler 19, 9 und 5) sich signifikant vom Mittelwert von Fehler 15 unterscheiden und es sich somit um tendenziell eher niedriger akzentuiert bewertete und somit leichtere Fehler handelt.

- (29) Fehler 19: Einfügung [ə], fehlende Realisierung überlanger Vokale in betonten Silben
uge ‚Woche‘ [‘u:ə] statt [u:]

Fehler 9: Lautersatz [ɛ] für [ä]
træt ‚müde‘ [tʰɛɔ̯] statt [tʰɛ̥äɔ̯]

Fehler 5: Lautersatz [ʃ] für [ç]
chok ‚Schock‘ [ʃɔ̯] statt [çʌ̯]

Im Gegensatz dazu sind Fehler 1, 20, 3, 15, 13, 11, 17, 7 und 12, die in Gruppe B außerhalb der Schnittmenge der Gruppe A liegen, eher tendenziell negativ bewertete Fehler.

(30) Fehler 1: Lautersatz [z] für [s]
rose ‚Rose‘ [ʰɔ:zə] statt [ʰɔ:sə]

Fehler 20: Einfügung [ə], fehlende Realisierung überlanger Vokale in unbetonten Silben
lærere ‚Lehrer (PL)‘ [ʰɛ:ɛɛə] statt [ʰɛ:ɛ:]

Fehler 3: Lautersatz [l] für [ð]
nødder ‚Nüsse‘ [nøɫ] statt [nøðʰɛ]

Fehler 15: Lautersatz [ɪɹ] für [iɹ]
hvirvler ‚wirbeln (PRÄS)‘ [ʰviɹɫ] statt [ʰviɹɫ]

Fehler 13: Lautersatz [ø̥] für [ʌ̥]
løj ‚lügen (PRÄT)‘ [lø̥] statt [lʌ̥ʰ]

Fehler 11: Lautersatz [ɔ̥] für [ö̥]
låne ‚leihen‘ [ʰlɔ:nə] statt [ʰl̥ö:nə]

Fehler 17: Lautersatz [ɪ] für [i]
milde ‚mild (DEF)‘ [ʰmilə] statt [ʰmilə]

Fehler 7: Lautersatz [ä:] für [ɛ:]
maler ‚malen (PRÄS)‘ [ʰmä:lə] statt [ʰmɛ:lə]

Fehler 12: Lautersatz [o̯] für [ɔ:̯]
vores ‚unser‘ [vo̯s] statt [vɔ:̯s]

Einen Sonderfall in der niedrigen Bewertung stellt Fehler 14 dar (vgl. (30)). Dieser wird in Item 2 signifikant niedriger bewertet als die Fehler der anderen Gruppen. Er stellt inhaltlich einen besonderen Fall dar, da er das Zielmerkmal zwar stark verändert, es allerdings eine phonotaktisch mögliche Form beibehält.

(31) Fehler 14: Lautersatz [eg] für [ü̥]
regner ‚regnen (PRÄS)‘ [ʰɛ:ɛnə] statt [ʰɛ̥ü̥nə]

Die L1-Rater nehmen diesen Fehler also als nicht sonderlich akzentuiert wahr. Da dem Wort mit der fehlerhaften Realisierung trotz phonotaktischer Normentsprechung aller-

dings die lexikalische Entsprechung fehlt, ist zu erwarten, dass der Fehler innerhalb der Bewertungsdimension *comprehensibility* negative Werte erhält.

Generell lässt sich für die Bewertungsdimension *accentedness* festhalten, dass es sich bei dieser vermutlich eher um eine dichotom strukturierte Variable handelt und Akzent entweder wahrgenommen wird oder nicht. So ist die Abstufung durch die Mittelwerte von Fehler 14 und 23 vermutlich nicht auf einen graduellen Unterschied zurückzuführen. Bei Fehler 14 kann davon ausgegangen werden, dass ein Teil der Probanden sich von der phonotaktischen Normentsprechung leiten lässt und entsprechend den Stimulus als eher wenig akzentuiert bewertet, während der andere Teil den Fehler als solchen wahrnimmt. Ähnlich verhält es sich mit Fehler 23, bei dem ein Teil der Probanden den Fehler wahrnimmt, während dies für den anderen Teil der Probanden nicht zutrifft. Auf dieses Antwortverhalten und die dichotome Struktur der Variable deuten ebenfalls die weiten Konfidenzintervalle bei den beiden Fehlern hin (wenn die Standardabweichung groß ist, liegt Streuung im Antwortverhalten der L1-Rater vor).

Comprehensibility

Die signifikanten Mittelwertunterschiede werden innerhalb der aV *comprehensibility* im ersten Analyseschritt ebenfalls in zwei Tabellen dargestellt. Abb. 4.8 zeigt die Mittelwertunterschiede aller Fehler in Item 1, während Abb. 4.9 die Mittelwertunterschiede aller Fehler in Item 2 darstellt. Die Ausprägungen der Mittelwertunterschiede der einzelnen Fehler verteilen sich über beide Items sehr ähnlich. Die beiden Items bilden die Dimension also relativ gleich ab.

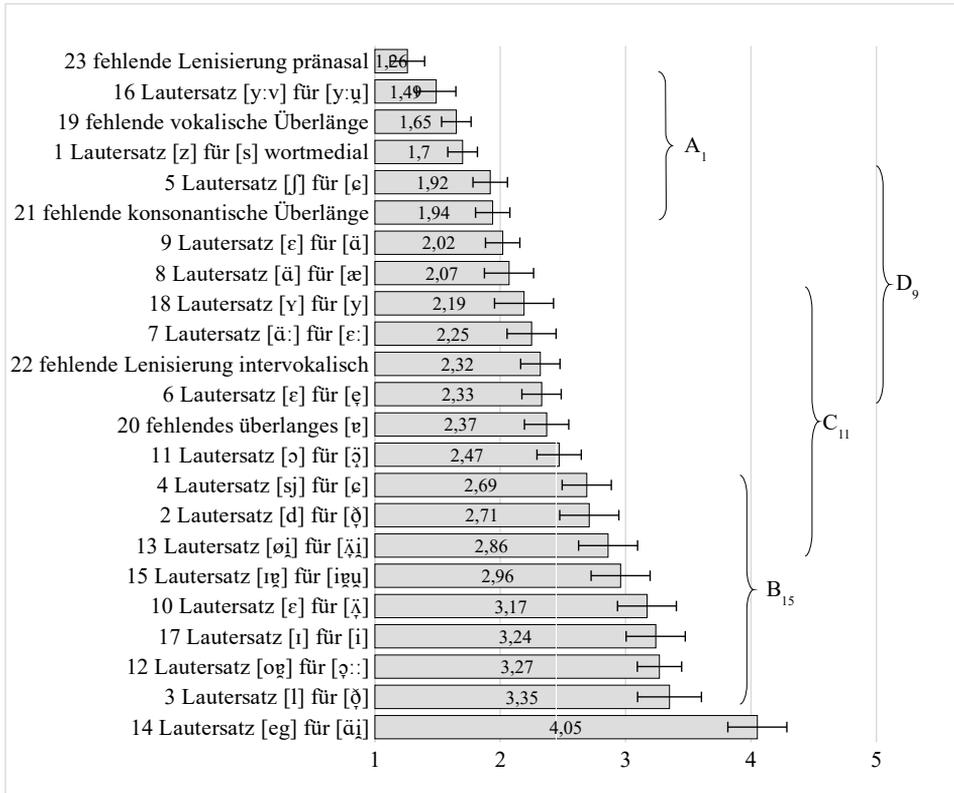


Abb. 4.8: *Comprehensibility* Item 1: *Det er svært at forstå, hvad han/hun siger* ‚Es ist schwer zu verstehen, was er/sie sagt‘ (1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Aus der Betrachtung der Konfidenzintervalle der Mittelwerte geht hervor, dass sich im Gegensatz zur Dimension *accentedness* das Verteilungsbild der Fehlermittelwerte diverser darstellt. Anhand der Konfidenzintervalle werden 4 Gruppen festgelegt. Fehler 23 mit der niedrigsten und Fehler 14 mit der höchsten Bewertung werden keiner Gruppe zugeordnet. Sie werden als Ausreißer gesondert betrachtet.

Gruppe A umfasst alle Fehler innerhalb der Grenzen des Konfidenzintervalls von Fehler 1 ($M = 1,7$). Die Mittelwerte dieser Gruppe sind im Verhältnis zu den anderen Gruppen eher niedrig, weshalb die zugehörigen Fehler (16, 19, 1, 5, 21) als eher leichte Fehler betrachtet werden können. Die Fehler der Gruppe B (Fehler 4, 2, 13, 15, 10, 17, 12, 3) stellen die am negativsten bewertete Gruppe dar und ergeben sich aus dem Konfidenzintervall von Fehler 15 ($M = 2,96$). Die mittleren Gruppen werden anhand des Mittelwerts der Gesamtverteilung der Fehler von Gruppe A und B ($2,42$) festgelegt. So dient als Referenzwert für Gruppe C (Fehler 18, 7, 22, 6, 20, 11, 4, 2, 13) Fehler 11 ($M = 2,47$). Grup-

pe D (Fehler 5, 21, 9, 8, 18, 7, 22, 6) umschließt die noch nicht verteilten Fehler und orientiert sich als Referenzwert am Fehler 9 ($M = 2,02$). Eher leicht bewertete Fehler stellen diejenigen Fehler aus Gruppe A dar, die sich nicht innerhalb der Schnittmenge von Gruppe D befinden. Diese sind in Beispiel (32) dargestellt.

(32) Fehler 1: Lautersatz [z] für [s]
rose ‚Rose‘ [ˈʁo:zə] statt [ˈʁo:sə]

Fehler 19: Einfügung von [ə] mit fehlender Realisierung überlanger Vokale in betonten Silben

uge ‚Woche‘ [ˈu:ə] statt [u:]

Fehler 16: Lautersatz [y:v] für [y̥]
flyver ‚fliegen (PRÄS)‘ [ˈfly:və] statt [ˈfly̥v̥]

Die Fehler lassen sich sprachlich strukturell nicht einer gemeinsamen Kategorie zuordnen. Während Fehler 16 die Polyphthonge betrifft, handelt es sich bei Fehler 19 um einen nicht umgesetzten phonetischen Prozess und bei Fehler 1 (Lautersatz [z] für [s]) um einen konsonantischen Fehler. Hinsichtlich der Fehlerursache besteht zwischen den Fehlern, die aufgrund ihrer Fehlerschwere derselben Gruppe zugeordnet wurden, nur bedingt Gemeinsamkeit. Während Fehler 16 ein artikulatorisches Problem der L2-Lerner darstellt, resultieren Fehler 19 und Fehler 1 aus der Nichtanwendung einer lautlichen Regel.⁵¹ Zur Interpretation des Höreindrucks der L1-Rater lässt sich für zwei der Fehler feststellen, dass es sich um Realisierungen der erweiterten L1-Norm handelt. So können Fehler 16 (Lautersatz [y:v] für [y̥]) und Fehler 19 auch im Dänischen auftreten. Zu bedenken ist, dass der Fehler auf Grund des Untersuchungssettings besser bewertet wird, als dies in der Wirklichkeit der Fall wäre. So ist vorstellbar, dass die L1-Rater die veraltete oder nicht umgangssprachliche Realisierung in Form von [ˈfly:və] für [ˈfly̥v̥] (*flyver* ‚fliegen (PRÄS)‘ oder [ˈli:ə] statt [li:]) (*lige* ‚eben‘) innerhalb eines Stimulus tolerieren. Würde der Fehler in mehreren Sätzen auftreten, ließe sich vorstellen, dass die Bewertung negativer ausfallen würde. Interessant ist außerdem, dass der recht auffällige Fehler 1 (statt *næse* ‚Nase‘ [ˈnɛ:sə] produziert der Sprecher [ˈnɛ:zə]) sich nicht schwerwiegend auf die Bewertung der *comprehensibility* auswirkt.

Gruppe C und D stellen die im Vergleich mittelschwer bewerteten Fehler dar. Die Fehler der Gruppe D können im Vergleich zu den anderen als mittelschwer bis leicht eingestuft werden. Sie lassen sich ebenfalls nicht auf eine sprachlich strukturelle Kategorie festlegen. Während Fehler 5 (Lautersatz [ʃ] für alveolopalatalen Frikativ [ç]) die Kategorie der Konsonanten betrifft, stellt Fehler 21 (Einfügung von [ə] mit fehlender Realisierung überlanger Konsonanten) einen nicht umgesetzten phonologischen Prozess dar. Fehler 9 (Lautersatz [ɛ] für [ä]) und 8 (Lautersatz [ü] für [æ]) betreffen die Kategorie der

⁵¹ Die Regel für Fehler 19 lautet, dass in den meisten Kontexten nachtoniges [ə] wegfällt, einhergehend mit kompensatorischen Längungen (Höder 2016: 61). Zur Darstellung der lautlichen Regel zu Fehler 1 vgl. (25).

Vokale sowie Fehler 18 (Lautersatz [ʏ] für [y]) zu der vokalischen Quantität. Auch in dieser Gruppe gestalten sich die Ursachen der Fehler unterschiedlich. Während Fehler 5 und Fehler 8 auf artikulatorischen Schwierigkeiten beruhen, liegt das Problem bei den restlichen Fehlern der Gruppe in der Nichtanwendung einer lautlichen Regel.⁵²

Zur Interpretation des Höreindrucks der L1-Rater lässt sich für Fehler 21 feststellen, dass es sich um eine Realisierung innerhalb der erweiterten L1-Norm handelt. So kann [fɛn:] (*finde* ‚finden‘) auch normsprachlich unproblematisch mit [ə] als [ˈfɛnə] realisiert werden.

Die Fehler der Gruppe C werden in Relation zu den anderen Gruppen als mittelschwer bis schwer bewertet. Zu ihnen gehören die Fehler 7, 6, 22, 20, 4, 2, 13 und 11.

(33) Fehler 7: Lautersatz [ü:] für [ɛ:]
maler ‚malen (PRÄS)‘ [ˈmü:lɐ] statt [ˈmɛ:lɐ]

Fehler 6: Lautersatz [ɛ] für [ɐ]
melde ‚melden‘ [ˈmɛlə] statt [ˈmɐlə]

Fehler 22: nicht realisierte Lenisierung intervokalisch
pakke ‚Paket‘ [ˈpükʰə] statt [ˈpügə]

Fehler 20: Einfügung [ə], fehlende Realisierung überlanger Vokale in unbetonten Silben
lærere ‚Lehrer (PL)‘ [ˈlɛ:ɐvɛɐə] statt [ˈlɛ:vɛ::]

Fehler 4: Lautersatz [sj] für [ç]
sjældent ‚selten‘ [sjɛ:nd̥] statt [çɛ:nd̥]

Fehler 2: Lautersatz [d] für [ð]
mødes ‚sich treffen (PRÄS)‘ [ˈmø:dəs] statt [ˈmø:ð:s]

Fehler 13: Lautersatz [øj] für [ʌj]
løj ‚lügen (PRÄT)‘ [løj] statt [lʌjʔ]

Fehler 11: Lautersatz [ɔ] für [ö]
låne ‚leihen‘ [ˈlɔ:nə] statt [ˈlö:nə]

Ebenso wie die anderen Gruppen enthält diese Gruppe mehrere unterschiedliche sprachlich strukturelle Kategorien. Während Fehler 6, 7 und 11 zur Kategorie der Vokale, Fehler 13 zur Kategorie der Polyphthonge und Fehler 2 und 4 zur Kategorie der Konsonanten gehören, sind Fehler 22 und 20 den phonologischen Prozessen zuzuordnen. Die Ursachen der Fehler 6 und 11 sind artikulatorische Schwierigkeiten, während Fehler 13

⁵² Zur Erklärung der lautlichen Regel von Fehler 21 vgl. die Erklärung zu Fehler 19. Fehler 9 resultiert aus der Realisierung eines falschen Lauts, der den Lernern eigentlich bekannt sein sollte ([ɛ] statt [ü]). Die lautliche Regel von Fehler 18 besagt, dass die Qualität der Konsonanten nicht an Länge gekoppelt ist und der vordere geschlossene Vokal [y] entsprechend kurz gebildet werden muss (Höder 2016: 58).

auf die rein orthographische Umsetzung zurückzuführen ist. Fehler 4 stellt eine Variante der erweiterten L1-Norm dar, deren Realisierung sich in einigen Soziolekten des Dänischen finden kann. Die restlichen Fehler (20, 22, 7) beruhen auf der Nichtanwendung einer phonetischen Regel.⁵³ Zur Interpretation des Höreindrucks der L1-Rater ist für Fehler 11 wahrscheinlich, dass die L1-Rater Schwierigkeiten beim Dekodierungsprozess erfahren, da die Realisierung [ˈlɔ:nə] (von *låne* [ˈlɔ:nə] ‚leihen‘) einem phonotaktisch möglichen Wort entspricht, das im dänischen Lexikon allerdings nicht vorkommt.

Eher schwer bewertete Fehler stellen die Fehler in Gruppe B dar, die sich nicht in der Schnittmenge von Fehler C befinden. Darunter fallen die Fehler 15, 10, 17, 12 und 3. Diese sind in (34) im Detail dargestellt.

- (34) Fehler 15: Lautersatz [ɪʁ] für [iʁu]
hvirvler ‚wirbeln (PRÄS)‘ [ˈviʁlɐ] statt [ˈviʁuɫɐ]
- Fehler 10: Lautersatz [ɛ] für [ɛ̃]
hånd ‚Hand‘ [hɛn] statt [hɛ̃nʔ]
- Fehler 3: Lautersatz [l] für [ø̃]
nødder ‚Nüsse‘ [ˈnøɫɐ] statt [ˈnøðʔɐ]
- Fehler 17: Lautersatz [ɪ] für [i]
milde ‚mild (DEF)‘ [ˈmɪɫə] statt [ˈmiɫə]
- Fehler 12: Lautersatz [oʁ] für [ɔ::]
vores ‚unser‘ [voʁs] statt [vɔ::s]

Auch innerhalb von Gruppe B sind die sprachlichen Kategorien übergreifend vertreten. Fehler 10 und 12 gehören zu den Vokale, Fehler 3 ist ebenfalls ein segmentaler Fehler im Bereich der Konsonanten und Fehler 14 und 15 gehören in den Bereich der Polyphthonge. Fehler 17 stellt eigentlich einen Fehler innerhalb der vokalischen Quantität dar, dieser wirkt sich jedoch so aus, dass der Laut sich qualitativ ändert und somit ebenfalls ein Lautersatz vorliegt.

Eine weitere relevante Beobachtung für diese Gruppe ist, dass zwei der fünf Fehler (3 und 10) zwar nicht der normgerechten Umsetzung des Zielworts, aber doch einer phonotaktisch möglichen Realisierung entsprechen. So gibt es kein dänisches Wort, dass der lautlichen Realisierung [ˈnøɫɐ] (Fehler 3 *nødder* ‚Nüsse‘, richtige Realisierung [ˈnøðʔɐ]) entspricht, phonotaktisch wäre dies allerdings möglich. Ebenso verhält es sich mit Fehler 10 ([hɛn] statt [hɛ̃nʔ] *hånd* ‚Hand‘).

⁵³ Für die Darstellung der lautlichen Regeln von Fehler 20 vgl. Beispiel (30). Die Regel zu Fehler 22 lautet, dass im Dänischen stimmlose Plosive nur vor dem akzentuierten Vokal vorkommen können (Höder 2016: 64). Fehler 7 beruht nicht auf einer artikulatorischen Schwierigkeit. So kommt der Vokal [ɛ:] auch im Standarddeutschen Lautinventar vor. Die Lerner müssen ihn allerdings richtig einsetzen.

Die Gruppenaufteilung in Item 2 stimmt annähernd mit derjenigen von Item 1 überein. In die Gruppe B, die im Verhältnis negativ bewertete Fehler umfasst, wird allerdings noch Fehler 13 aufgenommen.

(35) Fehler 13: Lautersatz [øj̥] für [ʎ̥i̥]
løj ‚lügen (PRÄT)‘ [løj̥] statt [lʎ̥i̥ʎ̥]

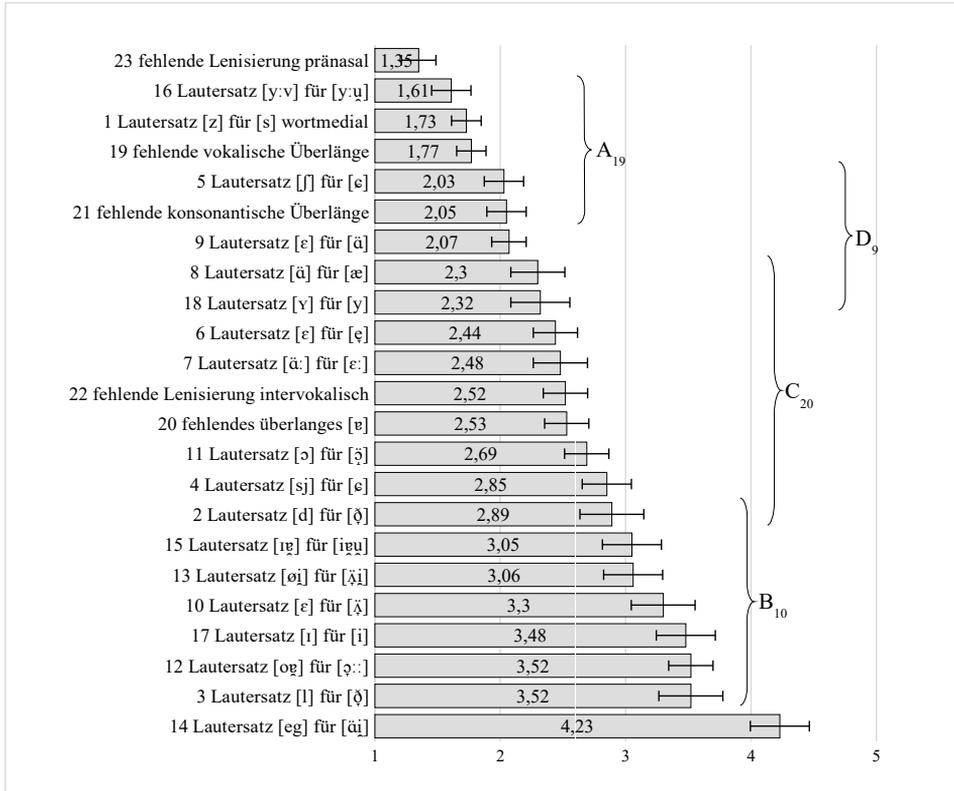


Abb. 4.9: *Comprehensibility* Item 2: *Jeg skal koncentrere mig for at forstå, hvad han/hun siger* ‚Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt‘ (1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Fehler 14 und 23 werden aus der Gruppierung in beiden Items ausgeschlossen, da diese signifikant besser bzw. schlechter als die anderen Fehler bewertet werden. Fehler 23 beeinträchtigt die wahrgenommene Verständlichkeit anscheinend im Grunde überhaupt nicht. Vermutlich kann davon ausgegangen werden, dass die L1-Rater diesen Fehler nicht wahrnehmen. Fehler 14 wirkt sich im Vergleich zu den anderen Fehlern stark negativ auf die wahrgenommene Verständlichkeit aus. Dies wurde oben bereits als Erwartung formuliert. Strukturell ist dieser Fehler in die Kategorie der fehlerhaften Polyphthonge

einzuordnen. Höchstwahrscheinlich handelt es sich bei dem Fehler um ein Artefakt aus der Methode der Stimuluserstellung, d. h. es handelt sich um einen Vorleseeffekt bzw. orthographischen Fehler. Fraglich ist deshalb, ob dieser Fehler in spontaner Sprache überhaupt auftreten würde. Auf die L1-Rater hat der Fehler vermutlich den Effekt, dass er das intendierte Wort unverständlich macht, da die lautliche Realisierung von [e:g], die in dem Stimulus den Fehler darstellt, in prototypisch phonologischen Wörtern des Dänischen phonotaktisch möglich ist, allerdings zu keiner lexikalischen Entsprechung führt. Es ist davon auszugehen, dass die hohen Werte für beispielsweise Item 2 (*Jeg skal koncentrere mig for at forstå, hvad han/hun siger* ‚Ich muss mich konzentrieren, um zu verstehen, was er/sie sagt‘) die kognitiven Anstrengungsprozesse widerspiegeln, die die L1-Rater bei dem Versuch aufbringen, ein passendes Lexem für [ˈkø:ŋnø] zu finden.

Die Analyse der Ergebnisse zum Vergleich der Mittelwertunterschiede innerhalb der aV *comprehensibility* lässt sich im Wesentlichen in folgenden Punkten zusammenfassen:

- a. Anhand der Konfidenzintervalle ergeben sich unterschiedlich schwer bewertete Fehlergruppen, die sich neben ihren unterschiedlichen Mittelwerten (Fehler-schwere) auch teilweise strukturell voneinander abgrenzen lassen.
- b. Die meisten Fehler, die der erweiterten L1-Norm entsprechen, gehören zu den leichter bewerteten Fehlern. Eine Ausnahme bildet Fehler 4 (Lautersatz [sj] für [ç]), der im Vergleich zur Bewertung der anderen Fehler einen mittelschweren bis schweren Fehler darstellt.

Acceptability

Die signifikanten Mittelwertunterschiede aller Fehler innerhalb der aV *acceptability* sind in Abb. 4.10 (alle Mittelwertunterschiede in Item 1) und Tab. 4.11 (alle Mittelwertunterschiede in Item 2) dargestellt. Die Ausprägungen der Mittelwertunterschiede der einzelnen Fehler verteilen sich über die beiden Items ungefähr gleich, d. h. die Fehlerrangfolge ist bis auf kleine Unterschiede bei beiden Items einstimmig. Anders als bei den Dimensionen *accentedness* und *comprehensibility* verteilen sich die Ausprägungen der Mittelwertunterschiede der einzelnen Fehler über die beiden Items unterschiedlich. So liegen die Mittelwerte der Fehlerbewertung in Item 1 (Tab. 4.10) deutlich unter den Mittelwerten in Item 2 (Tab. 4.11). Dieses Ergebnis bildet die Operationalisierung des Konzepts *acceptability* und die inhaltliche Ausrichtung der Items ab (vgl. Kapitel 4.2.2.3). Das Item mit der erklärten Absicht, langsamer zu sprechen (Item 2), erzielt höhere Zustimmungswerte als das Item mit der erklärten Absicht die Sprache zu wechseln (Item 1). Diese Ergebnisse entsprechen auch zuvor formulierten Erwartungen, die sich aus der inhärenten inhaltlichen Abstufung der Items ergeben.

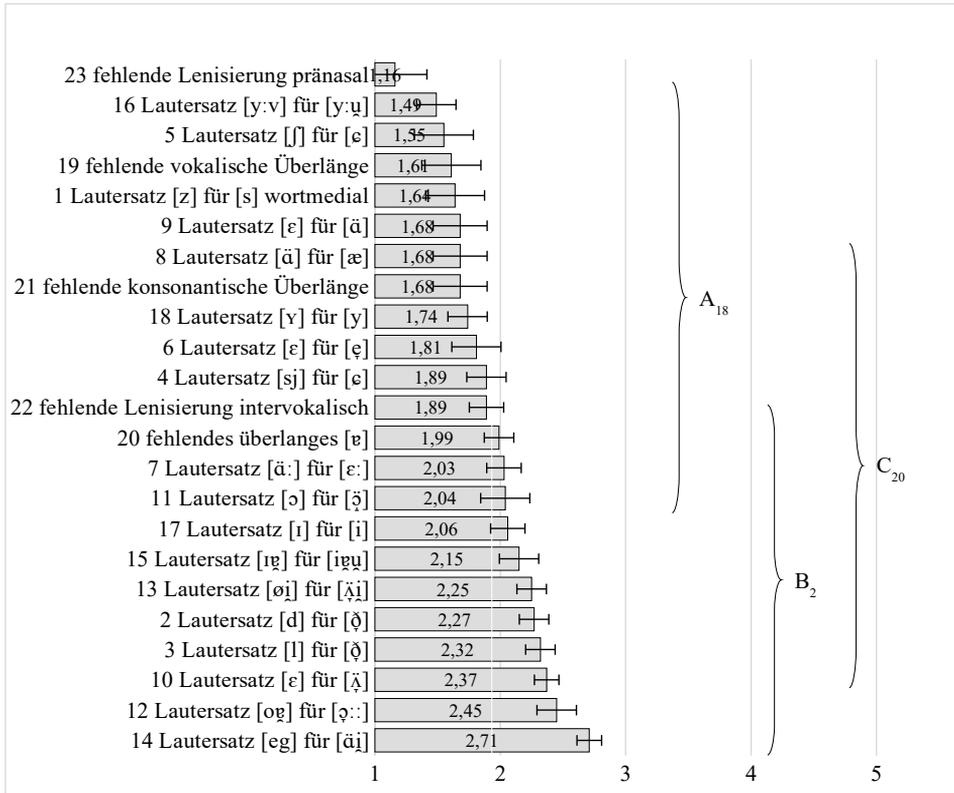


Abb. 4.10: *Acceptability* Item 1: *Jeg ville tale et andet sprog end dansk med ham/hende* ‚Ich würde eine andere Sprache als Dänisch mit ihm/ihr sprechen‘ (1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Ebenso wie innerhalb der anderen Bewertungsdimensionen verteilen sich die Mittelwerte in ihrer Gesamtheit als nicht klar voneinander in Gruppen abtrennbares Kontinuum. Die Einteilung der Gruppen beruht deshalb – wie bei den anderen Dimensionen auch – immer auf einem Referenzwert, der sich aus der Datenverteilung ergibt. Zwischen Mittelwerten unterschiedlicher Gruppen können sich allerdings Überschneidungen ergeben. In Item 1 lassen sich aus der Einteilung anhand der Konfidenzintervalle zwei Gruppen bilden, die alle Mittelwerte abdecken (A und B). Anhand des Konfidenzintervalls von Fehler 20 ($M = 1,99$), der dem Mittelwert des gesamten untersuchten Spektrums (2,1) sehr nahekommt, wird eine mittlere Fehlergruppe (C) erstellt. Diese stellt eine Vereinigungsmenge aus Fehlern von Gruppe A und B dar und bezeichnet die im Vergleich als mittelschwer bewerteten Fehler. Fehler 23 mit der niedrigsten Bewertung wird keiner Gruppe zugeordnet, sondern als Spezialfall/Ausreißer behandelt. Fehler 23 wird auch in den anderen beiden *av accentedness* und *comprehensibility* auffällig niedrig bewertet. Die

restlichen Fehler werden innerhalb ihrer Gruppen im Folgenden beschrieben und inhaltlich interpretiert. Gruppe A basiert auf dem Konfidenzintervall von Fehler 18, da dieser den höchsten Mittelwert darstellt, dessen Konfidenzintervall noch den am niedrigsten bewerteten Fehler (16) mit einschließt. Zu Gruppe A werden somit die Fehler 16, 5, 19, 1, 9, 8, 21, 18, 6, 4, 22, 20, 7 und 11 gezählt. Diejenigen dieser Fehler, die sich nicht innerhalb der Schnittmenge von Gruppe B und C befinden, können als signifikant leichter als die mittelschweren und schweren Fehler eingestuft werden. Diese somit als verhältnismäßig leichte Fehler beschriebenen Fehler sind Fehler 9, 1, 19, 5 und 16.

(36) Fehler 9: Lautersatz [ɛ] für [ä]
træt ‚müde‘ [tʰɛɐ̯] statt [tʰäɐ̯]

Fehler 1: Lautersatz [z] für [s]
rose ‚Rose‘ [ʰɔ:zə] statt [ʰɔ:sə]

Fehler 19: Einfügung [ə], fehlende Realisierung überlanger Vokale in betonten Silben
uge ‚Woche‘ [ʰu:ə] statt [u:]

Fehler 5: Lautersatz [ʃ] für [ç]
chok ‚Schock‘ [ʃɔŋ] statt [çʌŋ]

Fehler 16: Lautersatz [y:v] für [yʊ]
flyver ‚fliegen (PRÄS)‘ [ʰfly:vɐ] statt [ʰflyʊʔɐ]

Insgesamt lassen sich die Fehler sprachlich-strukturell keiner gemeinsamen Kategorie zuordnen. Bei Fehler 1 und 9 handelt es sich um Fehler auf segmentaler Ebene, während es sich bei Fehler 19 um einen phonologischen Prozess handelt. Auch bezogen auf die Fehlerursachen lassen sich keine gemeinsamen Tendenzen ausmachen. So liegen die Fehler sowohl im Bereich artikulatorischer Schwierigkeiten (Fehler 5) als auch in der Nichtanwendung lautlicher Regeln (Fehler 19, 1 und 21).⁵⁴ Für die Fehler 19 und 16 lässt sich feststellen, dass es sich um eine Realisierung der erweiterten L1-Norm handelt. So stellt Fehler 16 eine prinzipiell mögliche Realisierung dar, die zwar eher wenig frequent ist, im Prinzip allerdings der Norm entspricht.

Gruppe B definiert sich über das Konfidenzintervall von Fehler 2. Dessen Konfidenzintervall schließt den innerhalb des Gesamtratings am höchsten bewerteten Fehler noch mit ein und ist somit ein geeignetes Maß für die am schwersten bewerteten Fehler. Diese liegen in der Mittelwertspanne zwischen 1,99 und 2,71 und umfasst somit die Fehler 20, 7, 11, 17, 15, 3, 2, 10, 12 und 14. Diejenigen Fehler, die nicht in die Schnittmenge von Gruppe C fallen, werden als besonders schwere Fehler gewertet. Dazu gehören Fehler 12 und 14:

⁵⁴ Zur lautlichen Regel von Fehler 1 vgl. Beispiel (25). Zur Darstellung der lautlichen Regel von Fehler 19 und 21 vgl. die Erklärung zu Fehler 19 in Beispiel (29).

(37) Fehler 12: Lautersatz [oɤ] für [ɔ:]

vores ‚unser‘ [voɤs] statt [vɔ:s]

Fehler 14: Lautersatz [eg] für [äi]

regner ‚regnen (PRÄS)‘ [ˈɛ:gnɐ] statt [ˈɛä:ɲɐ]

Beide Fehler stellen sich als segmentale Fehler heraus. Fehler 12 ist auf Nichtanwendung einer lautlichen Regel zurückzuführen.⁵⁵ Fehler 14 zeigt wahrscheinlich die rein orthographische Umsetzung. Fehler 14 resultiert in einer phonotaktisch normkonformen Äußerung. Die Fehler, die sich als im Verhältnis mittelschwer bezeichnen lassen (Gruppe C) ergeben sich durch das Konfidenzintervall von Fehler 20. Davon eingeschlossen werden die Fehler 8, 21, 18, 6, 4, 22, 20, 11, 17, 15, 13, 2, 3 und 10.

⁵⁵ Fehler 12 resultiert nicht aus der artikulatorischen Schwierigkeit. Der Laut [ɔ] kommt auch im deutschen Lautinventar vor. Aus unterschiedlich vorstellbaren Gründen setzten die Lerner allerdings hier die falsche Lautkombination [oɤ] ein.

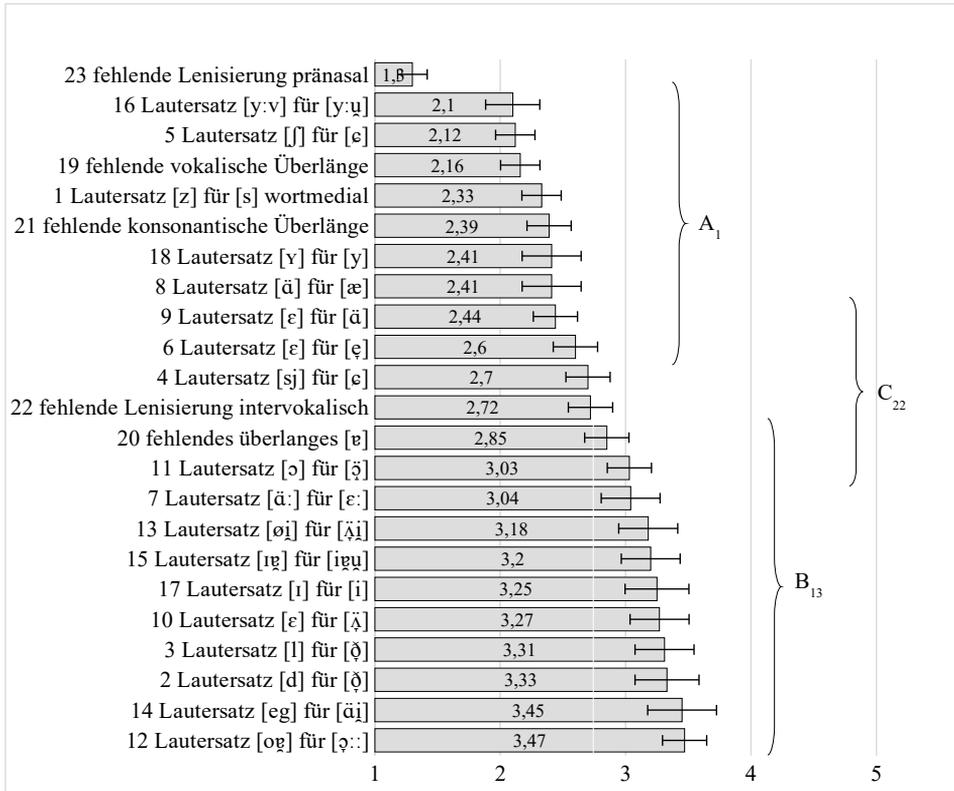


Abb. 4.11: *Acceptability* Item 2: *På dansk ville jeg tale langsomt med ham/hende* ‚Auf Dänisch würde ich langsam mit ihm/ihr sprechen‘ (1 = *helt uenig* ‚stimme nicht zu‘, 5 = *helt enig* ‚stimme voll zu‘; durchschnittliche Fehlerbewertung)

Die Gruppeneinteilung in Item 2 ist insofern anders als die Einteilung in Item 1, als sich eine kleinere Gruppe von mittelschwer gewichteten Fehler ergibt und somit eine klarere Einteilung in leichtere und schwerere Fehler möglich ist. Die Gruppe der im Verhältnis am leichtesten bewerteten Fehler (A) ergibt sich aus dem Konfidenzintervall um Fehler 1 und umfasst somit die Fehler 16, 5, 19, 1, 21, 18, 8, 9 und 6. Die Gruppe der im Verhältnis schwer bewerteten Fehler (B) bildet sich um das Konfidenzintervall von Fehler 13 und umfasst die Fehler 20, 11, 7, 13, 15, 17, 10, 3, 2, 14 und 12. Die mittelschweren Fehler (C) liegen um den Fehler 22, dessen Mittelwert ($M = 2,72$) auch dem Mittelwert der Gesamtwertung entspricht ($M = 2,74$). Die mittelschwer gewichteten Fehler sind somit Fehler 9, 6, 4, 22, 20 und 11. Die besonders leichten und schweren Fehler ergeben sich wieder aus denjenigen Fehlern der Gruppe A und B, die keine Überschneidung mit Gruppe C haben. Als Unterschied zu Item 1 gehören in Item 2 Fehler 21, 18 und 8 zu der leichten Fehlergruppe und Fehler 9 zur mittelschweren Fehlergruppe. Als besonders schwere

Fehler kommen in Gruppe B in Item 2 im Gegensatz zu Item 1 Fehler 7, 13, 15, 17, 10, 3 und 2 noch als schwere Fehler hinzu:

- (38) Fehler 7: Lautersatz [ü:] für [ε:]
maler ‚malen (PRÄS)‘ [ˈmü:lɐ] statt [ˈmɛ:lɐ]
- Fehler 13: Lautersatz [øj] für [ʌj̃]
løj ‚lügen (PRÄT)‘ [løj] statt [lʌj̃]
- Fehler 15: Lautersatz [ɪɐ] für [iɐu]
hvirvler ‚wirbeln (PRÄS)‘ [ˈvɪɐlɐ] statt [ˈviɐu]
- Fehler 17: Lautersatz [ɪ] für [i]
milde ‚mild (DEF)‘ [ˈmɪlɐ] statt [ˈmilɐ]
- Fehler 10: Lautersatz [ɛ] für [ʌ]
hånd ‚Hand‘ [hɛn] statt [hʌn]
- Fehler 3: Lautersatz [ɪ] für [ø]
nødder ‚Nüsse‘ [ˈnøɐ] statt [ˈnøøɐ]
- Fehler 2: Lautersatz [d] für [ð]
mødes ‚sich treffen (PRÄS)‘ [ˈmø:dəs] statt [ˈmø:ð:s]

Fehler 15 und 13 sind als Fehler im Bereich der Polyphthonge einzuordnen. Bei den Fehlern 3 und 2 handelt es sich um konsonantische Fehler, Fehler 10 liegt im Bereich der Vokale. Fehler 17 resultiert zwar aus Nichtumsetzung der vokalischen Quantität und ist somit eigentlich der suprasegmentalen Ebene zugeordnet, stellt aber in der Realisierung einen segmentalen Lautersatz dar. Die Fehlerursachen liegen überwiegend bei den artikulatorischen Schwierigkeiten (Fehler 3, 2, 15), ein Fehler beruht allerdings auch auf orthographischer Umsetzung (Fehler 13). Relevant ist, dass Fehler 3 zu einer Realisierung eines phonotaktisch möglichen Wortes führt, das im Wortschatz des Dänischen allerdings nicht vorkommt.

Die Analyse der unterschiedlichen Fehlerbewertung in der aV *acceptability* lässt sich in den folgenden wesentlichen Punkten zusammenfassen:

- a. Auch für die aV *acceptability* lassen sich unterschiedlich schwer bewertete Fehlergruppen anhand von Konfidenzintervallen ermitteln, die sich neben der Abgrenzung durch die Fehlerschwere, teilweise auch strukturell voneinander abgrenzen lassen.
- b. So werden Fehler, die auch als Realisierungen der erweiterten L1-Norm betrachtet werden könnten, in der aV *acceptability* relativ leicht bewertet.
- c. Fehler aus dem Bereich der phonologischen Prozesse werden unterschiedlich bewertet. Die fehlenden [ə]-Assimilationsprozesse werden einerseits leicht (Fehler 19 ([ˈu:ə] statt [u:]), andererseits mittelschwer bewertet (Fehler 20 ([ˈlɛ:ɐʊə] statt [ˈlɛʊ:])) und Fehler 21 ([ˈfɪnə] statt [ˈfɛn:])). Zu den mittelschwer bewerteten Feh-

- lern der phonologischen Prozesse gehört außerdem die fehlende Lenisierung nachtoniger intervokalischer Plosive (Fehler 22 ([^hmet^hə] statt [^hmɛdə])).
- d. Generell finden sich unter den sehr negativ bewerteten Fehlern nur segmentale Fehler. Besonders negativ werden Fehler beim weichen *d* (Fehler 2 und 3) bewertet sowie Fehler, die phonotaktisch möglich sind und ein anderes Wort ergeben könnten (Fehler 3 und 14). Die nicht durchgeführte Realisierung von kurzem [i] und stattdessen eine Realisierung von kurzem [ɪ] wird ebenfalls als schwerer Fehler eingeordnet (Fehler 17).
 - e. Segmentale Fehler, die nicht so stark ins Gewicht fallen und als leicht bewertet werden, sind der Lautersatz von [ʃ] für [ç] (Fehler 5), der Lautersatz [z] für [s] (Fehler 1), der Lautersatz [ä] für [æ] (Fehler 8), der Lautersatz [ɣ] für [y] (Fehler 18) sowie der Lautersatz [ɛ] für [ü] (Fehler 9).
 - f. Fehler, die auf Umsetzung der Orthographie beruhen, werden besonders negativ bewertet (Fehler 14 und 13).

Zusammenfassung

Die übergeordnete Fragestellung der oben durchgeführten Analysen lautete, ob L1-Rater die untersuchten Fehler unterschiedlich bewerten. Die Hypothese H_{2.1} geht davon aus, dass die unterschiedlichen Fehler in den jeweiligen Bewertungsdimensionen verschieden bewertet werden. Zur Überprüfung der Hypothese wurde die Unterschiedlichkeit zwischen den Mittelwerten der Bewertung der L1-Rater anhand von Konfidenzintervallen für alle Fehler einzeln und in jeder abhängigen Variable auf Signifikanz geprüft. Die Ergebnisse zeigen, dass die Nullhypothese der Hypothese (H_{2.0}) verworfen werden und mit großer Wahrscheinlichkeit davon ausgegangen werden kann, dass H_{2.1} zutrifft und somit die verschiedenen Merkmale von den L1-Sprechern signifikant unterschiedlich bewertet werden. Die Ergebnisse der Analysen werden im Folgenden zusammengefasst.

Innerhalb der Dimension *accentedness* besteht zwischen den Fehlern relative Gleichwertigkeit bezogen auf die Fehlerschwere. So wurden fast alle Fehler in beiden Items im Mittel zwischen 4,13 und 4,92 bewertet. Fehler 23 stellt eine Ausnahme dar, indem er von den L1-Ratern als signifikant weniger (im Mittel 2,62) akzentuiert bewertet wird.

- (39) Fehler 23: nicht realisierte Lenisierung nachtoniger Plosive pränasal
nitten ‚neunzehn‘ [^hnɛtɪn] statt [^hnɛdn]

Neben Fehler 23 wird außerdem Fehler 14 signifikant niedriger bewertet, obwohl der Fehler einen stark von Ziellaut abweichenden Lautersatz darstellt.

- (40) Fehler 14: Lautersatz [eg] für [üi]
regner ‚regnen (PRÄS)‘ [^hʁe:gnɐ] statt [^hʁüi:nɐ]

Es zeigt sich, dass diese niedrige Bewertung mit großer Wahrscheinlichkeit darauf zurückzuführen ist, dass die fehlerhafte Realisierung einem phonotaktisch möglichen Wort

ähnelt, das von den L1-Ratern lexikalisch nicht entschlüsselt werden kann. So stellt sich der Fehler zwar als relativ wenig akzentuiert heraus, in den anderen beiden Bewertungsdimensionen zeigt sich allerdings, dass er kaum verständlich und auch wenig akzeptabel wahrgenommen wird.

Innerhalb der Dimensionen *comprehensibility* und *acceptability* zeigen sich deutlichere Unterschiede zwischen den Fehlern. Allerdings stellen sich diese weniger als klar voneinander unterscheidbare Gruppen dar, sondern bilden eher ein Kontinuum von Mittelwerten. Aus diesem Grund werden einzelne Fehler als Referenzwerte festgelegt. Deren Konfidenzintervalle dienen als Grenzen zur Festlegung der unterschiedlich schwer gewichteten Gruppen. Die Gruppen staffeln sich von leichten über mittelschwere bis schwere Fehler. Die Einteilung der Gruppen ermöglicht es, diejenigen Fehler zusammenzustellen, die sowohl in der Dimension *comprehensibility* als auch in der Dimension *acceptability* als leichte bzw. schwere Fehler eingestuft wurden (vgl. Tab. 4.17).⁵⁶

	<i>comprehensibility</i>	<i>acceptability</i>
leichte Fehler	23, 16, 19, 1, 5, 21, 9, 8, 18	23, 16, 19, 1, 5, 21, 9, 8, 18
mittelschwere Fehler	7, 22, 6, 20, 11, 4, 2, 13, 15	7, 22, 6, 20, 11, 4, 17
schwere Fehler	10, 12, 3, 14, 17	10, 12, 3, 14, 2, 13, 15

Tab. 4.17: Verteilung der Fehler in den Dimensionen *comprehensibility* und *acceptability*

Die meisten Fehler ähneln sich in beiden Dimensionen in ihrer Bewertung. Lediglich die in Tab. 4.6 fett markierten Fehler werden unterschiedlich bewertet. So beeinträchtigen Fehler 2, 13 und 15 die *comprehensibility* nur mittelmäßig, wirken sich allerdings sehr negativ auf die *acceptability* aus. In (41) werden die drei als nicht akzeptabel aber mittelmäßig verständlich eingestuften Fehler dargestellt.

(41) Fehler 2: Lautersatz [d] für [ð]
mødes ‚sich treffen (PRÄS)‘ [ˈmøːdəs] statt [ˈmøːðːs]

Fehler 13: Lautersatz [ø̃] für [ʌ̃]
løj ‚lügen (PRÄT)‘ [lø̃] statt [lʌ̃]

Fehler 15: Lautersatz [ɪ̃] für [ĩ]
hvirvler ‚wirbeln (PRÄS)‘ [ˈvĩɫɐ] statt [ˈvĩɫɐ]

Bei Fehler 17 verhält die Unterschiedlichkeit zwischen *comprehensibility* und *acceptability* umgekehrt. So wirkt sich der Fehler nur mittelmäßig negativ auf die *acceptability* aus,

⁵⁶ In der Tabelle wurden die 5 Gruppen der Bewertungsdimension zusammengekürzt, indem die Gruppen A (leichte Fehler) und B (leichte bis mittelschwere Fehler) sowie die Gruppen C (mittelschwere Fehler) und D (mittelschwere bis schwere Fehler) zusammengefasst wurden.

obwohl er als schwer verständlich wahrgenommen wird. Bei dem Fehler handelt es sich eigentlich um die fehlerhaft umgesetzte Vokalquantität im Dänischen. So sind Vokalquantitäten nicht an Qualitäten gekoppelt und der gespannte [i] Laut ist, anders als im Deutschen, auch in kurzer Vokalquantität vorhanden. Die L2-Lerner mit Deutsch als L1 führen bei der Umsetzung dieses Lautes einen Qualitätswechsel durch, wodurch sich ein Lautersatz zu [ɪ] ergibt, wie (42) zeigt.

- (42) Fehler 17: Lautersatz [ɪ] für [i]
milde ‚mild (DEF)‘ [ˈmɪlə] statt [ˈmilə]

Der Fehler in der Vokalquantität stellt sich somit ebenfalls als segmentaler Fehler dar. Die übrigen Fehler verteilen sich über beide Dimensionen einstimmig auf die Gruppen leichter, mittelschwerer und schwerer Fehler. Die Gemeinsamkeiten können darauf hinweisen, dass die beiden Dimensionen *comprehensibility* und *acceptability* eventuell inhaltliche Überschneidungen aufweisen bzw. sich gegenseitig beeinflussen. Die Unterschiede verdeutlichen gleichzeitig, dass *comprehensibility* und *acceptability* trotz eventueller Nähe nicht dasselbe Konstrukt abbilden.

Die Fehlerschwere der sprachlich-strukturellen Kategorien variiert. Während unter den leichten und mittelschweren Fehler sowohl segmentale Fehler als auch Fehler innerhalb der phonologischen Prozesse vertreten sind, stellen die schweren Fehler ausschließlich segmentale Phänomene dar. Darüber hinaus zeigt sich, dass die Fehler, die aus rein orthographischer Umsetzung resultieren alle als schwere Fehler bewertet werden. Beispiel (43) zeigt die aus orthographischer Umsetzung resultierenden Fehler in der Übersicht.

- (43) Fehler 13: Lautersatz [ø̃] für [ʌ̃]
løj ‚lügen (PRÄT)‘ [lø̃] statt [lʌ̃]²
 Fehler 14: Lautersatz [eg] für [ü̃]
regner ‚regnen (PRÄS)‘ [ˈʁe:gñ] statt [ˈʁü̃ñ]

Diejenigen Fehler, die sich als phonotaktisch mögliche Realisierungen darstellen, werden ebenfalls überwiegend negativ bewertet. Diese werden in Beispiel (44) in der Übersicht dargestellt.

- (44) Fehler 10: Lautersatz [ɛ] für [ʌ̃]
hånd ‚Hand‘ [hen] statt [hʌ̃ñ]
 Fehler 3: Lautersatz [ɪ] für [ø̃]
nødder ‚Nüsse‘ [ˈnø̃] statt [ˈnø̃²ɛ]
 Fehler 14: Lautersatz [eg] für [ü̃]
regner ‚regnen (PRÄS)‘ [ˈʁe:gñ] statt [ˈʁü̃ñ]

Die fehlerhaften Realisierungen in (44) sind phonotaktisch umsetzbar, eine lexikalische Entsprechung würde allerdings fehlen. Die segmentalen Fehler lassen sich nicht weiter

kategorial differenzieren. So sind sowohl Konsonanten, Polyphthonge als auch Vokale in allen Fehlerschweregraden vertreten. In Studien, die die sprachlichen Kategorien als Ausgangslage zur Untersuchung von unterschiedlichen Bewertungen anlegen, zeigt sich häufig, dass suprasegmentale Fehler negativer auf die Bewertung von Aussprache wirken als segmentale Fehler (Anderson-Hsieh, Johnson & Koehler 1992; Hahn 2004; Derwing & Munro 2005). Für die vorliegende Studie ergibt sich auf Grund der Konstellation von Ausgangs- und Zielsprache allerdings der Fall, dass mögliche suprasegmentale Fehler, zumindest im Bereich fehlerhaften Wortakzents, zu vernachlässigen sind. So sind Fehler innerhalb dieses Bereichs im Kontext L2-Erwerb Dänisch mit Ausgangssprache Deutsch nicht zu erwarten, da sprachenpaarspezifisch keine, bis nur geringfügige Unterschiede bestehen (Höder 2016: 47). Es ist vorstellbar, dass auch im L2-Dänischen Fehler im prosodischen Bereich negativeren Einfluss auf die Bewertung von Aussprache haben, als dies ein einzelnes fehlerhaftes Segment hat. In den vorweggehenden Überlegungen der vorliegenden Arbeit wird allerdings davon ausgegangen, dass prosodische Fehler im Sprachenpaar Deutsch-Dänisch weniger ein Problem schulischen L2-Ausspracheerwerbs als vielmehr anderweitiger Sprachprobleme sind.

Die unterschiedlichen Kategorien der Segmente (Konsonanten, Monophthonge, Polyphthonge) sind in allen nach der Fehlerschwere eingeteilten Gruppen vertreten. Die untersuchten phonologischen Prozesse sind lediglich in den mittelschwer und leicht bewerteten Fehlergruppen zu finden. Als besonders schwerwiegende segmentale Fehler ergeben sich:

- a. Lautersatz für weiches *d* [ð] (Fehler 2 und 3);
- b. Lautersatz bei verschiedenen Vokalqualitäten (Fehler 10 und 12);
- c. Lautersatz bei verschiedenen Polyphthongen (Fehler 13, 14 und 15);
- d. Nicht umgesetzte Vokalquantität, die zu einem Qualitätswechsel führt (Fehler 17).

Der von allen Fehlern insgesamt am niedrigsten bewertete (auch in der Dimension *accentedness*), ist der Fehler 23, der einen phonologischen Prozess darstellt. Interessant ist, dass dieser in seiner Beschaffenheit mit Fehler 22 zu vergleichen ist. Fehler 22 wird allerdings mittelschwer bewertet.

- (45) Fehler 22: nicht realisierte Lenisierung intervokalisch

pakke ‚Paket‘ [ˈpækʰə] statt [ˈpägə]

Fehler 23: nicht realisierte Lenisierung nachtoniger Plosive pränasal

nitten ‚neunzehn‘ [ˈnɛt̪n] statt [ˈnɛdn]

Zu den phonologischen Prozessen insgesamt lässt sich feststellen, dass die nicht realisierte Lenisierung intervokalisch (Fehler 22) stärker ins Gewicht fällt, als die nicht realisierte Lenisierung pränasal (Fehler 23). Auch innerhalb der Bewertung der Fehler auf Grund von fehlenden [ə]-Assimilationsprozessen zeigen sich Unterschiede. So werden Fehler 19 und Fehler 21 niedriger bewertet als Fehler 20. Der Unterschied zwischen den drei Fehlern liegt darin, dass Fehler 19 und Fehler 21 normsprachlich korrekt im Sinne

der erweiterten L1-Norm sind, Fehler 20 hingegen keine normsprachlich korrekte Realisierungsform darstellt, auch nicht immer Sinne einer erweiterten Norm.

(46) Fehler 19: Einfügung [ə], fehlende Realisierung überlanger Vokale in betonten Silben

uge ‚Woche‘ [ˈu:ə] statt [u:]

Fehler 21: Einfügung [ə], fehlende Realisierung überlanger Konsonanten

finde ‚finden‘ [ˈfinə] statt [fɛn:]

Fehler 20: Einfügung [ə], fehlende Realisierung überlanger Vokale in unbetonten Silben

lærere ‚Lehrer (PL)‘ [ˈlɛ:ɐ̯ɐ̯ə] statt [ˈlɛ:ɐ̯:]

Generell ist zu der niedrigen Bewertung der Fehler der Schwa-Assimilation zu bedenken, dass diese in den Stimuli mit nur einem Wort abgefragt werden. Würde der Sprecher allerdings die Schwa-Assimilation nicht beherrschen, würde sich der Fehler in jedem Wort wiederfinden. Es ist wahrscheinlich, dass das Urteil der L1-Rater in diesem Fall schlechter ausfallen würde.

Insgesamt lassen sich die folgenden zentralen Schlussfolgerungen ziehen:

- a. Die Bewertungsdimensionen *comprehensibility* und *acceptability* sind wahrscheinlich verwandt bzw. beeinflussen sich gegenseitig. Unterschiede lassen dennoch darauf schließen, dass es sich um zwei voneinander unabhängige Dimensionen handelt, die unterschiedliche Konstrukte abprüfen.
- b. Die meisten der überprüften Fehler innerhalb der Prozesse sind weniger kommunikativ ausschlaggebend als förderlich. Eine Ausnahme stellen Fehler 20 und Fehler 22 dar (vgl. Beispiel (45)).
- c. Bei den segmentalen Fehlern sind insbesondere der Lautersatz für weiches *d* [ð] (Fehler 2 und 3), der Lautersatz bei verschiedenen Vokalqualitäten (Fehler 10 und 12), der Lautersatz bei verschiedenen Polyphthongen (Fehler 13, 14 und 15) und die nicht umgesetzte Vokalquantität, welche zu einem Qualitätswechsel führt (Fehler 17), zentral.
- d. Fehler innerhalb der Vokalquantität resultieren in segmentalen Fehler und werden entsprechend hoch bewertet.
- e. Befunde aus Studien können immer nur sprachenpaarspezifische Relevanz haben. Allgemeingültige, sprachenübergreifende Auslegung von Studienergebnissen und der daraus folgenden Etablierung von Haltungen, z. B. suprasegmentale oder segmentale Phänomene als wichtiger für L2-Lerner einzustufen, sollten generell kritisch betrachtet werden.

4.2.4 Diskussion und Fazit

Im Folgenden werden die Ergebnisse beider Analyseschritte (Bewertung der unterschiedlichen Stimulustypen in Kapitel 4.2.3.2 und Bewertung der einzelnen Fehler in Kapitel 4.2.3.3) zusammengebracht und in Bezug auf die Ausgangsfrage diskutiert, wie sich die Bewertung unterschiedlicher phonetischer Abweichungen durch dänische L1-Sprecher darstellt. Dazu wurden zwei zu untersuchende Hypothesen aufgestellt. Die erste Hypothese $H_{1,1}$ fokussiert die Unterschiede innerhalb der Stimulustypen, während die zweite Hypothese $H_{2,1}$ die einzelnen Fehler genauer betrachtet. Aus der Überprüfung der ersten Hypothese geht hervor, dass die zur ersten Hypothese bestehende Nullhypothese ($H_{1,0}$) verworfen werden kann. So bewerten die L1-Rater fehlerhafte L2-Sprecher signifikant schlechter als fehlerfreie L2-Sprecher oder L1-Sprecher.

Die Hypothesenprüfung wurde in allen drei abhängigen Variablen (Bewertungsdimensionen) durchgeführt. In der Dimension *accentedness* zeigen sich große Unterschiede zwischen der Bewertung der einzelnen Stimulustypen. So werden von den L1-Ratern nicht nur die Stimuli der L2-Lerner, in denen das zu untersuchende Merkmal in fehlerhafter Realisierung vorliegt, als akzentuiert wahrgenommen, sondern auch diejenigen der L2-Kontrollsprecher, in denen das zu untersuchende Merkmal in normsprachlicher Realisierung vorhanden ist. Auch wenn sowohl die L2-Lerner als auch die L2-Kontrollsprecher signifikant schlechter als die L1-Kontrollsprecher bewertet werden, werden die L2-Kontrollsprecher noch signifikant besser als die L2-Lerner bewertet. Die L1-Kontrollsprecher werden insgesamt als nicht akzentuiert bewertet. Die Ergebnisse bestätigen, dass die L1-Rater die L2-Sprecher als nicht-muttersprachliche Sprecher erkennen, egal ob das zu untersuchende Merkmal vorliegt oder nicht. Innerhalb der Dimensionen *acceptability* und *comprehensibility* erzielen die L2-Kontrollsprecher zwar ähnliche wie die L1-Kontrollsprecher, aber es bestehen dennoch signifikante Unterschiede. Die L2-Lerner werden in beiden Dimensionen signifikant schlechter bewertet. Die zusammenfassende Betrachtung der Ergebnisse aller Bewertungsdimensionen zeigt, dass die Überwindung zentraler Aussprachefehler trotz bleibenden L2-Rauschens, das sich in der Bewertung eines wahrgenommenen Akzents niederschlägt, zu einer besseren wahrgenommenen Verständlichkeit und Akzeptabilität führen kann und bestätigen zumindest teilweise eine Unabhängigkeit der Bewertungsdimensionen. So grenzt sich *accentedness* deutlich von den anderen beiden Dimensionen ab. Dass *comprehensibility* und *acceptability* näher zusammenhängen, zeigt sich auch im zweiten Analyseschritt.

In diesem wird die zweite Hypothese $H_{2,1}$ bearbeitet, indem die Fehler einzeln auf die Frage hin untersucht werden, welche Phänomene in der Bewertung durch die L1-Rater besonders hohe Werte aufweisen. Dies wurde in jeder Bewertungsdimension anhand des Vergleichs der Mittelwertunterschiede der Stimuli des L2-Lerners untersucht. Auch die Nullhypothese der zweiten Hypothese ($H_{2,0}$) kann, zumindest für bestimmte Fehlergruppen, verworfen werden. So sind Mittelwertunterschiede zwar nicht zwischen allen einzelnen Fehlern, aber zwischen Gruppierungen von Fehlern signifikant nachweisbar.

Die Gruppen lassen sich anhand ihrer Mittelwerte nach Fehlerschwere beschreiben. Es ergeben sich im Hinblick auf die Bewertung durch die L1-Rater leichte, mittelschwere und schwere Fehler. Die Gruppierung konnte für die Dimensionen *comprehensibility* und *acceptability* durchgeführt werden. In der Dimension *accentedness* unterschied sich die Bewertung der einzelnen Fehler nicht. Dass sich für unterschiedliche Fehler in der Dimension *accentedness* keine unterschiedlich hohen Bewertungen ergeben, resultiert darin, dass sich die Fehler innerhalb der Dimension *accentedness* nicht nach ihrer Fehlerschwere einteilen lassen. Ausschlaggebend für die Bewertung innerhalb dieser Dimension sind also weniger die Unterschiedlichkeit der einzelnen Fehler, sondern vielmehr, ob ein Fehler überhaupt vorhanden ist oder nicht.

Für die *comprehensibility* und die *acceptability* stellt sich die Unterschiedlichkeit der Fehler allerdings als entscheidend dar. So bestimmen unterschiedliche Fehler im unterschiedlichen Maße darüber, ob eine Aussage als mehr oder weniger verständlich oder akzeptabel wahrgenommen wird. Die einzelnen nach Fehlerschwere eingeteilten Gruppen wurden inhaltlich analysiert. Aus dieser Analyse ergibt sich, dass sich besonders Fehler, die auf orthographischer Interferenz beruhen, negativ auf die Bewertung von Aussprache auswirken. Tatsächlich ist allerdings zu bedenken, dass diese Fehler in frei gesprochener Sprache wahrscheinlich nicht auftreten würden. Für Fehler, die auf artikulatorischen Schwierigkeiten oder der Nichtanwendung einer lautlichen Regel beruhen, kann festgestellt werden, dass bei den segmentalen Fehlern besonders die Lautersetzungen von weichem *d*, unterschiedlichen Vokalqualitäten und Polyphthongen eine wichtige Rolle spielen. Bei den phonologischen Prozessen ist besonders die Lenisierung nachtoniger Plosive ein bedeutsames Phänomen, ebenso wie Schwa-Assimilierungsprozesse, deren Nichtumsetzung in einer nicht-normsprachlichen lautlichen Umsetzung resultiert.

Die in den Dimensionen *comprehensibility* und *acceptability* relativ gleich erzielten Mittelwerte und dementsprechend relativ gleich eingeteilten Fehlergruppen lassen darauf schließen, dass die Bewertungsdimensionen untereinander zusammenhängen bzw. korrelieren. So ist es vorstellbar, dass eine besonders niedrige wahrgenommene Verständlichkeit auch eine niedrige Akzeptabilität voraussagt. Diese Vermutung kann mit den durchgeführten Analyseschritten allerdings nicht ausreichend beantwortet werden und bietet Raum für nachfolgende Untersuchungen. Solche Untersuchungen würde Ergebnisse von Studien stützen, die Korrelationen zwischen den Bewertungsdimensionen nachweisen (z. B. Trofimovich & Isaacs 2012; Pilott 2016) und deren Spektrum um eine weitere, wenig untersuchte L2 erweitern. Eine interessante weiterführende Fragestellung ist, ob und wie die beiden Bewertungsdimensionen miteinander zusammenhängen und sich gegebenenfalls untereinander beeinflussen.

Trotz möglicher Einschränkungen kann die vorliegende Studie die Relevanz einzelner Fehlergruppen nachweisen, was besonders im Zusammenhang mit der Untersuchung der Fehlerhäufigkeit aus Kapitel 3 den eingangs gegebenen Fokus auf Fragen zur Ausspracheförderung im schulischen Dänischunterricht abrundet.

5 Zusammenfassung

Die zwei Hauptstudien dieser Arbeit haben verschiedene Forschungsbereiche zusammengebracht. So wurden typische Fragestellungen und Methoden der L2-Forschungsbereiche der Fehleranalyse und der Fehlerschwerebestimmung bzw. der Bewertung durch Muttersprachler auf den Bereich Dänisch als L2 für deutschsprachige Lerner angewendet. Das hauptsächliche Ziel bestand darin, empirische Belege für Hypothesen im Bereich Aussprache deutschsprachiger Dänischlerner zu erbringen, um somit eine theoretisch und empirisch fundierte Grundlage für diesen Bereich zu bereiten. Dazu wurden zunächst in einer kontrastiven phonetischen Analyse lautliche Merkmale erarbeitet, die deutschen Dänischlernern schwerfallen könnten. Diese wurden anschließend in einer empirischen Erhebung mit Schülersprache abgeglichen. Nachdem die Probleme für deutschsprachige Dänischlerner auf diese Weise evidenzbasiert bestätigt waren, wurde deren Auswirkung auf die Kommunikation mit L1-Sprechern untersucht. Die Ergebnisse zeigen, dass L1-Sprecher verschiedene Fehler unterschiedlich bewerten.

Die Studien und deren Ergebnisse sind nicht nur relevant für den konkreten praktischen Bereich Ausspracheförderung im L2-Dänischerwerb, sondern stellen auch für den Forschungsbereich generell einen Zugewinn dar, insofern als dessen Fragestellungen und Methoden auf eine Sprache angewandt wurden, die bislang noch nicht viel Aufmerksamkeit in diesem Kontext erhalten hat. Darüber hinaus wurde der Forschungsbereich zur Untersuchung von *acceptability* erweitert.⁵⁷ Im Folgenden werden die Ergebnisse beider Studien zusammengefasst und in Relation zueinander diskutiert (Kapitel 5.1). Im Anschluss daran findet die Übertragung der Erkenntnisse auf den unterrichtlichen Kontext statt, indem mögliche relevante Erkenntnisse dargestellt werden (Kapitel 5.2). Zuletzt wird in einem Ausblick die Arbeit abgeschlossen (Kapitel 5.3).

5.1 Übergreifende Zusammenfassung der Hauptstudien

In der ersten Studie wurden Häufigkeitsuntersuchungen zu Aussprachefehlern deutschsprachiger Dänischlerner gemacht. Diese dienen zum empirischen Abgleich von theoretischen Vorannahmen, die aus kontrastiven phonetischen Analysen gewonnen wurden, und bestätigen außerdem anekdotische Lehrererfahrungen, die bislang lediglich qualitativ festgehalten wurden (Fredsted 1989; Jacob 2004). In den Darstellungen relevanter sprachenpaarspezifischer Arbeiten zu Merkmalen oder Fehlern (Fredsted 1989, 2011; Jacob 2004) wird außerdem nicht zwischen orthographisch bedingten und rein lautlichen Problemen unterschieden. Eine solche Unterscheidung nimmt diese Arbeit vor (detaillierter hierzu in Kapitel 5.2). Die Studien ergeben somit nicht nur empirische Belege zu Fehlerhäufigkeiten, sondern analysieren die typischen Fehler nach Erscheinungsform und Ursachen. Aus der Analyse der Häufigkeiten ergibt sich, dass dieselben Laute in verschiedenen Wortpositionen unterschiedlich häufig vorkommen. Dieses Ergebnis bestätigt Studien, in denen die Position des Lauts innerhalb des Wortes als unterschiedlich

57 Dies wird von Forschern des Bereichs gefordert (z. B. Thomson 2018: 26).

schwierig für Lerner angegeben wird (Flege 1991). So zeigt sich, dass stimmloses *s* [s] für die Lerner wortmedial zwar in mehreren Fällen auftritt, sich in wortinitialen Kontexten allerdings als unproblematisch darstellt. Dies ist insofern interessant, als wortinitiales stimmloses *s* [s] im Deutschen nur in Lehnwörtern aus dem Englischen auftritt (vgl. *Duden* 2015: 120) und somit eigentlich erwartet wird, dass hier mehr Fehler gemacht werden. Es treten allerdings keinerlei Fehler auf, sodass sich zusammenfassen lässt, dass den Lernern der Erwerb dieses Phänomens keine Probleme bereitet. Dieser Befund geht mit der von Flege (1995) konstatierten Verbindung zwischen hoher Kontrastivität und leichterem Erwerb einher (vgl. dazu die Beschreibung von neuen Lauten in Kapitel 3.2.2).

In der zweiten Studie wurden die ermittelten Fehler einem muttersprachlichen Ratingpanel unterzogen. Die L1-Sprecher bewerteten die Fehler in den Bewertungsdimensionen *accentedness*, *comprehensibility* und *acceptability*. Es zeigt sich, dass die verschiedenen Fehler unterschiedlich bewertet werden, die Bewertungen innerhalb der Dimensionen *comprehensibility* und *acceptability* allerdings relativ ähnlich ausfallen. Von diesen beiden Dimensionen wird ausgegangen, dass sie einander entweder stark ähneln, auf zu untersuchende Weise miteinander verbunden sind oder sich gegenseitig bedingen. Die Fehlerbewertungen in der Dimension *accentedness* sind über alle Fehler hinweg relativ gleich. Die Fehler wurden nach der Bewertung in den Dimensionen *comprehensibility* und *acceptability* in leichte, mittelschwere und schwere Fehler eingeteilt (vgl. Tab. 4.17). Hieraus ergibt sich, dass die überprüften Fehler der Prozesse eher den leichten und mittelschweren Fehlern zugeordnet werden können und nie unter den schweren Fehlern auftreten. Zu den schweren Fehlern gehören der Lautersatz für weiches *d* [ð] (Fehler 2 und 3), der Lautersatz bei verschiedenen Vokalqualitäten (Fehler 10 und 12), der Lautersatz bei verschiedenen Polyphthongen (Fehler 13, 14 und 15) sowie die nicht umgesetzte Vokalquantität bei geschlossenem [i] (Fehler 17).

Die Ergebnisse beider Studien lassen sich zusammenfassen, sodass typische Fehler deutschsprachiger Dänischlerner sowohl in ihrer Häufigkeit als auch in ihrer Fehler-schwere beschrieben werden können. Dadurch lassen sich Fehler herausarbeiten, die gleichzeitig besonders schwerwiegende kommunikative Konsequenzen haben und dazu häufig von Lernern produziert werden. Darüber hinaus lassen sich Einstufungen besonders häufig gemachter Fehler insofern relativieren, als sie sich nicht besonders negativ auf die Kommunikation auswirken. Ergebnisse aus der Zusammenfassung beider Studien werden in Abb. 5.1 dargestellt.

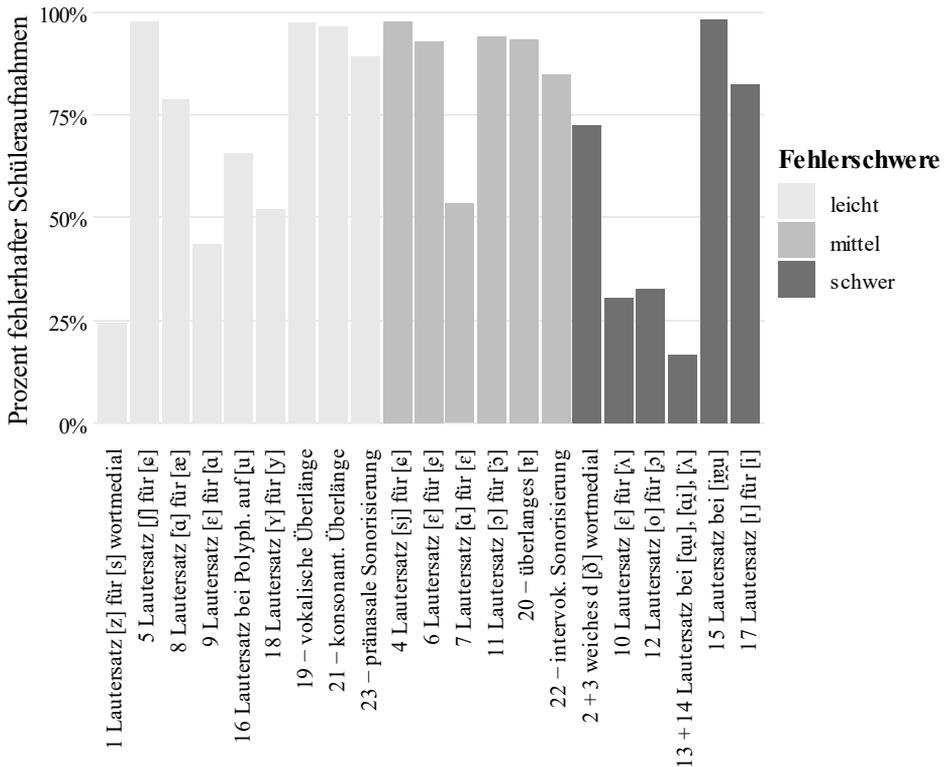


Abb. 5.1: Zusammenfassung der Ergebnisse aus den Studien 1 und 2: Fehlerschwere und Fehlerhäufigkeiten

Die Darstellung zeigt, dass diejenigen Fehler innerhalb der phonologischen Prozesse, die als besonders häufige ermittelt wurden, in der Untersuchung zur Fehlerschwere als kommunikativ eher nebensächlich einzustufen sind. In der Gruppe der schweren Fehler stellen sich Fehler 17 (Lautersatz [ɪ] für [i]), Fehler 15 (Lautersatz bei [iɥu]) und Fehler 2 und 3 (Lautersatz bei weichem *d*) ebenfalls als häufig auftretende Fehler heraus. Die Häufigkeit kann dabei auf unterschiedliche Gründe zurückgeführt werden. So ist im Falle von Fehler 2 und 3 (Lautersatz bei weichem *d*) wahrscheinlich, dass der Fehler im Unterricht zwar geübt wird, allerdings Probleme bei der Artikulation bereitet. Für Fehler 17 hingegen lässt sich annehmen, dass dieser keine artikulatorischen Probleme bereitet, sondern im Unterricht wenig fokussiert wird. Von den weniger häufig auftretenden schweren Fehlern lässt sich annehmen, dass diese im Unterricht fokussiert wurden oder den Lernern keine Probleme bei der Artikulation bereiten. So sollten beispielsweise bei Fehler 13 und 14 (beide Lautersatz bei bekannten Polyphthongen) keine Probleme bei der Artiku-

lation bestehen, da die hier abgefragten Diphthonge auch im Deutschen vorkommen. Lediglich die lautliche Umsetzung der Orthographie kann problematisch sein und ist in Anbetracht der vorliegenden Erhebungsmethode anhand von Vorlesetexten eine wahrscheinliche Ursache der hier auftretenden Fehler. Deren relativ niedrige Anzahl deutet allerdings darauf hin, dass dieses Merkmal im Unterricht behandelt wird. Wenn der Laut trotzdem fehlerhaft realisiert wird, wirkt er sich kommunikativ stark negativ aus. Die mittelschweren Fehler kommen in den untersuchten Schüleraufnahmen alle häufig vor. Ebenso häufig vertreten sind die meisten der leichten Fehler. Lediglich Fehler 1 (stimmloses [s] wortmedial) wird relativ selten falsch realisiert. Diese sich aus der gleichzeitigen Betrachtung von Fehlerschwere und -häufigkeit gewonnenen Erkenntnisse sind besonders für den unterrichtlichen Kontext von Relevanz. Im folgenden Kapitel werden diese aufgeführt und diskutiert.

5.2 Erkenntnisse für die Ausspracheförderung im schulischen Dänischunterricht

Auf den Anfang dieser Arbeit (Kapitel 1 und 2) zurückgreifend kann zusammengefasst werden, dass die gewonnenen Ergebnisse wesentlich zur Präzisierung relevanter Phänomene im Kontext Dänischunterricht als Fremdsprache beitragen. Die empirisch belegten Erkenntnisse ermöglichen evidenzbasierte Formulierungen zu curricularen Anforderungen hinsichtlich einer realistischen Zielnorm im Sinne eines mindestens zu erreichenden Sprachlevels (vgl. Kapitel 1.1). Diese bislang fehlende hinreichend präzise Festlegung, stellt sich auch in den Ergebnissen der eingangs vorgestellten explorativen Vorstudie dar (vgl. Kapitel 2). So geht aus dieser hervor, dass der Bereich der Ausspracheförderung im Dänischunterricht Leerstellen aufweist, die es den Lehrkräften erschweren, Aussprache explizit zu unterrichten. Neben dem offensichtlichen Mangel an sprachspezifischen Materialien fehlt es an einer Aufarbeitung der grundlegenden sprachenpaarspezifischen Merkmale. Die vorliegende Studie liefert einen ersten empirischen Beitrag zur Aufarbeitung des Bereichs. Die Befunde der zwei Einzelstudien erlauben es, evidenzbasierte Aussagen darüber zu tätigen, welche lautlichen dänischen Merkmale für deutschsprachige Dänischlerner problematisch sind und welche von diesen, in Anbetracht des Unterrichtsziels ungehindert mit L1-Sprechern kommunizieren zu können, unbedingt gelernt werden sollten. So stellen sich als besonders schwerwiegende Fehler, die auch noch besonders häufig von Lernern falsch realisiert werden, der Lautersatz beim weichen *d* sowie Fehler innerhalb der Vokalquantität mit einhergehendem Lautwechsel heraus. Diese Fehler gilt es im Unterricht besonders zu fokussieren. Auch die Diphthonge und Triphthonge stellen sich als kommunikativ relevant heraus. Die lautliche Umsetzung der Orthographie von aus dem Deutschen bekannten Diphthongen wird von den Lehrkräften anscheinend bereits im Unterricht fokussiert. Bei den Triphthongen bestehen allerdings noch Probleme.

Diese Arbeit unterscheidet in ihren Darstellungen zwischen rein lautlichen und orthographisch basierten Fehlern. Ein solcher Ansatz ist für den unterrichtlichen Bedarf insofern von Relevanz, als zwischen Ausspracheproblemen unterschieden werden muss, die auf Grund der orthographischen Umsetzung zu Stande kommen und denen, die auch bzw. zuallererst lautliche (artikulatorische oder perzeptorische) Probleme bereiten. Schwierigkeiten in Perzeption und Artikulation müssen, bevor die lautliche Umsetzung der Orthographie fokussiert werden kann, zunächst geklärt werden. Diese Unterscheidung wird in verfügbaren Lehrmaterialien bislang nicht genügend vorgenommen. Hauptsächlich werden Probleme aufgegriffen, die aus der orthographischen Umsetzung hervorgehen. Darüber hinaus wird der Erwerb der neuen bzw. ähnlichen oder unterschiedlichen Laute häufig mit der Orthographie verknüpft, wodurch das Problem der lautlichen Umsetzung der Orthographie erneut in den Vordergrund rückt und eine gedankliche bzw. didaktische Trennung erschwert wird. Diese Arbeit erhofft sich, in dieser Hinsicht ein gezielteres Problembewusstsein zu schaffen und so eine Entwicklung innerhalb der Materialerstellung und Fehlerkorrektur voranzubringen, die zwischen rein lautlichen und orthographisch basierten Problemen trennt. Erst dadurch wird es möglich, problemspezifische Fehlerkorrektur vorzunehmen bzw. entsprechend nützliche Materialien für den Unterricht zu entwickeln. Der kontrastierende Vergleich zwischen Orthographie und Aussprache sollte erst den zweiten Schritt darstellen. Die Ergebnisse der Hauptstudien geben Hinweise auf Merkmale, die in diesem Rahmen relevant sein könnten. Darüber hinaus stellen sich einige Phänomene heraus, die rein artikulatorisch keine Probleme bereiten sollten und durch Bewusstmachung relativ leicht behoben werden könnten. Hier ließen sich beispielsweise Hinweise dazu geben, dass stimmhaftes *s* [z] nie im Dänischen auftritt (Fehler 1), der Wechsel innerhalb der Vokalquantität im Dänischen nicht gleichzeitig an einen Qualitätswechsel gebunden ist (Fehler 17 und 18) und die Lenisierung nachtoniger Plosive (Fehler 22) der Realisierung im Norddeutschen ähnelt.

Weiterführende Forschung müsste an dieser Stelle ansetzen und untersuchen, ob die Bewusstmachung und Übung der dargestellten Phänomene zu einer Verbesserung der Aussprache führt. Darüber hinaus sollte geprüft werden, ob die präzise Aufbereitung Lehrkräfte darin unterstützt, Aussprache in Hinblick auf das Erreichen eines mindestens notwendigen sprachlichen Levels im Unterricht explizit zu fördern.

5.3 Ausblick

Diese Arbeit hat zu wichtigen Erkenntnissen in Hinblick auf die L2-Aussprache deutschsprachiger Dänischlerner geführt und dazu den Forschungsbereich der Fehlerschwerebestimmung nicht nur um eine bislang wenig erforschte Fremdsprache ergänzt, sondern auch die Forschungsarbeit zur Bewertungsdimension *acceptability* erweitert. Dabei haben sich interessante weiterführende Fragestellungen und Aufgabenbereiche ergeben. So sollte an diese Arbeit anknüpfend untersucht werden, ob die Bewertungsdimensionen

acceptability und *comprehensibility* miteinander korrelieren oder sich gegenseitig beeinflussen. Dazu sollten zunächst die für die Studie entwickelten Indikatoren bzw. Items zur Messung der Bewertungsdimensionen einer Reliabilitätsprüfung unterzogen werden und Korrelationen zwischen den beiden Variablen errechnet werden.

Für Dänisch als Fremdsprache legen die Ergebnisse der Analyse typischer Fehler deutschsprachiger Dänischlerner die Grundlage für die merkmals- und fehlerspezifische Ausspracheförderung. Die Arbeit zeigt in dieser Hinsicht leichte Einschränkungen, da sie die Fehler lediglich in vorgelesenen Texten empirisch erhebt. Hier sollte ein Abgleich der ermittelten Fehler mit frei gesprochener Sprache stattfinden. Darüber hinaus haben sich weitere relevante Fragen zum Einfluss der Lehrkraft ergeben. Zur Untersuchung dieses Bereichs zeigt diese Arbeit bereits einige Ansätze, die in einem angepassten Forschungssetting aufgegriffen und erweitert werden sollten. In diesem Zusammenhang könnten einige bereits erhobene Daten genutzt werden, beispielsweise um die Aussprache der Lehrkräfte derjenigen der jeweils zusammenhängenden Schüler vergleichend gegenüberzustellen. Zum Schluss dieser Arbeit erweisen sich die gesammelten Ergebnisse als Wegbereiter für weiterführende Forschung, indem sie das Feld für experimentelle Forschung im Bereich Verbesserung spezifischer lautlicher Phänomene öffnen. So ist es von Interesse zu überprüfen, ob eine problemspezifische fehlertypenbezogene Korrektur gegenüber der Korrektur durch reine Imitation Vorteile in Hinblick auf die Ausspracheförderung liefert. Die Vorgehensweise dieser Arbeit stellt die angemessene Herangehensweise für das bislang recht unbearbeitete Feld der Aussprache und Aussprachefehlern im Dänischunterricht für deutschsprachige Lerner dar. Sie legt die theoretische Grundlage für anschließende praktische Implementierungen und weiterführende Forschung.

Literatur

- Anderson-Hsieh, Janet, Ruth Johnson & Kenneth Koehler. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42, 529–555.
- Audacity = Audacity Team. 2018. Audacity®: Free audio editor and recorder [Software]. Version 2.3.0. o. O. <audacityteam.org; Stand vom 20. 12. 2018>.
- Basbøll, Hans. 2005. *The phonology of Danish*. Oxford: Oxford University Press.
- Bates, Douglas et al. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1–48.
- Beinhoff, Bettina. 2014. What is acceptable. The role of acceptability in English. In Edith Esch & Martin Solly (Hgg.), *Language education and the challenges of globalisation. Sociolinguistic issues*, 155–174. Newcastle Upon Tyne: Cambridge Scholars Publishing.
- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (Hg.), *Speech perception and linguistic experience. Issues in cross-language research*, 171–204. Baltimore: York Press.
- Best, Catherine T., Gerald W. McRoberts & Nomathemba M. Sithole. 1988. Examination of perceptual reorganization for nonnative speech contrasts. Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14, 345–360.
- Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception. In Murray J. Munro & Ocke-Schwen Bohn (Hgg.), *Language experience in second language speech learning. In honor of James Emil Flege* (Language Learning and Language Teaching 17), 13–34. Amsterdam: Benjamins.
- Bohn, Ocke-Schwen. 2002. On phonetic similarity. In Petra Burmeister, Thorsten Piske & Andreas Rohde (Hgg.), *An integrated view of language development. Papers in honor of Henning Wode*, 119–216. Trier: Wissenschaftlicher Verlag.
- Bohnsteffen, Markus. 2011. *Fehler-Korrektur. Lehrer- und lernerbezogene Untersuchungen zur Fehlerdidaktik im Englischunterricht der Sekundarstufe II*. Frankfurt am Main: Lang.
- Bortz, Jürgen & Nicola Döring. 2006. *Forschungsmethoden und Evaluation*. Berlin: Springer.
- Bose, Ines et al. 2016. *Einführung in die Sprechwissenschaft: Phonetik, Rhetorik, Sprechkunst*, 2. Aufl. Tübingen: Narr Francke Attempto.
- Böttger, Katharina. 2008. *Die häufigsten Fehler russischer Deutschlerner. Ein Handbuch für Lehrende*. Münster: Waxmann.
- Bühner, Markus. 2010. *Einführung in die Test- und Fragebogenkonstruktion*. 2. Aufl. München: Pearson Studium.
- Bühner, Markus & Matthias Ziegler. 2012. *Statistik für Psychologen und Sozialwissenschaftler*. 3. Aufl. München: Pearson Studium.
- Celce-Murcia, Marianne et al. 2012. *Teaching pronunciation. A course book and reference guide*. 2. Aufl. Cambridge: Cambridge University Press.

- Chang Li-Ann, Diane. 2008. Understand me or not? Accent, acceptability and intelligibility in International English. The case of Singapore. Singapore: National University of Singapore [master's thesis] <scholarbank.nus.edu.sg/handle/10635/13223; Stand vom 15. 11. 2018>.
- Corder, Pit. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5, 161–170.
- Couper, Graeme. 2016. Teacher cognition of pronunciation teaching amongst English language teachers in Uruguay. *Journal of Second Language Pronunciation* 2, 29–55.
- Crowther, Dustin et al. 2015. Second language comprehensibility revisited. Investigating the effects of learner background. *TESOL Quarterly* 49, 814–837.
- Darcy, Isabelle, Doreen Ewert & Ryan Lidster. 2012. Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation 'toolbox'. In John Levis & Kimberly LeVelle (Hgg.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, 93–108. Ames: Iowa State University.
- DDO = *Den danske ordbog*. 2003–2005. København: Gyldendal <www.ordnet.dk/ddo>.
- Derwing, Tracey & Murray Munro. 1997. Accent, intelligibility, and comprehensibility. Evidence from four L1s. *Studies in Second Language Acquisition* 19, 1–16.
- Derwing, Tracey & Murray Munro. 2005. Second language accent and pronunciation teaching. A research-based approach. *TESOL Quarterly* 39, 379–397.
- Derwing, Tracey & Murray Munro. 2009. Putting accent in its place. Rethinking obstacles to communication. *Language Teaching* 42, 476–490.
- Derwing, Tracey & Murray Munro. 2015. *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research* (Language Learning & Language Teaching 42). Amsterdam: Benjamins.
- Dieling, Helga & Ursula Hirschfeld. 2000. *Phonetik lehren und lernen* (Fernstudieneinheit 21). Berlin: Langenscheidt.
- Dretzke, Burkhard. 1985. *Fehlerbewertung im Aussprachebereich. Objektive Fehlerbeurteilung versus subjektive Fehlerbewertung*. Hamburg: Buske.
- DUDA = Jensen, Erik & Ana Jensen. DUDA <www.duda.dk/grundfag/dansk/dansk.html; Stand vom 23. 1. 2019>.
- Duden 2015 = *Duden. Aussprachewörterbuch*. 2015. Bearb. von Stefan Kleiner u. Ralf Knöbl. 7. Aufl. Berlin: Dudenverlag.
- Ellis, Rod. 2006. *The study of second language acquisition*. Oxford: Oxford University Press.
- Faulstich-Wieland, Hannelore. 2008. *Schule und Geschlecht*. In Werner Helsper & Jeanette Böhme (Hgg.), *Handbuch der Schulforschung*, 2. Aufl., 673–695. Wiesbaden: Springer VS.
- Ferris, Dana R. 2004. The 'grammar correction' debate in L2 writing. Where are we, and where do we go from here? (and what do we do in the meantime ...?). *Journal of Second Language Writing* 13, 49–62.

- Flege, James E. 1987. Effects of equivalence classification on the production of foreign language speech sounds. In Allan James & Jonathan Leather (Hgg.), *Sound patterns in second language acquisition*, 9–39. Dordrecht: Foris.
- Flege, James E. 1991. The interlingual identification of Spanish and English vowels. Orthographic evidence. *The Quarterly Journal of Experimental Psychology* 43, 701–731.
- Flege, James E. 1995. Second language speech learning. Theory, findings, and problems. In Winifred Strange (Hg.), *Speech perception and linguistic experience. Issues in cross-language research*, 233–277. Baltimore: York Press.
- Flege, James E., Murray J. Munro & Robert Allen Fox. 1994. Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America* 95, 3623–3641.
- Flege, James E., Murray J. Munro & Ian R. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97, 3125–3134.
- Fredsted, Elin. 1989. Variable og invariable fejl i tysksprogede elevs udtale af dansk. In Mette Kunøe & Erik Vive Larsen (Hgg.), 2. *Møde om Udforskningen af Dansk Sprog*, 99–109. Aarhus: Nordisk Institut.
- Fredsted, Elin. 2011. *Dansk segmental fonologi i og prosodi. En lærebog i dansk udtale for sprogstuderende* (Studia Danica 2). Flensburg: Flensburg University Press.
- Fries, Charles C. 1945. *Teaching & learning English as a foreign language*. Ann Arbor: University of Michigan Press.
- Grønnum, Nina. 2009. *Fonetik og Fonologi. Almen og dansk*. 3. Aufl. København: Akademisk Forlag.
- Hahn, Laura D. 2004. Primary stress and intelligibility. Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38, 201–223.
- Hall, T. Alan 2000. *Phonologie. Eine Einführung*. Berlin: de Gruyter.
- Handbook = Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet*. 1999. Cambridge: Cambridge University Press.
- Hansen, Peter M. 1990. *Udtaleordbog*. København: Gyldendal.
- Harden, Theo. 2006. *Angewandte Linguistik und Fremdsprachendidaktik*. Tübingen: Narr Francke Attempto.
- Helmke, Andreas et al. 2008. Selbstkonzept, Motivation und Englischleistung. In Eckhard Klieme (Hg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie*, 244–258. Weinheim: Beltz.
- Hirschfeld, Ursula. 2001. Der ‚fremde Akzent‘ in der interkulturellen Kommunikation. In Margret Bräunlich (Hg.), *Gesprochene Sprache – transdisziplinär. Festschrift zum 65. Geburtstag von Gottfried Meinhold* (Hallesche Schriften zur Sprechwissenschaft und Phonetik 5), 3–91. Frankfurt am Main: Lang.
- Höder, Steffen. 2014. Low German. A profile of a word language. In Javier Caro Reina & Renata Szczepaniak (Hgg.), *Syllable and word languages* (Linguae & litterae 40), 305–326. Berlin: de Gruyter.

- Höder, Steffen. 2016. Dänische Phonetik im Kontrast zu norddeutschen Ausgangsvarianten beim schulischen Zweitspracherwerb (Kieler Arbeiten zur skandinavistischen Linguistik 1). 2. Ausg. Kiel: Christian-Albrechts-Universität zu Kiel.
- Jacob, Renate. 2004. Udtale. Et bogstav om dagen. Kiel: Institut für Qualitätsentwicklung an Schulen Schleswig-Holstein <faeher.lernnetz.de/faeherportal/index.php?key=98&auswahl=177; Stand vom 9. 12. 2018>.
- Jansing, Barbara et al. 2013. Professionelles Wissen im Studienverlauf. Lehramt Englisch. In Sigrid Blömeke et al. (Hgg.), *Professionelle Kompetenzen im Studienverlauf. Weitere Ergebnisse zur Deutsch-, Englisch- und Mathematiklehrerausbildung aus TEDS-LT*, 77–107. Münster: Waxmann.
- James, Carl. 1971. The exculpation of contrastive linguistics. In Gerhard Nickel (Hg.), *Papers in Contrastive Linguistics*, 53–68. Cambridge: Cambridge University Press.
- Järventausta, Marja. 2008. Kontrastive Linguistik und Fremdsprachenunterricht. In Johanna Laakso (Hg.), *Ungarischunterricht in Österreich. Perspektiven und Vergleichspunkte = Teaching Hungarian in Austria* (Finno-Ugrian Studies in Austria 6), 10–32. Wien: LIT-Verlag.
- Jespersen, Otto. 1890–1892. Danias lydskrift. *Dania* 1, 33–79.
- Jespersen, Otto. 1901. *Sprogundervisning*. København: Det Schuboeske Forlag.
- Johansson, Stig. 1973. The Identification and evaluation of errors in foreign languages. A functional approach. In Jan Svartvik (Hg.), *Errata: Papers in error analysis*, 102–114. Lund: Gleerup.
- Johansson, Stig. 1975. *Papers in contrastive linguistics and language testing* (Lund Studies in English 50). Lund: Gleerup.
- Johansson, Stig. 1978. *Studies of error gravity. Native reactions to errors produced by Swedish learners of English* (Gothenburg Studies in English 44). Göteborg: Acta Universitatis Gothoburgensis.
- Jonkisz, Ewa, Helfried Moosbrugger & Holger Brandt. 2012. Planung und Entwicklung von Tests und Fragebogen. In Helfried Moosbrugger & Augustin Kelava (Hgg.), *Testtheorie und Fragebogenkonstruktion*, 2. Aufl., 27–72. Berlin: Springer.
- Juhász, János. 1970. *Probleme der Interferenz*. München: Hueber.
- Kennedy, Sara & Pavel Trofimovich. 2008. Intelligibility, comprehensibility, and accentedness of L2 speech. The role of listener experience and semantic context. *The Canadian Modern Language Review/La revue canadienne des langues vivantes* 64, 459–489.
- Khalil, Aziz. 1985. Communicative error evaluation. Native Speakers' Evaluation and Interpretation of Written Errors of Arab EFL Learners. *TESOL Quarterly* 19, 335–351.
- Kirk, Katrine. 2008. *Dansk udtale: En undervisningsvejledning*. København: Ministeriet for Flygtninge, Indvandrere og Integration.
- Kleppin, Karin. 1997. *Fehler und Fehlerkorrektur*. München: Goethe-Institut.
- Kleppin, Karin. 2001. Formen und Funktionen von Fehleranalyse, -korrektur und -therapie. In Gerhard Helbig et al. (Hgg.), *Deutsch als Fremdsprache. Ein internationales*

- Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 19), Bd. 2, 986–994. Berlin: de Gruyter.
- Kleppin, Karin. 2009. Zum Umgang mit Fehlern im Fremdsprachenunterricht. In Udo O. H. Jung & Heidrun Jung (Hgg.), *Praktische Handreichung für Fremdsprachenlehrer* (Bayreuther Beiträge zur Glottodidaktik 2), 5. Aufl., 64–70. Frankfurt am Main: Lang.
- Kleppin, Karin. 2013. Fehler, Fehlerkorrektur, Fehlerbewertung. In Wolfgang Hallet & Frank G. Königs (Hgg.), *Handbuch Fremdsprachendidaktik*, 2. Aufl., 224–228. Seelze-Velber: Klett/Kallmeyer.
- Kleppin, Karin & Frank G. Königs. 1997. *Der Korrektur auf der Spur. Untersuchungen zum mündlichen Korrekturverhalten von Fremdsprachenlehrern* (Manuskripte zur Sprachlehrforschung 34). 2. Aufl. Bochum: Brockmeyer.
- Klusmann, Uta & Dirk Richter. 2014. Beanspruchungserleben von Lehrkräften und Schülerleistung. Eine Analyse des IQB-Ländervergleichs in der Primarstufe. *Zeitschrift für Pädagogik* 60, 202–224.
- Koch, Peter & Wulf Oesterreicher. 1985. Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte. *Romanistisches Jahrbuch* 36, 15–43.
- Koster, Cor & Ton Koet. 1993. The Evaluation of Accent in the English of Dutchmen. *Language Learning* 43, 69–92.
- Kraut, Rachel & Stefanie Wulff. 2013. Foreign-accented speech perception ratings: A multifactorial case study. *Journal of Multilingual and Multicultural Development* 34, 249–263.
- Kristensen, Marius. 1924. *Vejledning til brugen af Danias lydskrift*. København: Thiele.
- Kristiansen, Tore. 2003. Language attitudes and language politics in Denmark. *International Journal of the Sociology of Language* 159, 57–71.
- Kristiansen, Tore. 2009. The macro-level social meanings of late-modern Danish accents. *Acta linguistica Hafniensia* 41, 167–192.
- Krosnick, Jon A. & Leandre R. Fabrigar. 2012. Designing rating scales for effective measurement in surveys. In Lars E. Lyberg et al. (Hgg.), *Survey measurement and process quality*, 141–164. Hoboken: Wiley.
- Kunter, Mareike et al. (Hgg.). 2011. *Professionelle Kompetenz von Lehrkräften: Ergebnisse des Forschungsprogramms COACTIV*. Münster: Waxmann.
- Kurtz, Jürgen. 2013. Der Kommunikative Ansatz und seine Bedeutung für die Theorie und Praxis des Fremdsprachenunterrichts zu Beginn des 21. Jahrhunderts. *Fremdsprachen lehren und lernen* 42, 80–93.
- Lado, Robert. 1957. *Linguistics across cultures. Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.
- Lambert, William E. et al. 1960. Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60, 44–51.
- LamaPoll 2011 = Langner, Lars, Maik Maibaum & Stoyko Notev. 2011. Umfragetool LamaPoll [Software] <www.lamapoll.de; Stand vom 25. 7. 2019>.

- Lee, Jieun, Dong Jin Kim & Hanyong Park. 2019. Native listeners' evaluations of pleasantness, foreign accent, comprehensibility, and fluency in the speech of accented talkers. In John Levis, Charles Nagle & Erin Todey (Hgg.), *Proceedings of the 10th Annual Pronunciation in Second Language Learning and Teaching Conference*, 168–178. Ames: Iowa State University.
- Levis, John. 2016. Accent in second language pronunciation research and teaching. *Journal of Second Language Pronunciation* 2, 153–159.
- Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam: Benjamins.
- MacDonald, Shem. 2002. Pronunciation. Views and practices of reluctant teachers. *Prospect* 17, 3–18.
- Maddieson, Ian. 2013. Vowel quality inventories. In Matthew S. Dryer & Martin Haspelmath (Hgg.), *The World Atlas of Language Structures Online*, Kap. 2. Leipzig: Max Planck Institute for Evolutionary Anthropology <wals.info/chapter/2; Stand vom 28. 12. 2018>.
- Marx, Nicole & Grit Mehlhorn. 2016. Analyse von Lernersprache. In Daniela Caspari et al. (Hgg.), *Forschungsmethoden in der Fremdsprachendidaktik. Ein Handbuch*, 297–305. Tübingen: Narr Francke Attempto.
- Matsuura, Hiroko, Reiko Chiba & Miho Fujieda. 1999. Intelligibility and comprehensibility of American and Irish Englishes in Japan. *World Englishes* 18, 49–62.
- McBride, Kara. 2015. Which features of Spanish learners' pronunciation most impact listener evaluations? *Hispania* 98, 14–30.
- Meindl, Claudia. 2011. *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. Tübingen: Narr Francke Attempto.
- Menold, Natalja & Kathrin Bogner. 2015. Gestaltung von Ratingskalen in Fragebögen. München <www.gesis.org/fileadmin/upload/SDMwiki/Archiv/Ratingskalen_MenoldBogner_012015_1.o.pdf; Stand vom 12. 7. 2017>.
- Ministerium für Bildung und Wissenschaft des Landes Schleswig-Holstein (Hg.). 2014. *Fachanforderungen Englisch. Allgemeinbildende Schulen. Sekundarstufe I. Sekundarstufe II*. Kiel.
- Ministerium für Bildung und Wissenschaft des Landes Schleswig-Holstein (Hg.). 2015. *Fachanforderungen Französisch. Allgemeinbildende Schulen. Sekundarstufe I. Sekundarstufe II*. Kiel.
- Ministerium für Schule und Berufsbildung des Landes Schleswig-Holstein (Hg.). 2016. *Fachanforderungen Dänisch. Allgemeinbildende Schulen. Sekundarstufe I. Sekundarstufe II*. Kiel.
- Ministerin für Bildung, Wissenschaft und Kultur. 2018. Dänischunterricht in Schleswig-Holstein. Kleine Anfrage der Abgeordneten Jette Waldinger-Thiering (SSW) und Antwort der Landesregierung (Ministerium für Bildung, Wissenschaft und Kultur). Kiel: Schleswig-Holsteinischer Landtag, 19. Wahlperiode, Drucksache 19/481 <www.

- landtag.ltsh.de/infotek/wahl19/drucks/00400/drucksache-19-00481.pdf; Stand vom 25. 11. 2018>.
- Munro, Murray J. & Tracey M. Derwing. 1995a. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45, 73–97.
- Munro, Murray J. & Tracey M. Derwing. 1995b. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289–306.
- Munro, Murray J. & Tracey M. Derwing. 2008. Segmental acquisition in adult ESL learners. A longitudinal study of vowel production. *Language Learning* 58, 479–502.
- Munro, Murray J. & Tracey M. Derwing. 2011. The foundations of accent and intelligibility in pronunciation research. *Language Teaching* 44, 316–327.
- Munro, Murray J., Tracey M. Derwing & Susan L. Morton. 2006. The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition* 28, 111–131.
- Munsberg, Klaus-Simon. 2001. Bedeutung und Formen von Ausspracheschulung. In Karin Aguado & Claudia Riemer (Hgg.), *Wege und Ziele: Zur Theorie, Empirie und Praxis des Deutschen als Fremdsprache (und anderer Fremdsprachen)*. Festschrift für Gert Henrici zum 60. Geburtstag (Perspektiven Deutsch als Fremdsprache 15), 341–353. Baltmannsweiler: Schneider-Verlag Hohengehren.
- Nimz, Katharina. 2015. *Sound perception and production in a foreign language* (Potsdam Cognitive Science Series 9). Potsdam: Universitätsverlag Potsdam.
- Norell, Pia. 1991. *Native-speaker reactions to Swedish pronunciation errors in English. Recognition, intelligibility and attitude* (Stockholm Studies in English 79). Stockholm: Acta Universitatis Stockholmiensis.
- Nossok, Swetlana. 2009. *Kontrastive phonologische und phonetische Analyse Weißrussisch-Deutsch und Analyse interferenzbedingter Ausspracheabweichungen* (Hallesche Schriften zur Sprechwissenschaft und Phonetik 29). Frankfurt am Main: Lang.
- Odlin, Terence. 1996. *Language transfer. Cross-linguistic influence in language learning*. 5. Aufl. Cambridge: Cambridge University Press.
- Phan, Hoa & Sonca Vo. 2012. Pronunciation errors and perceptual judgements of accented speech by native speakers of English. *TESOL in Context Special Edition* 3, 1–9.
- Pilott, Marty. 2016. *Migrant pronunciation. What do employers find acceptable?* Wellington: Victoria University of Wellington.
- Piske, Thorsten, Ian R. MacKay & James E. Flege. 2001. Factors affecting degree of foreign accent in an L2. A Review. *Journal of Phonetics* 29, 191–215.
- Politzer, Robert L. 1978. Errors of English speakers of German as perceived and evaluated by German natives. *The Modern Language Journal* 62, 253–261.
- Pompino-Marschall, Bernd. 2009. *Einführung in die Phonetik*. 3. Aufl. Berlin: de Gruyter.
- Pude, Angelika. 2007. *Vi snakkes ved. Der Dänischkurs*. Ismaning: Hueber.
- Quirk, Randolph & Jan Svartvik. 1966. Investigating linguistic acceptability. *Janua linguarum* 54.

- R = R Core Team. 2018. *R. A language and environment for statistical computing* [Software]. Vienna: R Foundation for Statistical Computing <www.r-project.org; Stand vom 19. 12. 2018>.
- Revelle, William. 2018. psych. Procedures for personality and psychological research. <cran.r-project.org/web/packages/psych/index; Stand vom 12. 1. 2019>.
- Richards, Jack C. 1972. Social factors, interlanguage and language learning. *Language Learning* 22, 159–188.
- Riemer, Claudia & Julia Settinieri. 2010. Empirische Forschungsmethoden in der Zweit- und Fremdsprachenerwerbsforschung. In Hans-Jürgen Krumm et al. (Hgg.), *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch* (Handbücher zur Sprach- und Kommunikationswissenschaft 35), Bd. 1, 764–781. Berlin: de Gruyter.
- Rifkin, Benjamin & Felicia D. Roberts. 1995. Error gravity. A critical review of research design. *Language Learning* 45, 511–537.
- Saito, Kazuya et al. 2016. Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition* 19, 597–609.
- Saito, Kazuya, Pavel Trofimovich & Talia Isaacs. 2017. Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness. A validation and generalization study. *Applied Linguistics* 38, 439–462.
- Schaden, Stefan. 2006. *Regelbasierte Modellierung fremdsprachlich akzentbehalteter Aussprachevarianten. Grundlagen, Entwurf und Implementierung eines Regelsystems für sprachtechnologische Anwendungen*. Duisburg/Essen: Universität Duisburg-Essen.
- Selinker, Larry. 1972. Interlanguage. *International Review of Applied Linguistics in Language Teaching* 10, 209–231.
- Settinieri, Julia. 2011. Soziale Akzeptanz unterschiedlicher Normabweichungen in der L2-Aussprache Deutsch. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 16, 66–80.
- Sewell, Andrew. 2012. The Hong Kong English Accent. Variation and Acceptability. *Hong Kong Journal of Applied Linguistics* 13, 1–21.
- Sheldon, Amy & Winifred Strange. 1982. The acquisition of /r/ and /l/ by Japanese learners of English. Evidence that speech production can precede speech perception. *Applied Psycholinguistics* 3, 243–261.
- Snefstrup, Kirsten & Lars Schmidt Møller. 2010. *Det lyder godt: Grundbog i udtale*. København: Alfabet.
- SoSci Survey = Leiner, Dominik. 2018. SoSci Survey [Software]. Version 3.1.06-i. München <www.sosicisurvey.com; Stand vom 6. 9. 2017>.
- Spillner, Bernd. 2009. Was der Fremdsprachenunterricht von der Fehleranalyse erwarten darf. In Udo O. H. Jung & Heidrun Jung (Hgg.), *Praktische Handreichung für Fremdsprachenlehrer* (Bayreuther Beiträge zur Glottodidaktik 2), 5. Aufl., 548–555. Frankfurt am Main: Lang.
- Szpyra-Kozłowska, Jolanta. 2015. *Pronunciation in EFL Instruction. A Research-Based Approach* (Second Language Acquisition 82). Bristol: Multilingual Matters.

- Tekin, Özlem. 2012. *Grundlagen der kontrastiven Linguistik in Theorie und Praxis* (Staufenburg Linguistik 64). Tübingen: Stauffenburg.
- Ternes, Elmar. 1976. Probleme der kontrastiven Phonetik (Forum phoneticum 13). Hamburg: Buske.
- Thomson, Ron. 2018. Measurement of accentedness, intelligibility, and comprehensibility. In Okim Kang & April Ginther (Hgg.), *Assessment in second language pronunciation*, 11–29. London: Routledge.
- Tomokiyo, Laura, Alan Black & Kevin Lenzo. 2005. Foreign accents in synthetic speech. Development and evaluation. *Interspeech 2005 – Eurospeech. 9th European Conference on Speech Communication and Technology, Lisbon, Portugal, September 4–8*, 1469–1472.
- Trofimovich, Pavel & Talia Isaacs. 2012. Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition* 15, 905–916.
- Trubetzkoy, Nikolaj. 1971 [1937]. *Grundzüge der Phonologie*. 7. Aufl. Göttingen: Vandenhoeck & Ruprecht.
- Truscott, John. 1996. The case against grammar correction in L2 writing classes. *Language Learning* 46, 327–369.
- Truscott, John. 2007. The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing* 16, 255–272.
- Völkl, Kerstin & Christoph Korb. 2018. *Deskriptive Statistik. Eine Einführung für Politikwissenschaftlerinnen und Politikwissenschaftler*. Wiesbaden: Springer VS.
- Wagner, Johannes & Hans Basbøll. 1985. *Kontrastive Phonologie des Deutschen und Dänischen. Segmentale Wortphonologie und -phonetik* (Linguistische Arbeiten 160). Berlin: de Gruyter.
- Weinreich, Uriel. 1953. *Languages in contact. Findings and problems* (Publications of the Linguistic Circle of New York 1). New York: o. V.
- Wickham, Hadley. 2016. ggplot2: Create elegant data visualisations using the grammar of graphics. R package version 3.1.0 (cran.r-project.org/web/packages/ggplot2/index.html; Stand vom 25. 10. 2018).
- Wickham, Hadley et al. 2018. dplyr. A grammar of data manipulation. R package version 0.7.6 (cran.r-project.org/package=dplyr; Stand vom 11. 7. 2018).
- Zaubauer, Anna C. M., Jan Retelsdorf & Jens Möller. 2009. Die Vorhersage von Englischleistungen am Anfang der Sekundarstufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 41, 153–164.

Anhang A: Studie 1

A.1 Fragebogen zur Erhebung sprecherbiographischer Daten



CODE:

wird am Erhebungstag ausgefüllt

GESCHLECHT:

MÄNNLICH WEIBLICH

ALTER

Dein Geburtsjahr:

SCHULE

KLASSENSTUFE

Welche Sprachen sprichst du meistens zu Hause?

Deutsch

Plattdeutsch

Dänisch

andere Sprachen:

HERKUNFT

Bitte kreuze Zutreffendes an!

	JA	NEIN
Ich bin in Schleswig-Holstein geboren.	<input type="radio"/>	<input type="radio"/>
Ich bin in Deutschland geboren.	<input type="radio"/>	<input type="radio"/>
Mein Vater ist in Deutschland geboren.	<input type="radio"/>	<input type="radio"/>
Meine Mutter ist in Deutschland geboren.	<input type="radio"/>	<input type="radio"/>

DÄNISCHKENNTNISSE

Seit welcher Klasse lernst du Dänisch in der Schule?

Seit Klasse _____.

Hast du Dänisch auch außerhalb der Schule gelernt? Wenn ja, wo?

NEIN

JA _____

Sprichst du Dänisch auch außerhalb des Unterrichts?
Wenn ja, in welchem Rahmen und mit wem?

NEIN

JA _____

A.2 Analysebogen zur effizienten quantitativen Auswertung von Lernerfehlern

dans-Projekt

Auswertungsbogen für Sprecher (deutsche Dänischlerner)

Code: _____



Sprecher: _____

Datum der Aufnahme: _____

Auswerter: _____

Datum der Auswertung: _____

Zutreffende Realisierungen werden angekreuzt oder farblich markiert.

Satz		Kodierung							
1	Jeg har kun to kroner.	K_ts_ts	K_ts_th				K_ts_mx	K_ts_?	
24	Jeg er træt .	V_a_a	V_a_ä				V_a_mx	V_a_?	
2	Det er Tines bil.	K_ts_ts	K_ts_th				K_ts_mx	K_ts_?	
28	Må jeg låne din bil?	V_gaa_gaa	V_gaa_o	V_gaa_c	V_gaa_aa		V_gaa_mx	V_gaa_?	
5	Tak for den dejlige rose .	K_sm_s	K_si_z				K_sm_mx	K_sm_?	
31	Det var alt for kort .	V_c_c	V_c_o				V_c_mx	V_c_?	
61	Hvad skal det betyde ?	Pro_QuKbS_wd:	Pro_QuKbS_e				Pro_QuKuS_mx	Pro_QuKuS_?	
8	Jeg elsker nødder .	K_mwd_wd	K_mwd_l	K_mwd_d			K_mwd_mx	K_mwd_?	
9	Vi skal gå ud i haven.	K_fwd_wd	K_fwd_l	K_fwd_d			K_fwd_mx	K_fwd_?	
3	Han synger gerne.	K_si_s	K_si_z				K_si_mx	K_si_?	
12	Det var sjætte gang nul	K_sj_1K [c]	K_sj_sj	K_sj_sk	K_sj_sch		K_sj_mx	K_sj_?	
38	Der sidder en spurv på taget.	PolyTr_uau_uau	PolyTr_uau_uav	PolyTr_uau_ua			PolyTr_uau_mx	PolyTr_uau_?	
17	Vi maler vores hus.	V_3_3	V_3_e	V_3_ä	V_3_a		V_3_mx	V_3_?	
58	Jeg skal finde min bil.	Pro_QuKuS_n:	Pro_QuKuS_e				Pro_QuKuS_mx	Pro_QuKuS_?	
40	Vi flyver til Mexico.	Poly_yu_yu	Poly_yu_yv	Poly_yu_y			Poly_yu_mx	Poly_yu_?	

A.2 Analysebogen zur effizienten quantitativen Auswertung von Lernerfehlern

dans-Projekt

Auswertungsbogen für Sprecher (deutsche Dänischlerer)

Code.: _____



20	Lines mand er venlig.	V_ä_ä	V_ä_3	V_ä_a		V_ä_?	V_ä_mx
64	Må jeg se mappen ?	Pro_SonPpn_b	Pro_SonPpn_p			Pro_SonPpn_?	Pro_SonPpn_mx
37	Sneen hvirvler rundt.	PolyTr_jau_jau	PolyTr_jau_jav	PolyTr_jau_irv		PolyTr_jau_?	PolyTr_jau_mx
22	Jeg elsker den sang .	V_a_a	V_a_ä			V_a_?	V_a_mx
54	Den her vej er kortere .	Pro_QuVuS_a:	Pro_QuVuS_e			Pro_QuVbS_?	Pro_QuVbS_mx
7	Hvornår mødes vi?	K_mwd_wd	K_mwd_l	K_mwd_d		K_mwd_mx	K_mwd_?
25	Du har ret .	V_a_a	V_a_ä			V_a_?	V_a_mx
14	Vi skal melde os mere i timen.	V_e_e	V_e_3			V_e_?	V_e_mx
26	Den gamle mands ben er lam .	V_a_a	V_a_ä			V_a_?	V_a_mx
66	Må jeg få jakken ?	Pro_SonPpn_g	Pro_SonPpn_k			Pro_SonPpn_?	Pro_SonPpn_mx
27	Giv mig din hånd !	V_huV_aa	V_huV_o	V_huV_a	V_huV_ä	V_huV_?	V_huV_mx
43	De besøgte milen ved Skagen.	S_ji_i:	S_ji_j	S_ji_1		S_ji_?	S_ji_mx
57	De siger vi skal komme i aften.	Pro_QuKuS_m:	Pro_QuKuS_e			Pro_QuKuS_?	Pro_QuKuS_mx
29	Hän skal gå!	V_gaa_gaa	V_gaa_o	V_gaa_c	V_gaa_aa	V_gaa_?	V_gaa_mx
13	Det sker bare sjældent!	K_sj_1K [c]	K_sj_sj	K_sj_sk	K_sj_sch	K_sj_mx	K_sj_?
34	Det regner .	Poly_ai_ai	Poly_ai_eg			Poly_ai_?	Poly_ai_mx
63	Mette er min vens kone.	Pro_SonPiv_d	Pro_SonPiv_t			Pro_SonPiv_?	Pro_SonPiv_mx
19	Lise har en kat .	V_ä_ä	V_ä_3	V_ä_a		V_ä_?	V_ä_mx
4	Det er sødt af dig.	K_si_s	K_si_z			K_si_mx	K_si_?

2

A.3 R-Output Chi-Quadrat-Tests⁵⁸

A.3.1 Ergebnisse zu sprachspezifischen Unterschieden (Niederdeutsch)

Alle Fehler (übergreifend)

```
> table(quant_i.nd $value, quant_i.nd $Niederdeutsch)
```

```
stdt nd
kf 45 13
F 79 41
```

```
> tbl.quant_i.nd <- table(quant_i.nd $value, quant_i.nd $Niederdeutsch)
> chisq.test(tbl.quant_i.nd, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.quant_i.nd
```

⁵⁸ kf = kein Fehler; F = fehlerhaft

X-squared = 2.5557, df = 1, p-value = 0.1099

```
> assocstats(tbl.quanti.nd)
  X^2 df P(> X^2)
Likelihood Ratio 2.6394 1 0.10424
Pearson 2.5557 1 0.10990
```

```
Phi-Coefficient : 0.12
Contingency Coeff.: 0.119
Cramer's V : 0.12
```

Zentrales a [ä]

```
> table(zentr.a.nd $value, zentr.a.nd $Niederdeutsch)

      stdt  nd
kF  115  46
F   11  8
> tbl.zentr.a.nd <- table(zentr.a.nd $value, zentr.a.nd
$Niederdeutsch)
> chisq.test(tbl.zentr.a.nd, corr=F)
```

Pearson's Chi-squared test

```
data:  tbl.zentr.a.nd
X-squared = 1.4823, df = 1, p-value = 0.2234
```

```
> assocstats(tbl.zentr.a.nd)
  X^2 df P(> X^2)
Likelihood Ratio 1.4044 1 0.23598
Pearson 1.4823 1 0.22342
```

```
Phi-Coefficient : 0.091
Contingency Coeff.: 0.09
Cramer's V : 0.091
```

Vorderer ungerundeter Vokal [æ]

```
> table(ä.nd$value, ä.nd$Niederdeutsch)

      stdt  nd
kF   25  13
F   99  41
> tbl.ä.nd <- table(ä.nd$value, ä.nd$Niederdeutsch)
> chisq.test(tbl.quanti.nd, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.quanti.nd
X-squared = 2.5557, df = 1, p-value = 0.1099
```

```
> assocstats(tbl.quanti.nd)
  X^2 df P(> X^2)
Likelihood Ratio 2.6394 1 0.10424
Pearson 2.5557 1 0.10990
```

```
Phi-Coefficient : 0.12
Contingency Coeff.: 0.119
Cramer's V : 0.12
```

Hinterer gerundeter Vokal [ɔ]

```
table(c.nd$value, c.nd$Niederdeutsch)
```

```
stdt nd
kF 80 36
F 40 17
```

```
> tbl.c.nd <- table(c.nd$value, c.nd$Niederdeutsch)
> chisq.test(tbl.quanti.nd, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.quanti.nd
X-squared = 2.5557, df = 1, p-value = 0.1099
```

```
> assocstats(tbl.quanti.nd)
  X^2 df P(> X^2)
Likelihood Ratio 2.6394 1 0.10424
Pearson 2.5557 1 0.10990
```

```
Phi-Coefficient : 0.12
Contingency Coeff.: 0.119
Cramer's V : 0.12
```

Vokalquantität

```
table(quant_i.nd$value, quant_i.nd$Niederdeutsch)
```

```
stdt nd
kF 45 13
```

```
F    79    41
> tbl.quanti.nd <- table(quant_i.nd $value, quanti.nd $Niederdeutsch)
> chisq.test(tbl.quanti.nd, corr=F)
```

Pearson's Chi-squared test

```
data:  tbl.quanti.nd
X-squared = 2.5557, df = 1, p-value = 0.1099
```

```
> assocstats(tbl.quanti.nd)
  X^2 df P(> X^2)
Likelihood Ratio 2.6394 1 0.10424
Pearson 2.5557 1 0.10990
```

```
Phi-Coefficient   : 0.12
Contingency Coeff.: 0.119
Cramer's V       : 0.12
```

Lenisierung nachtoniger Plosive intervokalisch

```
table(snpiv.nd $value, snpiv.nd $Niederdeutsch)
```

```
      stdt  nd
      kF  21 10
      F   109 62
> tbl.snpiv.nd <- table(snpiv.nd $value, snpiv.nd $Niederdeutsch)
> chisq.test(tbl.snpiv.nd, corr=F)
```

Pearson's Chi-squared test

```
data:  tbl.snpiv.nd
X-squared = 0.18297, df = 1, p-value = 0.6688
```

```
> assocstats(tbl.snpiv.nd)
  X^2 df P(> X^2)
Likelihood Ratio 0.18527 1 0.66688
Pearson 0.18297 1 0.66883
```

```
Phi-Coefficient   : 0.03
Contingency Coeff.: 0.03
Cramer's V       : 0.03
```

Lenisierung nachtoniger Plosive pränasal

```
table(snpnp.nd $value, snpnp.nd $Niederdeutsch)
```

```

      stdt  nd
kF 19 5
F 130 72
> tbl.snppn.nd <- table(snppn.nd $value, snppn.nd $Niederdeutsch)
> chisq.test(tbl.snppn.nd, corr=F)

```

Pearson's Chi-squared test

```

data: tbl.snppn.nd
X-squared = 2.0947, df = 1, p-value = 0.1478

```

```

> assocstats(tbl.snppn.nd)
  X^2 df P(> X^2)
Likelihood Ratio 2.2550 1 0.13318
Pearson 2.0947 1 0.14781

```

```

Phi-Coefficient : 0.096
Contingency Coeff.: 0.096
Cramer's V : 0.096

```

A.3.2 Ergebnisse zu den anwendungshäufigkeitsbedingten Unterschieden

Alle Fehler (übergreifend)

```

> table(alle.fehler.dan$value,
alle.fehler.dan$Dänisch.sprechen.außerhalb.der.Schule)

```

```

aadS kD
kF 370 1235
F 662 2378
> tbl.dan <- table(alle.fehler.dan$value,
alle.fehler.dan$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.dan)

```

Pearson's Chi-squared test with Yates' continuity correction

```

data: tbl.dan
X-squared = 0.91817, df = 1, p-value = 0.338

```

```

> assocstats(tbl.dan)
  X^2 df P(> X^2)
Likelihood Ratio 0.98661 1 0.32057

```

Pearson 0.99066 1 0.31958

Phi-Coefficient : 0.015

Contingency Coeff.: 0.015

Cramer's V : 0.015

Weiches *d*

```
table(wd.m.dan$value, wd.m.dan$Dänisch.sprechen.außerhalb.der.Schule)
```

```
aadS kD
```

```
kF 14 40
```

```
F 23 94
```

```
> tbl.wd.m.dan <- table(wd.m.dan $value, wd.m.dan
$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.wd.m.dan, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.wd.m.dan
```

```
X-squared = 0.85605, df = 1, p-value = 0.3548
```

```
> assocstats(tbl.wd.m.dan)
```

```
X^2 df P(> X^2)
```

```
Likelihood Ratio 0.83708 1 0.36023
```

```
Pearson 0.85605 1 0.35485
```

Phi-Coefficient : 0.071

Contingency Coeff.: 0.071

Cramer's V : 0.071

hinterer ungerundeter Vokal [ɤ]

```
> table(V.huV.dan$value,
V.huV.dan$Dänisch.sprechen.außerhalb.der.Schule)
```

```
aadS kD
```

```
kF 15 43
```

```
F 3 22
```

```
> tbl.V.huV.dan <- table(V.huV.dan $value, V.huV.dan
$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.V.huV.dan, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.V.huV.dan
X-squared = 1.9766, df = 1, p-value = 0.1598
```

```
> assocstats(tbl.V.huV.dan)
  X^2 df P(> X^2)
Likelihood Ratio 2.1509 1 0.14249
Pearson 1.9766 1 0.15975
```

```
Phi-Coefficient : 0.154
Contingency Coeff.: 0.153
Cramer's V : 0.154
```

Bekannte Polyphthonge [äy]

```
table(poly.au.dan$value,
poly.au.dan$Dänisch.sprechen.außerhalb.der.Schule)
```

```
aadS kD
kF 17 49
F 2 20
```

```
> tbl.poly.au.dan <- table(poly.au.dan $value, poly.au.dan
$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.poly.au.dan, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.poly.au.dan
X-squared = 2.7073, df = 1, p-value = 0.09989
```

Warning message:

```
In chisq.test(tbl.poly.au.dan, corr = F) :
Chi-Quadrat-Approximation kann inkorrekt sein
```

```
> assocstats(tbl.poly.au.dan)
  X^2 df P(> X^2)
Likelihood Ratio 3.1051 1 0.078046
Pearson 2.7073 1 0.099887
```

```
Phi-Coefficient : 0.175
Contingency Coeff.: 0.173
Cramer's V : 0.175
```

Bekannte Polyphthonge [äi]

```
> table(poly.eu.dan$value,
poly.eu.dan$Dänisch.sprechen.außerhalb.der.Schule)
```

```
aadS kD
kF 14 44
F 5 25
> tbl.poly.eu.dan <- table(poly.eu.dan $value, poly.eu.dan
$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.poly.eu.dan, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.poly.eu.dan
X-squared = 0.65195, df = 1, p-value = 0.4194
```

```
> assocstats(tbl.poly.eu.dan)
  X^2 df P(> X^2)
Likelihood Ratio 0.67313 1 0.41196
Pearson 0.65195 1 0.41941
```

```
Phi-Coefficient : 0.086
Contingency Coeff.: 0.086
Cramer's V : 0.086
```

Diphthonge auf [u]

```
> table(diph.u.dan$value,
diph.u.dan$Dänisch.sprechen.außerhalb.der.Schule)
```

```
aadS kD
kF 28 59
F 28 145
> tbl.diph.u.dan <- table(diph.u.dan $value, diph.u.dan
$Dänisch.sprechen.außerhalb.der.Schule)
> chisq.test(tbl.diph.u.dan, corr=F)
```

Pearson's Chi-squared test

```
data: tbl.diph.u.dan
X-squared = 8.768, df = 1, p-value = 0.003066
```

```
> assocstats(tbl.diph.u.dan)
  X^2 df P(> X^2)
Likelihood Ratio 8.4243 1 0.0037024
Pearson 8.7680 1 0.0030655
```

```
Phi-Coefficient : 0.184
```

Contingency Coeff.: 0.181
 Cramer's V : 0.184

A.3.3 Ergebnisse zu geschlechtsspezifischen Unterschieden

	m	w
kein Fehler	501	1168
Fehler	1063	2080

Pearson's Chi-squared test with Yates' continuity correction

data: tbl.geschl
 X-squared = 7.0151, df = 1, p-value = 0.008082

```
> assocstats(tbl.geschl)
  X^2 df P(> X^2)
Likelihood Ratio 7.2361 1 0.0071454
Pearson 7.1875 1 0.0073415
```

Phi-Coefficient : 0.039
 Contingency Coeff.: 0.039
 Cramer's V : 0.039

A.4 Mittelwerte (nicht näher untersuchte Fehler der verschiedenen Klassen)

Untersuchtes Merkmal	Lehrkräfte und relative Fehlerhäufigkeiten in %							
	LK1	LK2	LK3	LK4	LK5	LK6	LK7	LK8
affriziertes t [tʰ]	100	100	100	100	100	100	100	100
stimmloses [s] wortinitial	0	0	0	0	0	0	0	6
zentraler a-Laut [ä], orthographisch leicht	4	11	18	12	6	14	5	17
vorderer ungerundeter Vokal [æ]	85	75	89	81	88	57	62	89
hinterer ungerundeter Vokal [ɔ̃]	30	54	36	33	0	14	18	33
hinterer gerundeter Vokal [ɔ̈]	96	96	100	92	94	93	91	88
hinterer gerundeter Vokal [ɔ̄]	0	7	0	0	0	0	0	14
Triphthonge	100	93	96	100	100	100	100	100
überlange Vokale in Folge von [ə]-Tilgungsprozessen, betonte Silben	100	100	96	100	93	93	95	100

Untersuchtes Merkmal	Lehrkräfte und relative Fehlerhäufigkeiten in %							
	LK1	LK2	LK3	LK4	LK5	LK6	LK7	LK8
überlange Vokale in Folge von [ə]- Tilgungsprozessen, unbetonte Silben	100	75	96	95	100	93	86	89
lange Konsonanten in Folge von [ə]- Tilgungsprozessen: [l:], [m:], [n:], [ŋ:]	100	100	98	98	96	96	100	94
lange Konsonanten in Folge von [ə]- Tilgungsprozessen: Längung weiches <i>d</i> ([ð])	100	100	92	92	100	86	86	94
Lenisierung nachtoniger Plosive pränasal	69	100	79	97	95	100	94	100

Anhang B: Studie 2

B.1 Onlinefragebogen

 **CAU** < 0% færdig

Kære deltager,

tak for din interesse i vores undersøgelse og at du vil være med!

I vores projekt prøver vi at finde de specifikke sproglige fænomener, der gør, at danskerne ret hurtigt kan høre, at en person har lært dansk som et fremmedsprog.

Det tager ca. en halv times tid at udfylde spørgeskemaet. Du vil lytte til forskellige sprogoptagelser og besvare nogle spørgsmål om det hørte.

Det er vigtigt, at du bruger **firefox-browsersen**. Ellers kan det ske, at lyden ikke virker.

Deltagelsen er frivilligt, du må altid stoppe, hvis du ikke længere vil deltage i undersøgelsen. Resultaterne er anonyme og ikke tilknyttet til dit navn, e-mail eller andre oplysninger.

Med venlig hilsen

Lisa Tulaja

CAU

Christian-Albrechts-Universität zu Kiel

[dansk-Projekt, CAU Kiel – 2017](#)

 **CAU** < 1% færdig

Fortæl først lidt om dig selv!

Jeg er... år gammel.

Jeg læser/ har læst... / Jeg er uddannet...

Jeg er [Vælg]

Jeg bor i

Jeg er opvokset i

Min højeste uddannelse er

Folkeskole

Efterskole

Gymnasiet

Erhvervsakademiuddannelse

Bachelor

Kandidat

PhD

andet

Jeg taler de følgende sprog:

På modersmål niveau:

Ubesværet:

Kan kommunikere:

Kan forstå, men taler ikke:

Har kendskab til (kender elementer):

Det/ de første sprog, jeg talte, var...
(Det kan være, at du er opvokset med mere end et sprog som første sprog. I så fald skriv alle sprog og sæt komma.)

Det er nemt for mig at lære nye sprog.
helt uenig helt enig

Det er nemt for mig at kommunikere på andre sprog end dansk.
helt uenig helt enig

Jeg kan godt lide at lære nye sprog.
helt uenig helt enig

[dansk-Projekt, CAU Kiel – 2017](#)

 **CAU** < 2% færdig

OBS!

Du vil komme til at høre mange forskellige sprogoptagelser. Dette er vigtigt for resultaternes bevismagt, men det kan også blive lidt trættende.

Du må altid gerne holde en pause undervejs, hvis du føler dig til det!

Det er vigtigt at afslutte spørgeskemaet helt for at være med i konkurrencen.

Før du klikker på "næste", check lige, om lyden er tændt og at du har høretelefonerne på.

[dansk-Projekt, CAU Kiel – 2017](#)

B.2 Ergebnisse der Regressionsanalysen zu den Unterschiedlichkeiten zwischen den Stimulustypen

B.2.1 *Accentedness*

```
> #Accentedness Item 1 (A_01)
> #Muttersprachler als Intercept
> model1A_01 <- lm(A_01 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1A_01)
```

Call:

```
lm(formula = A_01 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-3.5420 -0.4044 0.4580 0.4580 3.5956
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.40440 0.02373 59.19 <2e-16 ***
Sp.Kl.F 3.13763 0.02961 105.97 <2e-16 ***
Sp.Kl.K 1.90626 0.03350 56.90 <2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 8108 degrees of freedom

(4993 observations deleted due to missingness)

Multiple R-squared: 0.5811, Adjusted R-squared: 0.581

F-statistic: 5623 on 2 and 8108 DF, p-value: < 2.2e-16

```
> Accentedness Item 2 (A_02)
> model1A_02 <- lm(A_02 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1A_02)
```

Call:

```
lm(formula = A_02 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-3.6363 -0.3857 0.3637 0.3637 3.6143
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.38566 0.02393 57.91 <2e-16 ***
```

```
Sp.Kl.F  3.25063 0.02986 108.88 <2e-16 ***
Sp.Kl.K  2.01350 0.03379  59.60 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.105 on 8108 degrees of freedom
(4993 observations deleted due to missingness)
Multiple R-squared:  0.594, Adjusted R-squared:  0.5939
F-statistic: 5930 on 2 and 8108 DF, p-value: < 2.2e-16
```

```
> #Kontrollsprecher als Intercept
> model1bA_01 <- lm(A_01 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
> summary(model1bA_01)
Item 1
Call:
lm(formula = A_01 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

```
Residuals:
  Min   1Q   Median   3Q   Max
-3.5420 -0.4044  0.4580  0.4580  3.5956
```

```
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.31067  0.02365  139.96 <2e-16 ***
Sp.Kl.F      1.23137  0.02955   41.67 <2e-16 ***
Sp.Kl.KM    -1.90626  0.03350  -56.90 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.096 on 8108 degrees of freedom
(4993 observations deleted due to missingness)
Multiple R-squared:  0.5811, Adjusted R-squared:  0.581
F-statistic: 5623 on 2 and 8108 DF, p-value: < 2.2e-16
```

```
> Item 2 (A_02)

> model1bA_02 <- lm(A_02 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
> summary(model1bA_02)
Call:
lm(formula = A_02 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

```
Residuals:
  Min   1Q   Median   3Q   Max
-3.6363 -0.3857  0.3637  0.3637  3.6143
```

```
Coefficients:
```

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.39916  0.02385  142.50  <2e-16 ***
Sp.Kl.F      1.23713  0.02980   41.52  <2e-16 ***
Sp.Kl.KM    -2.01350  0.03379  -59.60  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.105 on 8108 degrees of freedom
(4993 observations deleted due to missingness)
Multiple R-squared:  0.594, Adjusted R-squared:  0.5939
F-statistic: 5930 on 2 and 8108 DF, p-value: < 2.2e-16
```

B.2.2 Comprehensibility

```
> #Comprehensibility Item 1 (S_03)
> #Muttersprachler als Intercept
> model1S_03 <- lm(S_03 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1S_03)
```

Call:

```
lm(formula = S_03 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
  Min    1Q  Median    3Q   Max
-1.3789 -0.4634 -0.0717  0.5366  3.9283
```

Coefficients:

```
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.07170  0.02224   48.18  <2e-16 ***
Sp.Kl.F      1.30719  0.02776   47.09  <2e-16 ***
Sp.Kl.K      0.39174  0.03141   12.47  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.027 on 8105 degrees of freedom
(4996 observations deleted due to missingness)
Multiple R-squared:  0.237, Adjusted R-squared:  0.2368
F-statistic: 1258 on 2 and 8105 DF, p-value: < 2.2e-16
```

```
> Comprehensibility Item 2 (S_04)
> model1S_04 <- lm(S_04 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1S_04)
```

Call:

```
lm(formula = S_04 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-1.5276 -0.5276 -0.0665 0.4724 3.9335
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.06654 0.02334 45.69 <2e-16 ***
Sp.Kl.F 1.46103 0.02913 50.15 <2e-16 ***
Sp.Kl.K 0.44860 0.03296 13.61 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.078 on 8105 degrees of freedom

(4996 observations deleted due to missingness)

Multiple R-squared: 0.2595, Adjusted R-squared: 0.2593

F-statistic: 1420 on 2 and 8105 DF, p-value: < 2.2e-16

#Kontrollsprecher als Intercept

```
> model1bS_03 <- lm(S_03 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
```

```
> summary(model1bS_03)
```

Item 1

Call:

```
lm(formula = S_03 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-1.3789 -0.4634 -0.0717 0.5366 3.9283
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.46344 0.02217 66.00 <2e-16 ***
Sp.Kl.F 0.91545 0.02771 33.04 <2e-16 ***
Sp.Kl.KM -0.39174 0.03141 -12.47 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.027 on 8105 degrees of freedom

(4996 observations deleted due to missingness)

Multiple R-squared: 0.237, Adjusted R-squared: 0.2368

F-statistic: 1258 on 2 and 8105 DF, p-value: < 2.2e-16

```
> Item 2
```

```
> model1bS_04 <- lm(S_04 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
```

```
> summary(model1bS_04)
```

Call:

```
lm(formula = S_04 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-1.5276 -0.5276 -0.0665 0.4724 3.9335
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.51514 0.02327 65.11 <2e-16 ***
Sp.Kl.F 1.01243 0.02908 34.82 <2e-16 ***
Sp.Kl.KM -0.44860 0.03296 -13.61 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 1.078 on 8105 degrees of freedom
(4996 observations deleted due to missingness)
Multiple R-squared: 0.2595, Adjusted R-squared: 0.2593
F-statistic: 1420 on 2 and 8105 DF, p-value: < 2.2e-16

B.2.3 Acceptability

```
#Multilevel regression: Muttersprache Kontroll-Intercept und Fehler und
Fremd Kontroll Pred
>
> #Acceptability Item 1 (S_01)

> #Muttersprachler als Intercept
> model1S_01 <- lm(S_01 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1S_01)
```

Call:

```
lm(formula = S_01 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-0.8999 -0.3116 -0.0450 0.1001 3.9550
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.04499 0.01840 56.80 <2e-16 ***
Sp.Kl.F 0.85494 0.02296 37.23 <2e-16 ***
Sp.Kl.K 0.26661 0.02598 10.26 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.8499 on 8105 degrees of freedom
(4996 observations deleted due to missingness)
Multiple R-squared: 0.1614, Adjusted R-squared: 0.1612
F-statistic: 780.2 on 2 and 8105 DF, p-value: < 2.2e-16
```

```
> Acceptability Item 2 (S_02)
> model1S_02 <- lm(S_02 ~ Sp.Kl.F + Sp.Kl.K ,data_1)
> summary(model1S_02)
```

Call:

```
lm(formula = S_02 ~ Sp.Kl.F + Sp.Kl.K, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-1.6998 -0.6698 -0.0773 0.3302 3.9227
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.07732 0.02382 45.22 <2e-16 ***
Sp.Kl.F 1.62245 0.02973 54.57 <2e-16 ***
Sp.Kl.K 0.59245 0.03364 17.61 <2e-16 ***
---
```

```
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 1.101 on 8105 degrees of freedom
(4996 observations deleted due to missingness)
Multiple R-squared: 0.2856, Adjusted R-squared: 0.2854
F-statistic: 1620 on 2 and 8105 DF, p-value: < 2.2e-16
```

```
> #Kontrollsprecher als Intercept
> model1bS_01 <- lm(S_01 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
> summary(model1bS_01)
```

Item 1

Call:

```
lm(formula = S_01 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

Residuals:

```
Min 1Q Median 3Q Max
-0.8999 -0.3116 -0.0450 0.1001 3.9550
```

Coefficients:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.31160 0.01834 71.50 <2e-16 ***
Sp.Kl.F 0.58832 0.02292 25.67 <2e-16 ***
Sp.Kl.KM -0.26661 0.02598 -10.26 <2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8499 on 8105 degrees of freedom
 (4996 observations deleted due to missingness)
 Multiple R-squared: 0.1614, Adjusted R-squared: 0.1612
 F-statistic: 780.2 on 2 and 8105 DF, p-value: < 2.2e-16

```
>
> model1bS_02 <- lm(S_02 ~ Sp.Kl.F + Sp.Kl.KM ,data_1)
> summary(model1bS_02)
Item 2
Call:
lm(formula = S_02 ~ Sp.Kl.F + Sp.Kl.KM, data = data_1)
```

Residuals:
 Min 1Q Median 3Q Max
 -1.6998 -0.6698 -0.0773 0.3302 3.9227

Coefficients:
 Estimate Std. Error t value Pr(>|t|)
 (Intercept) 1.66977 0.02375 70.30 <2e-16 ***
 Sp.Kl.F 1.02999 0.02968 34.71 <2e-16 ***
 Sp.Kl.KM -0.59245 0.03364 -17.61 <2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.101 on 8105 degrees of freedom
 (4996 observations deleted due to missingness)
 Multiple R-squared: 0.2856, Adjusted R-squared: 0.2854
 F-statistic: 1620 on 2 and 8105 DF, p-value: < 2.2e-16

B.3 Mittelwerte und Konfidenzintervalle

Accentedness Item 1

Fehlernummer	M	SD	SE	95% KI
1	4,59	0,70	0,05	[4,5; 4,68]
2	4,70	0,70	0,07	[4,57; 4,83]
3	4,74	0,58	0,05	[4,63; 4,85]
4	4,57	0,82	0,05	[4,46; 4,68]

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	4,35	0,98	0,07	[4,22; 4,48]
6	4,65	0,65	0,04	[4,56; 4,73]
7	4,79	0,59	0,06	[4,68; 4,9]
8	4,63	0,77	0,07	[4,49; 4,77]
9	4,50	0,92	0,06	[4,38; 4,61]
10	4,71	0,61	0,06	[4,59; 4,82]
11	4,73	0,62	0,04	[4,65; 4,81]
12	4,84	0,53	0,03	[4,77; 4,91]
13	4,72	0,60	0,06	[4,61; 4,83]
14	4,32	1,23	0,12	[4,09; 4,54]
15	4,70	0,65	0,06	[4,58; 4,82]
16	4,53	0,77	0,07	[4,39; 4,68]
17	4,82	0,45	0,04	[4,73; 4,9]
18	4,37	0,92	0,09	[4,2; 4,54]
19	4,39	0,86	0,06	[4,29; 4,5]
20	4,72	0,67	0,04	[4,63; 4,81]
21	4,47	0,89	0,06	[4,36; 4,59]
22	4,63	0,73	0,05	[4,53; 4,72]
23	2,62	1,39	0,13	[2,37; 2,87]

Accentedness Item 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
1	4,77	0,53	0,04	[4,7; 4,84]
2	4,76	0,75	0,07	[4,62; 4,9]
3	4,81	0,51	0,05	[4,72; 4,91]
4	4,69	0,71	0,05	[4,6; 4,78]

Anhang B: Studie 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	4,45	0,97	0,07	[4,32; 4,58]
6	4,71	0,67	0,04	[4,62; 4,8]
7	4,88	0,40	0,04	[4,81; 4,96]
8	4,69	0,66	0,06	[4,57; 4,81]
9	4,53	0,96	0,06	[4,4; 4,65]
10	4,71	0,68	0,07	[4,58; 4,83]
11	4,84	0,47	0,03	[4,78; 4,91]
12	4,92	0,34	0,02	[4,87; 4,96]
13	4,83	0,61	0,06	[4,72; 4,95]
14	4,13	1,34	0,13	[3,88; 4,38]
15	4,83	0,48	0,04	[4,75; 4,92]
16	4,67	0,62	0,06	[4,55; 4,78]
17	4,85	0,56	0,05	[4,74; 4,95]
18	4,54	0,85	0,08	[4,39; 4,7]
19	4,53	0,85	0,06	[4,42; 4,64]
20	4,79	0,58	0,04	[4,72; 4,87]
21	4,70	0,70	0,05	[4,61; 4,79]
22	4,70	0,66	0,04	[4,61; 4,78]
23	2,77	1,45	0,13	[2,51; 3,03]

Comprehensibility Item 1

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
1	1,70	0,92	0,06	[1,58; 1,82]
2	2,71	1,22	0,12	[2,49; 2,94]
3	3,35	1,33	0,13	[3,1; 3,6]
4	2,69	1,45	0,10	[2,5; 2,88]

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	1,92	1,07	0,07	[1,78; 2,06]
6	2,33	1,27	0,08	[2,17; 2,5]
7	2,25	1,05	0,10	[2,05; 2,44]
8	2,07	1,05	0,10	[1,88; 2,26]
9	2,02	1,12	0,07	[1,88; 2,17]
10	3,17	1,29	0,12	[2,92; 3,41]
11	2,47	1,29	0,09	[2,3; 2,65]
12	3,27	1,33	0,09	[3,1; 3,44]
13	2,86	1,26	0,12	[2,63; 3,09]
14	4,05	1,27	0,12	[3,82; 4,29]
15	2,96	1,30	0,12	[2,73; 3,2]
16	1,49	0,82	0,08	[1,33; 1,64]
17	3,24	1,29	0,12	[2,99; 3,48]
18	2,19	1,25	0,12	[1,96; 2,42]
19	1,65	0,90	0,06	[1,53; 1,76]
20	2,37	1,30	0,09	[2,2; 2,54]
21	1,94	1,08	0,07	[1,79; 2,08]
22	2,32	1,19	0,08	[2,16; 2,47]
23	1,26	0,77	0,07	[1,12; 1,4]

Comprehensibility Item 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
1	1,73	0,96	0,06	[1,6; 1,86]
2	2,89	1,35	0,13	[2,64; 3,14]
3	3,52	1,35	0,13	[3,27; 3,77]
4	2,85	1,52	0,10	[2,65; 3,05]

Anhang B: Studie 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	2,03	1,21	0,08	[1,87; 2,19]
6	2,44	1,36	0,09	[2,26; 2,62]
7	2,48	1,16	0,11	[2,26; 2,69]
8	2,30	1,15	0,11	[2,09; 2,51]
9	2,07	1,13	0,07	[1,93; 2,22]
10	3,30	1,31	0,13	[3,06; 3,55]
11	2,69	1,35	0,09	[2,52; 2,87]
12	3,52	1,36	0,09	[3,34; 3,69]
13	3,06	1,26	0,12	[2,83; 3,29]
14	4,23	1,26	0,12	[3,99; 4,46]
15	3,05	1,25	0,12	[2,82; 3,28]
16	1,61	0,88	0,08	[1,45; 1,78]
17	3,48	1,28	0,12	[3,24; 3,72]
18	2,32	1,29	0,12	[2,09; 2,56]
19	1,77	0,97	0,06	[1,64; 1,89]
20	2,53	1,39	0,09	[2,35; 2,72]
21	2,05	1,21	0,08	[1,89; 2,21]
22	2,52	1,31	0,09	[2,35; 2,7]
23	1,35	0,80	0,07	[1,2; 1,49]

Acceptability Item 1

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
1	1,64	0,96	0,06	[1,51; 1,76]
2	2,27	1,25	0,12	[2,04; 2,5]
3	2,32	1,22	0,12	[2,09 ; 2,55]
4	1,89	1,08	0,07	[1,75; 2,03]

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	1,55	0,82	0,05	[1,44; 1,66]
6	1,81	1,00	0,07	[1,68; 1,94]
7	2,03	1,07	0,10	[1,83; 2,22]
8	1,68	0,91	0,08	[1,52; 1,85]
9	1,68	0,98	0,07	[1,55; 1,8]
10	2,37	1,26	0,12	[2,13; 2,6]
11	2,04	1,16	0,08	[1,88; 2,19]
12	2,45	1,24	0,08	[2,29; 2,61]
13	2,25	1,19	0,11	[2,03; 2,46]
14	2,71	1,38	0,13	[2,46; 2,97]
15	2,15	1,12	0,11	[1,94; 2,36]
16	1,49	0,83	0,08	[1,33; 1,64]
17	2,06	1,14	0,11	[1,85; 2,28]
18	1,74	1,02	0,10	[1,55; 1,92]
19	1,61	0,92	0,06	[1,49; 1,72]
20	1,99	1,15	0,08	[1,84; 2,14]
21	1,68	0,93	0,06	[1,55; 1,8]
22	1,89	0,97	0,06	[1,76; 2,01]
23	1,16	0,51	0,05	[1,07; 1,25]

Acceptability Item 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
1	2,33	1,24	0,08	[2,17; 2,5]
2	3,33	1,34	0,13	[3,08; 3,58]
3	3,31	1,32	0,12	[3,07; 3,56]
4	2,70	1,37	0,09	[2,52; 2,88]

Anhang B: Studie 2

Fehlernummer	<i>M</i>	<i>SD</i>	<i>SE</i>	95% KI
5	2,12	1,20	0,08	[1,96; 2,27]
6	2,60	1,35	0,09	[2,42; 2,77]
7	3,04	1,26	0,12	[2,8; 3,27]
8	2,41	1,32	0,12	[2,17; 2,65]
9	2,44	1,31	0,09	[2,27; 2,61]
10	3,27	1,27	0,12	[3,03; 3,5]
11	3,03	1,34	0,09	[2,86; 3,21]
12	3,47	1,31	0,09	[3,31; 3,64]
13	3,18	1,33	0,12	[2,94; 3,43]
14	3,45	1,44	0,14	[3,18; 3,72]
15	3,20	1,25	0,12	[2,97; 3,43]
16	2,10	1,14	0,11	[1,89; 2,31]
17	3,25	1,32	0,13	[3,01; 3,5]
18	2,41	1,31	0,12	[2,17; 2,65]
19	2,16	1,22	0,08	[2; 2,31]
20	2,85	1,40	0,09	[2,66; 3,03]
21	2,39	1,26	0,09	[2,22; 2,55]
22	2,72	1,29	0,09	[2,55; 2,89]
23	1,30	0,68	0,06	[1,17; 1,42]

Akademischer Werdegang

Oktober 2015 – Mai 2019	Promotion (Nordische Philologie) Christian-Albrechts-Universität zu Kiel
Oktober 2012 – September 2015	Studium (Germanistik, Biologie, Dänisch) Master of Education (Lehramt Gymnasium) Christian-Albrechts-Universität zu Kiel, Aarhus Universitet
Oktober 2011 – Februar 2014	Zertifikatsstudium Deutsch als Fremdsprache
Oktober 2008 – April 2012	Studium (Germanistik, Biologie, Dänisch) Bachelor of Arts (Lehramt Gymnasium) Christian-Albrechts-Universität zu Kiel
Juni 2007	Abitur Auguste-Viktoria-Gymnasium Flensburg