

# Alternation preferences and focus marking

Von der Philosophisch-Historischen Fakultät der Universität Stuttgart  
zur Erlangung der Würde einer Doktorin der Philosophie (Dr. phil.)  
genehmigte Abhandlung

Vorgelegt von  
Nadja Schauffler  
aus Tübingen

Hauptberichterin Prof. Dr. Sabine Zerbian  
Mitberichterin Prof. Dr. Petra Wagner

Tag der mündlichen Prüfung: 13. April 2021

Institut für Linguistik/Anglistik  
der Universität Stuttgart

2021

---

**Erklärung (Statement of Authorship)**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst habe und dabei keine andere als die angegebene Literatur verwendet habe. Alle Zitate und sinngemäßen Entlehnungen sind als solche unter genauer Angabe der Quelle gekennzeichnet.

I hereby declare that this text is the result of my own work and that I have not used sources without declaration in the text. Any thoughts from others or literal quotations are clearly marked.

(Nadja Schauffler)

# Danke

Viele Menschen haben zum Gelingen dieser Arbeit beigetragen.

Herzlichen Dank an meine “Doktormutter” Sabine Zerbian, für die Betreuung, die vielen guten Ratschläge und die wundervolle Fähigkeit zum richtigen Zeitpunkt positives Feedback zu geben! Vielen Dank auch Petra Wagner, für das Interesse an meiner Arbeit, Gespräche über *Clashes* und das Treffen mitten auf der Stuttgart-21-Baustelle! Ein Dank geht auch an Grzegorz Dogil, für das Vertrauen in mein Potenzial als Wissenschaftlerin.

Ein großes Dankeschön geht an Petra Augurzky für die Einführung in die Psycholinguistik, die lustigen Konferenzbesuche, die wertvollen Gespräche und Ideen zu meiner Arbeit, das Mutzusprechen und die vielen schönen Spaziergänge!

Danke danke, Kati Schweitzer, für den Hiwi-Job mit dem alles begann und die wunderbare Freundschaft, die daraus gewachsen ist. Außerdem für all die Hilfe mit großen und kleinen Problemen, insbesondere für die Geduld und das Coaching in Linux- und R-Angelegenheiten (und für die Kleidertauschbörse).

Herzlichen Dank, Antje Schweitzer, für wertvolle Ratschläge und die schnelle Hilfe bei allem was anstand, insbesondere die Anpassung des PaIntE-Models für meine Analyse und das Korrekturlesen!

Ganz herzlichen Dank an Katharina Zahner, Ann-Kathrin Grohe und Heather Kember für die Hilfe bei der Datenerhebung in Sydney!

A heartfelt thanks goes to Diego Frassinelli, for all the pep talks, the best workshop-organising collaboration and for proof-reading!

Vielen Dank an Jörg Mayer, für die geduldige Hilfe mit Praats Demo-Window (und für die Praatpfanne) und an Fabian Schubö, für gute Ratschläge und all die phonologischen Diskussionen. Außerdem vielen Dank an Arndt Riestler für die informativen Gespräche über Topik und Fokus!

Vielen Dank an Judith Tonhauser und an die Expsy-Gruppe für das offene Ohr, für wertvolle Tipps und Kommentare und die Erinnerung daran, dass es auch ein akademisches Leben nach der Doktorarbeit gibt. Die Arbeit hat von den Diskussionen insbesondere zur Statistik in großem Maße profitiert! In Sachen Statistik danke ich außerdem Bodo Winter und Martijn Wieling für ihre Hilfsbereitschaft.

---

Ein herzliches Dankeschön geht an mein E-Team für's 'Im-gleichen-Boot-sitzen', das Verständnis und die Motivation und insbesondere Wiltrud Kessler für die Hilfe wenn es mit L<sup>A</sup>T<sub>E</sub>X mal geklemmt hat.

A warm thank you to Mike Walsh, Giusy Turco, Meg Zellers and Rebecca Carroll for everything I have learnt from you and for being excellent office mates!

Herzlichen Dank, Ralf Bothner, für die freundliche Atmosphäre am Institut und für's Schokotariat! Und ein aufrichtiges Danke an das gesamte IfLA-Team, dass ich gerne ins Büro gehe.

Vielen Dank, Svea Schauffler, dafür ein Vorbild zu sein und natürlich für das gründliche Korrekturlesen. Danke auch für die Erinnerung an die Existenz von Kommas und Relativsätzen...

Zu guter Letzt geht mein Dank an meine Familie, der diese Arbeit gewidmet ist: Vielen Dank Svea Schauffler und Katharina Gartner für die "Sistahood", die mich durch's Leben trägt, komme was wolle. Ein besonderer Dank geht an meine Eltern Gudrun und Wolfgang Schauffler, für die bedingungslose Unterstützung von Kindesbeinen an, für die zahlreichen Oma-und-Opa-Einsätze, besonders im Endspurt und für den tollen Doktorhut! Ebenso danke ich Achim und Bettina Freund für ihren Einsatz, die mir die "Diss-Bootcamp"-Wochen ermöglichten. Ein warmes Dankeschön nicht zuletzt an Jürgen Freund für die Unterstützung und Liebe und an Jorin, für den wundervollsten Grund, dass diese Arbeit etwas länger gedauert hat als ursprünglich geplant.

Teile dieser Arbeit wurden im Rahmen des Projekts "Prosodische Phrasierung in der auditorischen und visuellen Satzverarbeitung" und des Projekts "Cross-linguistic interactions in second language prosody" des SFB 732 von der DFG gefördert. Vielen Dank auch dafür.

---

*“Rhythm is a dancer  
it’s a soul’s companion  
you can feel it everywhere”  
Snap!*



# Contents

<b>List of abbreviations</b>	<b>xiii</b>
<b>Abstract</b>	<b>xv</b>
<b>Deutsche Zusammenfassung</b>	<b>xvii</b>
<b>1. Introduction</b>	<b>1</b>
<b>2. Background</b>	<b>7</b>
2.1. Pitch accents and information structure . . . . .	7
2.1.1. The autosegmental-metrical model of intonation . . . . .	7
2.1.1.1. GToBI(S) . . . . .	11
2.1.1.2. Other models of intonation . . . . .	12
2.1.2. Information-structural concepts used in this thesis . . . . .	13
2.1.3. The prosodic marking of focus . . . . .	16
2.1.4. Variation in the prosodic marking of information structure . . . . .	20
2.2. Alternation preferences influence sentence production and processing . . . . .	24
2.2.1. The Principle of Rhythmic Alternation . . . . .	25
2.2.2. The function of rhythmic alternation . . . . .	28
2.2.3. Experimental investigations on rhythmic alternation . . . . .	29
2.2.3.1. Studies investigating alternation effects on pitch accenting	30
2.2.3.2. Opposites attract - Alternation of tonal events . . . . .	33
2.3. Second language prosody . . . . .	36
2.3.1. Models of categorical perception and production of L2 speech sounds	36
2.3.2. Models of language learning and their predictions concerning prosody acquisition . . . . .	38
2.3.3. Beyond L1 transfer – L2 universals . . . . .	40
2.3.4. Experimental investigations of L2 prosody . . . . .	42
2.4. Summary of research questions . . . . .	46

---

<b>3. Alternation preferences and focus marking in German</b>	<b>49</b>
3.1. The intonation of double-focus in German	50
3.2. Research questions and hypotheses	53
3.3. The experiment	55
3.3.1. Participants	55
3.3.2. Stimulus design	55
3.3.3. Procedure	58
3.3.4. General remarks on the statistical analyses of this chapter	59
3.4. Analysis of pitch accent placement	60
3.4.1. Procedure	60
3.4.2. Statistical analysis	62
3.4.3. Results	63
3.5. Analysis of pitch accent type	65
3.5.1. Procedure	66
3.5.2. Contours used by the speakers	67
3.5.3. Statistical analysis and results	67
3.5.4. Parametric representation of the contours	70
3.6. Analysis of pitch accent relation	74
3.6.1. Procedure	75
3.6.1.1. Inter-annotator and intra-annotator agreement	75
3.6.2. Statistical analysis	76
3.6.3. Results	77
3.7. Analysis of phrasing	78
3.7.1. Boundary annotation	79
3.7.1.1. Statistical analysis and results	81
3.7.2. Word duration	82
3.7.2.1. Procedure	82
3.7.2.2. Statistical analysis and results	83
3.8. Summary of the results and discussion	83
3.8.1. Pitch accent placement	84
3.8.2. Pitch accent type	86
3.8.3. Pitch accent relation and phrasing	89
3.8.4. Phonological structure of the double-focus realisations	92
3.9. Conclusion of this chapter	93



<b>4. Alternation preferences and focus marking in English</b>	<b>95</b>
4.1. The intonation of double-focus in English . . . . .	96
4.2. Experiment 1: Alternation preferences and corrective focus in L1 English	98
4.2.1. Stimulus design . . . . .	98
4.2.2. Research questions and hypotheses . . . . .	100
4.2.3. Participants . . . . .	101
4.2.3.1. The use of Australian English speakers . . . . .	102
4.2.4. Procedure . . . . .	102
4.2.5. General remarks on the statistical analyses of this chapter . . . . .	103
4.2.6. Analysis of pitch accent placement . . . . .	104
4.2.6.1. Intra-annotator and inter-annotator agreement . . . . .	105
4.2.6.2. Statistical analysis . . . . .	106
4.2.6.3. Results . . . . .	106
4.2.7. Analysis of pitch accent type . . . . .	108
4.2.7.1. Procedure . . . . .	109
4.2.7.2. The problem of comparing pitch accent labels of different labelling conventions . . . . .	109
4.2.7.3. Statistical analysis and results . . . . .	112
4.2.8. Analysis of pitch accent relation . . . . .	113
4.2.8.1. Procedure . . . . .	113
4.2.8.2. Inter-annotator and intra-annotator agreement . . . . .	113
4.2.8.3. Statistical analysis . . . . .	114
4.2.8.4. Results . . . . .	115
4.2.9. Analysis of phrasing . . . . .	115
4.2.9.1. Boundary annotation . . . . .	116
4.2.9.2. Word duration . . . . .	117
4.2.10. Summary and discussion of results . . . . .	118
4.2.10.1. Pitch accent placement . . . . .	120
4.2.10.2. Pitch accent type . . . . .	122
4.2.10.3. Pitch accent relation . . . . .	122
4.2.10.4. Phrasing . . . . .	123
4.2.10.5. Phonological structure of the English double-focus real- isations . . . . .	124
4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts . . . . .	125
4.3.1. Method . . . . .	129

4.3.2.	Stimulus design . . . . .	129
4.3.3.	Participants . . . . .	130
4.3.4.	Procedure . . . . .	131
4.3.5.	Stress-shift annotation . . . . .	131
4.3.5.1.	Inter- and intra-annotator agreement . . . . .	131
4.3.6.	Pitch-accent annotation . . . . .	133
4.3.6.1.	Inter- and intra-annotator agreement . . . . .	133
4.3.7.	Statistical analyses . . . . .	133
4.3.8.	Results . . . . .	134
4.3.9.	Summary and discussion of results . . . . .	135
4.4.	Conclusion of this chapter . . . . .	138
<b>5.</b>	<b>Alternation preferences and focus marking in L2 English</b>	<b>141</b>
5.1.	Comparison of German and English results in light of LILt, and predic- tions for L2 speakers . . . . .	142
5.2.	General remarks on the statistical analyses of this chapter . . . . .	146
5.3.	Experiment 1: Alternation preferences and corrective focus in L2 English	147
5.3.1.	Stimulus design . . . . .	147
5.3.2.	Participants . . . . .	147
5.3.3.	Procedure, annotations and analyses . . . . .	148
5.3.4.	Analysis of pitch accent placement . . . . .	148
5.3.4.1.	Intra-annotator and inter-annotator agreement . . . . .	148
5.3.4.2.	Statistical analysis . . . . .	149
5.3.4.3.	Results . . . . .	149
5.3.5.	Analysis of pitch accent type . . . . .	150
5.3.5.1.	Procedure . . . . .	150
5.3.5.2.	Statistical analysis and results . . . . .	151
5.3.6.	Analysis of pitch accent relation . . . . .	152
5.3.6.1.	Procedure . . . . .	153
5.3.6.2.	Inter-annotator and intra-annotator agreement . . . . .	153
5.3.6.3.	Statistical analysis . . . . .	154
5.3.6.4.	Results . . . . .	154
5.3.7.	Analysis of Phrasing . . . . .	154
5.3.7.1.	Boundary annotation . . . . .	155
5.3.7.2.	Word duration . . . . .	156

5.3.8.	Summary and discussion . . . . .	157
5.3.8.1.	Pitch accent placement . . . . .	157
5.3.8.2.	Pitch accent type . . . . .	159
5.3.8.3.	Pitch accent relation . . . . .	160
5.3.8.4.	Phrasing . . . . .	161
5.3.8.5.	Phonological structure of the L2 English double-focus realisations . . . . .	162
5.3.8.6.	Comparing the results of all language groups – overview	163
5.4.	Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts . . . . .	163
5.4.1.	Method . . . . .	165
5.4.2.	Stimulus design . . . . .	165
5.4.3.	Participants . . . . .	165
5.4.4.	Procedure and analyses . . . . .	166
5.4.5.	Statistical analyses . . . . .	166
5.4.6.	Results of stress shift annotation . . . . .	167
5.4.7.	Results of pitch accent annotation . . . . .	168
5.4.8.	Summary and discussion of results . . . . .	169
5.5.	Conclusion of this chapter . . . . .	172
<b>6.</b>	<b>General conclusions, contributions and outlook</b>	<b>175</b>
6.1.	Discussion of the main findings and general outlook . . . . .	175
6.2.	Summary . . . . .	179
6.3.	Open questions for future research . . . . .	180
	<b>Bibliography</b>	<b>185</b>
<b>A.</b>	<b>Materials for the L1 German experiment (chapter 3)</b>	<b>209</b>
A.1.	Context story . . . . .	209
A.2.	Experimental stimuli . . . . .	209
A.3.	Filler items . . . . .	215
<b>B.</b>	<b>Materials for the English experiments (chapters 4 and 5)</b>	<b>219</b>
B.1.	Context story . . . . .	219
B.2.	Experimental stimuli for the experiments in 4.2 and 5.3 . . . . .	219
B.3.	Experimental stimuli for the experiments in 4.3 and 5.4 . . . . .	225
B.4.	Filler items . . . . .	231



# List of abbreviations

**F<sub>0</sub>** Fundamental frequency

**GToBI** German Tones and Break Indices

**GToBI(S)** German Tones and Break Indices (Stuttgart version)

**ip** intermediate phrase boundary

**IP** intonation phrase boundary

**L1** First language

**L2** Second language

**LILt** L2 intonation learning theory

**NLM** Native Language Magnet Theory

**NP** noun phrase

**PA** Pitch accent

**PaIntE** Parametric Representation of Intonation Events

**PAM** Perceptual Assimilation Model

**SLM** Speech Learning Model

**ToBI** Tones and Break Indices



# Abstract

This thesis presents a crosslinguistic investigation on the interplay between alternation preferences and the prosodic marking of focus in German, English and German learners of English. It does so by means of five production experiments investigating the realisation of double-focus sentences with two directly adjacent foci eliciting pitch accent clashes.

The general aim of this thesis is to investigate whether alternation preferences are found at the sentence accent level, and whether they interfere with the prosodic marking of focus. Contrary to what has been claimed before, results obtained from my experiments suggest that rhythmic adjustment strategies do take place under focus marking. The thesis shows, however, that despite their similarity, the two languages rely on different strategies when alternation and focus marking are working in opposite directions. While English speakers often omit the first focus accent in clash contexts, German speakers often create a melodic alternation of high and low by realising the first of two adjacent focus accents with a rising pitch accent (L\*H).

Evidence obtained from a production experiment with L2 English speakers suggests that these differences matter in language acquisition. German strategies in clash contexts are transferred but used to a lesser extent than found with the L1 German group. Unlike in the L1 English group, however, focus accents are rarely omitted. L2 speakers seem less guided by alternation preferences than native speakers of both languages. This finding is corroborated by a second experiment investigating pitch accent clashes in rhythm rule contexts under different focus environments.

In view of the results obtained from this thesis, I conclude that: (i) the preference for alternation can influence the prosodic marking of focus and is a source for variation in the realisation of information-structure categories. The rhythmic context should therefore be taken into account when assigning semantic to phonological categories. (ii) even though German and English share the preference for alternation, it affects prosodic focus marking differently in the two languages. (iii) strategies emanating from alternation preferences are transferred and can cause misproductions in a second language. L2 speech itself is, however, less driven by rhythmic factors.





# Deutsche Zusammenfassung

Diese Dissertation basiert auf einer sprachübergreifenden Untersuchung der Wechselwirkung zwischen rhythmischer Alternation und der prosodischen Markierung von Fokus im Deutschen, im Englischen, sowie bei Sprechern mit Deutsch als Erst- und Englisch als Zweitsprache. Die Studie besteht aus fünf Produktionsexperimenten, mithilfe derer die Realisierung von Doppelfokussätzen untersucht wird, in denen zwei adjazente Foki einen so genannten Pitchakzentzusammenstoß (*pitch accent clash*) provozieren. Eine direkte Abfolge zweier Pitchakzente ist zwar erlaubt, aber markiert, während eine Alternation starker und schwacher Silben für deutsche und englische Sprecher wohlgeformt und damit präferiert ist.

Ziel der Dissertation ist es festzustellen, ob sich die Vorliebe für rhythmische Alternation auf die Satzakkzentebene auswirkt, und ob sie die prosodische Markierung von Fokus beeinflusst.

Entgegen vorherrschender Annahmen zeigen die Ergebnisse meiner Studie, dass Strategien zur rhythmischen Anpassung durchaus auch unter Fokusmarkierung zu beobachten sind. Die Dissertation zeigt zudem, dass, aller Ähnlichkeit zum Trotz, die Sprecher beider Sprachen auf jeweils unterschiedliche Strategien zurückgreifen wenn Alternation und Fokusmarkierung gegeneinander arbeiten. Während englische Muttersprachler oft den ersten Fokusakzent in einem solchen Kontext auslassen, präferieren deutsche Sprecher hingegen eine melodische Alternation von prosodischen Hoch- und Tieftönen, indem sie den ersten der beiden aufeinanderfolgenden Fokusakzente mit einem steigenden Pitchakzent realisieren (L\*H).

Die Ergebnisse eines Produktionsexperiments mit Probanden, die Englisch als Zweitsprache sprechen, legen nahe, dass diese Unterschiede beim Zweitspracherwerb eine Rolle spielen. Deutsche Strategien werden zum Teil in die Zweitsprache übertragen, aber in geringerem Umfang genutzt als es für deutschen L1-Sprecher gefunden wurde. Anders als für die englischen Muttersprachlern gefunden wurde, wurden Fokusakzente von den Englischlernenden dagegen nur selten übergangen. L2-Sprecher scheinen weniger von rhythmischen Alternationsbestrebungen beeinflusst zu sein als Muttersprachler beider Sprachen. Dieses Ergebnis lässt sich durch ein zweites Experiment untermauern, in dem

---

Akzentzusammenstöße in “Rhythm Rule”-Kontexten unter verschiedenen Fokusbedingungen erforscht wurden.

In Anbetracht der hier vorliegenden Ergebnisse komme ich zu dem Schluss, dass: (i) die Präferenz für rhythmische Alternation die prosodische Fokusmarkierung beeinflussen kann und damit Auslöser für Variabilität in der Realisierung von Informationsstrukturkategorien ist. Der rhythmische Kontext sollte daher bei der Zuordnung von phonologischen zu semantischen Kategorien berücksichtigt werden. (ii) obwohl Deutsch und Englisch die Präferenz für rhythmische Alternation teilen, Fokusmarkierung in beiden Sprachen unterschiedlich beeinflusst wird. (iii) Realisierungen, die auf der Einhaltung von rhythmischer Alternation gründen, übertragen werden und zu Fehlern in der Zweitsprache führen können. Die Zweitsprache selbst scheint jedoch weniger von rhythmischen Faktoren bestimmt.

# 1. Introduction

At the International Congress of Phonetic Science 2015 in Glasgow, Frank H. Guenther gave a keynote talk on a neural network model able to produce learned phonemes and syllables, and its role in the development of a brain-computer interface aimed at restoring speech output for patients with locked-in syndrome (Guenther, 2015). At the end of this very impressive talk he addressed the linguistic audience: “With respect to prosody – it’s difficult, it’s hard, please help!” This rather desperate appeal reflects the issues still under investigation in the suprasegmental realm referred to as prosody.

We can see that it has to be quite a complex matter, for example, by looking at speech synthesis systems which are impressively good on the segmental level but still give themselves away by suddenly using just a slightly unnatural intonation. We can also see that in second language learners, who may be perfect in other areas of grammar but the attainment of these last prosodic nuances seems incredibly hard to master. What contributes to the complexity certainly is the multi-faceted character and the multiple functions of prosody. To cite Dogil (2003, p. 546) (bracketing in the original):

“The systematic variation of prosody is used to direct the listener’s attention to the more and less significant elements of the message [linguistic focus], to signal the type of speech act (such as question, declaration, imperative)[linguistic modus], and to convey an emotional state of a speaker [para-linguistic affect], just to mention the most familiar prosodic functions.”

Phonetically, these functions are coded in duration, intensity and fundamental frequency, and these phonetic cues will be perceived as length, loudness and pitch. Compared to segmental categories (phonemes), it is, however, not (yet) as obvious what the meaningful categories are, where we should look for categorical boundaries in a continuous  $F_0$  and where for gradience. And how does this differ between languages? As a consequence, it is not at all straightforward to map one prosodic form to one particular function. And in fact, while probabilities have been identified we often find various prosodic realisations of one function and one prosodic realisation for various functions.

---

The aim of this thesis is to investigate whether there are prosody-inherent factors that contribute to variation in the prosodic implementation of the function [linguistic focus].<sup>1</sup> Specifically, it is explored whether a preference for rhythmic alternation affects the prosodic marking of focus in German, English, and German learners of English.

With German and English, two well-investigated stress-timed languages are chosen that share a number of similarities. These similarities are the starting points for my investigation:

**German and English prosodically mark focus by means of pitch accent placement and pitch accent type, but there is variation.**

The first starting point is the prosodic marking of information-structure categories in German and English. Both languages use pitch accenting to indicate focus structure, and they have been shown to do so by similar means. A corrective focus, such as investigated in the study at hand, is typically associated with a falling pitch accent with a high tonal target on the accented syllable (H\*L) (e.g. Büring, 1997; Féry, 2017; Kügler and Gollrad, 2015; Selkirk, 1995; Gussenhoven, 2004).

However, the actual choice of pitch accent type, as well as the distribution of pitch accents in a phrase, can vary beyond what can be explained by semantic or pragmatic factors, and the prosodic realisation of information structure categories cannot necessarily be predicted: Sometimes there is variation in the actual choice of pitch accent types, that is, we find a different category than expected (cf. Féry, 2007; Baumann, 2006; Féry and Kügler, 2008; Schweitzer et al., 2009), and sometimes there is variation in the placement of pitch accents, in other words, a pitch accent may not be realised even though there should be one from an information-structural perspective (e.g. Riester and Piontek, 2015; Kentner, 2012a).

**German and English have a preference for rhythmic alternation.**

The second starting point of the study is the finding that speakers of both German and English prefer an alternation of strong and weak beats. This means they prefer an alternation of stressed and unstressed syllables and try to avoid stress clashes, meaning two directly adjacent stresses, wherever possible. This preference has been shown to affect sentence production, perception and processing (e.g. Breen and Clifton, 2011; Magne et al., 2010; Tilsen, 2012; Kentner, 2012a; Bohn et al., 2011; Rothermich et al., 2012). It manifests

---

<sup>1</sup>Prosody-inherent factors have been identified before to affect sentence production and processing, such as, for example, prosodic balance (Fodor, 1998; Augurzky, 2008; Schweitzer et al., 2012; Schauffler et al., 2012).

itself, for example, in rhythmic repair strategies in what has been called the “rhythm rule” (Liberman and Prince, 1977; Selkirk, 1984). The rhythm rule describes a process in which a primary stress is “shifted” towards an earlier secondary-stressed syllable in order to avoid a stress clash. The phrase *thirtéen wómen* is, thus, typically realised as *thírteen wómen*. It also appears in the preference of constructions such as *ín the nót so dístant fúture* over *ín the nót dístant fúture* (cf. Schlüter, 2005). In a corpus study preceding this thesis, I also found that German speakers prefer an alternation of falling and rising pitch accents, that is  $F_0$  peaks and valleys, when the pitch accents are closer together. It seems as if speakers prefer both an alternation of strong and weak, and an alternation of high and low.

Against this backdrop, this thesis investigates the role of alternation preferences in the variation found in the prosodic marking of focus.

Previous studies mostly address alternation effects with respect to lexical stress and stress shift. Lexical stress and sentence accent are often conflated (e.g. in Grabe and Warren, 1995; Horne, 1990, cf. also Tilsen (2012), Shattuck-Hufnagel (1995) for a critique of these studies). Therefore, it remains unclear how these preferences interact with other prosodic functions such as the prosodic marking of information-structure categories. The experiments in this thesis are designed to investigate whether alternation preferences are found at the sentence accent level and whether they interfere with the prosodic marking of focus.

For this purpose, I investigate sentences with two directly adjacent foci such as in (1). The realisation of focus accents (marked by small caps in the example) on each of the two foci (marked by square brackets) creates a dispreferred pitch accent clash.

(1) Did Carl say that Clara gave the boys horror stories to read?

No, he said that she gave [the GIRLS]<sub>Focus1</sub> [ROMance novels]<sub>Focus2</sub>

Looking at German and English native speakers, I pursue the question of whether the two languages differ in their approach to conflicts between focus marking and rhythmic alternation. Possible differences may have consequences for language learners.

It is unclear what role alternation preferences have in a second language, considering that repair strategies in clash contexts are optional and meaning is not affected when clashes are not repaired but may be compromised when speakers pertain to the ideal of alternation (e.g. by omitting focus accents or using unlicensed contours). Models of language learning such as the Speech Learning Model (Flege, 1995) and the L2 intonation learning theory (LILt, Mennen, 2015) predict that similar categories will be more difficult to acquire. Given the similarity between the two languages, this thesis investi-

---

gates in which dimension alternation preferences operate in an L2, and whether different strategies are transferred or attained.

**The architecture of this thesis** The following chapter sets out to motivate and develop the research questions. I will start by establishing the relationship between prosody and information structure in section 2.1. This section will first provide an overview of the autosegmental-metrical framework in which the analysis is set (2.1.1), and the information-structural concepts addressed in this thesis (2.1.2). We will look at how these categories are typically prosodically realised in German and English in section 2.1.3, and the variation that has been found in 2.1.4. Some of this variation may be attributed to alternation preferences. What exactly is meant by that and what has been found by other researchers regarding this preference will be addressed in 2.2. Section 2.3 will give an overview of L2 prosody and what has been found with respect to focus marking and alternation preferences for language learners. It will introduce the L2 Intonation Learning theory (LILt, Mennen 2015) in 2.3.2 as the theoretical framework within which the results will be interpreted. Section 2.4 summarises the research questions derived from this chapter.

Chapters 3 to 5 comprise the experimental part of this thesis. Chapters 3 and 4 address the first major research question on whether and how alternation preferences affect the prosodic realisation of focus in German and English. In order to compare the two language groups in a controlled manner, double-focus sentences were constructed in both languages. Chapter 3 will present the results for the German group and chapter 4 for the English speakers. A second experiment with English speakers was designed in order to investigate stress shift phenomena under focus marking whose results will be presented in 4.3.

Chapter 5 will address the second major research question, namely what is the role of alternation preferences in an L2. For one thing, are these preferences and the strategies they may generate subject to L1 transfer (albeit optional) interfering with prosodic functions, and also, are optional repair strategies acquired in a second language? To answer the former, a comparative analysis was conducted which repeats the exact same experiment conducted with English speakers, but now with German learners of English. To answer the latter, the second English experiment was repeated with the same advanced learners in order to investigate the interplay of rhythmic preferences and focus marking in so-called rhythm rule contexts. Design and results of these L2 experiments will be presented in 5.3 and 5.4 respectively.

Chapter 6 will summarise and discuss the main findings of this thesis and give a short

overview of ongoing and future work (6.3).

The experimental chapters (3 to 5) are designed in a way that they can be understood by their own. The remainder of this introductory chapter will therefore concentrate on the more basic aspects pertinent to this study. Relevant details of previous research will be revisited at the respective points of interest in the following chapters.





## 2. Background

In this chapter I will develop and motivate the research questions by addressing important theoretical concepts and previous research. It starts off with the relationship between pitch accents and information structure as the first starting point for the investigation, and moves then to previous research on the preference for rhythmic alternation. The third section comments on relevant findings from L2 research. The chapter concludes with the research questions derived from the observations discussed in the chapter.

### 2.1. Pitch accents and information structure

As outlined above, the first starting point of this investigation is that both German and English use pitch accents to mark information structure. Therefore, the aim of the following sections is to clarify firstly what pitch accents are and how they can be modelled, and secondly what information structure refers to. With respect to the former the thesis follows the assumptions made by the autosegmental-metrical framework which will be briefly introduced in section 2.1.1. The latter side of that relationship will be explored in section 2.1.2 attempting to clarify which information-structural concepts are investigated in this thesis.

#### 2.1.1. The autosegmental-metrical model of intonation

When talking about the prosodic realisation of information-structure categories, it is necessary to establish what the assumed prosodic categories are, or rather how these categories are represented. Different suggestions have been made as to how to model intonation, attempting to mediate between detailed acoustics (phonetically) and the linguistic function with which they are associated (phonologically). This means that a phonological model of intonation must be able to describe the intonation of an utterance by means of categorical entities, and these entities have to be depictable in a physical description of the utterance in form of varying parameters (c.f. e.g. Ladd, 2008). There are two major ways in which these categories are represented. This thesis follows the

assumptions made by the autosegmental-metrical approach to intonation as based on the reflections by Pierrehumbert (1980) in which intonation contours are analysed as a sequence of high and low *tonal targets*. This is different from models assuming pitch *movements* rather than targets, as, most prominently, in the British school (e.g. Palmer, 1922; Kingdon, 1958; O'Connor and Arnold, 1973; Crystal, 1969). The British school, also referred to as the nuclear tone tradition, is often used in the context of second language learning given its intuitive description and depiction of intonation. The main assumptions are that each intonation phrase has one most prominent syllable - the nucleus - which is associated with a distinct pitch movement, such as for example a (*high or low*) *rise*, a *rise-fall* or a (*high or low*) *fall*.<sup>1</sup> These tunes are associated with particular meanings.

Contrary to these pitch configurations assumed by the British school, the autosegmental-metrical framework describes intonation as a sequence of categorically different high (H) or low (L) tones which may be combined to, for example, a rise - a low tonal target followed by a high tonal target, or a fall - a high tonal target followed by a low tonal target. According to Dogil (2003), the development of the autosegmental-metrical framework was a breakthrough for phonological models of phonology:

"Autosegmental phonology provided a new descriptive language for prosody in which its phonetic cues and linguistic frames could be treated separately from the rest of the phonology (building so-called auto[nomous]-segments), and at the same time, it provided a set of well-formedness conditions (tune-to-text association conventions), which established clearly defined mappings and relations between the prosodic and the other linguistic categories." (Dogil, 2003, p. 545)

This model has its roots in Janet Pierrehumbert's dissertation (Pierrehumbert, 1980) where she introduces an abstract representation of English intonation which has become highly influential for the description of intonations of various other languages. Due to the assumption that an utterance's intonation can be described as a sequence of tones, these models are also referred to as *tone sequence models*. Ladd (1996) then introduced the term *autosegmental-metrical model*, referring to, on the one hand, the metrical aspect by dividing the utterance in phrases and assigning relative prominence to elements within the utterance, following Liberman and Prince (1977), and, on the other hand, the autosegmental aspect, referring to the autonomy of tones and tunes from the text,

---

<sup>1</sup>See for example Féry (2017) for an overview or Ladd (2008) for a comparison between the British school and the autosegmental-metrical framework.

which is mirrored by their respective representation on different independent tiers. A third component then specifies how the text and the tune are aligned at specific anchor points (often referred to as *text-to-tune*). Pierrehumbert's inventory of tones for English comprises three different categories of tones: Pitch accents, phrase accents and boundary tones.

**Pitch accents** are tones that mark the syllable with which they are associated as prominent. They are either monotonal or bitonal. The two monotonal pitch accents are  $H^*$  and  $L^*$  where H and L indicate the tonal target as high (H) or low (L) respectively in the speakers' register, and the \* indicates the association with the stressed syllable (which is the anchor point with the text, in a manner of speaking). The bitonal pitch accents describe a tonal event by two tonal levels, for example  $H^*+L$ , where the starred tone (here  $H^*$ ) represents the tonal target on the accented syllable and the unstarred tone represents the pitch height on the adjacent syllable. An unstarred tone can either follow a starred tone, as in the example just mentioned (then it is referred to as trailing tone), or precede the accented syllable, such as for example in the bitonal pitch accent  $H+L^*$  (then it is referred to as leading tone). As an example, the difference between these two accents  $H^*+L$  and  $H+L^*$  is therefore a matter of alignment: both represent falling contours but while  $H^*+L$  is high on the accented syllable and the fall follows it, the fall is earlier in  $H+L^*$  where the accented syllable is low.<sup>2</sup>

**Boundary tones** are aligned with the edge of the prosodic domain, that is at either the beginning or at the end of the intonation phrase.  $\%H$  and  $\%L$  are the labels for phrase-initial boundary tones, where the % indicates the boundary and the H or L indicates the pitch level at the boundary, and  $H\%$  and  $L\%$  are the phrase-final boundary tones.

**Phrase accents** describe the  $F_0$  movement between the pitch accent and the boundary tone at the right edge of the phrase, they can be either H- or L-. This was changed in Beckman and Pierrehumbert (1986) where phrase accents were reinterpreted as boundary tones at *intermediate* phrases, which are demarcated by weaker boundaries than intonation phrases.

These three categories of tones are hierarchically sorted in a grammar of well-formed tunes, as illustrated in figure (2.1): Each intonation phrase consists of one or more intermediate phrases, whose boundaries are indicated by a phrase accent. And each intermediate phrase consists of one or more pitch accents (here symbolised by black

---

<sup>2</sup>The labels for all bitonal pitch accents in Pierrehumbert's inventory are  $H^*+L$ ,  $H+L^*$ ,  $H^*+H$  (later eliminated from the inventory by Beckman and Pierrehumbert (1986),  $L^*+H$  and  $L+H^*$ . Note that the + can be absent in some descriptions.

## 2.1. Pitch accents and information structure

---

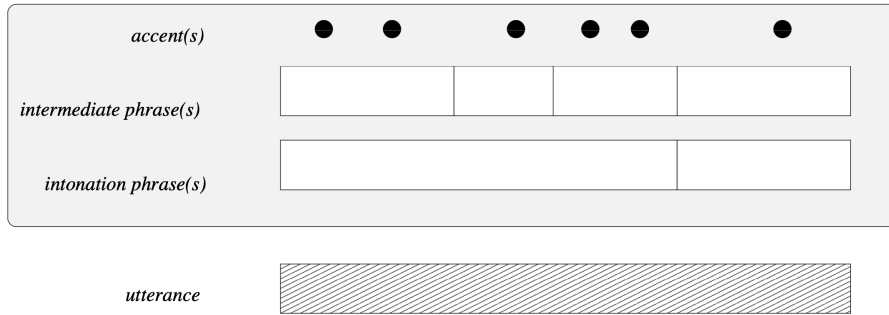


Figure 2.1.: The hierarchy of tones in the tone-sequence-model and the ToBI conventions as depicted by Katrin Schweitzer (teaching material).

dots). The contour between the respective pitch accents and boundary tones can be constructed by interpolation.

It is important to emphasise that *high* and *low* are abstract categories and can only be interpreted in relation to each other and on a local level (cf. Pierrehumbert, 1980, p.2). It is therefore possible that, for example, a low tonal target at the beginning of the phrase is acoustically higher than a high tonal target at the end of the phrase. This has also to do with downtrends, a term that describes the decrease of pitch height towards the end of the phrase, either via declination (a gradual fall of  $F_0$  over the course of the utterance), downstep (the controlled lowering of the second pitch accent in a sequence of two high-starred pitch accents), and final lowering (the finding that in a sequence of downstepped high tones the last one is lower than expected (see e.g. Féry, 2017, for an overview of views on downstep phenomena).

As already mentioned in passing, Pierrehumbert’s original approach was modified in the following years (Lieberman and Pierrehumbert, 1984; Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988) and was incorporated in the formalisation of the *Tones and Break Indices (ToBI)*, a prosodic labelling system for the transcription and analysis of Mainstream American English, General Australian and southern Standard British English (Silverman et al., 1992; Beckman and Hirschberg, 1999; Beckman and Ayers, 1997). ToBI has become a widely accepted annotation standard for intonation that has been adopted and adapted for the analysis of several other languages (see, e.g. Jun, 2005, 2014, for an overview). (In order to specify the ToBI as developed for Mainstream American English from the concept of ToBI in general, the system is now referred to as MAE\_ToBI.)

Based on the work of Pierrehumbert (1980) and Beckman and Pierrehumbert (1986) the system comprises labels for pitch accents and boundary tones. Pitch accents are

associated with the lexically-stressed syllable of the accented word and boundary tones align with the left edge (optional) or right edge (obligatory) edge of the phrase.<sup>3</sup> There are two levels of phrasing, with intermediate phrases (ip), marked by phrase accents, and intonation phrases (IP), marked by boundary tones (cf. figure 2.1 above.) Additionally, it comprises break-strength symbols from 0 (very close interword juncture) to 4 (intonation phrase break) to indicate the type of phonological breaks. Contrary to the assumptions by the British school, differences in prominence between pitch accents within a phrase are not annotated, meaning all pitch accents are annotated with a starred tone (the nuclear accent is not marked differently in any way).

### 2.1.1.1. GToBI(S)

The ToBI standard has also been adopted for German in different developments: GToBI (German ToBI, Grice et al. (see 1996); Baumann et al. (see 2001) and the Stuttgart version of the German ToBI: GToBI(S), Mayer (1995)). The Stuttgart system modifies GToBI including Féry's work (1993) within the autosegmental-metrical approach. It has a smaller inventory of labels, with the intention to describe only those tonal events with categorical labels that are associated with different discourse functions. The two standard pitch accents are H\*L and L\*H. Contrary to MAE\_ToBI or GToBI, GToBI(S) for example only assumes trailing tones and does not label leading tones, that is the tonal target preceding the accented syllable. There are monotonal versions of these pitch accents (H\* and L\* respectively) when the trailing tone is realised later in the (intermediate) phrase, namely either just before or with the next pitch accent - a phenomenon referred to as *linking*. Mayer (1995) adds three labels not found in GToBI, namely two tritonal accents (L\*HL and HH\*L) and a bitonal pitch accent that described the use of a middle-range tonal level (M): H\*M.<sup>4</sup> With respect to boundaries, intermediate phrases are not associated with phrase accents, since it is assumed that the tonal contour at the end of phrases is modelled by the last pitch accent (the nuclear accent) in the case of intermediate phrases, and by the nuclear accent and the edge tone at the end of intonation phrases. The Stuttgart system is used to annotate the data in this thesis. The labels used in the study are therefore described in more detail in 3.5.1.

<sup>3</sup>In terms of the inventory, it has some changes in comparison to Pierrehumbert's model. The pitch accent H\*+L and the initial boundary tone %L were not adopted, and the diacritic ! was introduced to mark a downstepped pitch accent. Additionally, there are labels for uncertainty and delayed and early peaks.

<sup>4</sup>These three pitch accents only occur very infrequently, this is why L\*H and H\*L are referred to as the *standard* accents.

### 2.1.1.2. Other models of intonation

Phonological models such as the British school or the autosegmental-metrical framework (and the taxonomies based on it) are not the only approaches to intonation.<sup>5</sup> *Superposition models*, mainly based on the Fujisaki-model (Fujisaki, 1988), consider the intonation of an utterance not as a sequence of tonal events, but interpret the  $F_0$  contour hierarchically as being composed of different components - local components operating on accented syllables and global components relating to larger units such as phrases or sentences (Möbius, 1995). These components are superimposed (hence the name) and added to a base  $F_0$ -value. Given that both components are based on the physiology of the human larynx they can be considered phonetic models of intonation.

Another phonetic approach to intonation is the PaIntE-model (**P**arametric Representation of **I**ntonation **E**vents, Möhler and Conkie (1998); Möhler (2001)), which models the shape of the  $F_0$  contour in the vicinity of intonation events (originally for  $F_0$ -generation in speech synthesis). The model approximates stretches of the  $F_0$ -contour by means of six parameters. These parameters are linguistically meaningful in that they depict phonetic cues on and around the accented syllable (e.g. peak height, peak alignment, amplitude of the rise). This model was used to acoustically evaluate the manual labels for the German data in the study at hand. More details on the PaIntE-model, the parameters investigated and the procedure in general are, therefore, to be found in the respective chapter (3.5.4).

Another model of intonation that is not phonological is the Parallel ENcoding and Target Approximation model (PENTA) developed by Xu (2005; 2004). A major component of this model is that intonation results from communicative functions that each instruct the articulators in a defined way via encoding schemes. Unlike in AM models, this relationship between communicative function and phonetic realisation is modelled more directly via encoding schemes that add to the specification of the target (target approximation), and not via an abstract phonological level. Unlike in ToBI, the tone target is established for every syllable. (See Féry (2017) for an overview and critical examination of this assumption).

---

<sup>5</sup>Other accounts of German and English intonation within the autosegmental-metrical framework apart from the ones mentioned above were developed by e.g. Gussenhoven (1984) for British English, and Wunderlich (1988), Uhmman (1991) and Féry (1993) for German, and Grabe (1998) for a cross-linguistic comparison of both languages.

### 2.1.2. Information-structural concepts used in this thesis

This investigation zooms in on the interface of prosody and information structure. Having specified the framework in which the intonation analysis takes place as the autosegmental metrical approach, let us now turn to the information-structural concepts that are relevant to this endeavour. But what actually is information structure? The term was first used by Halliday (1967) emphasising that an utterance consists of “information units” that are independent of syntactic constituents.<sup>6</sup> According to Chafe (1976), information structure is how we “package” information with respect to the immediate communicative needs of the interlocutor (cf. also Krifka, 2007; Féry and Krifka, 2008). The ‘immediate communicative needs’ are met when the form of the message is optimised such that it is well understood by the addressee “in the current attentional state” (Féry and Krifka, 2008, p. 123). Information structure, thus, is about organising the constantly changing knowledge that is shared between the interlocutors (the Common Ground); what is background information and what is the answer to the currently relevant question. Different parts of the sentence usually have different informational functions or roles. One of the ‘techniques’ for information packaging facilitating communication is the highlighting of constituents which are in *focus*. Other important notions of information structure, which have different informational roles, are *Topic*, and *Givenness*. They will be briefly addressed at the end of this section.

**Focus** The notion of focus is a concept that has been subject to extensive theoretical considerations and observations (e.g. Halliday, 1967; Rooth, 1992; Krifka, 2007; Büring, 2003; Schwarzschild, 1999; Féry and Krifka, 2008) and it is the information-structural concept crucial to this investigation. Following the Alternative Semantics account of focus<sup>7</sup> (Rooth, 1985, 1992, 1995), Krifka defines focus as follows: “Focus indicates the presence of alternatives relevant for the interpretation of linguistic expressions” (2007:18).

Focus can be identified as the answer to the current wh-question - the answer to an implicit or explicit Question under Discussion (QUD) (e.g. Riester and Baumann, 2013; Riester et al., 2018).<sup>8</sup>

There are different kinds of sets of alternatives which focus can indicate. Krifka dis-

---

<sup>6</sup>Information structure as part of linguistics in fact goes back to the Prague school linguistic circle. Even though the concept has been modified over the decades their notion of focus and topic as well as the connection of these categories to intonation are still central to the field (cf. Dogil, 2003).

<sup>7</sup>The central idea of *Alternative Semantics* is to add a focus semantic value to the ordinary semantic value (cf. Rooth, 1992)

<sup>8</sup>Note that constituents that answer questions that are not in the discourse (implicitly or explicitly) may also be focused. Riester and Baumann (2013) refer to such instances as *not-at-issue foci*.

tinguishes between expression focus indicating alternatives of form, such as in example (2) taken from Krifka (2007:20), on the one hand,

- (2) A: They live in BERlin.  
 B: They live in [BerLIN]<sub>F</sub>.

and denotation focus indicating alternatives in denotation (such as, for example, in (3), on the other hand. He also differentiates between different uses of focus: between pragmatic uses of focus with focus as the answer to an underlying question, (such as in answers, corrections, confirmations, parallels and delimitation), on the one hand, and semantic uses of focus where focus changes the truth condition of the sentence and alternatives are used for semantic purposes<sup>9</sup>, on the other hand.

Note that while different terms are used to differentiate between different types of foci, for example wide focus and narrow focus, or contrastive focus and presentational focus<sup>10</sup>, all foci have in common that they are defined by implying a set of alternatives, only the scope of this set varies, with closed alternatives as for example in (3) where the set of alternatives is very small (*Mary, Peter*), and open alternatives in answers to questions like *what happened?*, where the alternatives are uncountable.

- (3) A: Mary stole the cookie.  
 B: (No,) [PEter]<sub>F</sub> stole the cookie!

The size of the focus domain can therefore also vary, from the whole sentence to single words or even syllables (and sometimes even phonemes, for example in *not helpED but help[S]<sub>F</sub>*).

The type of focus under investigation in the study at hand belongs to the pragmatic use of focus in the sense that overt focus alternatives in the immediately preceding context are corrected, such as in (3), taken from Krifka (2007:24), bracketing and capital letters in the original, *F* indicates focus. The alternatives are always alternatives of denotation.

**Givenness** The second central information-structural category is Givenness, which indicates that the denotation of an expression is present in the immediate common ground, that is shared by both speaker and listener (Krifka, 2007). Given information therefore

---

<sup>9</sup>Such as, for example, in Krifka (2007:26), with operators like *fortunately* which necessarily contrast alternatives with each other:

- (1) Fortunately, Bill spilled [WHITE]<sub>F</sub> wine on the carpet.

<sup>10</sup>See Gussenhoven (2008) for a discussion of different types of focus in English.



is “information [the speaker] believes the listener already knows and accepts as true” (Clark and Haviland 1977:3 as cited in Féry 2017:138). Anaphoric expressions, such as personal pronouns, clitics, person inflection, demonstratives, definite articles always indicate givenness. But using deaccentuation, deletion or a non-canonical word order can also mark that the denotation of an expression is currently present in the common ground. The material in the study at hand is constructed such that it is expected that given expressions are deaccentuated as in the example (4) cited from Krifka (2007:39), bracketing and capital letters in the original. This example also shows that unlike focus, givenness always refers to denotations and never to expressions (ibid.) - even though the expression *the shed* is new it refers to the *old farm* which is already present in the common ground. While focus and givenness seem to be direct opponents it needs to be pointed out that they are not necessarily mutually exclusive. It is possible, and not unusual, that given expressions, like for example pronouns, are focused (e.g. Mary only saw [HIM]<sub>F</sub>, Krifka, 2007:39). The complement of given therefore is not focus, but *new*. A referent is new when it is not activated in the current attentional state of the interlocutors at the point of their introduction into the discourse (Féry, 2017).<sup>11</sup>

- (4) Ten years after John inherited an old farm, he SOLD [the shed]<sub>GIVEN</sub>.

**Topic** The term topic refers to the aboutness of the utterance, that is to that part of the sentence about which the rest of the sentence, usually referred to as *comment*, comments, see example (5) as taken from (Krifka, 2007)<sup>12</sup>.

- (5) [Aristotle Onassis]<sub>TOPIC</sub> [married Jacqueline Kennedy]<sub>COMMENT</sub>.

A special sub-category is contrastive topic, which is a combination between topic and focus in so far as the topic implies alternatives, such as in example (6), also taken from Krifka (2007:44). Looking at it from a questions-under-discussion account we can say that a contrastive topic is focal to the overall question and topical to the subquestion - one of the questions from the alternative set (*what does your sister do, what does your brother do*) is still under consideration.

<sup>11</sup>Note that some authors assume different degrees of givenness, for example Baumann and Riester (2012).

<sup>12</sup>*Topic* and *comment* are sometimes also referred to as *theme* and *rheme*, for example in Halliday (1967).

- (6) A. What do your siblings do?  
B. [My [SISter]<sub>FOCUS</sub>]<sub>TOPIC</sub> [studies MEDicine]<sub>FOCUS</sub>, and [my [BROther]<sub>FOCUS</sub>]<sub>TOPIC</sub> is [working on a FREIGHT ship]<sub>FOCUS</sub>.

Note that sentences may not have a topic at all, such as for example in (7) as taken from Féry (2017, p. 147), where the whole sentence is the comment as indicated by the subscripted COM. These kind of sentences are called *thetic* (ibid.).

- (7) [The HOUSE is on fire]<sub>COM</sub>

Topic is mentioned here for the sake of completeness but will only be touched upon in the discussion of the first experiment.

### 2.1.3. The prosodic marking of focus

Having defined what focus is we can say that marking the focus structure helps the listener to interpret which parts of the utterance are the most relevant from the perspective of the speaker. There are different ways in which languages induce focus. Common strategies include: changing the word order, as in, for example, Russian, Hungarian, Czech, Catalan, or Turkish (cf. e.g. Zimmermann and Onea, 2011, and the references therein), by using morphological markers, as for example in Gùrùntùm (Hartmann and Zimmermann, 2009) by lexical means (e.g. by particles like *only*, *also*) or by using prosodic cues: either by prosodic phrasing, as for example in the tone language Chichewa (Kanerva, 1990, as cited in Zimmerman and Onea 2011), or by the placement of a pitch accent over a stressable category, such as in German and English, and in many other languages (see Kügler and Calhoun, 2020, and the references therein). Note that there are also languages that do not have a grammatical marker to systematically indicate focus in any way (apart from context), as, for example, Northern Sotho (Zerbian, 2006).

**Focus realisation by pitch accent placement** Being so-called intonation-only languages (Gussenhoven, 2004), German and English use pitch accenting to highlight the most informative parts in an utterance: the focus. Pitch accents are tonal events, meaning changes in pitch which give prominence to the syllable with which they are associated (cf. e.g. Ladd 2008 and 2.1.1. (If the pitch-accented word consists of more than one syllable, it is the lexically-stressed syllable that carries the pitch accent.)

Consider the following example (8). The exact same sentence may change with respect to what is interpreted as the most informative part according to what is highlighted by

means of pitch accents. The focus of the answer to the question ‘what does Julia like’ implies that there are alternatives for *novels* as indicated by *F*, a grammatical marker indicating focus material (it is not *comics*, not *dramas* etc. no, it is *novels* that she likes). The focus in example (9) is on a different constituent, implying alternatives to *Julia*: it is not *Denise*, or *Mary* or *John*. Pitch-accenting the verb as in example (10) indicates alternatives to the verb: not *hate* but *like*, and in a context such as the correction in example (11) the alternative is likewise overt and explicit: not *comics* but *novels*.

- (8) What does Julia like?  
Julia likes [NOVELS]<sub>F</sub>
- (9) Who likes novels?  
[JULIA]<sub>F</sub> likes novels
- (10) Does Julia hate novels?  
(No,) Julia [LIKES]<sub>F</sub> novels
- (11) Does Julia like comics?  
(No,) Julia likes [NOVELS!]<sub>F</sub>

While the relationship between pitch accent and focus seems to be pretty straightforward as presented in the examples above (focus constituents are pitch accented), it is important to note, however, that there is no exclusive one-to-one relationship between focus and the prosodic realisation of a sentence. In a narrow-focus such as for example in (8), focus is on one syntactic constituent while in an all-new focus (also referred to as wide focus or broad focus) as in (12) the focus is on the whole sentence - that is *Julia likes novels* is a unit to which countless alternatives are possible. Not every word in an all-new focus can be prominent (this would be counter to the function of highlighting). Once the focus constituent is clear (and it is larger than one word), which word within the focus constituent is realised with a pitch accent is defined by (language-specific) rules of default accent placement, such as Gussenhoven’s Sentence Accent Assignment Rule, which states that a focus accent on the argument can project to the whole verb phrase leaving the predicate unaccented when argument and predicate are adjacent (Gussenhoven, 1983a, 1992).<sup>13</sup>

<sup>13</sup>In fact, there has been much debate on how to analyse which word in the all-new focus is realised

In the syntax and semantics literature this process is referred to as *focus projection*: a grammatical marker *F* marking focus material can project the focus accent up the syntactic tree to a hierarchically higher node marking the entire constituent as focused (see e.g. Selkirk, 1984; Truckenbrodt, 2007; Schmerling, 1976; Féry and Samek-Lodovici, 2006, see also Ladd, 2008:218ff and the references therein).

In effect, we can see that the prosodic realisation of two different focus structures as in (9) and (12) (evoking different sets of alternatives, see above) is the same, namely with a pitch accent on the rightmost constituent.<sup>14</sup>

- (12) What did you say?  
       [Julia likes NOVELS]<sub>Focus</sub>

All-new focus, however, is not investigated in the study at hand and therefore not discussed further.

Another important aspect that should be addressed with respect to the relationship between pitch accent and focus is that there are pitch accents which are not associated with focus, as for example in (13) taken from Ladd (2008:260, 265). Here, the focus accent is on *diaries* as indicated by the capital letters, but there is an additional pitch accent on *mother's* (indicated by H\*) which is not associated with focus. Examples like this have resulted in the common distinction between prenuclear and nuclear accents, with the latter being the most prominent accent in the phrase associated with focus and the former being any (optional) pitch accents preceding the nuclear accent.<sup>15</sup>

- (13)       H\*       !H\* LL%  
           my mother's DIARIES?

**Focus realisation by pitch accent type** Coming back to the idea from the phonological models of intonation presented in 2.1.1 that a particular tonal category can be linked to a specific linguistic function, it has been claimed that focus is not only marked by the presence of a pitch accent but also by a specific pitch accent type.<sup>16</sup>

---

more prominent than the others, ranging from what Ladd refers to as the normal stress view to the highlighting view to the Focus-To-Accent Approach in a way combining the two perspectives. See (Ladd, 2008, ch. 6.1) for an overview.

<sup>14</sup>Note that phonetically they may differ with respect to emphasis, that is, for example, how large the pitch excursion on the accented word is. While *pitch accent* is categorical, *emphasis* is gradient (cf. Ladd, 2008).

<sup>15</sup>There is a debate whether there are also postnuclear accents, see for example Kügler and Féry (2017).

<sup>16</sup>Note that the term phonological model may be misleading since ToBI annotation convention include labels approximating the contour without them corresponding to a linguistic function in any clarified or one-to-one way.

In general, focus in declaratives “has a raising effect on the  $F_0$  of the focused word” (Féry, 2017:152). In effect, it is often marked by a pitch accent with a high tonal target on the accented syllable within the focus domain. This is the case not only for intonation languages, but also several other languages, such as for example Swedish, where focus is marked with an additional high tone added to the lexical pitch accent (see Dogil 2003 and the references therein). Likewise in German and English, focus is typically marked with a falling pitch accent with a high tonal target on the accented syllable ( $H^*L$ ) (see e.g. Büring, 1997; Féry, 1993, 2017; Uhmann, 1991; Kügler and Gollrad, 2015 for German, and Selkirk, 1995; Gussenhoven, 2004 for English). A topic, on the other hand, is typically marked with a rising accent ( $L^*H$ ) (Büring, 1997, 2003; Gast, 2010; Braun, 2006) which, however, is not the main accent in the sentence.<sup>17</sup> This partitioning has led to the formulation of the composed meanings of pitch accents (e.g. Pierrehumbert and Hirschberg, 1990; Hobbs, 1990) arguing that  $H^*$  marks the respective element as salient and new, and  $L^*$  marks the respective element as salient but not new. The low trailing tone in  $H^*L$  is said to add the meaning that the information is not specified concerning its completeness (consistent with the interpretation of focus), while the high trailing tone in  $L^*H$  is said to imply incompleteness (consistent with the interpretation of topic) (Dogil, 2003). Example (14), taken from Mayer (1997) who adapts the example from Büring (1997) is a German example of a contrastive topic prosodically realised with a rising accent ( $L^*H$ ) and a focus marked with a falling accent ( $H^*L$ ).<sup>18</sup>

- (14) A: Welches Buch würde Fritz kaufen?  
 B: [Ich]<sub>T</sub> würde [“Das Hotel New Hampshire”]<sub>F</sub> kaufen.  
            $L^*+H$      $H^*+L$

Phonetically, the pitch accent in a contrastive focus was found to have a high and late  $F_0$  peak, a great pitch excursion and a long duration (see, e.g. Kügler and Gollrad, 2015; Baumann et al., 2007 for German, and Breen et al., 2010; Gussenhoven, 2004; Rietveld and Gussenhoven, 1985 for English). The height of the pitch peak correlates with the level of emphasis - the more contrastive the focus is, the higher the perceived pitch (Gussenhoven, 2004; Kügler and Gollrad, 2015).<sup>19</sup> In MAE\_ToBI and GToBI, allowing

<sup>17</sup>Falling accent types for foci and rising accent types for topics are also found in other languages, such as Russian and Arabic (Alter and Junghanns 2002, and El Zarka 2017 respectively, as cited in Kügler and Calhoun, 2020).

<sup>18</sup>Note that some authors use a + in bitonal pitch accents while others do not. This, however, does not imply any differences. The pitch accents in this example are cited from Mayer (1997) who uses the + in this case.

<sup>19</sup>Whether the distinctions based on peak scaling are gradient or categorical is not entirely clear (cf. Kügler and Calhoun, 2020).

bitonal pitch accents with leading tones (see above), the distinction between contrastive and new is reflected in the inventory offering two distinct pitch accents, L+H\* and H\*, with the former being typically associated with contrastive foci and the latter with non-contrastive foci (e.g. Pierrehumbert and Hirschberg, 1990; Baumann et al., 2007; Nanette Veilleux and Brugos, 2006; Ito et al., 2004). But even within the category of contrastive focus, it may be necessary to distinguish between different levels of emphasis. This means that in a corrective focus context, such as in (11), and in the contexts used in the experiments in the following chapters, the  $F_0$  peak is expected to be higher than in, for example, informational focus contexts and may be higher than in a selective contrastive focus, such as in (15) taken from Féry (2013).

(15) Where are your sons going to?

TOM is going to VIENNA, and ALAN to BERLIN.

From a phonetic perspective it is therefore important to clarify what kind of focus is under investigation. Corrections are often used in experimental investigations, so sentences in which something is directly contradicted, under the notion of contrastiveness. We can assume that this is done more emphatically than a non-corrective contrast. Even though correction is also a contrastive focus we cannot assume that what it elicits prosodically is representative for all contrastive foci.<sup>20</sup>

In the experiments of this thesis, double-focus sentences with corrections are used and will therefore be referred to as corrective foci. Double-focus sentences are sentences involving two foci (therefore double-focus, sometimes also referred to as dual-focus) giving answers to two questions such as “*what* did she give?” and “*to whom* did she give it?”. The answer then involves two foci: *He gave [the piano]<sub>Focus1</sub> to [teachers]<sub>Focus2</sub>*. In the stimuli design of the experiments at hand such an answer is elicited in a context of correction, such as in *She didn't give the drums to students, she gave [the piano]<sub>F</sub> to [teachers]<sub>F</sub>*. For more information on double focus in German and in English the reader is referred to (3.1) and to (4.1) respectively.

### 2.1.4. Variation in the prosodic marking of information structure

While the relationship between information structure and prosody outlined above is quite persuasive, it is, as just pointed out, not always completely reliable: sometimes

---

<sup>20</sup>Wagner (1999) notes that correction contrast and selection contrast (or coordinate contrast as she calls it) also differs pragmatically since in the case of correction the speaker not only introduces a contrast to the discourse but additionally wants to “actively override an element in the listener’s informational state” (Wagner, 1999, p. 1529).

there is variation in the actual choice of pitch accent types; that is, we find different types than expected (Féry, 2007; Baumann, 2006; Féry and Kügler, 2008; Schweitzer et al., 2009). And sometimes there is variation in the placement of pitch accents, and pitch accents may not be realised even though we would expect them from a semantic or pragmatic point of view, or the other way round, we find one where we would not have expected it from an information-structural perspective (Riester and Piontek, 2015; Kentner, 2012a).

**Variation in pitch accent type** With respect to the partitioning of rising accents for topics and falling accents for foci, Féry (2007) argues that this is not a reflex of their respective informational role in the discourse but of the languages' general syntactic and phonological structure. Topics are typically realised at the beginning of the sentence and foci at the end of the sentence. The typical accent pattern thus arises from their position in the sentence, since "in a sequence of two accents the first one has a rising and the second one a falling contour, independently of the role of the constituent" (2007:9).<sup>21</sup>

Irrespective of position in the sentence, there are numerous studies which demonstrate that the same accent type may be used for different discourse functions and the same information-structural category may be marked by different accent types. Several studies showed, for example, that the two pitch accents H\* and L+H\* are, on the one hand, difficult for labellers to distinguish and on the other hand not a reliable cue for speaker intention, since both accent types are often used for either context. This was found for English (e.g. Watson et al., 2008) and for German (e.g. Kügler and Gollrad, 2015; Grice et al., 2017).

Grice et al. (2017), for example, investigated the production of broad, narrow and contrastive focus and found that while speakers prefer specific accent types for particular focus types this relationship is rather probabilistic and not exclusive - there is variation both within and across speakers.<sup>22</sup>

Not only focus accents, but also pitch accent types that have been associated with information status, that is, the degree of givenness of a referent in the discourse (cf. e.g. Baumann and Grice, 2006), have also been found to show variation. While there is the tendency that novelty is expressed with falling accents and given or accessible information with rising accents or no accents, these categories cannot be unambiguously

---

<sup>21</sup>In a corpus study I conducted preceding this study (Schauffler et al., 2015a) such a pattern could, however, not be confirmed as a general rule.

<sup>22</sup>Regardless of the phonological categories, they found that speakers showed all the same tendencies in relative phonetic terms, such as higher peak, later peak alignment, and greater tonal onglide for contrastive focus compared to narrow focus, and for narrow focus compared to broad focus.

mapped to these specific accent types (Brown, 1983; Baumann, 2006; Féry and Kügler, 2008; Baumann and Grice, 2006; Schweitzer et al., 2009).

These findings have brought Roettger et al. (2019) to the conclusion “[...] that there is no one-to-one-mapping between intonational events and speaker intentions; any assumed mapping is probabilistic at best (systematic but not deterministic). (2019:2)”

Also Kügler (2011) concludes that the function of focus accents is to highlight the focus and that this function is fulfilled by a deviation from neutral register, regardless of which direction the deviation takes.

The aim of this thesis is to identify factors that are responsible for some of the variation that we find in the prosodic marking of focus in terms of pitch accent type AND pitch accent placement.

**Variation in pitch accent placement** While the variability of pitch accent types is a well-accepted finding, the marking of focus constituents by the placement of a pitch accent should be a more stable matter. It is important to note, however, that not every focus constituent is realised with a pitch accent and not every pitch accent marks a focus (or topic). With respect to accent without focus, prenuclear accents, for example, are relatively common and do not affect the focus. With respect to focus without accent, I already mentioned the case of broad focus, where not every part of the focus is accented. And there are also other regular instances where focus is typically not marked with a pitch accent. One of these cases is the so-called Second Occurrence Focus which constitutes a focus associated with a focus operator but on the same time given material, such as *only vegetables* in example (16b) as taken from Féry and Krifka (2008).

- (16) a. Everyone already knew that Mary only eats [vegetables]<sub>F</sub>  
 b. If even [Paul]<sub>F</sub> knew that Mary only eats [vegetables]<sub>SOF</sub>, then he should have suggested a different restaurant.

Féry and Krifka (2008) list other examples of regularly missing accents, which are given in example (17). In (17a), the usually accented intensifier *herself* is deaccented, due to the adjacent narrow focus on *rice* in order to avoid a stress clash; in (17b) (taken from Rooth, 1992), the focus associated with *only* is again deaccented due to the parallel structure; in (17c) (taken from Reis and Rosengren, 1997), *also* carries the pitch accent while the rest of the sentence is unaccented, and in (17d) (taken from Krifka, 1999), the topic *Peter* is not accented due to the more prominent other topic *Gauguin*.

- (17) a. Marie-Luise even grows RICE herself.



- b. People who GROW rice only EAT rice.
- c. John said that Mark is coming, but what did Sue say?  
She ALSO said that Mark is coming.
- d. Boy, Paul possesses a Gauguin.  
Einen GAUGUIN besitzt Peter AUCH  
'Peter also owns a Gauguin'

Apart from these instances reported in the (theoretical) literature, various studies found variation with respect to accent placement in experimental as well as corpus data. Speakers sometimes place pitch accents where we would not expect them, for example on given information (e.g. Nootboom and Kruyt, 1987; Yule, 1980; Bard and Aylett, 1999) or they omit an expected one for example on new information (Terken and Hirschberg, 1994).

In their investigation of a German radio news corpus (DIRNDL, Eckart et al., 2012) Riester and Piontek (2015), for instance, found several cases in which an accent was missing where it would have been expected and instead is realised on an earlier syllable. Example (18) and Figure 2.2 demonstrate one of these cases:<sup>23</sup> given the context, the accent would normally fall on *Bank* ('bank'), since this is the default stress pattern here. Instead, it is realised earlier, on *Dresdner* ('Dresden'). Note that there is no other 'bank' mentioned in the discourse which could elicit a contrastive interpretation and justify a focus accent on *Dresdner* on pragmatic grounds. The authors explain this shift with rhythmic preferences similar to the ones illustrated in (17a): the pitch accent was shifted from *Bank* to *Dresdner* in order to avoid two pitch accents directly following each other since the following word *Ponto* carries a pitch accent on the initial syllable. Such a pitch accent clash is demonstrated in example (19) (notation and highlighting in the original).

- (18) Nach 24 Jahren Haft ist die frühere RAF-Terroristin Mohnhaupt freigelassen worden. [...] Unter anderem war sie an der Ermordung von Generalbundesanwalt Buback, des Vorstandssprechers der DRESdner Bank, PONto, und von Arbeitgeberpräsident Schleyer beteiligt.

*'After 24 years in prison, former RAF terrorist Mohnhaupt has been released. Amongst other crimes, she was involved in the murder of Federal Public Prosecutor Buback, the Board Spokesman of the DRESdner Bank, PONto, and the President of the Employer Association, Schleyer.'*

<sup>23</sup>Example (19) and Fig. 6 in the original (Riester and Piontek, 2015, p. 244f.), I omitted the information-status labels for better readability.

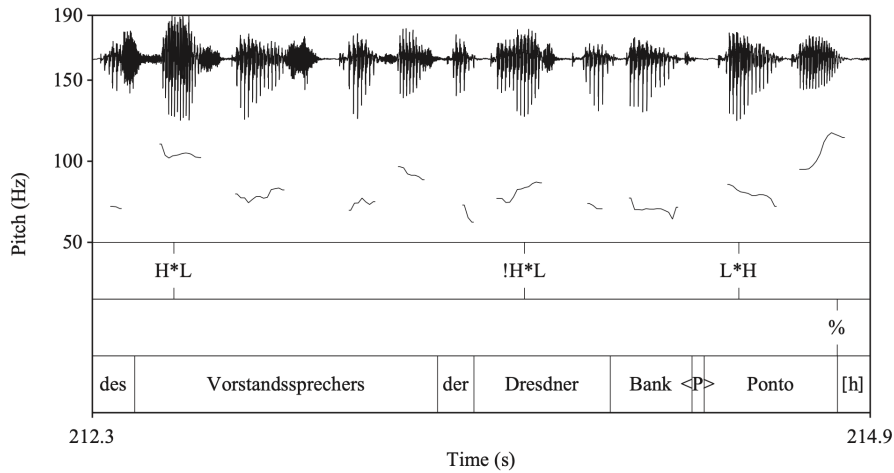


Figure 2.2.: Example from Riester and Piontek (2015:244) for a rhythmic accent placement within a broad focus.

(19) ?[. . . des VORstandssprechers der Dresdner BANK, PONto, . . .]<sub>F</sub>

Other studies made similar findings, which will be presented in more detail in the following sections. Kentner (2012a), for example, found that in unprepared oral reading, speakers avoided producing an information-structurally required accent when the production would have resulted in a stress clash.

To sum up, in German and English focus is prosodically marked by means of placement and type of pitch accent. However, there is variation in the mapping of form and function in everyday production so that we cannot assume a one-to-one relationship. This thesis sets out to investigate factors responsible for some of the variation that we find in the prosodic marking of focus. It thereby contributes to identifying factors that shape the use of intonation in German and English and investigates language-specific tendencies. The factors investigated are prosody-inherent in nature, namely the preference for rhythmic alternation that both languages share. The following section will give some background on this preference.

## 2.2. Alternation preferences influence sentence production and processing

At the heart of this thesis is the question whether alternation preferences can account for variation in the prosodic marking of focus. Looking for rhythmic factors to explain unexpected pitch accenting, such as in (18) above, is motivated by numerous studies

that have shown that speakers - and listeners - prefer an alternation of strong and weak beats and avoid structures that violate this principle (e.g. Kentner, 2012a; Bohn et al., 2011; Geiser et al., 2006; Rothermich et al., 2012; Tilsen, 2012; Breen and Clifton, 2011; Quené and Port, 2003; Kimbell and Cole, 2014). These studies will be outlined in the following after some aspects of rhythmic alternation preference are introduced.

### 2.2.1. The Principle of Rhythmic Alternation

This preference has long been known as the Principle of Rhythmic Alternation describing a tendency not only in language but also in music to create rhythmically well-formed structures (e.g. Sweet, 1876; Selkirk, 1984; Liberman and Prince, 1977; Schlüter, 2005). What is considered well-formed in language may be language-specific - for some languages it may be the avoidance of stress clashes, for others it may be the preference to stress the first or the last syllable in a phrase (cf. Wagner, 2008). In either case, well-formedness emanates from a certain structure which results in regularly recurring events, thereby creating expectations.

It is reasonable to assume that German and English share similar influences of rhythmic alternation on the language, since they both belong to the group of “stress-timed” languages. In stress-timed languages the distance between stressed syllables is assumed to be isochronous, whereas in languages which belong to the group of “syllable-timed” languages, such as, for example, French, Spanish or Italian, it is the syllables that are isochronously distributed (Pike, 1945; Abercrombie, 1965, 1967). In a third group of languages the crucial interval is ascribed to morae, which is why this group is referred to as “mora-timed” languages, such as, for example, Japanese. While neat isochrony of these respective intervals was not found experimentally in any measurable way (e.g. Pointon, 1980; Roach, 1982; Dauer, 1983) these terms are still regarded as extreme points on a continuum correlating with, for example, the degree of vowel reduction and phonotactic complexity (e.g. Dauer, 1983; Auer and Uhmman, 1988). The different classes are also reflected in poetry, as Wagner (2008) points out. Languages prefer different metres depending on their rhythm class categorisation: While in German poetry the important unit is the number of stresses per line, it is the number of syllables in, for example, Italian and Spanish, and the number of morae in Japanese, such as in the famous poetic form, the *haiku* (see *ibid.* p. 9f and the references therein).

In her investigation on the Principle of Rhythmic Alternation in English corpora, Schlüter (2005) found numerous instances where rhythmic alternation preferences affected grammatical variation. To name but one, she found effects of alternation prefer-

ences on the use of variants of multiword attributive structures, such as *quite a* versus *a quite*, where the former was more often used preceding an initially-stressed adjective (e.g. *quite a different view*) thereby avoiding a stress clash, compared to the latter (cf. also Bolinger, 1965).

For German, Vogel et al. (2015) could show that function words are frequently used in a way which optimises “the rhythmic flow in speech”. In one study they found, for instance, that pronouns are used in their strong, weak or reduced form depending on the rhythmic context, with the long version in contexts where a strong syllable is rhythmically preferred, and with the reduced version in contexts where an unstressed syllable is preferred. With another study (ibid.) they showed that the optional word order in three-verb clusters can be influenced by the rhythmic context.

The minimal domain of the Principle of Rhythmic Alternation is the syllable (Schlüter, 2005).<sup>24</sup> Rhythmic alternation was also dealt with by Chomsky and Halle (1968, p.114-117) in *The Sound Pattern of English* in the form of three phonological rules that aimed at separating primary and secondary stress by intervening unstressed syllables. These rules, however, are restricted to the word level, hence only accounting for word-level stress (as cited in Schlüter, 2005). Alternation preferences above the single word level found their formalisation within Metrical Phonology (Lieberman and Prince, 1977; Selkirk, 1984; Hayes, 1984).

**The rhythm rule** A major contribution to Metrical Phonology was the paper “On stress and linguistic rhythm” by Liberman and Prince (1977), where they highlight the relational nature of stress - the prominence of a syllable can always only be determined in relation to its surrounding syllables. They also provided a formal account for the so-called “rhythm-rule” - an instantiation of the Principle of Rhythmic Alternation, that until then had only been discussed informally as the exception to a rule. The rhythm rule operates in order to avoid stress clashes and to ensure an alternation of strong and weak beats, i.e. stressed and unstressed syllables, by shifting one of the clashing stresses to an earlier syllable. Consider a phrase like “fifteen women”. *Fifteen* in isolation has its primary stress on *-téen*, but when adjacent to a word with initial stress (*wómen*) primary stress is shifted to the first syllable onto *fif-* resulting in *fifteen wómen* - a phrase with an alternating stress pattern of strong and weak. Liberman and Prince (1977) refer to this (optional) process of stress retraction as *Iambic Reversal*.

Different graphical representations have been used to depict these prominence rela-

---

<sup>24</sup>In some languages alternative or additional units may be the mora, the foot or the phrase (cf. Wagner, 2008).

tions, and there has been some discussion on whether metrical trees (Giegerich, 1985; Hayes, 1995), metrical grids (e.g Selkirk, 1984; Nespor and Vogel, 1986), a combination of both (Lieberman and Prince, 1977), or bracketed grids as postulated by Idsardi (1992, as presented in Archibald, 1994) best serve to present prosodic constituent structure. I will not go into the details of this debate here since it does not serve the purpose of my investigation. In order to illustrate examples and stimuli used in this study, I will rely on the metrical grid as the representation that is most commonly used, and better suited to represent and visualise prominence relations than the metrical tree.<sup>25</sup> The grid treats stress in relation to stress of the surrounding syllables represented by columns where the height of the columns - represented by more or fewer marks - indicate more or less prominent syllables. It is therefore well suited to visualise when and how the rhythm rule may apply. See (20) as an example, a column height of three represents the lexical stress level (level 1 refers to the syllable level and level 2 to the foot - the fundamental beats that each syllable usually receives in the grid are left out for better readability.)

		×				×	
	×	×			×		
×	×	×	×		×	×	×

(20) thirt e e n wom e n → thirt e e n wom e n

The rhythm rule is also applied in German, where secondary stress is shifted away from primary stress when the two syllables happen to be adjacent. This can occur when a word carrying primary stress on the initial syllable, as in *sichtbar*, is embedded into a larger constituent which in turn requires primary stress on the first syllable and “pushes” the original stress, now secondary stress, away to a later syllable, as in *unsichtbar*. Shift to the right may also happen across word boundaries, like in phrases which involve prefixed verbs that carry the primary stress on the prefix, like in *den Róck anziehen* (Kiparsky, 1966). There is a difference between German and English, however, since unlike German, English does not allow a shift to the right, *spórts còntest* will never become *spórts contèst* (Lieberman and Prince, 1977, p. 311). Gussenhoven (1983) refers to the shift to the right, which is also possible in Dutch, as *trochaic reversal*.<sup>26</sup>

Experimentally, the rhythm rule in German has been less investigated until now. Among the studies conducted, Wagner and Fischenbeck (2002), for example, found that

<sup>25</sup>The metrical grid is also flawed, as Grabe and Warren (1995) point out: It bears no reference as to whether the marks on the grids refer to production, perception or both, it does not specify acoustic correlates, it mostly offers no surrounding context, and “matters of intonational phrasing such as nucleus placement tend to be implied rather than overtly incorporated into the account”. (p. 96)

<sup>26</sup>Gussenhoven notes, however, that the trochaic reversal differs phonetically from the iambic reversal discussed above since it results in a durational restructuring of the postaccentual syllables but not in a shift of pitch peaks (1983b, p. 316).

stress shift in German trisyllabic and quadrisyllabic compounds is an optional process that is not very regularly used. Bohn et al. (2011) on the other hand found that the rhythm rule operates not only on the word but also on the phrasal level in German (e.g. *Termín ábsagen* → *Termín abságen*). They found that stress shifts are regularly perceived in both German compounds (e.g. *Hauptbahnhof*) and phrasal verbs. Additionally, their acoustic analysis revealed that the rhythm rule was also present in production where speakers applied it frequently in order to avoid a stress clash. The strategy that speakers used to resolve the stress clash was mostly stress reduction by shortening the primary stressed syllable of phrasal verbs (Deletion Analysis) rather than a complete stress shift (Reversal Analysis).

Chapter 4.3 will present a production experiment which investigates how the rhythm rule operates under focus marking in English. The rhythm rule and previous studies investigating stress shift will be introduced in more detail there in order keep the relevant information closer for the integration of said experiment.

The rhythm rule, however, is not the only outcome of the Principle of Rhythmic Alternation. Various studies showed other effects resulting from the preference for alternation on both language production and processing (see below). Alternation preferences have also been argued to have influenced the development of English grammar such as for example the stress shift in words after derivational affixes are added (e.g. *solid* vs. *solidity*), or the different stress patterns of English nouns (mainly trochaic) and English verbs (mainly iambic) as a result of the different rhythmic environments they occur in (Kelly, 1988; Kelly and Bock, 1988). It has also been shown to affect the choice of dative constructions (Anttila et al., 2010), genitive construction (Shih et al., 2015) or word order (Vogel et al., 2015), and the historical loss of certain word forms (Schlüter, 2005). And as already mentioned above, in her book “Rhythmic grammar”, Schlüter (2005) names many more examples of grammatical phenomena affected by the Principle of Rhythmic Alternation.

But what is the function of such processes as stress shift – why the effort to adhere to a principle that at first glance does not seem to have any functional load?

### 2.2.2. The function of rhythmic alternation

Forming rhythmically agreeable utterances is, as it happens, not only a matter of aesthetics as the term *eurhythm* may suggest. Rhythm in language has a particular guiding function and makes speech understanding more effective: It supports the guidance of the listener’s attention to relevant events in the utterance – on the one hand, different levels

of boundaries and, on the other hand, prominently marked units relevant for sentence meaning (cf. Wagner, 2008; Kohler, 2009). Particular attention will be drawn to where the rhythmic beat is strongest, for example in stressed syllables (Quené and Port, 2003; Pitt and Samuel, 1990). Stressed syllables carry the greatest amount of information and this is why a rhythm which guides the attention to these units is particularly economical (Lehiste as cited in Kohler, 2009). Breen and Clifton (2010) attest stressed syllables a “privileged” status. They are used by listeners as anchor points to decode word onsets, phonemes in stressed syllables are detected faster and mispronunciation of stressed syllables affects lexical access stronger than erroneous unstressed syllables (see Cutler and Norris, 1988, Culter, Dahan and Van Donselaar, 1997, Shields, McHugh and Martin, 1974 and Mattys and Samuel, 1997 as cited in Breen and Clifton 2010). It has also been shown that words with two stressed syllables take longer to read than words with one stressed syllable, even when the number of syllables are the same (cf. Ashby and Clifton, 2005 also as cited in Breen and Clifton, 2010).

An alternation of strong and weak syllables helps to highlight these points in time additionally, by making the stressed syllables in direct contrast to the adjacent weak syllables as salient as possible (cf. Schlüter, 2005). Maintaining maximal contrast between adjacent units or units within short distances is not only beneficial for perception but also facilitates language production, as has been shown for example by Cutler in an experiment on speech errors (1980, as cited in Schlüter 2005:33).

According to Abercrombie (1967:96), rhythm “in speech as in other human activities, arises out of the periodic recurrence of some sort of movement, producing an expectation that the regularity will continue”. It is this expectation that is the crucial benefactor of rhythmic well-formedness – a regular pattern of recurring events, like strong and weak beats, facilitates sentence processing since it creates expectations on when the next event occurs, that is when the next stressed syllable might appear. And thus we can use our cognitive capacities more economically. This expectation, fuelled by the rhythmic structure, is also responsible for the effect rhythmic alternation has on memory, as countless mnemonics impressively demonstrate (cf. Wagner, 2008).

### **2.2.3. Experimental investigations on rhythmic alternation**

Several studies have demonstrated the effect on language production and processing if these expectations are not met, even in silent reading. Breen and Clifton (2011) for example investigated expectations about certain stress patterns that were established by the rhythmic environment in an eye-tracking experiment. They compared limericks

with regular meter, thus conforming to expectations like in (21a) (capital letters mark lexical stress), to limericks in which the last word was inconsistent with the metrical environment and therefore unexpected, like in 21b.<sup>27</sup> They found that whenever readers encountered a weak-strong word in a strong-weak context, they experienced difficulty, as shown by, for example, slowed down reading times and a lowered probability to skip the critical word. This means that shifting stress to the right was costly.<sup>28</sup>

- (21) a. There once was a clever young gent  
           Who had a nice talk to preSENT
- b. There once was a penniless peasant  
           Who went to his master to **preSENT**

In a second experiment, testing material with less salient stress patterns than usually present in poetry, the authors found a similar effect: Readers had difficulty when a noun with a strong-weak stress pattern had to be reanalysed as a verb with a weak-strong stress pattern (e.g. the noun *ABstract* vs. the verb *abSTRACT*.)

Similarly, Rothermich et al. (2012) could show in an ERP experiment that semantically unexpected words elicit a smaller N400 component in a metrically regular than a metrically irregular sentence context, suggesting that perceptual regularities facilitate the integration of lexico-semantic information.

Magne et al. (2010) had participants read lists of five bisyllabic words in which the last word diverted from the either trochaic or iambic stress pattern (e.g. *body*, *level*, *study*, *woman*, *idea*). Measuring ERPs, he could also show that the rhythmic context affected semantic processing since metrically unexpected fifth words elicited a larger negative component than metrically expected fifth words.

### 2.2.3.1. Studies investigating alternation effects on pitch accenting

As we have seen, various studies found evidence for the preference for rhythmic alternation (see also studies on the rhythm rule presented in 4.3). But how do these tendencies

---

<sup>27</sup>I only present the two first lines that were important to the experiment. The limericks, however, always consisted of 5 verses.

<sup>28</sup>They did not find an effect for the reverse pattern, that is speakers had no problem shifting stress to the left when they encountered a strong-weak word in a weak-strong context. The authors explain that with the relative acceptability and frequency of left- and rightward stress shifts in English: Speakers frequently shift stress to the left to the initial syllable in order to avoid stress clashes, so “shifting stress from a later to an earlier syllable is a well-practised skill for English readers” (ibid:158). Shifting stress to the right, however, is not something that English speakers are used to do.



interact with other prosodic functions above the word, such as the prosodic marking of information-structure categories? There are only few studies which consider information-structurally induced prominences.

The findings by Riester and Piontek (2015), as presented in section 2.1.4, suggest that pitch accents may be omitted, or shifted in order to avoid a pitch accent clash. However, this investigation does not provide us with insights as to whether such an omission or shift will be a strategy that speakers take in more marked contexts, such as contrastive focus.

Henrich et al. (2015) included information structure in an ERP experiment on the processing of rhythmic irregularities in German phrases. See (22) for an example (emphasis in the original). The questions were presented visually and followed by the pre-recorded answer with either a stress shift in the target word (*abságen*) conforming to rhythmic expectations, or without stress shift constituting a stress clash (*ábsagen*).

(22) Question: **WAS** soll sie **absagen**?

*WHAT is she supposed to cancel?*

Answer: Sie soll den **TERMIN absagen**, wie besprochen.

*She is supposed to cancel the APPOINTMENT, as discussed.*

They found that irregularities elicit weaker responses when the attention is shifted towards the preceding word (by contextually inducing a narrow focus, in this case on *Termin*). The target words themselves were not investigated under narrow focus marking.

In a study on oral and silent reading, Kentner (2012a) could show that the placement of pitch accents is also affected by the rhythmic environment. His investigation centres around the ambiguous item *mehr*, which can have two readings with different prosodic realisations. The first one is the temporal reading of *nicht mehr* (in the sense of ‘no longer’), as in example (23a) and (23b). In this reading *mehr* is not expected to be accented.

(23) Der Polizist sagte, dass man...

*The policemen said that one...*

a. ... nicht mehr NACHweisen kann, wer der Täter war.

*...couldn't prove anymore who the culprit was.*

b. ...nicht mehr erMITteln kann, wer der Täter war.

*...couldn't determine anymore who the culprit was.*

- c. ...nicht MEHR nachweisen kann, als die Tatzeit.  
...*couldn't prove more than the date of the crime.*
- d. ...nicht MEHR ermitteln kann, als die Tatzeit.  
...*couldn't determine more than the date of the crime.*

The other reading of *mehr* is comparative (in the sense of ‘more than’) such as in (23c) and (23d). In this usage *mehr* is required to be accented. He manipulated the rhythmic environment by changing the stress pattern of the following verb with either initial stress, as in (23a) and (23c), or medial stress, as in (23b) and (23d). A pitch accent on the comparative *mehr* followed by an initially stressed verb such as in (23c) would then lead to a dispreferred stress clash.

He found that in unprepared oral reading speakers more often did not produce the required accent on comparative *mehr* when the following word was stressed on the initial syllable (i.e. in the stress clash condition 23c); and in silent reading they had longer reading times in this clash condition. However, with respect to focus marking, Kentner assumes that the main pitch accent in the phrase may not be altered by the desire for rhythmic alternation (cf. Kentner, 2012b).

To sum up the findings so far, we have seen that many studies have provided evidence for an alternation preference. How these preferences interact with other prosodic functions such as the prosodic marking of information structure categories, however, remains unclear. In fact, in several studies lexical stress and nuclear stress, that is, lexically stressed syllables and pitch accented syllables, are confounded, and studies compare varying levels of prominence.

In the rhythm-rule studies the information-structural context often is not considered and, lacking a phonological analysis, the prosodic structure of the stimuli is left unrepresented – nuclear accent and lexical stress are often conflated (e.g. in Grabe and Warren, 1995; Horne, 1990, cf. also Tilsen (2012), Shattuck-Hufnagel (1995) for a critique of these studies). In the study by Henrich et al. (2015) focus accent is included in the design, but not on the target word; and in Kentner’s (2012) study two levels of prominences are compared, namely a pitch accent followed by a lexical stress.

In order to understand the effect alternation preferences may have on the prosodic marking of focus, it is, however, important to distinguish between lexical stress (part of the phonology of the word) and sentence accent (part of the prosodic structure of the utterance). In this thesis I aim to be more clear in the types of clashes I look at, namely pitch accent adjacent to pitch accent and pitch accent adjacent to lexical stress. This thesis will thereby contribute to understanding the effect of alternation preferences on

the prosodic marking of focus.

### 2.2.3.2. Opposites attract - Alternation of tonal events

So far we have talked about studies investigating the preference for an alternation of strong and weak beats. Rhythmic well-formedness is, however, not only about the regular occurrence of beats, but according to Jun (2012) refers to the temporal organisation of speech by the regular occurrence of events in general. These events may be aural or visual, the acoustic medium may be timing, fundamental frequency or amplitude (ibid.) Looking at fundamental frequency, this means that not only prominence and phrasing but also variation in pitch contributes to the perception of rhythm and rhythmic well-formedness, with an alternation of high and low being “more rhythmic” than sequences of level tones (Jun, 2012, 2014). Jun terms these aspects of the tonal melody of languages “macro-rhythm” to differentiate it from the traditional speech rhythm based on weak and strong syllables, which she refers to as micro-rhythm (ibid.). The link between macro and micro rhythm is that languages differ in how weak or strong their macro rhythm is depending on their use of prominence-lending cues. Stress-timed languages may have weaker macro rhythm since they use duration rather than tonal rhythm to cue word prominence, while syllable-timed or mora-timed languages may depend more on tonal cues for word prominence given that duration is more stable and not a reliable cue for word prominence (ibid.) This kind of tonal rhythm serves a similar function as has been described for the preference for alternating strong and weak beats: it facilitates the segmentation and acquisition of words and has a guiding function in the processing of syntactic, semantic, and pragmatic meaning of a sentence (see Jun 2012, 2014 and the references therein). In order to highlight elements and guide the listener’s attention, it makes sense to create maximal contrasts between adjacent elements whether this contrast is in strong and weak, or high and low.<sup>29</sup>

In order to investigate whether alternation preferences on the tonal level play a role in German, Schauffler et al. (2015a) conducted a study on a corpus of German radio news recordings (Eckart et al., 2012). Motivated by the findings that speakers prefer an alternation of strong and weak and try to avoid stress clashes if possible (see above), the research question was whether speakers also avoid melodic clashes, expecting that speakers prefer the succession of alternating pitch accent types (e.g. L\*H H\*L) to the

<sup>29</sup>Similarly, Berg (2004:1089-1095 as cited in Schlüter, 2005) points out that “elements that stand in a syntagmatic relationship tend to contrast maximally if they are adjacent or separated at short distances.”

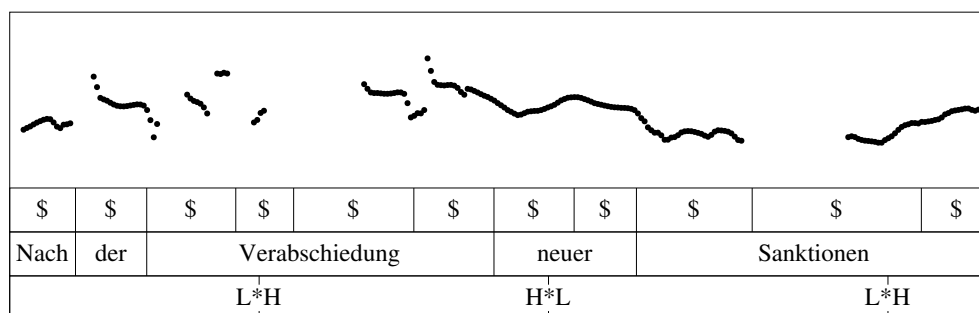


Figure 2.3.: Example from the corpus for alternating pitch accent types; \$ marks syllables.

succession of non-alternating pitch accent types (e.g. H\*L H\*L).<sup>30</sup>

Pitch accents in the corpus were categorised into *falling* pitch accents comprising all accents with a high starred tone (H\*L, HH\*L, H\*), and *rising* pitch accents, comprising accents with a low tonal target on the accented syllable (L\*H, L\*).<sup>31</sup> Each accent was then classified as either *alternating* or non-alternating, considering the preceding accent type, so that a *falling* accent following a *rising* accent was labeled *alternating* and a *falling* accent following a *falling* accent as *non-alternating*.

See Figure 2.3 for an example of *alternating* pitch accent types and Figure 2.4 for an example of *non-alternating* pitch accent types.

While a general tendency for alternation in the corpus was not found, it could be shown that the closer pitch accents are together, the more likely they alternate in type. This means that speakers indeed seem to prefer adjacent pitch accents to be different, and disprefer a succession of pitch accents with an identical contour. See Figure 2.5

<sup>30</sup>Note that there are different types of tones in sentence intonation, as outlined above: starred tones, trailing tones, phrasal tones and boundary tones. The starred tones are the ones which occur on the accented syllable and are thus the ones that evoke prominence and structure the utterance rhythmically. The trailing tones mark the trajectory of the contour after the accented syllable. For instance, an H\*L accents marks a falling contour, i.e. the contour is composed of a high starred tone (i.e. a high accented syllable which is perceived as prominent) followed by a low trailing tone on the post accented syllable(s). Therefore starred tones and trailing tones are weighted differently in terms of their relevance for sentence intonation. We therefore consider adjacent starred tones of different types as alternating.

<sup>31</sup>The pitch accent type L\*HL was excluded from the analysis due to its special perceptive character. (The L\*HL has a rising contour on the accented syllable, but due to the characteristic arrangement of the trailing tones gives the auditory impression of a fall. This accent is often found in very emphatic contexts and occurred in only about three percent of the radio news at hand.)

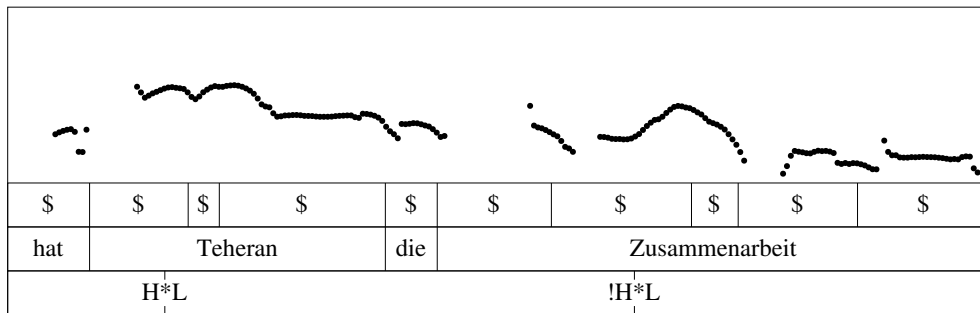


Figure 2.4.: Example from the corpus for non-alternating pitch accent types; \$ marks syllables.

for the effect of distance on the probability of alternation on the left, with alternation probability significantly higher the fewer syllables are between two pitch accents.<sup>32</sup> As can be seen on the right, the effect does not hold across intonation phrase boundaries (IP) or intermediate phrase boundaries (ip).

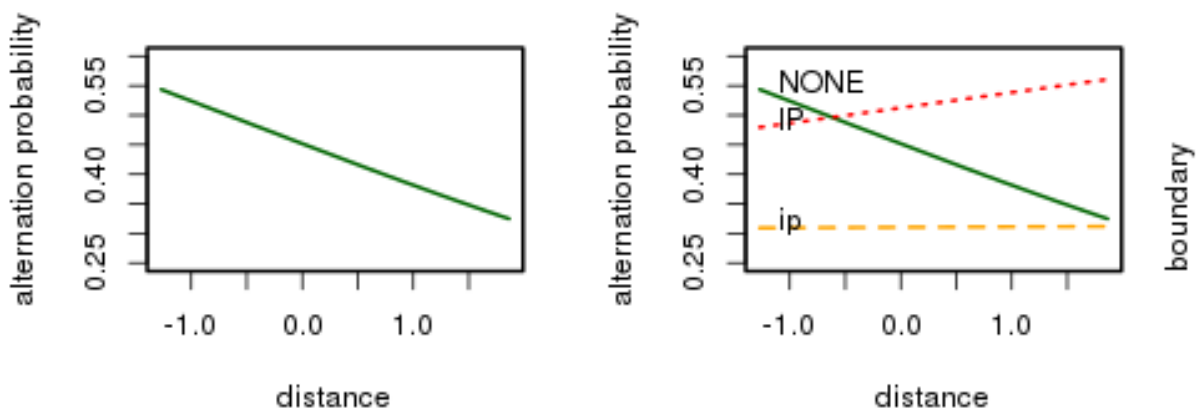


Figure 2.5.: The effect of distance on alternation probability as main effect (left panel) and in the interaction with boundary type (right panel).

This study provides some evidence that there is a tendency for tonal alternation in German and that speakers try to avoid melodic clashes. What we do not know, however,

<sup>32</sup>Distance was defined as the number of intervening syllables between the pitch accent and the preceding one. For statistical analyses, this number was log-transformed and centred, and this is the number given in Figure 2.5 on the x-axis.

is how this preference interacts with the prosodic realisation of information structure categories. Given that focus is not only realised by pitch accent placement, but is also often associated with a specific pitch accent type (see above), it is plausible that the rhythmic environments affect the prosodic structure in both - pitch accent placement and pitch accent type.

This thesis aims at gaining insight into the role alternation preferences play with respect to focus marking, looking at both the preference to alternate strong and weak beats and the preference to alternate high and low. Can alternation preferences account for some of the variation that we find in the prosodic marking of focus? Do speakers of German and English differ in their strategies to navigate between prosodic function and rhythmic well-formedness? And what is the role of alternation preferences under focus marking in an L2? The following section provides a brief overview of previous studies on L2 prosody before the next chapters are set about to answer these questions.

## 2.3. Second language prosody

Different strategies and implementations of alternation preferences may contribute to learners' difficulties in attaining accent-free speech. For example, acquiring the correct form for a specific function in the target language may not be enough when this function is embedded in an environment that adapts the required intonation category. The third part in this thesis presents two experiments conducted with German learners of English designed to investigate learners' strategies at the interface of rhythm and focus marking. This section will first establish the theoretical framework in which this investigation is carried out, and then outline relevant studies which provide insight into L2 phonology with respect to pitch accent placement, pitch accent type and rhythmic aspects.

### 2.3.1. Models of categorical perception and production of L2 speech sounds

Most theories of L2 learning have focused on the relative difficulty of learning L2 segments. Prominent models of segmental acquisition are certainly Flege's Speech Learning Model (henceforth SLM, Flege (1995)), the Perceptual Assimilation Model (henceforth PAM and PAM-L2, Best 1995, Best and Tyler 2007) and the Native Language Magnet Theory (henceforth NLM, Kuhl and Iverson, 1995). These models predict the relative difficulty for learners to acquire native-like pronunciation by means of similarity between the categories encountered in the foreign language and categories in their L1. Whether

a non-native (phonetic) contrast is perceived - and consequently produced authentically - depends on the degree of similarity.

The NLM Theory concentrates on the internal organisation of phonetic categories and how listeners distinguish between “good instances” and less good exemplars of prototypes, which were established in the course of L1 acquisition. Prototypes are those instances of a category that are representative of the category as a whole. The categorisation of L2 sounds is then based on the prototypical L1 sounds which function as a perceptual magnet, pulling the sounds within the category towards itself, reducing the perceptual distance to sounds within that category and stretching the perceived distance between sounds at the category boundaries (cf. Kuhl and Iverson, 1995). For L2 acquisition, this means that L2 sounds which are similar to sounds of an L1 category will be difficult to discriminate, while sounds which are dissimilar to an L1 category are relatively easy to discriminate (ibid.).

The SLM is less concerned with the insides of phonetic categories but rather with the boundaries between those categories. It predicts that when being confronted with unknown L2 sounds that are similar to L1 sounds, adults will sort them into the categories established for the respective L1 sounds. For new L2 sounds, adult speakers will form a new category, making native-like pronunciation possible (Flege, 1995). The greater the perceived distance of the L2 sound to the next L1 sound, the more likely it is that a new category will be established for that sound. Age of Learning (AoL) plays a crucial role in L2 acquisition since the earlier L2 learning begins, the smaller the perceived phonetic distance has to be in order to trigger the process of category formation (Flege, 1995, p. 264). The SLM thus implies that the phonetic system is able to reorganise itself by adding new phonetic categories. The production of an L2 sound will change when the learner has established a new category for it. L1 categories and newly established L2 categories exist in a common phonological space. In order to maintain contrast within this phonological space, it may happen that a newly established category for an L2 sound may be deflected away from an L1 sound category thereby hampering native-like pronunciation despite the separate category (ibid. p. 239ff.).

Similarly, also in the PAM (Best, 1995) and the extended version PAM-L2 (Best and Tyler, 2007) the concept of cross-language phonetic similarity plays a major role in determining how well L2 learners will perceptually differentiate L2 sounds. Depending on whether contrasting L2 sounds are perceived as good exemplars of one single L1 category, as unequally good exemplars of one L1 category or as exemplars of two different categories, discrimination is predicted to be poor, intermediate or excellent respectively (ibid.)

### 2.3.2. Models of language learning and their predictions concerning prosody acquisition

Adapting these models to suprasegmentals poses several challenges. For one thing, it is much more difficult to establish what the categories are that have to be compared, as is mirrored by the different intonation models (see above). It is further unclear whether intonational phenomena are indeed perceived categorically or in a gradient way. In addition to that, suprasegmentals are much more prone to variability than is the case with segments both within and across speakers (cf. Polyanskaya, 2015). Compared to the rather elaborate modelling of acquiring L2 speech sounds, there has not yet been as much work on modelling L2 intonation. This is also mirrored by the rather limited number of experimental investigations of L2 prosody. While the field of research on L2 has certainly grown in the last ten years, Gut (2009) notes that only nine out of 172 studies published between 1969 and 2008 investigated aspects of L2 intonation. This may be due to the common idea that used to prevail in earlier second language research, as pointed out by Dogil (2003), that prosody is processed differently compared to other linguistic aspects. Prosody models therefore focused more on extralinguistic, emotional and attitudinal functions of prosody, like for example by Lieberman (1967) and Bolinger (1978) (in Dogil, 2003). According to Dogil, it was the introduction of phonological models of prosody, like the autosegmental-metrical approach, that changed the universalist-extralinguistic perception of prosody by regarding prosody as a linguistic system with its own categories, which interact with categories of other linguistic levels. AM phonology assumes that intonation consists of a limited number of phonological categories that are phonetically implemented in some way (see 2.1.1). In fact, studies have showed that L2 learners have difficulties in like manner acquiring and using these prosodic categories as they have with categories of other linguistic levels, and prosodic deviations have been shown to be contributors to foreign accentedness (e.g. Jilka, 2000; Boula de Mareüil and Vieru-Dimulescu, 2006; Munro, 1995). It is therefore reasonable to investigate L2 prosody just as any other linguistic aspect.

There are two models, by Archibald (1994) and Özçelik (2016) mainly attending to formalising a model of L2 phonology from a Universal Grammar perspective. Özçelik (2016) proposes the Prosody Acquisition Path Hypothesis, which predicts the stages learners go through based on the differences between their L1 and L2. Archibald (1994) assumes a parameter-setting model, looking at parameter settings such as, for example, extrametricality, that may vary across languages, leading to predictions about likely places of transfer errors. Since both attempts, however, focus on word stress parameters



and stress acquisition on the lexical level only, I will not discuss them in more detail.

More interesting for the work at hand is Mennen (2015) who proposes a model of L2 intonation which draws parallels to models of segmental acquisition, like the SLM or the PAM, introduced above. In order to make predictions on the difficulties L2 learners may have it is essential to establish similarity between two languages. And, as Mennen points out, in order to establish similarity - segmental or intonational- one would have to measure similarity. This seems to be more straightforward for segmental comparisons, although even for segments there is no widely accepted measure so far. For intonation, however, this seems even more difficult. It is not at all clear what actually is categorical and what is gradient in the different functions that intonation serves. With the introduction of the AM framework, so Mennen states, comparison between languages in terms of intonation has become easier. Within this framework cross-language similarity can be established with respect to both phonological categories and phonetic shape.

In her L2 Intonation Learning theory (henceforth LILt) she suggests that we differentiate between four dimensions (modified from Ladd 1996) along which similarities and differences between languages can be described:

1. **The systemic dimension:** comprising the inventory and distribution of categorical phonological elements; that is, what structures and tunes are permitted and how the different structural elements combine with one another, for example, the ‘rise-plateau-slow’ in Belfast and Glasgow accents which is not found in RP or American English (Cruttenden, 1986:139ff, Ladd, 1996:126, as cited in Mennen (2015). The tune-text association also belongs to this dimension, for example, accents in Italian and Greek may be realised on unstressed syllables (Ladd, 1996:119, Arvaniti et al., 2006, as cited in Mennen).
2. **The realisational dimension:** comprising the phonetic implementation of these categories, like, for example, how pitch accents are aligned to segments (e.g. pre-nuclear rises occur later in German than in English (Atterer and Ladd 2004)).
3. **The semantic dimension:** the functionality of the categorical elements or tunes, like, for example, languages may differ in how they mark focus.
4. **The frequency dimension:** meaning the frequency of use of the categorical elements, like, for example, rises are more commonly used by female speakers of Northern Standard German than by female speakers of Southern British English (Mennen et al., 2012).

Differentiating between these dimensions helps to systematically predict the relative difficulty learners would experience with certain L2 intonational parameters or dimensions. Additionally, it may give insights into how and for example in which order parameters of intonation are acquired in an L2. In a review of L2 production studies, Mennen demonstrates how deviations from the native form are found in each of the four dimensions (see p. 175ff for further examples and an overview). It is important to note that a deviation in one dimension may lead to a deviation in another dimension, so that it is not always easy to keep the dimensions apart.

While LILt shares the assumption of SLM and PAM-L2 that many difficulties are perceptually motivated, it also points out that in order to determine perceptual similarity, we have to refer to the semantic dimension of intonation. Additionally, LILt accounts for the fact that there may be other reasons that can explain deviations in production aside from perception, for example speakers may have articulatory difficulties to produce L2 sounds or difficulties to store them in acoustic memory.

Another shared assumption with models of segmental learning such as the SLM is that L1 and L2 categories exist in a common phonological space. This may cause languages to interact and this interaction is bi-directional in nature (cf. Flege 1995, Mennen 2004). This means that assimilation or merging as well as dissimilation or polarisation of L1 and L2 properties is possible.

### 2.3.3. Beyond L1 transfer – L2 universals

Interference of the L1 sound system with L2 pronunciation in one way or the other is assumed in most research on L2 production. The L1 system may either facilitate or hinder accurate L2 production. However, L1 interference is, of course, not sufficient to explain all there is to L2 acquisition and not all difficulties speakers may have with producing L2 sounds and intonation can be attributed to it.

For one thing, there are various extralinguistic factors which affect L2 acquisition that were investigated to quite some degree by different studies. Jilka (2009) classifies these factors into developmental or neuropsychological issues (such as the age of learning), sociopsychological factors (like motivation, and the attitude towards the L2), and practice and experience (the use of L1 and L2). In her elaborations on LILt, Mennen (2015) also considers these aspects and agrees that age of arrival or age of learning is a crucial factor in predicting overall success in acquiring L2 intonation (based on Mennen's research on Dutch learners of Greek 2004, and the work by others). The earlier the exposure takes place, the better intonation acquisition will be. However, it appears that age of arrival

impacts different aspects of intonation to varying degrees, and the effects may not be the same for all of the dimensions of LILt. With respect to practice and experience, production of L2 intonation parameters will approximate L2 norms more closely as learners gain experience in the L2. As with L2 segments, learners will rely on their L1 in the production of L2 intonation when they have limited experience with the L2, so that at an early stage more transfer can be expected. Over time, learners will improve at least in some of the dimensions.<sup>33</sup>

There is no doubt that extralinguistic factors are crucial in both segmental and suprasegmental learning. Trofimovich and Baker (2006), for example, found that the age at the time of L2 learning significantly affects speakers' performance on the suprasegmentals speech rate, pause frequency and pause duration.

Coming back to linguistic factors, there is some evidence for L2 universals that go beyond L1 transfer. With respect to L2 segmental acquisition, Bohn (1995) for example, formulated the "Desensitization Hypothesis" to account for L1 independent strategies by Mandarin, Spanish and German speakers to differentiate non-native vowel contrasts for which there is only one category in the vowel space of their L1. The hypothesis states that L2 learners use a language-independent strategy whenever they are "linguistically desensitized" to spectral contrasts (*ibid.* p. 294ff.). He found that the use of duration cues is a strategy all language learners - irrespective of their L1 - revert to whenever spectral differences do not provide sufficient information.

Similarly, in terms of acquiring L2 intonation, studies often find the same kind of errors by L2 speakers with different L1s, suggesting universal patterns in the acquisition of L2 prosody (*cf.* Rasier and Hiligsmann, 2007). For example, Grosser (1993) and Wieden (1993) found an overuse of pitch accents in an early learning stage and overgeneralisation of accentuation rules in a later learning stage. According to Rasier and Hiligsmann (2007) this suggests that "marked" and "unmarked" accentuation rules are learned differently.<sup>34</sup>

Another universal in second language acquisition found its formulation in Eckman's Markedness differential hypothesis (Eckman, 1977) which states that marked structures are more difficult to acquire than unmarked ones. In his definition of typological markedness, he states that "[a] phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B; but the presence of B does not imply the presence of A." (1977:320). With his hypothesis, he argues that similarity

<sup>33</sup>Again, LILt hypothesises that not all intonation dimensions constitute the same amount of difficulty in L2 learning.

<sup>34</sup>Of course, universals can also be found in the acquisition of other linguistic levels. Several studies found, for example, that English learners with a wide variety of L1s seem to acquire function morphemes in similar orders (*cf.* Goldschneider and DeKeyser, 2005, and the references therein).

between languages is not enough to predict the success of attainment but that the relative degree of difficulty, which corresponds to the notion “typologically marked”, needs to be taken into account. Applied to the acquisition of intonation, different markedness scales have been suggested in order to derive predictions concerning learners’ difficulties (Rasier and Hiligsmann, 2007) or transfer of structures in language contact (Zerbian, 2015). With respect to the acquisition of rhythmic properties, van Maastricht et al. (2019), for example, found that Spanish learners of Dutch have more difficulties in attaining Dutch speech rhythm (measured in accentual and final lengthening) because speech rhythm, as well as its correlates syllable structure and lengthening effects, are more marked in Dutch (being a “stress-timed” language) than in Spanish (a “syllable-timed” language). Dutch learners of Spanish, on the other hand, were more successful in attaining the overall rhythm pattern of Spanish L1 speakers. The authors therefore argue, in line with Eckman (1977), that the directionality of acquisition is essential and needs to be implemented in models of L2 acquisition.

Other universal characteristics of L2 production described are, for example, that L2 learners make use of more pauses in their L2 compared to their L1 (cf. Trofimovich and Baker, 2006) and that they make more pronunciation errors in free speech tasks than in formal word reading tasks (Wu and Saito, 2014 as cited in Saito 2018). This is explained by the increased demands on linguistic processing since in free speech, such as, for example, in a picture description task, there is less planning time compared to a more controlled reading task. The task in the experiments in this thesis was formal reading with additional preparation time.

### 2.3.4. Experimental investigations of L2 prosody

Given the diverse use of language pairs, methodologies, and phenomena that studies on L2 prosody adopted, it is quite difficult to compare the results. In the following I will mention findings that are particularly interesting for the study at hand, without claiming that the overview is complete. I may refer to, for example, Gut (2009) for a more complete picture on L2 prosody investigations.

Since this thesis is about alternation preferences and focus marking, I will focus on studies that have investigated L2 production, with findings on prosody used for information structure on the one hand, and on rhythmic aspects on the other hand.

**Pitch accent placement** Concerning the acquisition of pitch accenting in general, several studies have found that learners typically overuse pitch accents (cf. Rasier and

Hilgsmann, 2007). This has been documented for learners especially in early stages of learning (Grosser, 1993; Wieden, 1993), but also in more advanced stages of the learning process (Hilgsmann and Rasier, 2002; Rasier, 2003; Juffs, 1990). The tendency to realise more pitch accents than L1 speakers was observed for a number of different learners in various L2s (cf. Rasier and Hilgsmann, 2007, Gut, 2009, and the references therein). An explanation could be that learners have difficulty recognising new and given information and therefore fail to mark the respective categories accordingly.<sup>35</sup> A number of studies with various language pairs in fact showed that L2 speakers often fail to deaccent given information (e.g. Ramirez Verdugo 2002; Nava and Zubizarreta 2009 for Spanish learners of English; Wennerstrom 1994, for Spanish, Japanese and Thai L2 speakers of English; Gut and Pillai 2014 for Malaysian speakers of English; and van Maastricht et al. 2015 for Spanish learners of Dutch), suggesting universal tendencies.<sup>36</sup>

Others observed that L2 speakers realise the same number of pitch accents as native speakers, but on different elements (e.g. Jilka, 2000; Gut, 2009). In a corpus comprising data from English L2 speakers with various L1s (the LeaP corpus), Gut (2009) observed that L2 speakers may, for example, accentuate pronouns where L1 speakers would not, and also often place pitch accents on given and accessible information. Investigating German L2 speakers of English, O'Brien and Gut (2011) found for their advanced learners, however, that pitch accent placement was not a major problem - only two speakers (out of ten) deviated from their L1 productions in that they placed no accent on new information and only two speakers produced a pitch accent on given material. Van Maastricht et al. (2015) investigated the use of pitch accents in the marking of focus in Dutch learners of Spanish and Spanish learners of Dutch. While in Dutch focus is expressed by pitch accenting and given material is typically deaccented, in Spanish the nuclear pitch accent is always placed at the end of the intonation group, irrespective of the information status (*ibid*, p. 49). The authors also found deviation with respect to pitch accenting (especially by Dutch learners of Spanish) and boundary tones (especially by Spanish learners of Dutch) and attribute it to L1 transfer.

**Pitch accent type** Beside pitch accent placement, also deviations in pitch accent types have been observed. Several studies with different L1s and target languages mainly

<sup>35</sup>Related to pitch accent placement may be non-native phrasing - it has been found that learners tend to have shorter phrases (e.g. Gut, 2009), which in turn certainly contributes to the realisation of more pitch accents compared to native speech.

<sup>36</sup>Rasier and Hilgsmann (2007), however, cite Barlow (1998) as counterevidence to this hypothesis who found that pitch accenting was not a problem for the Spanish, Italian and Chinese learners of English in his study.

observed falls instead of rises (e.g. Willems, 1983; Gut, 2009; O'Brien and Gut, 2011; Grosser, 1997; Jilka, 2000) and some also vice versa (e.g. Willems, 1983, such as rises at the end of WH-questions). In his study on Dutch learners of English, Willems (1983) found that L2 speakers often used falls where L1 speakers used continuation rises but on the other hand sometimes used rises where L1 speakers realised falls. Gut (2009), for example, found that while L1 speakers mostly use rising pitch accents pre-nuclearly, for L2 speakers pre-nuclear accents are mostly realised as falls.<sup>37</sup> This has also been found by O'Brien and Gut (2011) who compared pre-nuclear pitch-accent realisation in broad, narrow and contrastive focus of L1 German to L2 English. This finding is also in line with Grosser's longitudinal study with (Austrian) German learners of English who found that learners after one year of learning produced mainly falling accents, while the proportion of rising accents increased after two years of learning (Grosser, 1997).<sup>38</sup> Similarly, Ramirez Verdugo (2002) found that Spanish L2 speakers of English use falling pitch accents on given material while English L1 speakers use a low rise if they accent it at all.

Several studies have investigated deviations in phonetic details of these phonological categories, especially with respect to peak alignment (e.g. Atterer and Ladd, 2004; Trofimovich and Baker, 2006; O'Brien and Gut, 2011; Chen and Fon, 2008), or pitch range (e.g. Ramirez Verdugo, 2002; Jilka, 2000; Wennerstrom, 1994; Gut, 2009; Mennen, 1998), generally finding more narrow pitch range in non-native speech.<sup>39</sup>

**Rhythmic aspects** While rhythmic aspects have been subject to various L2 studies, alternation preferences in a second language have, to the best of my knowledge, not been investigated yet. Studies have often looked at language pairs of different rhythmic classes and investigated the respective L2 speech with respect to rhythmic properties. Different measurement techniques are used to measure such rhythmic properties, mostly based on interval measures such as, for example, the standard deviation of vocalic or consonantal intervals (cf. Ramus et al., 1999, 2003)<sup>40</sup>, pairwise variability indices such as, for example, the normalized Pairwise Variability Index of vowels (nPVI-V which is PVI

---

<sup>37</sup>She also points out that the speaking style has an effect on both the phrasing - with longer intonation phrases in read speech - and the boundary tones used.

<sup>38</sup>Although it has to be noted that school children at the very beginning of second language acquisition, as investigated by Grosser, may not be comparable to the advanced learners in many studies and also the one at hand.

<sup>39</sup>Zimmerer et al. (2015), however, did not find a reduced pitch range in productions of French and German speakers speaking German and French respectively.

<sup>40</sup>In order to reduce the influence of speech rate on interval measures, rate-normalised measures were suggested, for example by Dellwo (2006), see White and Mattys (2007:504).

normalised for speech rate) (cf. e.g. Low et al., 2000; Grabe and Low, 2002)<sup>41</sup>, and the percentage of vocalic material in a sentence (%V) (Ramus et al., 1999)(see White and Mattys (2007) for an overview and evaluation of these metrics). Other measurements do not consider distances or proportions between vocalic or consonantal material as indicators for the location of languages on a rhythm scale but the proportion between voiced and voiceless material (Dellwo et al., 2007), the duration ratio of stressed and unstressed syllables (Trofimovich and Baker, 2006), or the degree of vowel reduction as an indicator for rhythm acquisition, with stress-timed languages typically reducing more than syllable-timed languages (Gut, 2003b; Barry, 2012a).

The assumption behind these measurements is that speech rhythm arises from the phonotactic characteristics of a language. Stress-timed languages, for example, typically have more vowel reduction (leading to a smaller proportion of vocalic speech) and more complex consonant clusters (resulting in more variable consonantal interval durations) than languages that are typically classified as syllable-timed languages (cf. Dellwo et al., 2007; Wagner, 2008; Schiering, 2007).

Regardless of the measurement technique, what these studies mostly found is that the rhythmic metric values for L2 speech were somewhat in between the values for the L1 and the target language (e.g. Carter, 2005; Low et al., 2000; White and Mattys, 2007; Gut, 2003a).<sup>42</sup>

Aside from studies comparing learners of languages of a different rhythm class than their L1, learners with rhythmically similar L1s, such as in the study at hand, have also been investigated. Gut (2003), for example, looked at the degree of vowel reduction as an indicator for rhythmic L1 influence in, on the one hand, English learners of German (who reduced too much) and, on the other hand, Romance learners of German languages (who reduced too little). Ordin et al. (2011) investigated German learners of English with different levels of proficiency. Applying various rhythm metrics, they found Varco coefficients (normalised intervall measures) and nPVI-V increase with proficiency level suggesting that the speech of advanced learners is more stress-timed than the speech of lower-intermediate and intermediate learners. Given their results, they argue that rhythm acquisition develops from being more syllable-timed to being more stress-timed, even when the learners' L1 is also a stress-timed language (see also White and Mattys 2007, Ordin and Polyanskaya 2015, and Li and Post 2014 on the development of the

<sup>41</sup>Stress-timed languages, like English, are expected to have a high nPVI-V while syllable-timed language will have low nPVI-V values.

<sup>42</sup>But see White and Mattys (2007) p. 514. for an exceptional finding with respect to English learners of Spanish who had a higher %V than the native Spanish group.

acquisition of rhythmic properties at different levels of proficiency).

Studies applying rhythm metrics are only marginally relevant for the research question at hand since these measurements do not directly capture a certain preference speakers may have for rhythmic well-formedness such as the preference for alternation. Although there is no doubt that alternation preferences and rhythmic properties are interconnected since processes like vowel reduction are a consequence of reducing the interstress intervals in the strive for isochrony found in stress-timed languages (cf. Dauer, 1983; Wagner, 2008; Barry, 2012b). What we can also draw from these investigations is that German and English have very similar values in most of these measurements (e.g. Grabe and Low (2002), see also overview in Ordin and Polyanskaya (2015), p. 4).

**To sum up** Considering the findings reported above, we can conclude that deviations in pitch accent placements, especially in the shape of pitch accents on given information, as well as deviations in pitch accent type have been observed. For the study at hand this means that German learners may accentuate given information and/or may use inappropriate accent types. It has to be kept in mind, however, that the level of proficiency is essential (see above), so that the highly proficient learners in the study at hand may not be as prone to these deviations as less proficient learners. Concerning the role of alternation preferences in L2 speech, previous research does not give us any insights. This thesis will make a contribution to filling this gap. Considering the function of rhythmic alternation in terms of creating expectations, as described above, L2 learners may benefit less from this function, and consequently may be less prone to rhythmic adjustments, given that non-native speakers have a reduced ability to create expectations (cf. Grüter et al., 2014). More detailed predictions concerning the production of German learners of English are going to be discussed in chapter 5.

## 2.4. Summary of research questions

This background chapter set out to show that

- both German and English typically mark information structure by means of pitch accent placement and pitch accent type (see 2.1.3), but we often find variation in both the placement of pitch accents and in the choice of pitch accent type (see 2.1.4).
- both languages share the preference for an alternation of strong and weak beats and try to avoid stress clashes (see 2.2), and this has been shown to affect sen-



tence production, perception and processing (see 2.2.3). Additionally, speakers may prefer an alternation of  $F_0$  peaks and valleys (see 2.2.3.2).

- studies on L2 prosody often found deviations in both pitch accent placement and pitch accent type; the role of alternation preferences in an L2, however, is unclear.

The overall research questions derived from these starting points are

- How does the preference for rhythmic alternation interact with the prosodic marking of focus?
- Can alternation preferences account for variation in the prosodic marking of focus?
- What is the status of alternation preferences in a second language?

In order to address these questions, five experiments have been conducted which will be presented in the following chapters. They are designed to address the following sub-questions:

1. How do alternation preferences affect the realisation of corrective focus in German? (3)
2. How do alternation preferences affect the realisation of corrective focus in English? (4.2)
3. What is the status of the rhythm rule under corrective focus marking in English? (4.3)
4. Are alternation preferences transferred from L1 German to L2 English? (5.3)
5. Are rhythmic repair strategies acquired in the context of corrective focus? (5.4)



### 3. Alternation preferences and focus marking in German

This chapter presents a production study that was conducted in order to investigate the interplay of alternation preferences and the prosodic marking of information structure. The corpus study by Schauffler et al. (2015a) presented in the previous chapter (2.2.3.2) gave us some indications that there are prosody-inherent factors at play in the intonation of an utterance, that is that pitch accents may indeed influence each other.

However, pitch accents are not an end unto themselves, but accomplish crucial functions in German, such as the marking of information-structure categories by means of pitch accent placement and pitch accent type. Corrective focus, the information-structure category under investigation in this chapter, for example, is associated with a falling pitch accent with a high peak and a long duration, as outlined in section 2.1.3.

While the prosodic marking of focus is a strong constraint, it is not always straightforward and we do find variation in both pitch accent placement and pitch accent type (see section 2.1.4). There is evidence that at least part of the variation in pitch accent **placement**, that is whether or not a pitch accent is realised, or its specific position in a sentence, can be explained considering rhythmic preferences. As established in section 2.2, German and English speakers prefer an alternation of strong and weak, alias stressed and unstressed syllables, and they appear to use deaccentuation in order to avoid two accented or two stressed syllables directly following each other (e.g. Riestler and Piontek, 2015; Kentner, 2012a; Bohn et al., 2011).

Regarding variation in the use of pitch accent types, it has also been found that information-structure categories cannot be mapped to invariant pitch accent types (e.g. Grice et al., 2017; Féry, 2007; Baumann, 2006; Féry and Kügler, 2008; Schweitzer et al., 2009). Other aspects must play a role in assigning a phonological category, such as the position of the accented constituent in the sentence (Féry, 2007).

Considering the variation in the mapping of form and function on the one hand, and the above-mentioned findings with respect to alternation preferences on the other hand (both in terms of pitch accent placement and pitch accent type), this chapter sets out to

address the overall research question whether alternation preferences interact with the prosodic marking of corrective focus.

The framework of a controlled production experiment makes it possible to investigate whether alternation preferences have the power to interact or even override the strong focus-marking requirements in German. In order to elicit adjacent pitch accents, question-answer sentences were constructed involving double-focus sentences. While sentences involving one focus have been extensively studied in German (cf. section 2.1), the prosodic realisation of sentences involving multiple foci has not been much under investigation. Since it cannot be assumed that double focus is a mere concatenation of two single foci, the following subsection will give an overview of double focus in German.

### 3.1. The intonation of double-focus in German

The advantage of experimental stimuli, namely being able to control various variables (sentence length, segmental make-up, frequency etc.) and extract the effect of the variable in question, are ipso facto the disadvantage of experimental set-ups in language research: Experimental stimuli often suffer from being all too constructed and not exactly representative for daily-life communication. While a sentence involving two foci may not be the most prototypical utterance, it is not too far from reality that situations may come up requiring two foci in one sentence (at least often enough to deserve a more thorough reflection in research). Consider a question like *What did Tobi give to whom?* The answer sentence involves two foci (therefore double-focus, sometimes also referred to as dual-focus) giving answers to the two questions *what* and *to whom*, like for example - *He gave [the piano]<sub>Focus1</sub> to [teachers]<sub>Focus2</sub>* (square brackets indicate the size of the focus). In terms of the intonation of double-focus sentences where two constituents are corrected as in the experiment at hand, Büring (1997) predicts two falling pitch accents on both foci (H\*L). While there is a great body of research on German single focus, there are, to the best of my knowledge, only two studies that experimentally investigated double-focus in German. The first one is by Schauffler et al. (2015b) investigating pitch accent placement. This study served as the pilot study to the experiment at hand. The second one is by Wang and Féry (2018) investigating pitch accent type and pitch accent relation.

**Wang and Féry (2018)** looked at double-focus sentences in German and compared them to sentences involving only initial focus, only final focus, and all-new sentences. They manipulated the length of the focus constituents (short NP + short VP; short

NP + long VP; long NP + short VP; long NP + long VP) in order to see whether sentence length influences pitch accent type and phrasing. See (24) for an example of the short-short version answering the question *Who praises whom?*, small capital letters indicate stressed syllables in the focused words.

- (24) [Der LEHrer] lobt [MALI]  
           the       teacher praises Mali  
       ‘The teacher praises Mal’

All sentences were read by six female native speakers of Standard German. They found that there are three ways in which the first focus in double-focus sentences (they call them dual-focus) is prosodically realised in German, correlating with sentence length: Short sentences were preferably realised with a rising pitch accent on the first focus and a falling pitch accent on the second (“hat-pattern”). Longer sentences were mostly realised with a “two-peak pattern”, that is with two falling pitch accents on each of the foci, or with a “two-phrase pattern” which is the third option, where both foci were realised with falling accents but each in its own intonation phrase with a high boundary tone between them. Additionally, they found that double-focus sentences differ from initial focus sentences in that there is no post-focal compression after the first focus, that is  $F_0$  is not lowered after the first focus as it is the case in single-focus sentences. According to the authors, the two pitch accents are equally prominent in the two-peak pattern and in the two-phrase pattern.<sup>1</sup>

**Schauffler et al. (2015b)** looked at double-focus answer sentences involving two **corrective** foci, that is concrete alternatives for each of the focus constituents were presented in the context question and corrected by the speakers in the answer sentence. They manipulated the rhythmic environment in the same way as in the experiment at hand, by eliciting focus accents on directly adjacent syllables in the rhythmically dispreferred condition (*clash*) and separated by one syllable in the rhythmically regular answer sentence (*no clash*). In addition to the rhythmic environment, they manipulated the syntactic construction. One set of stimuli contained verb phrases consisting of an argument and a verb, such as in example (25). The rhythmically dispreferred condition

<sup>1</sup>With their analysis, Wang & Féry suggest that two nuclear pitch accents in one intonation phrase are possible in German in the context of double-focus sentences. With this claim they deviate from mainstream prosodic theory which only allows one nuclear head per prosodic domain, that is in this case only one nuclear pitch accent in one intonation phrase (Nespor and Vogel, 1986; Selkirk, 1995; Truckenbrodt, 1995). See also Ishihara (2011) for an analysis that allows non-culminative structures, that is phrases with more than one prominence.

is given in *clash*, and the rhythmically preferred condition in *no clash*. Small capitals indicate the stressed syllable of the focus word.

(25) Hat Konrad gesagt, dass Heinz in Bonn den Pfleger gefeuert hat?

‘Did Konrad say that Heinz in Bonn has fired the nurse?’

Nein, er hat gesagt,...

‘No he said...’

**clash:** ... dass Heinz in Bonn den [ARZT]<sub>F</sub> [ANGestellt]<sub>F</sub> hat.

that Heinz in Bonn the doctor hired has

‘... that H. in Bonn has hired the doctor.’

**no clash:** ... dass Heinz in Bonn den [ARZT]<sub>F</sub> [verHAftet]<sub>F</sub> hat.

that H. in Bonn the doctor arrested has

‘... that Heinz in Bonn has arrested the doctor.’

The other set of stimuli used two adjacent correctively focused arguments, such as in (26), the rhythmic manipulation is again given in *clash* eliciting two directly adjacent focus accents and in *no clash* where potential pitch accents are separated by one syllable according to alternation preferences.

(26) Hat Tom gesagt, dass Frank das Essen Pflegerinnen gegeben hat?

‘Did Tom say that Frank has given the food to nurses?’

Nein, er hat gesagt,...

‘No, he said...’

**clash:** ... dass Frank das [GESCHENK]<sub>F</sub> [Malerinnen]<sub>F</sub> gegeben hat.

that Frank the present painters given has

‘...that Frank has given the present to painters.’

**no clash:** ... dass Frank das [GESCHENK]<sub>F</sub> [Masseurinnen]<sub>F</sub> gegeben hat.

that Frank the present masseuses given has

‘... that Frank has given the present to masseuses.’

Regarding the manipulation of the rhythmic environment, their results showed that pitch accents on both focused constituents were more often produced when there was an unaccented intervening syllable between the pitch accented syllables of the respective constituents (condition *no clash*). In the *clash*-condition, speakers more often produced only the second focus accent. The difference, however, did not reach significance

( $p=0.07$ ), so that further testing is needed in order to find out whether the placement of focus accents is affected by the rhythmic environment.

Regarding the syntactic relation between the two foci, their results showed that when the argument and its adjacent predicate were both contrastively focused, German speakers produced only one contrastive accent on the argument and left the focused verb unaccented in the majority of cases. By contrast, in sentences in which two adjacent arguments were both contrastively focused, German speakers produced two accents significantly more often. The authors interpret the lack of accents on the verb as the generally limited accentability of predicates compared to arguments (cf. Ladd, 2008; Schmerling, 1976).

Due to this result, stimuli involving two correctively focused arguments were chosen for the experiment at hand in order to not have confounds with other factors such as syntactic category (since predicates seem to generally behave differently).

## 3.2. Research questions and hypotheses

The experiment aims at answering the following questions:

**1. Do alternation preferences affect pitch accent placement?** More specifically, do speakers realise two pitch accents in double-focus sentences with rhythmically regular and irregular environments, or are information-structurally required pitch accents omitted in rhythmic clash contexts? From a focus-marking point of view, we would expect that speakers realise each corrective focus with a pitch accent. If speakers are driven by a preference for an alternation of strong and weak they may not realise a pitch accent when this would lead to a pitch accent clash, similar to the findings by Kentner (2012a) or Riester and Piontek (2015).

**2. Do alternation preferences affect speakers' choice of pitch accent type?** More specifically, if speakers realise both focus accents in both, regular and irregular contexts, what pitch accent types are used? And is the choice of pitch accent affected by the rhythmic context? From a focus-marking perspective, we would expect to see two emphatic falling pitch accents with a high tonal target on the accented syllable, as the typically associated pitch accent with corrective focus (see 2.1.3). According to Wang & Féry, speakers have, however, various choices in the prosodic realisation of the first focus in double-focus sentences. As already mentioned, they manipulated sentence length and found that it affects the choice of realisation: short sentences were preferably realised

with a hat-pattern, that is with a rising pitch accent on the first focus, while longer sentences were realised with either a two-peak pattern or a two-phrase pattern. The experiment at hand investigates the effects of rhythmic context on speakers' choice of contour while keeping sentence length constant. Considering the findings of the corpus experiment presented in the previous chapter that speakers are more likely to alternate pitch accents in type when they are close together, I hypothesise that speakers prefer an alternating pattern when two directly-adjacent pitch accents are realised.

#### **3. Does the rhythmic environment affect the relation between pitch accents?**

Wang and Féry (2018) found that double-focus sentences in German can be realised with two equivalent pitch accents or with a weaker and a stronger one. The relation between the two pitch accents was also affected by sentence length. They found that the first focus in double-focus sentences may be realised as a weaker pre-nuclear accent in short sentences. In longer sentences they found in fact two accents equivalent in prominence. It is, however, unclear on what grounds they arrived at the conclusion that pitch accents were either equally prominent, or asymmetrical, whether this was derived from the authors' impression or based on acoustic measurements. Since prominence is a perceptual notion I assume that this statement is based on the authors' impression. In the experiment at hand, annotator judgements are used to investigate whether two equivalent pitch accents are more likely in rhythmically preferred sentences, and whether the relation between the pitch accents is asymmetrical in clash environments, that is, whether the first focus accent is reduced in prominence when it is directly adjacent to a second corrective focus accent.

#### **4. Does the rhythmic environment affect the placement of prosodic boundaries?**

Closely related to the aspect of the pitch accents' relation is the question of whether the rhythmic environment affects prosodic phrasing. More specifically, the question is: do speakers realise (a) no boundary, (b) an intermediate phrase boundary or (c) an intonation phrase boundary more often in clash environments? The effect of distance on the alternation of pitch accent types found by Schauffler et al. (2015a), was only present within intermediate phrases. It is conceivable that speakers realise a phrase boundary in order to resolve a pitch accent clash. We would then expect that speakers use a phrase boundary more often in a clash environment. This would be in line with optional repair strategies discussed by metrical accounts in the context of two adjacent lexically stressed syllables (stress clash) such as lengthening of the stressed syllable so that a beat is added,



and pause insertion (Lieberman and Prince, 1977; Selkirk, 1984).<sup>2</sup>

After some general remarks with respect to participants, material, procedure and statistical analyses in 3.3, the following sections will address these research questions in the order in which they were just presented. 3.4 is concerned with the question of pitch accent placement and presents the results of the first annotation of prominence by three annotators. The results of the annotation answering the question on possible effects on the choice of pitch accent type is presented in 3.5. In order to answer the sub-question about the relation between potential pitch accents in terms of their prominence, a second round of annotations was done, which will be presented in section 3.6. Results on phrasing will be presented in 3.7. All results will be summarised and discussed in 3.8.

### 3.3. The experiment

The data for the current study was elicited via a reading production experiment.

#### 3.3.1. Participants

Sixteen (5 men, 11 women) German native speakers, recruited at the university of Stuttgart, participated in the experiment. Their mean age was 27.25 years (range: 19 to 33) and none of them had known speech or reading disorders. All participants were naïve as to the purpose of the experiment. They were paid for their participation.

#### 3.3.2. Stimulus design

The stimuli were constructed as question-answer pairs that were designed in such a way that two noun phrases (NPs) introduced in the question needed to be corrected in the answer. To investigate influences of the rhythmic environment on the prosodic

---

<sup>2</sup>Wang and Féry (2018) found in their experiment that double-focus sentences were either realised in two prosodic phrases and one intonation phrase, or in two intonation phrases. They say that ‘each focus of a dual-focus sentence is phrased in a separate  $\phi$ -phrase (all realizations), and in the two-phrase pattern, each focus is in a separate  $\iota$ -phrase.’ When interpreting their results, it is however important to note that what they refer to as prosodic phrases ( $\phi$ -phrases) does not correspond to intermediate phrases as intended by the autosegmental-metrical approach. It is rather a theoretical domain for pitch accent allocation that is derived from syntax. All phrase boundaries derived from the actual data are referred to as intonation phrase boundaries marked by a high boundary tone. They do not report on any empirical evidence regarding intermediate phrase boundaries. Considering the fact that they report that the second focus was mostly realised as a downstepping movement to the first focus we can imply that their speakers mostly did not realise an intermediate phrase break since the intermediate phrase is the domain for downstep (Ladd, 1983; Beckman and Pierrehumbert, 1986; Kügler and Féry, 2017). At least in the longer sentences this seems, however, unlikely.

### 3.3. The experiment

---

realisation of the double-focus constructions, two conditions were designed: one eliciting pitch accents on successive syllables, which therefore directly follow each other (**clash** condition) and one condition in which the potential pitch accents are separated by an unaccented syllable (**no clash** condition). Examples are given in (27), lexically stressed syllables are underlined, small capital letters indicate focus expected to be marked by pitch accent.

(27) Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?

*Did Melli say that Tobi has given the drums to pupils?*

Nein, sie hat gesagt, dass Tobi...

*No, she said that Tobi...*

**clash:** ... das [KlaVIER]<sub>NP1</sub> [LEHrerinnen]<sub>NP2</sub> gegeben hat.

the piano teachers given has

*... has given the piano to teachers.*

**no clash:** ... das [KlaVIER]<sub>NP1</sub> [StuDENTinnen]<sub>NP2</sub> gegeben hat.

the piano students given has

*... has given the piano to students.*

Two control conditions were added to each sentence type, in order to see the tonal realisations of single-focus constructions under a corrective context. The control conditions also allow us to get an idea whether the participants generally processed the context question and understood the task. In condition *F1* only the first NP (NP1) was corrected while in condition *F2* only the second NP (NP2) was corrected (see examples in (28); *F1*=focus on NP1; *F2*=focus on NP2 ).

(28) Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?

Did Melli say that Tobi has given the drums to pupils?

Nein, sie hat gesagt, dass Tobi...

*No, she said that Tobi...*

**F1:** ... das [KlaVIER]<sub>NP1</sub> [Schülerinnen]<sub>NP2</sub> gegeben hat.

the piano pupils given has

*... has given the piano to pupils.*

**F2:** ... das [Schlagzeug]<sub>NP1</sub> [LEHrerinnen]<sub>NP2</sub> gegeben hat.

the drums students given has

... *has given the drums to teachers.*

Since phrase length matters in the distribution and possibly choice of accents (cf. Wang and Féry, 2018; Ladd, 2008), the stimuli were controlled for number of syllables (8 words and 13 syllables starting from the embedded clause). In the double-focus conditions, the first focused word (in NP1) was always a disyllabic iamb, therefore carrying the lexical stress on the final syllable. NP2 always had four syllables with lexical stress on the initial syllable in the *clash* condition and lexical stress on the second syllable in the *no-clash* condition.

In order to avoid a segmental influence on the tonal marking, which would be especially expected for stops, there were only continuants in the coda of NP1 and no voiceless stops in the onset of NP2.

Additionally, all “contrasting pairs” (the group of NP1s and the group of NP2s) were controlled for word form frequency which were taken from the Leipzig Wortschatz corpus (Quasthoff and Richter, 2005). Mean frequency of the contrasted words in the context question and the focused words in the target sentence, did not differ. The mean frequency of the first NP in the context question was 11.90 in the Leipzig Wortschatz corpus (i.e. “der”, which is the most frequent word in that corpus, is  $2^{11.90}$  more frequent than the first NP in the context question); the mean frequency of the first NP in the target sentence (which is contrasted to the first NP in the context question) was 12.35. A two sample t-test was conducted and revealed that this is not significantly different ( $t = -1.05$ ,  $df = 77.83$ ,  $p\text{-value} = 0.30$ ). The mean frequency of the second NP in the context question was 15.35, while the mean frequency value of the correcting second NP in the target sentence was 15.98 ( $t = -1.07$ ,  $df = 77.19$ ,  $p\text{-value} = 0.29$ ). The mean frequency of the target word was 15.90 in condition *clash* and 16.05 in condition *no clash*, the difference was also not significant ( $t = -0.19$ ,  $df=37.19$ ,  $p=0.85$ ). In summary, all words in position NP1 are equally frequent and all words in position NP2 are equally frequent.

To control for frequency is important because prominence perception is not only signal-driven, but also expectation-driven, which means that the less predictable a word the more prominent it is, as found by Cole and colleagues (Cole, Mo, and Hasegawa-Johnson 2010, cited in Baumann and Winter 2018). They operationalised unpredictability in terms of word frequency and discourse givenness and found that less frequent words are more prominent (they are also longer in duration).

#### 3.3.3. Procedure

Twenty sentences per condition (*clash*, *no clash*, *F1*, *F2*) were distributed over 4 lists using a Latin Square Design so that each participant read only one answer per context question. The experimental sentences in each list were pseudo-randomised for each participant so that the first 3 mini-dialogues were fillers and that sentences of the same condition were not successive. One list contained 20 experimental sentences and 40 filler sentences. The filler sentences did not involve double-focus constructions, but were dialogues of varying type, some involving one corrective focus, others no contrastive focus at all. See (29) for an example of a filler sentence, and the appendix for the whole collection.

- (29) Hat Sarah gesagt, dass Thomas den Kuchen gebacken hat?  
*Did Sarah say that Thomas has baked the cake?*

Nein, sie hat gesagt, dass Thomas den Kuchen gekauft hat.  
No, she has said that Thomas the cake bought has  
*No, she said that Thomas has bought the cake.*

The context questions of each question-answer pair had been previously recorded spoken by a female German native speaker who was instructed to read the questions neutral as well as natural. The recordings took place in a sound attenuated chamber. The mini-dialogues (both question and answer) were presented on a screen, preceded by instructions and a context story that was designed to make the question-answer pairs more plausible. The story introduced an aunt who had the whole very big family over. The family talked all day about private things and about the family business. Since the aunt's hearing as well as her memory was not the best any more she asked her grandchild in the evening what has been discussed all day. The aunt's question was the pre-recorded context-question and the participants were asked to click on a symbol which triggered playing of the question on loudspeakers.

In the pilot study (Schauffler et al., 2015b), one group of participants were instructed to first silently read both question and answer, and then produce it. The other half of the participants were instructed to read out the dialogue as soon as it appeared on the screen. Schauffler et al. did not find an effect of preparation in their experiment. Other studies, however, have found an influence of reading modality on accent distribution (Kentner, 2012a; Tilsen, 2012). In particular, effects of rhythmic repair strategies were predominantly found in prepared speech (Tilsen, 2012). Therefore, participants were instructed to first silently read the dialogue, listen to the context question and then

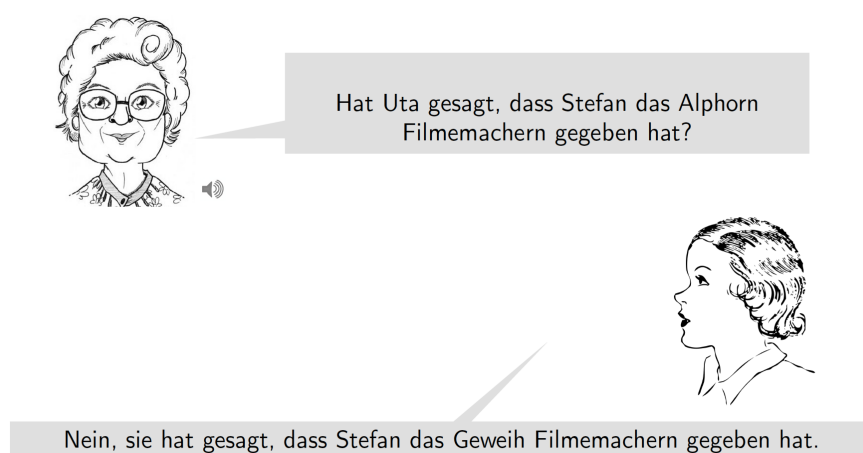


Figure 3.1.: Screenshot of the stimuli presentation. In this example condition *F1* is presented. Clicking on the loudspeaker played the recorded context question.

produce the answer. This means they were prepared for their answer sentence. This is also an attempt to get closer to spontaneous speech where speakers usually also know what they will utter. The speakers controlled the appearance of each new dialogue themselves by pressing a key on a keyboard. They were instructed to repeat their productions in case of misreadings. The instructions were also given verbally.

This resulted in 320 produced answers. Figure 3.1 gives an example of how the stimuli were presented. For male participants the presentation was adapted with the picture of a male grandchild.

### 3.3.4. General remarks on the statistical analyses of this chapter

Three recordings were excluded from the statistical analysis because they either included major hesitations or repetitions (2), or they were realised with the wrong stress pattern (*‘Beton* instead of *Be‘ton*), so that 317 recordings were statistically analysed.

**R packages used** All statistical analyses, including the statistical analyses in the following sections, were performed in R 3.5.0 (R Core Team, 2018), using the function `glmer` and `lmer` with the packages `lme4` (Bates et al., 2013). For data processing, and visual-

isation the R packages 'tidyverse' (Wickham, 2017), 'plyr' (Wickham, 2011), 'ggplot2' (Wickham, 2009), and 'jtools' (Long, 2018) were used.

I used an alpha level of 0.05 for all statistical tests.

## 3.4. Analysis of pitch accent placement

The collected recordings were first analysed with respect to pitch accent placement in order to answer the first research question, namely do alternation preferences affect pitch accent placement? More specifically, are both focus accents realised in a clash environment, that is when the pitch accent carrying syllables are successive, or are information-structurally required pitch accents omitted when their realisation would lead to a clash?

### 3.4.1. Procedure

If a word or syllable is perceived as prominent it stands out in relation to its local context. The acoustic correlates of prominence have been identified in various studies (see Wagner 2008, p.78ff for an overview), the most important among them being fundamental frequency ( $F_0$ ), duration and intensity, at least for German. However, it is still investigated which linguistic variables have the strongest impact on perceiving prominence, and in which way they interact (Baumann and Winter, 2018). Different acoustic cues may also have different effects on different listeners. I therefore decided to use listeners' judgements as the most representative marking for the perceptual entity "prominence", assuming that this is the most natural way to decide which word is prominent. A decision for specific cues which may not be representative for listeners' perception is then avoided. This is why all analyses are based on the judgements of annotators that are trained in annotating intonation.

All recordings were played in random order to three annotators using the Demowindow from praat (Boersma, 2001). All annotators have German as their L1. Two annotators were prosodically trained, but unaware as to the purpose of the experiment or the nature of the different conditions. One annotator was the author. Since prominence judgements are also highly biased by the listeners' expectation (e.g. Wagner, 2005), the answer sentences were played without the respective context question in order to gain a semantically (and theoretically) unbiased judgement. The annotators were asked to click on the word or words which sounded prominent to them from the determiner onwards (see figure 3.2). Each recording was played three times in the course of the annotation

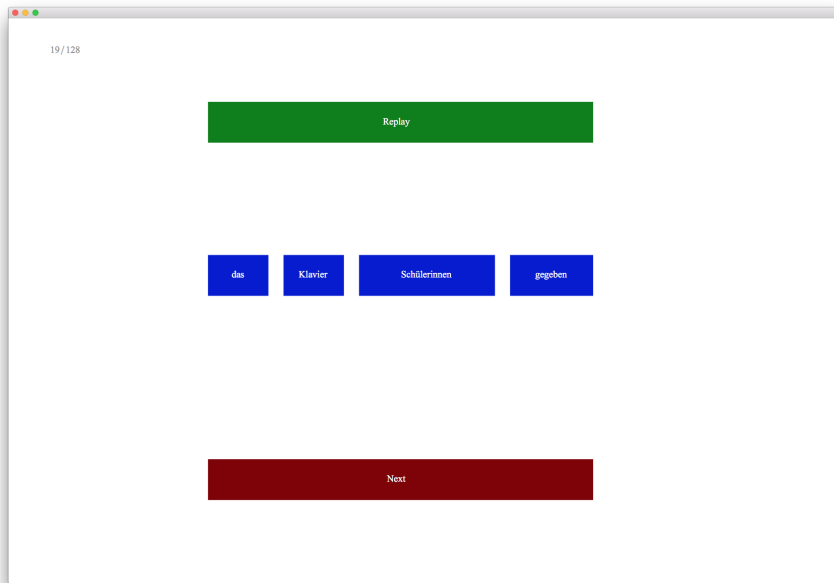


Figure 3.2.: Screenshot of the annotation task with praat's demo window

experiment. This resulted in a total of 957 recordings that were annotated (one of the three recordings which were removed from the statistical analysis (see 3.3.4) was already excluded at this point due to heavy hesitations). Since it is a time consuming and tiring task and there should be breaks in between, the annotation experiment was divided into eight blocks and could therefore be easily interrupted. It started with a test block so that the annotators could get acquainted with the task. They were able to replay each recording up to ten times before a decision was required.

**Determining pitch accents** I considered a constituent to be prominent when it was perceived as such at least 4 times. I chose this threshold because it means that more than one annotator perceived the prominence: either one annotator was so sure as to the presence of a prominence that they marked it three times and at least one other annotator perceived the prominence at least one time (see table 3.1, scenario 1) or two annotators perceived the prominence at least twice (scenario 2), or all three annotators perceived the prominence at least once (scenario 3). Given the inter- and intra-annotator agreement (see below) this threshold is rather theoretical, since there was substantial agreement between the annotators. In fact, there are only 12 (out of 317) cases that have indeed 4 out of 9 prominent judgements.<sup>3</sup>

<sup>3</sup>While this threshold is justified to the best of my beliefs, I can see that it is debatable. Therefore I did a second analysis keeping all differences in by creating a linear variable dividing the number of

### 3.4. Analysis of pitch accent placement

---

Table 3.1.: *Possible scenarios for how a 4-out-of-9-threshold for pitch accent judgement can come about. x stands for prominence. Each annotator judged the presence of prominences three times for each recording.*

	A1	A2	A3
scenario 1	x - x - x	x - 0 - 0	0 - 0 - 0
scenario 2	x - x - 0	x - x - 0	0 - 0 - 0
scenario 3	x - x - 0	x - 0 - 0	x - 0 - 0

**Inter-annotator and intra-annotator agreement** In order to see how similar the annotators behaved in judging prominences, both intra-annotator and inter-annotator agreement were calculated using the R package 'irr' (Gamer et al., 2012). For both calculations I applied Fleiss' Kappa to calculate the degree of agreement between multiple raters on a scale from 0 (chance agreement) to 1 (perfect agreement) (Fleiss, 1971). Kappa was calculated on the agreement on the prominence pattern (prominences on both objects, on NP1 or on NP2) given for each recording by one annotator compared to the other two annotations of the same recording by the same annotator (intra-annotator agreement) and to the annotations of this recording by the other annotators (inter-annotator agreement).

The intra-annotator agreements were  $\kappa=0.86$  for annotator one,  $\kappa=0.78$  for annotator two and  $\kappa=0.85$  for annotator three. This means that all annotators had substantial to almost perfect agreement within their own ratings (cf. Landis and Koch, 1977).

The inter-annotator agreements for the prominence judgement task was  $\kappa=0.77$ , which means that annotators agreed substantially on the prominence patterns of the recordings as played to them. This is comparable to other studies who found generally good agreement on whether or not a word carried a pitch accent. In Grice et al. (1996), for example, this agreement was reached in 87%.

#### 3.4.2. Statistical analysis

I performed a generalised linear mixed model using the logit link function. Since I wanted to investigate the relationship between prominence placement and the different condi-

---

“prominence votes” by the number of all votes per item (nine). This resulted in a number between zero and nine representing the probability of a prominence rating. Taking this variable instead of the yes-no-variable that resulted from the 4-out-of-9-threshold for the statistical analysis did not change the result in any meaningful way. I therefore chose the yes-no-variable since it is more representative of the task for the annotators who had to make a binary decision and could not judge prominence on a scale.



tions, the dependent variable was the number of prominences (binomial: one or two)<sup>4</sup>. As fixed factor I included *condition* (*clash*, *no clash*, *F1*, *F2*). As random effects I had intercepts for *subject*, and for *item*. The inclusion of a by-subject random slope for condition yielded in convergence problems, probably because the data set is too small for a such a complex model.<sup>5</sup> An overall effect for condition was tested for significance by using likelihood ratio tests comparing the model including *condition* to the model without it (cf. Baayen, 2008; Baayen et al., 2008; Winter, 2013). The model with an AIC value (Akaike’s information criterion) of at least two points smaller was considered the better model (c.f. Burnham and Anderson, 2002) when its p-value was <0.05.

### 3.4.3. Results

There is an overall effect of condition as determined by the likelihood ratio test ( $\chi^2(1) = 75.70$ ,  $p < 0.0001$ ). Table 3.2 gives an overview of the effects of the different levels as predicted by the final model. The intercept (first line) refers to condition *clash* to which the other conditions are compared. The estimates here are log-odds.

Table 3.2.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the first annotation task.*

	Estimate	Std. Error	z value	p value
(Intercept)	1.0464	0.3455	3.028	<0.01
cond. no clash	0.4500	0.3994	1.127	0.25988
cond. F1	-2.1736	0.3941	-5.515	<0.0001
cond. F2	-1.9028	0.3830	-4.968	<0.0001

Figure 3.3 shows the results for the number of prominences as annotated by condition. On the x-axis we see the four experimental conditions and on the y-axis the probability for two prominences is given.

Looking at the control conditions *F1* and *F2*, where the answer sentences have only one corrective focus, we can see that single-focus sentences are mostly produced with only one prominence, as expected. These are significantly different from the double focus

<sup>4</sup>In an additional model I kept all annotations in the model by using a linear variable predicting the probability of a perceived prominence by dividing the number of prominence judgements by the number of all judgements (9), see previous footnote. A linear mixed model using this variable did not reveal any different effects.

<sup>5</sup>When I manually changed the optimiser the model did converge, but gave estimated probabilities that were way off the actual data (at times more than ten percent) suggesting that the random structure is too complex for the small dataset. I therefore decided to refrain from the more complex random structure and to perform an intercept-only model.

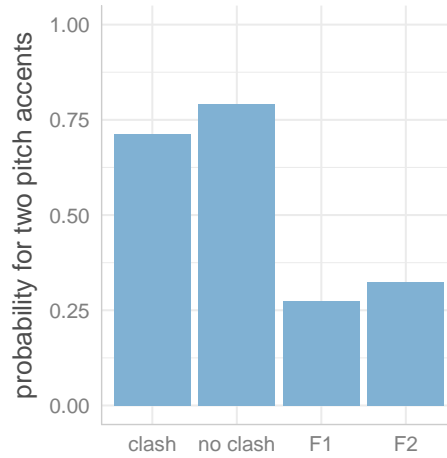


Figure 3.3.: Number of prominences by condition (x-axis). The probability for two pitch accents (on both foci) is on the y-axis.

condition *clash* (27.5% for *F1*,  $\beta=-2.17$ ,  $SE=0.39$ ,  $p<0.0001$ ; and 32.5% for *F2*,  $\beta=-1.90$ ,  $SE=0.38$ ,  $p<0.0001$ ).

In double-focus sentences (conditions *clash* and *no clash*), speakers mostly produced a prominence on **both** focused constituents, namely in 71% of *clash* and in 79% of *no clash* sentences. The difference between these two conditions does not reach significance ( $\beta=0.45$ ,  $SE=0.40$ ,  $p=0.26$ ), this means that the rhythmic manipulation did not affect the number of prominences.

Figure 3.4 shows which focus constituent speakers accented.<sup>6</sup> Let us look first at the single-focus sentences where we can see that speakers mostly realised a pitch accent on the first NP in *F1* and on the second NP in *F2*, leaving the respective given object mostly unaccented as expected. We can therefore assume that the participants understood the task and were paying attention to the context.

There are a few cases in which neither of the two focus constituents was marked as being prominent. These cases (4) appear in Figure 3.4 under the label “other”.

---

<sup>6</sup>Note that the numbers in this graph deviate slightly from the numbers as presented in Figure 3.3. This is because Figure 3.4 comprises the mean of all ratings by the annotators, while Figure 3.3 represents the 4 out of 9 approach thus having only one rating per sentence.

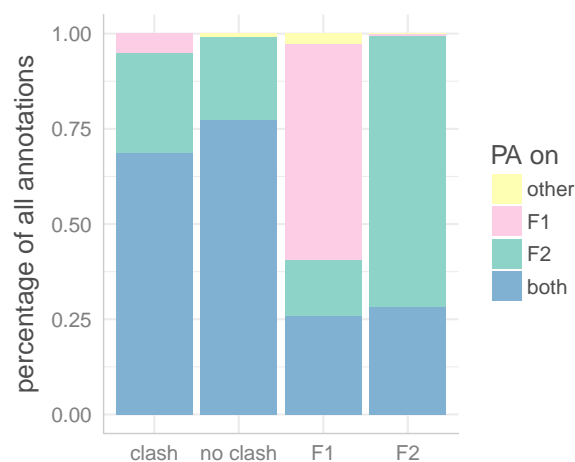


Figure 3.4.: Distribution of prominences by condition (x-axis) on NP1, NP2 or on both NPs.

### 3.5. Analysis of pitch accent type

The analysis of pitch accent placement revealed that in a double-focus context most speakers produce two prominences in most of the cases which I interpreted as instances of two pitch accents. What we do not know yet, however, is what types of pitch accents were chosen and what the relation between the two pitch accents was. The analysis presented here will therefore have a more detailed look at these two prominences and aims at answering the research question whether the rhythmic environment affects the choice of pitch accent.

Studies investigating corrective focus showed that the pitch accent typically used to mark a correction is a fall with a raised F0, that is an accent with a high tonal target on the accented syllable followed by a fall. From an information-structural point of view we would thus expect two emphatic falling accents with a high tonal target on the accented syllable - one on each focus constituent, regardless of condition (namely H\*L followed by H\*L). Now from a rhythmical point of view speakers may prefer alternating pitch accent types in rhythmically clashing contexts (i.e. L\*H followed by H\*L). An alternation of pitch accent types would comply with the corpus experiment by Schauffler et al. (2015a) which showed that pitch accents are more likely to alternate in type when they are close together. The corpus study, however, did not investigate pitch accent types in the light of information structure categories. This is why it is important to investigate preferences for alternation in environments that call for prosodic focus marking, like corrective focus.

### 3.5. Analysis of pitch accent type

---

In their study on single and double-focus sentences, Wang and Féry (2018), looked at the pitch accent types their speakers used to mark these foci. While their experiment confirmed that German speakers prefer a falling contour to mark a focused element in single-focus sentences, they found that in double-focus sentences speakers have various options to realise the first focus (while the second focus was always realised as in single-focus sentences). In their data, they identified three intonation contours that speakers used:

**two-peak pattern** the first focus is realised as a fall

**two-phrase pattern** the first focus is realised as a fall followed by a high intonation phrase boundary

**hat-pattern** the first focus is realised with a rise

In their experiment the authors were interested in the effect of sentence length. They found that the hat pattern is preferred in short sentences and the two-peak pattern is the preferred realisation in longer sentences.

The current experiment kept sentence length constant (see 3.3.2) and all sentences used here approximately correspond to the short sentences by Wang and Féry, in the sense that there is not more than one syllable between the foci.<sup>7</sup>

#### 3.5.1. Procedure

All double-focus sentences that were realised with two pitch accents according to the first annotation task, were annotated with respect to pitch accent type and phrase boundaries using GToBI(S) (see 2.1.1.1)<sup>8</sup>. The results concerning the phrase boundary will be presented in 3.7.

The following labels were used for the annotation of pitch accent type:

**H\*L** a high tonal target on the accented syllable followed by a fall on the same or on the post-accented syllable. This label also includes its downstepped version, that is when the tonal target on NP2 is high but lower than the tonal target of NP1 (!H\*L).

---

<sup>7</sup>An example for a short experimental sentences in Wang & Féry is 'Der LEHRer lobt MALi' - *the teacher praises Mali*, small capitals indicate the syllables that bear pitch accents when the words are in focus.

<sup>8</sup>The same annotation scheme that has been used to annotate the DIRNDL corpus used in the study by Schauffler et al. (2015a) presented in 2.2.3.2.

**H\*** a high tonal target on the accented syllable without a fall, but here the contour stays high until the next pitch accent's high tonal target (linked leveltone).

**L\*H** a low tonal target on the accented syllable followed by a rise on the same syllable or on the post-accented syllable. This label also includes L\*, that is a rising accent whose H-tone is reached with the next pitch accent.

In cases of uncertainty the data was played to a phonetician trained in the annotation of intonation and discussed until a decision was reached.

There was one case in which I did not perceive any pitch accent on the first NP even though there were at least 4 out of 9 judgements in favour of a prominence. Since this was a single case, I disregarded this instance from all further graphical and statistical investigation in order to improve readability. This case is from condition *clash* so that 54 recordings in the *clash* condition and 59 recordings in the *no clash* condition were analysed.

### 3.5.2. Contours used by the speakers

Being the last accent in a statement, NP2 was always realised with a fall (H\*L). I will refer to the different realisations of the first focus by looking at the whole contour encompassing NP1 and NP2. The following three realisations were identified:

**fall-fall pattern** The first focus accent is realised with an H\*L accent, this means that both foci were realised with a falling contour (this contour corresponds to the two-peak pattern described by Wang and Féry (2018)).

**high-fall pattern** The first focus is realised with an H\* accent, that is with raised F0, but there is a no low trail tone, that is no F0 lowering, between the two high targets (H\* on NP1 followed by H\*L on NP2; this contour is comparable to what Féry calls hat contour 1 (Féry, 1993)).

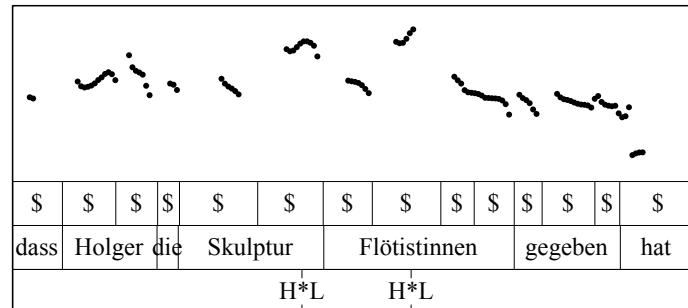
**rise-fall pattern** The first focus is realised with an L\*H, that is a rising contour. (This contour is comparable to what Féry refers to as hat contour 2 (Féry, 1993))

Figure 3.5 gives examples of the three different contours.

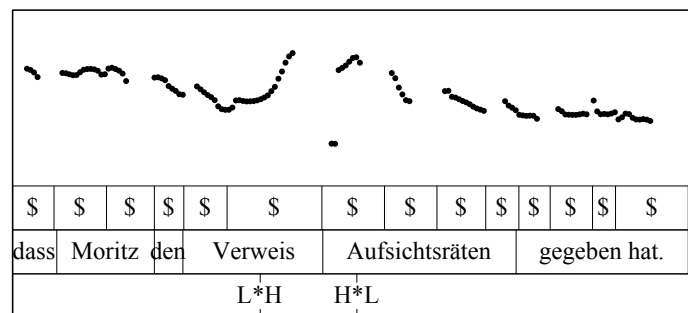
### 3.5.3. Statistical analysis and results

Figure 3.6 shows the distribution of pitch accent types across the two double-focus conditions. We can see that the contour expected for two successive corrective foci from an

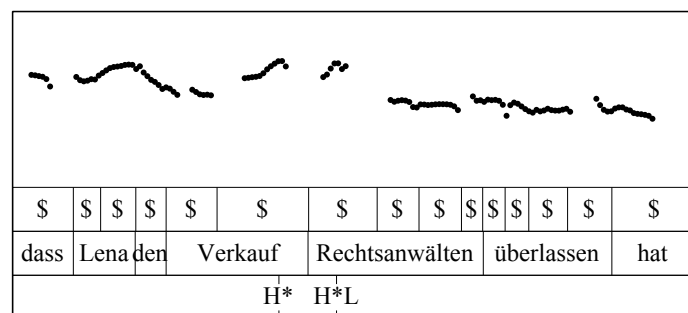
### 3.5. Analysis of pitch accent type



(a) An example of the fall-fall contour (condition *no clash*).



(b) An example of the rise-fall contour (condition *clash*).



(c) An example of the high-fall contour (condition *clash*).

Figure 3.5.: Examples of the three different contours when two pitch accents were used.

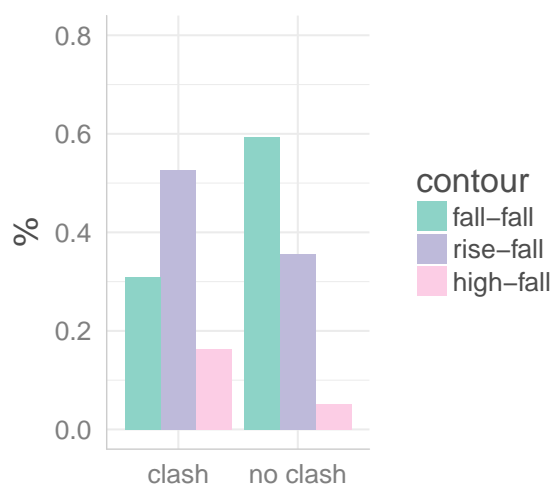


Figure 3.6.: The distribution of pitch accent types across conditions

information-structural view, namely the fall-fall pattern (H\*L on NP1 and H\*L on NP2), is the preferred contour in condition *no clash*: when speakers produced two prominences, they used this contour in about 58% of the *no clash* sentences. In the rhythmically dis-preferred condition *clash*, however, the preferred contour seems to be L\*H on NP1, that is a rise-fall with a rising pitch accent on the first focus. Whenever speakers produced two prominences, they used this contour in 54% of the *clash* sentences. The rise-fall was, however, also quite frequently used in the *no clash* condition (36 %) suggesting that this is an acceptable production also in rhythmically well-formed sentences.

In order to statistically test whether these differences are significant, I conducted a generalised linear mixed model using the logit link function. Since it looks like condition *clash* decreases the occurrence of fall-fall patterns and increases the occurrence rise-falls and high-falls, I used the probability for fall-fall as dependent variable (binomial). As fixed factor I included condition (*clash* and *no clash*) and as random factors an intercept for item and for subject. A model comparison using a likelihood ratio test comparing the final model to the null-model (without the fixed factor *condition*) revealed that there is an effect of condition ( $\chi^2(1)=13.42$ ,  $p<0.001$ ). There are significantly more fall-fall realisations in condition *no clash* ( $\beta=1.70$ ,  $SE=0.50$ ,  $p<0.001$ ) compared to condition *clash*. Table 3.3 gives an overview of the effects of the two conditions as predicted by the final model. Intercept is condition *clash*.

Table 3.3.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear model predicting the presence of the fall-fall contour by condition.*

	Estimate	Std. Error	z value	p value
(Intercept)	-1.3154	0.5015	-2.623	<0.01
cond. no clash	1.7002	0.5047	3.368	<0.001

### 3.5.4. Parametric representation of the contours

In order to phonetically evaluate the manual label categories, the contours were additionally automatically analysed with respect to their acoustic parameters. As a first step, the recordings were automatically segmented into words, syllables and phonemes (Rapp, 1995). The acoustic analyses were then carried out on the stressed (and accented) syllable of the target words. The details of the shape of the tonal contour on the syllables of interest were captured using the PaIntE model (**Parametric Representation of Intonation Events**, Möhler and Conkie 1998; Möhler 2001), which models the shape of the  $F_0$  contour in the vicinity of intonation events. This model approximates a peak in the smoothed  $F_0$  contour by employing a mathematical function term with 6 free parameters. The function is built by summing up two sigmoids with a fixed time delay. The sigmoids are subtracted from a basic value giving the function’s maximum value within the analysis window. In this way the upper bound for the function within the respective window is defined. The two sigmoids are defined each by 3 free parameters ( $a$ ,  $b$ , and  $c$ , where  $a$  and  $c$  are sigmoid-specific and hence indexed according to their belonging to the first or the second sigmoid, e.g. as  $a1$  or  $a2$ , respectively) and a constant alignment parameter  $\gamma$ . The function term is given in equation (3.1).

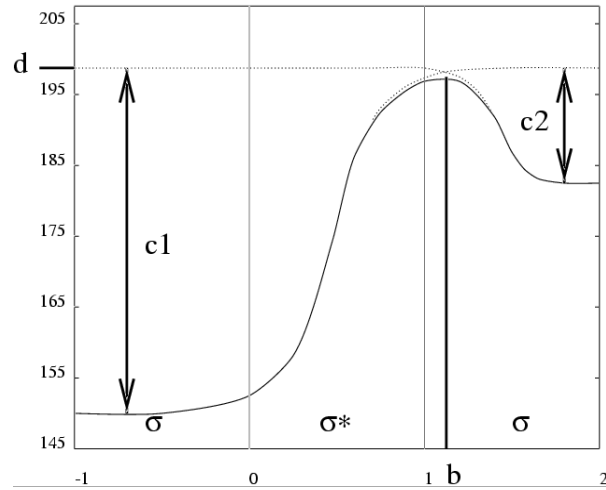
$$f(x) = d - \frac{c1}{1 + \exp(-a1(b - x) + \gamma)} - \frac{c2}{1 + \exp(-a2(x - b) + \gamma)} \quad (3.1)$$

All 6 free parameters in the function term are set by the model so that the actual  $F_0$  shape is fit best. They are linguistically meaningful: parameter  $b$  locates the peak within the three-syllable window, parameter  $d$  encodes its absolute height. The remaining parameters specify the amplitude and (amplitude normalised) steepness of the rise before, and the fall after the peak (parameters  $c1$  and  $c2$  for the amplitude and  $a1/a2$  for the steepness). Figure 3.7 illustrates the function.

A three syllable window can be problematic with the stimuli at hand since in the *clash* condition we expect pitch accents on successive syllables. Additionally we would not want the model to look for a peak across intermediate phrase boundaries. This is



Figure 3.7.: *The PaIntE model function operating on a 3-syllable window with the syllable for which the parametrisation is currently carried out ( $\sigma^*$ ) in the context of its immediate neighbours. The x-axis indicates time (normalised for syllable duration, i.e. the current syllable spans from 0 to 1) and the y-axis displays the fundamental frequency in Hertz. Parameters  $a1$  and  $a2$  are not displayed here.*



why the following rules for defining the approximation window were implemented for the current analysis:

- the window begins with the syllable preceding the accented syllable if
  1. there is a preceding syllable
  2. the forced alignment did not detect a silence between the preceding and the accented syllable
  3. there is no boundary (neither intermediate nor intonation phrase boundary) annotated between preceding and accented syllable
  4. there is no pitch accent annotated on the preceding syllable
- the window is extended to the following syllable unless
  1. there is no following syllable
  2. there is a silence before the following syllable
  3. there is an annotated boundary tone on the accented syllable
  4. there is a pitch accent annotated on the following syllable

Thus, the PaIntE parameters can be matched to the expectations for the shape of different GToBI(S)-accents in a straight-forward way: H\*L is defined as having a peak early in the accented syllable, followed by a fall of the pitch contour. That is, the  $b$  parameter should be located in the accented syllable which means its value should be between 0 and 1 (in the temporally normalised syllable). As for the  $c$ -parameters,  $c1$  (the amplitude of the rise) would be expected to be small, whereas the amplitude of the fall,  $c2$ , would be expected to be high, since the definition of the accent requires a fall of the contour into the lower range of the speakers register. Likewise, for L\*H, a higher  $b$  value would be expected (since there is a low target in the accented syllable, therefore a peak can only be very late in the accented syllable, or in one of the post-accented syllables), and the values for  $c1$  would be expected to be high, whereas the  $c2$  values are expected to be small. The H\* is defined by a high target on the accented syllable but without a fall. Therefore it is expected to have a  $b$  on the accented syllable like in H\*L but a  $c2$  like in L\*H.

To distinguish falling contours from rising contours the peak alignment is a good indicator, since in general falling contours have their peak typically at the beginning of the accented syllable (i.e. smaller  $b$ -values), while a peak at the end of the syllable (i.e. higher  $b$ -values) indicate a rising pattern (Eady and Cooper, 1986).

Figure 3.8 displays the average shape of the three different accent types on the first focus constituent in the current data. We can see that the assumptions are met: L\*H has a later  $b$  parameter and H\*L has a greater  $c2$  parameter. The statistical evaluation of the differences between the labels as annotated in the data at hand is described in the following paragraph. The pitch accent on the second focus constituent was not analysed further since it was always H\*L.

**Statistical analysis** For the statistical analysis I removed all values below the 1st and above the 99. percentile as outliers and normalised the parameters  $d$ ,  $c1$  and  $c2$  by speaker by calculating the z-score (the number of standard deviations the respective value is away from the mean). I used a linear mixed model with the respective PaIntE-parameter as dependent variable and pitch accent type ( $H^*L$ ,  $H^*$ ,  $L^*H$ ) as fixed factor. As random effects I included intercepts for *subject* and for *item*. An overall effect for pitch accent type was tested for significance by using likelihood ratio tests comparing the model including *pitch accent type* to the model without it (cf. Baayen, 2008; Baayen et al., 2008; Winter, 2013).

Since I expected differences in the amplitude of the fall between H\*L, and L\*H and H\* (see above), I used H\*L as the reference level in the  $c$ -parameter analyses. And since I

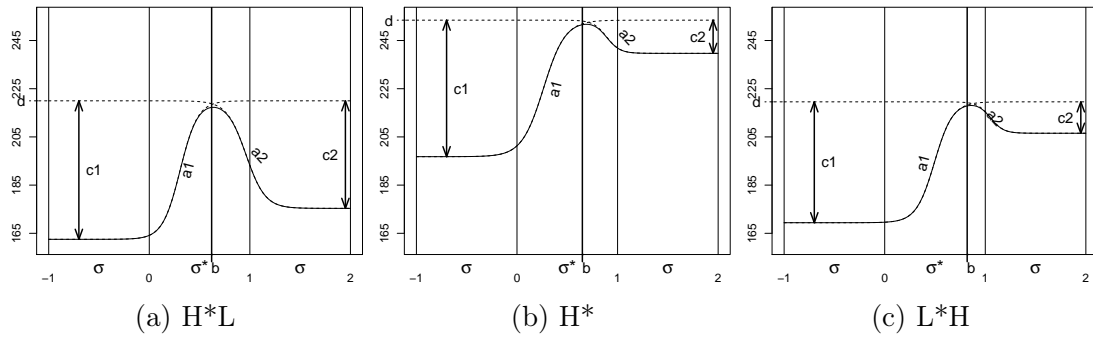


Figure 3.8.: The three accents types (H\*L, H\* and L\*H) generated using the mean values for each PaIntE dimension in the employed subset. The x-axis indicates time (normalised for syllable duration) and the y-axis displays the fundamental frequency in Hertz. Note that in cases with a pitch accent or an ip following the accented syllable PaIntE used only information up to 1 to estimate the contour.

expected a difference in peak alignment between L\*H, which is a rising accent where we expect a very late peak, and the two accents with a high peak on the accented syllable H\*L and H\*, I used L\*H as the reference level for the analysis with parameter  $b$ .

**Results** The statistical analyses confirmed that H\*L has a significantly greater parameter  $c2$  than both H\* ( $\beta = -1.0233$ , SE= 0.2608,  $t = -3.924$ ,  $p < 0.001$ ) and L\*H ( $\beta = -0.8721$ , SE= 0.1770,  $t = -4.928$ ,  $p < 0.0001$ ). This is expected since there is no fall in H\* and L\*H by definition.<sup>9</sup> For L\*H I expected a later peak than for the two high target accents H\*L and H\* which was confirmed by the statistical analysis predicting parameter  $b$ : taking L\*H as the reference level,  $b$  is significantly earlier in H\*L ( $\beta = -0.19990$ , SE= 0.05119,  $t = -3.905$ ,  $p < 0.001$ ) and also significantly earlier in H\* ( $\beta = -0.17011$ , SE= 0.07715,  $t = -2.205$ ,  $p < 0.05$ ). There was no significant difference between the three pitch accent types with respect to absolute peak height (parameter  $d$ ), and the normalised amplitude of the rise ( $a1$ ) and the fall ( $a2$ ). PaIntE always tries to model a peak and therefore all pitch accent types do have a peak. In order to know whether this peak is on the prominent syllable we have to look at parameter  $b$ . This means that also for pitch accents with a low tonal target (L\*H) a peak is modelled, and the peak is expected

<sup>9</sup>The absence of an effect on  $c1$  is not surprising considering that the pitch accents are corrective focus accents. In German, contrastive focus and narrow focus accents usually display a rising onglide to the syllable. Even though there is evidence that this rising onglide is not meaningful in terms of marking contrastiveness (Kügler and Gollrad, 2015) the rise will be phonetically reflected in a higher  $c1$  parameter than would be expected for H\*L accents in other contexts. In especially emphatic cases there is the option to use the label L\*HL in GToBI(S) to represent this onglide.

to be very close to the syllable boundary in two-window approximations and on the post-accented syllable in three-window approximations.

### 3.6. Analysis of pitch accent relation

In their study on dual focus, Wang and Féry (2018) found that in the specific context of having two focused constituents in the same sentence, speakers mark the two foci with asymmetrical prominence in short sentences with a hat pattern, that is with a rising pitch accent on the first focus and a falling pitch accent on the second focus. In longer sentences, however, speakers preferred to use a two-peak pattern, that is a fall on the first focus and a fall on the second focus, and these two pitch accents were equally prominent. In their study they examined sentence length and found that the number of syllables affect the speakers' choice for contour and phrasing.

The annotation of pitch accent type in the study at hand, revealed that the rhythmic environment affects the choice of contour in double-focus sentences (while sentence length was kept constant). Two falls is the preferred contour in rhythmically preferred sentences (condition *no clash*) and a rise and a fall is the dominant production strategy in rhythmically dispreferred sentences (condition *clash*). A second step in the analysis aims to take a look at the relation between the two pitch accents: does the rhythmic environment affect the relationship between the pitch accents in terms of prominence? Two options seem possible: the second focus accent is more prominent than the first, thus there is only one most prominent accent in the phrase; or the two focus accents are equally prominent, that is there are two equal prominences in one phrase. For the phonological analysis, we need to know what kind of phrase the two foci constituted - one intermediate phrase, two intermediate phrases and one intonation phrase, or two intonation phrases. Therefore, phrase boundaries had been annotated alongside pitch accent type. The results are presented in section 3.7. Example (30) illustrates the case of a subordinate (prenuclear) first focus accent in condition *clash*. The case of two equivalent pitch accents is given in example (31), also in condition *clash*.

		×	
	×	×	
	×	×	×

(30) ((...das Klavier)<sub>(ip)</sub> (Lehrerinnen gegeben hat)<sub>(ip)</sub>)<sub>IP</sub>

	×	×	
	×	×	
	×	×	×

(31) ((...das Klavier)<sub>(ip)</sub> (Lehrerinnen gegeben hat)<sub>(ip)</sub>)<sub>IP</sub>

The presence of an intermediate phrase boundary (ip) is given in parentheses since not all sentences were realised with two intermediate phrases. For example, all realisations with a linked monotone  $H^*$ , such as in the high-fall contour, by definition do not have an intermediate phrase boundary between the two focus accents (see above). The analysis concerning phrasing will be presented in the following section.

### 3.6.1. Procedure

The subset comprising the double-focus conditions that were realised with two prominences was again played to the annotators. They judged each recording three times in randomised order without the context question and blind to the condition. The two adjacent focused NPs were shown on the screen and the annotators were instructed to mark which of the words is more prominent or whether they are the same. They could choose the more prominent item by clicking on the respective NP. If they considered both NPs to be equivalent in prominence, they could click on the button “equal”. They could listen to the recordings up to ten times before making a decision and were instructed to take a break two times during the task. This task was implemented in praat using ExperimentMFC (Boersma, 2001), see Figure 3.9 for a screenshot of the annotation task. I considered an item to be more prominent when it was marked as more prominent by at least five out of nine annotations. The threshold here is different from the previous annotation task, because both task and research question are different. The previous task was designed to find prominences and the threshold therefore was set more liberal in order to detect all cases (for example cases which annotators perceive as “a little prominent”). Also the degree of prominence is gradient and annotators may differ as to when they mark an item as prominent. The current task, however, required a decision of which of the items is more prominent (or whether they are the same). So “a little more prominent” is still more prominent and a four-out-of-nine threshold would not be representative. This is why a five-out-of-nine threshold was chosen, that is an item was considered as more prominent (or the same) when the majority of annotations marked it as such.

#### 3.6.1.1. Inter-annotator and intra-annotator agreement

Agreement between the annotators as well as between the individual annotations within each annotator were again calculated using Fleiss’ Kappa in the same way as described in 3.4.1. The three labels on which agreement was measured are “prominences are equally strong”, “NP1 is stronger” and “NP2 is stronger”. The inter-annotator agreement for this



Figure 3.9.: Screenshot of the second annotation task with praat's ExperimentMFC.

task was  $\kappa=0.31$ , which, on a scale from 0 (chance agreement) to 1 (perfect agreement) means that there was only “fair agreement” between the annotators (cf. Landis and Koch, 1977). Intra-annotator agreement was substantial ( $\kappa=0.63$ ) for the first annotator, and moderate for annotator two and three ( $\kappa=0.48$  and  $\kappa=0.59$  respectively). The kappa values for this task are considerably lower compared to the previous task where annotators judged the presence of prominences. This issue will be addressed in the discussion.

#### 3.6.2. Statistical analysis

In order to analyse the relationship between rhythmic environment and prominence relation of the focus constituents, I conducted a generalised linear mixed model using the logit link function with prominence (*same*, or *not same*) as the dependent variable. As fixed factors I tested condition (*clash*, *no clash*), contour on the two focused constituents (*fall-fall*, *high-fall*, and *rise-fall*), and an interaction between condition and contour. As random factors I included intercepts for subject and for item. Main effects for the fixed factors were tested via likelihood ratio tests comparing the model with the factor in question to the model without the factor in question. If the model containing the factor had an AIC-value of at least 2 points smaller than the model without the factor and a p-value  $<0.5$  the factor was considered significant. Again, I excluded the one case where I could not perceive a particular pitch accent type from the statistical analysis.

### 3.6.3. Results

In Figure 3.10 we can see the effect of condition on the probability of having equivalent pitch accents. We can see that most of the time, the two prominences were not perceived as being equally prominent. The number of equivalent pitch accents is, however, greater in the rhythmically preferred context *no clash* (44%) compared to the rhythmically dispreferred condition *clash* (27.5%).

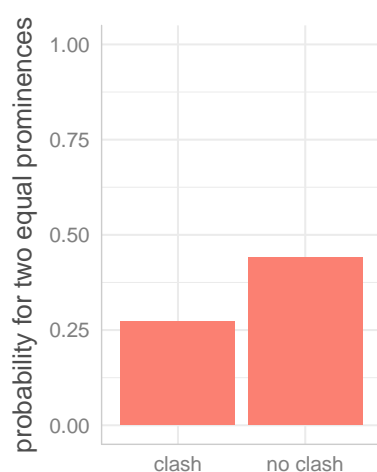


Figure 3.10.: The effect condition (x-axis) on the probability for two equally strong pitch accents (y-axis).

Figure 3.11 shows in addition to condition the effect of pitch accent type on the probability of equivalent pitch accents. As seen before, it appears that condition *no clash* generally increases the likelihood for two equally prominent pitch accents, although not when the first focus was realised with a rising accent (as in the rise-fall contour). This interaction, however, did not improve the model according to the criteria given above. Adding contour as fixed factor also did not improve the model, so that this plot has only descriptive power and may hint to open questions and future research. The fact that we cannot see the interaction in the data statistically, presumably is due to the fact that the dataset is too small when it is split up into 6 groups (2 condition and 3 contours) so that a statistical model is not able to make valid predictions.<sup>10</sup> I therefore selected the model containing condition as fixed factor, since my main interest is to see whether the rhythmic environment affects the relation between the two focus accents.

<sup>10</sup>There are only 12 high-fall realisations across both double-focus conditions, nine in condition *clash* and only three in condition *no clash*.

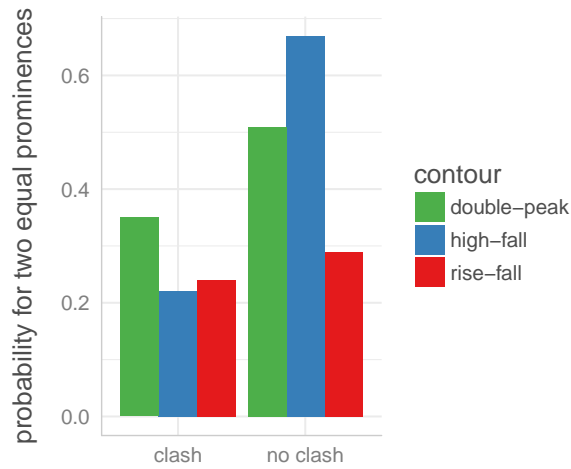


Figure 3.11.: The probability for two equally strong pitch accents (y-axis) by condition and pitch accent type (x-axis).

Table 3.4 gives an overview of the output of that model. The likelihood ratio test comparing the final model with the null model revealed a significant effect of condition ( $p < 0.05$ ) on the probability for two equivalent pitch accents. In condition *no clash* speakers realised two equally prominent pitch accents significantly more often compared to condition *clash* ( $\beta = 0.91$ ,  $SE = 0.45$ ,  $p < 0.05$ ).

Table 3.4.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model predicting the probability for two equivalent pitch accents with condition.*

	Estimate	Std. Error	z value	p value
(Intercept)	-1.2223	0.4296	-2.845	<0.05
no clash	0.9064	0.4526	2.003	<0.05

Whenever the two pitch accents were not the same, the second one was mostly the more prominent one, see Figure 3.12.

### 3.7. Analysis of phrasing

In order to understand whether speakers used phrase boundaries in the double-focus sentences, two approaches were taken. Phrase boundaries were manually annotated and word duration as acoustic cue was measured.



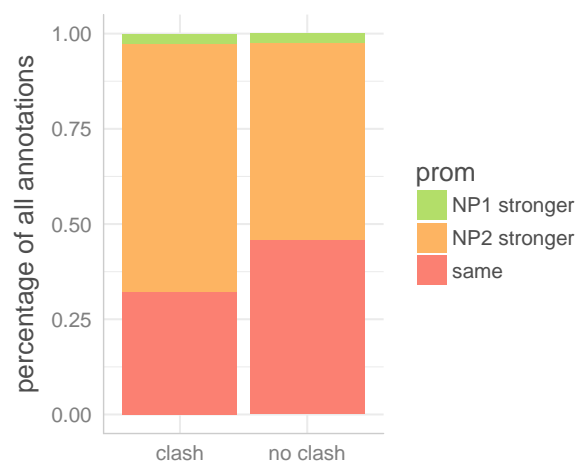


Figure 3.12.: Distribution of the strongest prominence by condition (x-axis): NP1, NP2 or both NPs are the same.

### 3.7.1. Boundary annotation

In addition to pitch accent type, all sentences were annotated as to the presence of a phrase boundary. Intermediate phrases (ip) and intonation phrases (IP) are ordered hierarchically in so far that an intonation phrase carries at least one intermediate phrase and an intermediate phrase carries at least one pitch accent. In GToBI(S), there is no tonal level assigned to intermediate phrases - it is assumed that the tonal contour at the end of the intermediate phrase is controlled by the trailing tone of the nuclear accent. The label for the ip-boundaries therefore is “-”. Intonation phrase boundaries can be labelled high (H%) or low (L%) or unspecified (the default case). The default intonation phrase boundary is tonally unspecified (%) because of the assumption that the nuclear accent’s trail tone spreads to the phrase boundary (cf. Mayer, 1995). The disjuncture at an intermediate phrase is smaller compared to an intonation phrase in terms of the phonetic cues accompanying the phrase break such as lengthening of the final syllable.<sup>11</sup> The two levels were annotated according to audio impression.

As outlined before, the case of double-focus sentences is not straightforward in terms of phrasing. Mainstream prosodic theory does not allow two main prominences in one prosodic domain (cf. Nespor and Vogel, 1986; Selkirk, 1995; Truckenbrodt, 1995), a principle known as *culminativity*. This requirement is met in single-focus sentences, where

<sup>11</sup>A good example where intermediate phrases can often be found are listings such as, for example, *bananas, apples, and pears*, where speakers mostly realise an intermediate phrase break at the positions of the commas and an intonation phrase break after *pears*.

a focus is typically marked with a pitch accent that is realised with a raise in F<sub>0</sub> and followed by post-focal F<sub>0</sub> compression (see 2.1), and this is the main prominence in the phrase. When a sentence involves two foci, the rule of having only one main prominence may be problematic since each focus may be realised with the same level of prominence, as found by Wang and Féry (2018) and in a not to be neglected number of cases in the experiment at hand.

According to Ladd (2008), in double-focus constructions such as *He didn't give me seven dollars, he gave me FIVE FRANC*, each focus constitutes its own intermediate phrase.<sup>12</sup>

Wang and Féry (2018), however, found that while all foci were realised within their own prosodic phrase, the two focus accents often were equivalent in prominence. A realisation with two intonation phrases was only one option that German speakers took in 23.5% of the time. They state that “each focus of a dual-focus sentence is phrased in a separate  $\phi$ -phrase (all realisations), and in the two-phrase pattern, each focus is in a separate  $\iota$ -phrase” (i.e. intonation phrase). When interpreting their results, it is, however, important to note that what they refer to as prosodic phrases ( $\phi$ -phrases) does not correspond to intermediate phrases as intended by the autosegmental-metrical approach. Even though the terms prosodic phrase, phonological phrase and intermediate phrase are often said to correspond to one another, they do not always refer to the same idea - as physical boundaries that are realised with boundary indicators (such as lengthening). So here, the  $\phi$ -phrase as referred to by Wang and Féry does not correspond to intermediate phrases as derived from the signal (by means of boundary indicators, such as lengthening), but is a theoretical domain for pitch accent allocation that is derived from syntax. All phrase boundaries derived from the actual data are referred to as intonation phrase boundaries marked by a high boundary tone.

Considering the assumptions and findings outlined above, the following strategies seem possible:

1. Given that the sentences are rather short, the two focused NPs may be realised in one intermediate phrase, and one pitch accent is realised less prominent than the other. From the pitch accent type annotation we already know that there are realisations within one intermediate phrase, namely at least all high-fall cases, because the H\* is a linked pitch accent that by definition cannot be phrase-final. See (32) for an illustration of a realisation with asymmetrical pitch accents within one ip.

---

<sup>12</sup>Having two equivalent focus accents in two intermediate phrases within one intonation phrase would still violate culminativity, unless the first focus accent is subordinate to the second.

2. Or the two foci are realised each in their own intermediate phrase, together constituting one intonation phrase, where the first focus accent is subordinate in prominence to the second one, as depicted in (33).
3. Or the two foci are realised each in their own intermediate phrase within one intonation phrase, but with equally emphatic pitch accents on the two foci. See (34) for an illustration.
4. Or each focus has their own intonation phrase in which it carries the nuclear pitch accent, also resulting in two equally prominent focus constituents accents, as depicted in (35). Wang and Féry found this option in some of their data, especially with longer focus constituents. They refer to this realisation as ‘two-phrase pattern’.

$$\begin{array}{cccc}
 & & \times & \\
 & \times & \times & \\
 & \times & \times & \times
 \end{array}$$

(32) ((...das Klavier Lehrerinnen gegeben hat)<sub>ip</sub>)<sub>IP</sub>

$$\begin{array}{cccc}
 & & \times & \\
 & \times & \times & \\
 & \times & \times & \times
 \end{array}$$

(33) ((...das Klavier)<sub>ip</sub> (Lehrerinnen gegeben hat)<sub>ip</sub>)<sub>IP</sub>

$$\begin{array}{cccc}
 & \times & \times & \\
 & \times & \times & \\
 & \times & \times & \times
 \end{array}$$

(34) ((...das Klavier)<sub>ip</sub> (Lehrerinnen gegeben hat)<sub>ip</sub>)<sub>IP</sub>

$$\begin{array}{cccc}
 & \times & \times & \\
 & \times & \times & \\
 & \times & \times & \times
 \end{array}$$

(35) ((...das Klavier)<sub>IP</sub> (Lehrerinnen gegeben hat)<sub>IP</sub>)

### 3.7.1.1. Statistical analysis and results

Table 3.5 shows the distribution of realisations with one intermediate phrase (ip), two intermediate phrases and two intonation phrases (IP) across the two double-focus conditions *clash* and *no clash*. In the majority of cases, namely in 78% of the *clash* condition and in 71% of *no clash* sentences, speakers used an intermediate phrase break between the two foci. In only one sentence (of condition *no clash*) the speaker realised the sentence with two intonation phrases, that is with a high boundary tone between the two foci. A generalised linear mixed model using the logit link function was conducted, with presence of intermediate boundary (yes/no) as dependent variable, condition (*clash*, *no*

*clash*) as fixed factor and an intercept for item and for subject as random factors. The model comparison (likelihood ratio test) comparing the model including condition to the null model confirmed that the difference between the two conditions in the probability of ips is not significant ( $\chi^2(1)=0.65$ ,  $p=0.42$ ).

Table 3.5.: *The distribution of one intermediate phrase (ip), two ips and two intonation phrases (IP) across the two double-focus conditions*

	one ip	two ips	two IPs
clash	21.8%	78.2	0
no clash	27.1	71.2	1.7

### 3.7.2. Word duration

The annotation of intermediate phrase boundaries is not always straightforward and often subject to discussion among annotators. Since in the current analysis the annotation of ip-boundaries is based on one person’s acoustic impression, word duration was additionally analysed as an indicator for the presence of a boundary (see e.g. Kentner and Féry, 2013). Additionally, it has been found that lengthening of the first syllable of two adjacent syllables in a stress clash is a major strategy in clash environments in Italian (Burroni and Tilsen, 2020). Such an effect may also play into a difference in word duration between the NP1 in the different conditions.

#### 3.7.2.1. Procedure

The utterances were automatically segmented into words (cf. 3.5.4, Rapp 1995), and word duration of the focused word in NP1 was extracted by means of Festival, a speech synthesis system for German (Festival, 2010).

In order to see a potential lengthening effect caused by an intermediate phrase boundary, the double-focus conditions were compared to the condition with single focus on NP1. The analysis was therefore conducted on a different subset of the data comprising all sentences in the double-focus conditions *clash* and *no clash*, and the single-focus condition *F1*. In all three conditions the target word (NP1) was focused and in the same position. While in the double-focus conditions a second focus follows on NP2, the single-focus condition *F1* only has one focus on NP1, and NP2 is given. Therefore there is no boundary expected after NP1 in *F1*. The three conditions are repeated in 3.7.2.1, the bold word is the target word whose duration is compared across conditions.

(36) Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?

'Did Melli say that Tobi has given the drums to pupils?'

Nein, sie hat gesagt, dass Tobi...

'No, she said that Tobi...'

**clash:** ... das [**KlavIER**]<sub>NP1</sub> [**LEH**rerinnen]<sub>NP2</sub> gegeben hat.

the piano teachers given has

'... has given the piano to teachers.'

**no clash:** ... das [**KlavIER**]<sub>NP1</sub> [**StuDEn**tinnen]<sub>NP2</sub> gegeben hat.

the piano students given has

'... has given the piano to students.'

**F1:** ... das [**KlavIER**]<sub>NP1</sub> [**Schü**lerinnen]<sub>NP2</sub> gegeben hat.

the piano pupils given has

'... has given the piano to pupils.'

### 3.7.2.2. Statistical analysis and results

I conducted a linear mixed model with word duration as dependent variable and condition as fixed factor. As random factors I included an intercept for subject and for item. A model comparison between this model and the null model excluding condition revealed that there is no difference in word duration between the three conditions ( $\chi^2(1)=1.21$ ,  $p=0.55$ ).

## 3.8. Summary of the results and discussion

The aim of this experiment was to investigate whether a preference for alternation affects the prosodic realisation of corrective focus. The following research questions were asked:

1. Do alternation preferences affect pitch accent placement? More specifically, are information-structurally required pitch accents omitted in rhythmic clash environments?
2. Do alternation preferences affect speakers' choice of pitch accent type used for focus marking?

3. Does the rhythmic environment affect the relation between pitch accents? More specifically, the question of the current investigation was whether two equally prominent pitch accents are more likely in rhythmically preferred sentences (*no clash*), and whether the relation between the pitch accents is asymmetrical in *clash* environment.
4. Does the rhythmic environment affect prosodic phrasing? More specifically, do speakers realise (a) an intermediate phrase boundary or (b) an intonation phrase boundary more often in clash environments?

The following will summarise and discuss the results in light of these research questions.

#### 3.8.1. Pitch accent placement

In most of the cases focus was realised with a pitch accent in both single focus conditions as well as in double-focus conditions. This means that in the data at hand an information-structurally required pitch accent was omitted in about 28% of *clash* sentences and in about 21% of *no clash* sentences. The difference between the two environments did not reach significance. Considering that the results of the pilot study (Schauffler et al., 2015b) point in the same directions future work may test this material with more data in order to see whether a tendency for more omitted pitch accents crystallises in condition *clash*, that is in the rhythmically dispreferred condition. It is also conceivable that the requirement to prosodically mark focus with a pitch accent is too strong in German to be overridden by rhythmic preferences. This is in line with results by Henrich et al. (2015) who showed in an ERP study that rhythmic irregularities are more tolerated or more acceptable when attention was guided towards a preceding context-induced narrow focus. A corrective focus, that is a word that is directly contrasted to an alternative given in a question, such as in the current experiment, certainly is quite on the far end of a scale of emphasis given to information-structural categories. It remains to be seen in future work whether other information-structure categories are equally stable in rhythmic dispreferred structures.

The examples found in the German radio news corpus by Riester and Piontek (2015) suggest that pitch accent placement was affected by the rhythmic environment in an all-new sentence. However, in these examples, the pitch accent was not omitted, but shifted to a constituent to the left. It may be the case that a pitch accent on the “wrong” part of the constituent is more acceptable, even though it indicates alternatives in the context

where there are none, than no pitch accent at all (which may be difficult to realise, also depending on phrase length - a longer phrase as in the examples given by Riester and Piontek, require a pitch accent at some point.)

Given the consideration that a corrective focus is typically realised with an emphatic pitch accent, the number of pitch accent omissions is relatively high in the condition that should be more in accordance with rhythmic preferences (*no clash*). One unaccented syllable may therefore not be enough to concur with some speakers preference for an alternation of weak and strong

On the other hand, with about 28% of double-accent productions in *F1* and 33% in *F2* these numbers are slightly higher than expected. A higher number of prominent NPs1 in condition *F2*, where NP1 constitutes given material, is less surprising since prefocal rising accents on given material are not uncommon (Féry and Kügler, 2008; Baumann, 2006; Baumann and Grice, 2006; Schweitzer et al., 2009). This does not apply to the instances in condition *F1* with a post-focal accent on the given NP2, though.

However, both numbers are higher than in the pilot study where only 15% of the single-focus conditions were realised with a pitch accent on the given NP (Schauffler et al., 2015b). The stimuli across the two experiments were the same in terms of syntactic construction, length and context. They only differed in the instruction: half of the participants in the pilot study were instructed to silently read through the dialogue so that they are prepared for their production, while the other half was instructed to read out the answer sentence as soon as it appeared on the screen. In the full study, all participants were instructed to read through the dialogue before producing the answer sentence. I can only speculate that the preparation may have led to a more pronounced articulation and resulted in the production of more pitch accents than in unprepared speech, as an explanation for the high number of double-accent productions. The manipulation of preparation did not significantly affect pitch accent placement in the pilot study. The number of participants in each group was, however, rather small (8) and further testing may be needed to investigate the effect of preparation.<sup>13</sup>

Looking again at the double-focus conditions we can see that whenever there was only one pitch accent realised, it was always realised on the second NP in condition *no clash* and almost always in condition *clash*. This realisation conforms with an all-new sentence where the nuclear accent would fall on the rightmost argument of the verb (NP2).

---

<sup>13</sup>As previously mentioned, rhythmic repair strategies were more often found in prepared speech (cf. Tilsen, 2011), post-focal accents on NP2 in condition *F1* therefore are all the more surprising given that this second accent leads to a pitch accent clash.

### 3.8.2. Pitch accent type

While the rhythmic environment did not significantly affect pitch accent placement, it was found that speakers preferred different pitch accent types in rhythmically preferred vs. rhythmically dispreferred sentences. The pitch accent type typically associated with contrastiveness, a fall, was more often found in the rhythmically regular condition, while a rising contour was preferred in clash environments. This means that the rhythmic environment affected how the first focus was prosodically realised. Finding the rise also quite frequently in regular contexts, it was in general confirmed that double-focus sentences are not merely composed of two single foci but that the first focus can be realised with different pitch accent types. Wang and Féry (2018) found that the rise is the preferred contour in short sentences. In the current experiment sentence length was controlled and both conditions can be considered short sentences in the sense that there was no, or only one syllable between the two focus constituents. This may explain the frequent use (about 36%) of rising contours in regular contexts.

Another explanation for the frequent use of rise-falls across the two conditions could be that speakers interpreted the double-focus sentences differently. There are two ways of how the dialogues theoretically could have been interpreted:

1. As a complex focus with two contrastive focus constituents, expecting two pitch accents, implying that the answer is exhaustive, that is there are no questions open after the answer sentence. This is the experiments' intended interpretation.
2. As a contrastive topic (NP1) and a contrastive focus (NP2), which is typically associated with a rising contour followed by a falling contour in German (Büring, 1997; Féry, 1993; Gast, 2010; Braun and Ladd, 2003). Such an interpretation would imply that the question is not exhaustively answered, something like “*she gave the piano to teachers, but the drums to pupils*”. The pitch accent on the contrastive topic would then be likely realised with a rise indicating the non-exhaustivity.

Given the context material such a topic-focus interpretation seemed to be only plausible if the speakers took all NP1 as one group of alternatives about which information is given, like in a list (“He gave the piano to teachers, the drums to pupils....”). A question-under-discussion analysis of the sentences is depicted in Figure 3.13 illustrating the non-exhaustive interpretation. *Klavier* would then be the focus for  $Q_0$  and the contrastive topic for the sub-question  $Q_1$ . The target sentence (A1) would then be an answer to one of several subquestions ( $Q_1, Q_2 \dots$ ).<sup>14</sup> See also example (37) that illustrates the parallel

---

<sup>14</sup>Thanks to Arndt Riester (p.c.) for pointing out this possibility.



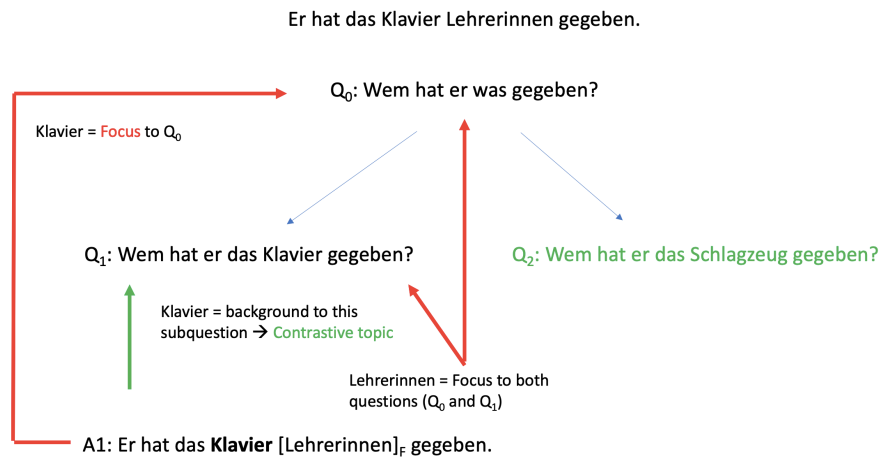


Figure 3.13.: QUD analysis of a possible topic-focus interpretation

in prosody between A<sub>1</sub> and A<sub>2</sub>.

(37) A1: Er hat [das Klavier]<sub>CT</sub> [Lehrerinnen]<sub>F</sub> gegeben.  
           L\*H        H\*L

A2: Er hat [das Schlagzeug]<sub>CT</sub> [Schülerinnen]<sub>F</sub> gegeben.  
           L\*H        H\*L

Given the concrete alternative in each context question which is corrected in the answer sentence, the different agents, and the extensive use of filler sentences I do not find that reading very likely (see appendix for all stimuli). Additionally, if a contrastive topic reading were the cause for the use of rising contours we would not expect a difference between the two conditions. The rise-fall contour, however, is more often used in clash contexts, suggesting that an alternating contour is preferable in cases of pitch accent clash. This is in line with the findings of the corpus experiment by Schauffler et al. (2015a), that pitch accents are more likely to alternate in type when they are close together. The current experiment showed that this alternation preference has an influence on the marking of corrective focus.

The question is whether the different pitch accent types conjured by rhythmic preferences result in different interpretations or acceptability on behalf of the listener. A perception study testing the different realisations in different contexts may shed light on this issue.

One of the contours that was less frequently used (more often in condition clash with 18.5% and 6.8% in condition no clash) was the high-fall pattern, that is a high pitch

accent H\* on the first focus followed by falling H\*L pitch accent on the second focus. In this contour both pitch accents are realised within the same intermediate phrase. The underlying pitch accent, according to GToBI(S), is an H\*L whose low trail tone is linked to the following pitch accent, that is the fall only happens later with the next falling accent (cf. Féry, 1993; Mayer, 1995). Féry (1993) refers to this contour as hat pattern 1, as opposed to hat pattern 2 with a rising L\*H accent followed by a falling H\*L accent, each of which within its own intermediate phrase. According to this reasoning the high-fall pattern should be a phonetic variant of the fall-fall contour. It needs to be tested whether this contour then is equally accepted in the context of corrective focus, since the pitch accent type typically associated with corrective foci is a fall with raised F<sub>0</sub> and post-focal F<sub>0</sub> compression (Grabe, 1998; Baumann and Grice, 2006; Braun, 2006; Féry and Kügler, 2008). However, the falling part is missing in a linked H\* accent.

According to Kügler and Gollrad (2015), the meaning of contrast is conveyed by the height of the peak. They manipulated peak height and the leading rise to the peak (L+H\* in GToBI), but did not consider the fall. In single-focus sentences the contrastive focus accent is typically the last pitch accent in the phrase and falls by the nature of being the last accent in the phrase (Féry, 2007).

In an experiment on Dutch, Chen (2012) found that the amount of sonorant material affected whether speakers used H\*L or H\*(L), irrespective of information structural categories (topic or focus). While speakers used an H\*L accent when there was enough sonorant segments to realise the bitonal contour, they used a monotonal H\* when less sonorant material was present. The author concludes from this finding that the distinction between the two accent types is rather phonetic in nature. The monotonal H\*(L) was realised with a wider span from the accented noun to the following unaccented segments in the focus condition than in the topic condition, as was found for the bitonal H\*L accent. While the monotonal H\* followed by a level-tone did not differ in span-fall across the conditions focus or topic, that is Chen did find a difference between a pre-final H\* followed by level tone and the just mentioned H\*(L) where the L-tone is realised in the following unaccented words.

In their investigation of double-focus sentences in English, Eady et al. (1986) found that the post-focal compression was mostly missing after the first of two foci (see section 4.1 for more details on this study).

In the current experiment the H\* variant was found more often in the clashing context, that is when there was no material between the two accented syllables which seems to support the assumption that it is a phonetic variant. However, only one syllable between the two accents does not seem to be enough material in order to result in such a low

number of high-fall patterns in the *no clash* condition. In fact, if H\* and H\*L were merely phonetic variants of the same contrastive pitch accent we would have expected a greater number of high-fall contours across conditions. There is, however, a crucial difference between the two variants: there is no phrase break after a monotonal H\* while a phrase break after an H\*L is optional. This means that all double-foci realised with a high-fall pattern were realised within one intermediate phrase, suggesting that the first focus accent was realised as a prenuclear accent. Given the low number of occurrences, it was unfortunately not possible to statistically evaluate a possible effect of the two accent types on the prominence relation between the two focus accents. While intuitively post-focal lowering seems essential in the interpretation of contrastive focus, further perception studies including multiple focus sentences are needed in order to investigate the role of the fall in focus accents in German.

### 3.8.3. Pitch accent relation and phrasing

While both foci were realised with a pitch accent most of the time in the double-focus conditions, the two prominences were not perceived as being equally prominent in the majority of cases. The rhythmic environment did, however, have an effect on the prominence relation between these two pitch accents. In rhythmically regular sentences realised with a high or falling accent on the first focus accent, the two pitch accents were judged more often to be equivalent in prominence than in rhythmically clashing sentences where the first focus was perceived more often to be realised with a weaker prominence.

The two foci were mostly realised within two intermediate phrases across conditions, and almost always within one intonation phrase. This means that there are indeed two main prominences within one phrase possibly violating culminativity. This is in line with findings by Wang and Féry (2018) for double-focus sentences realised with two falls, which was also the preferred realisation for the regular condition of the current experiment. The two-phrase pattern, that is a high intonation-phrase boundary after the first focus, found by Wang and Féry as an option in double-focus sentences, was only found once in the current experiment. However, also Wang and Féry found this pattern only in longer sentences that is where the two foci were further away than in the current experiment, where the two foci are adjacent.

It seems to be the case that the pitch accent type affects the prominence relation and that rises on NP1 are less prominent than falls on NP2 across conditions.<sup>15</sup> When

---

<sup>15</sup>According to Ladd, a high peak is acoustically more salient than an L\*, “at least to Western European ears” (2008:228).

the first focus was realised with a rise it was mostly subordinate in prominence in both conditions. This also replicates the finding by Wang and Féry that rises were mostly used in short sentences and then realised as prenuclear accents. Whether a rise-fall pattern is in general perceived as less prominent than a succession of two falls could not be statistically validated here, due to the sparse data problem, and further testing is needed.

In evaluating the findings just discussed it needs to be addressed that the inter-annotator agreement was only “fair” for this task. There are a number of acoustical cues that are typically associated with prominence, such as pitch movement, duration, loudness and voice quality (see e.g. Baumann and Winter, 2018, for an extensive study on various variables associated with prominence). However, there is no one-to-one relationship between perceived prominence and phonetic cues and it has been shown that listeners vary in which cues they use most for evaluating prominence (Baumann and Winter, 2018). Additionally, a very important factor in the perception of prominence seems to be the listeners’ top down expectations (cf. Wagner, 2005). In a study involving native as well as non-native listeners, Wagner (2005) showed that listeners stably rated prominences even in the absence of reliable acoustic cues. Concerning categorical variables it has been shown that pitch accent position and pitch accent type play major roles in the perception of prominence (cf. Baumann and Winter, 2018; Baumann and Röhr, 2015). In the analysis at hand, the three listeners were asked which of the two words were stronger or whether they are the same. The annotation task only comprised data with prominent foci because the relation between the prominences was investigated using the subset where only prominent foci were included. The stimuli were given without the respective context question. It seems as if the given parameters had different effects on the different annotators. It may be interesting to see whether the prominence perception changes when the context is included in the task.

Another question addressed was whether phrase boundaries were used to resolve the clash. A possible repair strategy in clash environments discussed in the framework of metrical accounts is that the clash may be repaired by inserting or augmenting a phrase boundary (Shattuck, 1994; Bolinger, 1965), similar to repairing the clash inserting an additional silent beat between words (Selkirk, 1984). It was therefore hypothesised that speakers make more often use of intermediate phrase boundaries in the presence of a pitch accent clash.

However, there was no difference between the two conditions in ip boundary production - in both conditions ip boundaries were mostly used. This suggests that either contrastive focus accent requires its own intermediate phrase irrespective of rhythmic en-

vironment. This speculation is supported by the findings by Wang and Féry (2018) that in double-focus sentences each focus was realised in one prosodic phrase<sup>16</sup> and the analysis by Ladd (2008) that each focus in double-focus sentences has its own intermediate phrase.

However, there are occasional realisations with only one intermediate phrase, albeit not very frequently, a finding that does not suggest that each focus in double-focus sentences needs its own intermediate phrase. Or it points to the fact that one unaccented syllable, as in the *no clash* condition, is not enough to dissolve the pitch accent clash and therefore an ip boundary is also inserted in this condition.

In order to substantiate the manual boundary annotation acoustically, word duration of the focused word preceding the potential boundary was measured and compared to the focused word that is followed by given material and thus not expected to precede a boundary. Unexpectedly, there was no difference in word duration between the double-focus conditions and the condition with single initial focus. When interpreting phrase boundaries it needs to be noted that the annotation of ip boundaries in this marked context is problematic. A possible boundary is preceded by a possibly emphatic pitch accent which may obscure cues that annotators use to determine the presence of an intermediate phrase boundary. Additionally, it has been shown that inter-annotator agreement is low when annotating specific boundary types even without such a specific context as the one here (Wagner and Watson, 2010). While it may seem to be advisable to revert to acoustic measurements in order to determine the use of phrase boundaries and to look at duration as a cue to boundary placement, the problem remains essentially the same. It is not trivial to disentangle effects on the lengthening of words and segments caused by stress, pitch accent or phrase boundaries since all phenomena are correlated with duration (see Tilsen, 2012, on this problem).<sup>17</sup>

A possible explanation as to why there is no difference in word duration between the first focus in double-focus sentences, even though an intermediate boundary was perceived in the majority of cases, may be the fact that in the single-focus sentences the pitch accent on the target word is the last target of the utterance. We know that the last pitch accent in an utterance is usually also the most prominent and thus also the longest. It may be that the effect of being the last pitch accent in the intonation phrase/utterance adds to the duration of NP1 to the same extent as an ip-boundary

---

<sup>16</sup>Although this may not necessary be the same thing as outlined before.

<sup>17</sup>A re-set of pitch range after the ip-boundary, which sometimes indicates a boundary is also not of help in the material at hand because the word following the potential ip-boundary is pitch accented overriding any pitch re-set due to the beginning of a new phrase.

adds to the duration of NP1 in the double-focus sentences. Additionally, there are cases (about 25%) without a boundary between the two foci which may obscure the effect on duration to the other direction.

I don't see how this problem can be resolved given the experimental stimuli designed for this experiment. Rhythmic effects on phrasing will have to be tested in some different design in future work.

### 3.8.4. Phonological structure of the double-focus realisations

In the following the phonological structure of the German double-focus realisations as discussed above are summarised in the form of an extended metrical grid. The metrical grid is used to indicate prominence relations between the two foci. In order to be consistent with the autosegmental-metrical framework in which the data were analysed, the three levels given here refer to the prosodic word, the intermediate phrase and the intonation phrase (see section 2.1.1 for a description of this hierarchy within the ToBI framework and page 27 for a description of the metrical grid in traditional metrical phonology.) A grid mark on the first level refers to the lexical stress, a grid mark on the second level to a prominence in the intermediate phrase (conforming to a prenuclear pitch accent here) and a third gridmark on level three indicates the strongest pitch accent in the intonation phrase (conforming to the nuclear accent). The ToBI labels below the respective syllables give the most common contour used. Both clash and no clash sentences were realised within two intermediate phrases within one intonation phrase. While in clashing environments speakers preferred to realise the first focus with a rising contour (38), in regular contexts speakers mostly chose a falling contour on the first corrective focus, a contour typically expected for corrective focus (39). So far the tonal realisation.

	×			Intonation phrase
	×	×		intermediate phrase
	×	×	×	prosodic word
(38)	[ (... das Kla vier <sub>ip</sub> (Lehrerinnen gegeben hat) <sub>ip</sub> ] <sub>IP</sub>			
	L*H - H*L			

		×		Intonation phrase
	×	×		intermediate phrase
	×	×	×	prosodic word
(39)	[ ( ...das Klavier <sub>ip</sub> Studentinnen gegeben hat) <sub>ip</sub> ] <sub>IP</sub>			
	H*L - H*L			







## 4. Alternation preferences and focus marking in English

German and English are very similar in various ways. They are both considered to belong to the group of stress-timed languages and they share the preference for rhythmic alternation (see section 2.2). They also have a very similar inventory in terms of pitch accents and how they are used for focus marking (e.g. Grabe, 1998). Nevertheless, L2 speakers of either language still show a foreign accent not only on the segmental but also on the suprasegmental level (see e.g. Jilka, 2000). This thesis contributes to the comparison of these two well-studied languages, investigating prosodic aspects where they may in fact differ.

This chapter comprises two experiments. The first one was designed to investigate how alternation preferences affect the realisation of corrective focus in English. The design of this experiment follows the production experiment conducted with German speakers (presented in chapter 3) as closely as possible. Thus, double-focus sentences are again used in order to investigate the effect of rhythmic alternation preferences on the marking of two adjacent foci.

The second experiment, again, uses double-focus constructions in order to elicit adjacent pitch accents. But this time the first focus consists of words prone to stress shift as a possible repair strategy in so-called rhythm rule contexts. The rhythm rule operates in order to avoid stress clashes and to ensure an alternation of strong and weak beats, that is stressed and unstressed syllables, by shifting one of the clashing stresses to an earlier syllable (see 2.2). Although the rhythm rule itself has been frequently investigated, it is still unclear how it behaves under corrective focus marking. The second experiment therefore aims to answer the question about the status of the rhythm rule under corrective focus marking - does focus marking override the rhythmic repair strategies in English?

The two experiments are thereby looking at the relationship between focus marking and alternation preferences from both sides of the coin, so to speak. Do alternation preferences affect the prosodic marking of focus (Experiment 1)? And does focus mark-

ing affect instantiations of alternation preferences, namely stress shift in rhythm rule contexts (Experiment 2)?

Since in both experiments double-focus constructions are used, I will spend a section on double-focus in English before delving into the details of the respective experiments. Similar to the situation regarding research on double-focus constructions in German, there are only few studies looking at double focus in English. Since we cannot assume that two adjacent foci in one sentence behave like a concatenation of two single foci (of whose likely realisation we are quite informed, see 2.1) the next section will review two studies on double-focus in English, before the more detailed research questions and analyses are presented in the following.

### 4.1. The intonation of double-focus in English

While single-focus sentences have been studied extensively in English (see 2.1), research on double-focus in English is not much more pronounced than it is for German double focus (see 3.1). And so there are, to the best of my knowledge, only few studies that investigated multiple focus constructions in English. Just to call it back to mind, a double-focus sentence (sometimes also referred to as dual-focus) involves two foci answering two questions, for example *What did Tobi give to whom? - He gave [the piano]<sub>Focus1</sub> to [teachers]<sub>Focus2</sub>.*

Eady et al. (1986) investigated the acoustic realisations of double-focus sentences of the kind given in example (41). The key content words are italicised as taken from the original, the two foci are indicated by me.

- (41) Who shot the puck to whom?  
*[Don]<sub>Focus1</sub> shot the *puck* to [*Kent*]<sub>Focus2</sub>.*

They compared these sentences to realisations of single-focus sentences (with focus in initial and final position respectively) and to all-new sentences. In terms of pitch accent placement, they found that focused items had a higher  $F_0$  peak and were longer in duration, regardless of whether the sentence contained one or two foci – pitch height and duration of the first focus in double-focus sentences was similarly increased to pitch and duration of single initial foci. And pitch and duration of the second focus was equally raised as pitch and duration of single final focus. In terms of pitch accent type, the contours of both single and double-focused words were falling, while it was rising for initial non-focus words in final focus and all-new sentences. There was no evidence that the two foci in a double-focus context affected each other in any way. Differences between

initial focus and the first focus in a double-focus sentence were only found regarding post-focal compression, which was absent in double-focus contexts. The authors interpret this as anticipatory effect caused by the presence of the additional focus item at the end of the utterance (*Kent*).

In a similar experiment, Liu (2010) tested the realisation of double-focus statements and double-focus yes/no questions by four native speakers of General American English. She extended on Eady et al.'s work by additionally manipulating sentence length and the position of the focus constituents (initial + medial, initial + final, and medial + final) and compared them to single-focus sentences with initial, medial, or final focus, as well as an all-new condition. See (42) for an example of a short sentence (initial + medial), capital letters in bold mark the two focus position (emphasis as given in the original).

- (42) **WHO** will read **WHOM** the new mail?  
**ANNE** will read **LEE** the new mail.

Looking at  $F_0$  and duration on all syllables, her results, too, showed that  $F_0$  and duration of the focused words in double-focus sentences was in all positions increased to the same degree as in single-focus sentences. She did not find an effect of sentence length.

These two studies show that double focus in English is in a way a composition of two single foci in terms of pitch, duration and contour. This is different from what has been found for German where the presence of the second focus caused various realisations of the preceding first focus (Wang and Féry, 2018).

The first experiment with English L1 speakers in the study at hand addresses the question of whether double-focus realisations are affected in any way by the rhythmic environment. Will speakers omit a pitch accent in clash contexts, or adjust their choice for pitch accent type, like the German group? Are both focus accents fully fledged as in the studies discussed above, or is one of the foci realised with reduced prominence in clash environments?

The second experiment in this chapter, as has been mentioned above, will present results on a second reading task also involving double-focus sentences but this time in 'rhythm rule' environments. This means that the first focus consists of a word with secondary stress to which primary stress can optionally be shifted, such as in phrases like 'thirteen women', where the stress on '-téen' potentially is shifted to the first syllable so rhythmic alternation is restored.

## 4.2. Experiment 1: Alternation preferences and corrective focus in L1 English

The data for the current experiment was elicited via a reading production experiment. It is designed to be maximally comparable to the German study presented in chapter 3 and aims to be as parallel as possible in terms of stimuli and analyses.

### 4.2.1. Stimulus design

In order to compare the prosodic realisation of double-focus sentences with preferred and dispreferred rhythmic context to the productions by the German native speakers (see 3.3.2) a comparable set of stimuli was constructed.

Question-answer pairs were designed in such a way that two noun phrases (NPs) introduced in the question needed to be corrected in the answer. This elicited two adjacent pitch accents to prosodically mark the corrective foci: on directly adjacent syllables in the rhythmically dispreferred condition (*clash*) and with an unaccented syllable between the potential pitch accents<sup>1</sup> in the rhythmically preferred condition (*no clash*). Examples are given in (43); lexically stressed syllables are underlined, small capital letters indicate focus expected to be marked by pitch accent. Square brackets indicate the target NPs.

(43) Did Carl say that Clara gave the boys horror stories to read?

No, he said that...

**clash:** ... she gave [the GIRLS]<sub>NP1</sub> [ROmance novels]<sub>NP2</sub>.

**no clash:** ... she gave [the GIRLS]<sub>NP1</sub> [adVENTURE books]<sub>NP2</sub>.

Parallel to the reading experiment with German L1 speakers (see 3.3.2), two control conditions were added in order to see the prosodic realisations of single-focus constructions under a corrective context by English speakers. In condition *F1* only the first NP was corrected while in condition *F2* only the second NP was corrected (see examples in 44).

(44) Did Carl say that Clara gave the boys horror stories to read?

No, he said that...

**F1:** ... she gave [the GIRLS]<sub>NP1</sub> [horror stories]<sub>NP2</sub>.

**F2:** ... she gave [the boys]<sub>NP1</sub> [ROmance novels]<sub>NP2</sub>.

---

<sup>1</sup>Again, I call it “potential” pitch accent here, since we do not know whether this pitch accent is going to be realised by the speakers.

Like in the German reading experiment, the stimuli were designed in such a way that the prosodic word of the first object was always a disyllabic iamb, therefore carrying the lexical stress on the final syllable (except for condition F2 where the given first NP was in 45% of the sentences a disyllabic name with initial stress repeating the first NP in the context question (e.g. *Did Ryan say that he lost Rory's master key? - No, he said that he lost Rory's access code*)). 45% of the sentences had proper names in the first NP position. The second NP had the lexical stress on the first syllable in *clash*, F1 and F2 and on the second in the *no-clash* condition. The number of syllables of the second NP ranged from two to seven but was kept constant within one item (that is across the four conditions with the same lexical material, as in (43) and (44)) except for few cases where there was a difference of no more than one syllable between the conditions. This means that there was no substantial difference in length between the conditions.<sup>2</sup>

Additionally, all target constituents were controlled for word form frequency (or compound frequency respectively) which were taken from the wackipediaEN corpus (Baroni et al., 2009). Frequencies were matched within item and position: words in first NP position are comparably frequent per item, and words in second NP position are comparably frequent per item. The mean frequency of NP1 in the context question was 13.60 and the mean frequency of the first focus in the **clash** sentence (which is contrasted to NP1 in the context question) was 13.65 ( $t = -0.0716$ ,  $df = 37.925$ ,  $p\text{-value} = 0.9433$ ). The mean frequency of NP2 in the context question was 19.2, while the mean frequency value of the correcting second focus in the **clash** sentence was 18.8 ( $t = 0.3812$ ,  $df = 38$ ,  $p\text{-value} = 0.7052$ ) and in the **no clash** sentence 19.3 ( $t = -0.1074$ ,  $df = 35.415$ ,  $p\text{-value} = 0.9151$ ). The frequency between conditions is not significantly different ( $t = -0.53703$ ,  $df = 35.415$ ,  $p\text{-value} = 0.5946$ ).

In order to avoid a segmental influence on the intonational marking, which would be especially expected for stops, there were only continuants in the coda of the first NP and no stops in the onset of the second NP.

Note that compared to the German stimuli for the production study, the subject in the answer sentences was exchanged with a pronoun in the English material. This was done in order to reduce the complexity of the sentences.

All English material was reviewed by one American English speaker and by one Australian English speaker in terms of grammaticality, acceptability and plausibility. Additionally, the material was tested in a pilot study.

---

<sup>2</sup>The mean number of syllables of the second NP in the context question was 3.8 and the mean number of syllables of the second object in the **clash** sentences was 3.9 ( $t = -0.2538$ ,  $df = 37.77$ ,  $p\text{-value} = 0.801$ ) and in the no-clash sentences also 3.9 ( $t = -0.2582$ ,  $df = 37.921$ ,  $p\text{-value} = 0.7976$ ).

### 4.2.2. Research questions and hypotheses

The aim of the first experiment conducted with English L1 speakers was

1. to get insights as to how alternation preferences interact with the prosodic marking of focus in English, and
2. to elicit control data for the experiment conducted with L2 speakers.

Proceeding with the first aim, the same research questions as in the German study were asked:

**1. Do alternation preferences affect pitch accent placement?** More specifically, do speakers realise two pitch accents in double-focus sentences with rhythmically regular and irregular environments, or are information-structurally required pitch accents omitted in rhythmic clash contexts? From a focus-marking point of view we would also expect for English that speakers realise each corrective focus with an emphatic falling pitch accent. Should alternation be a strong constraint in English, as for example the frequent use of the rhythm rule suggests (see below), it is conceivable that English speakers adapt their realisation of pitch accents according to their rhythmic preference. One strategy in rhythm rule contexts is to reduce the prominence of the weaker of the two clashing beats (Grabe and Warren, 1995; Horne, 1990; Vogel et al., 1995). Transferred to pitch accents in double-focus sentences, this would mean that speakers may omit one of the pitch accents in clashing environments in order to avoid a dispreferred pitch accent clash.

**2. Do alternation preferences affect speakers' choice of pitch accent type?** More specifically, if speakers realise both focus accents in both, regular and clash contexts, what pitch accent types are used? And is the choice of pitch accent affected by the rhythmic context? The results for German showed that the rhythmic environment affects the choice of pitch accent type used to mark double-focus sentences: speakers prefer an alternation of a rise on the first focus and a fall on the second focus when the two pitch accents are directly adjacent, while a pattern with two falls is preferred in regular contexts, that is when the pitch accents are separated by an unaccented syllable. The rise-fall production has also been found in other double-focus experiments on German (Wang and Féry, 2018). For English, however, a rising focus accent has not been shown to be used in previous research (Eady et al., 1986; Liu, 2010, see above), suggesting that for English speakers resolving a clash by pitch accent type alternation may not be available.

The rhythmic environment, however, was not investigated in these studies. Considering the similarity between the two languages in both their pitch accent inventory and in their preference for alternation, the question therefore is, whether they also share the preference for a *tonal* alternation. More precisely, does the alternation of pitch accent types become an available strategy for English speakers when information-structurally required pitch accents need to be directly adjacent?

### **3. Does the rhythmic environment affect the relation between pitch accents?**

The two studies on double-focus in English by Eady et al. (1986) and Liu (2010) showed that the two foci in double-focus sentences did not differ acoustically compared to the foci in the respective position in single-focus sentences. This suggests that the two foci are not affected in prominence by the presence of the second focus. The current experiment will investigate whether this holds true in clash environments which has not been under investigation before.

### **4. Do alternation preferences affect the phrasing of double-focus sentences?**

Unfortunately, we do not have detailed information about how speakers phrased the sentences in the two English studies on double focus mentioned above. The authors assume that the two foci were realised within one phrase but details on whether and how this was evaluated are not given. For German, Wang and Féry (2018) found that speakers inserted a high intonation phrase boundary between the two foci in about 23.5% of the cases. While, according to the authors, each focus constituted its own prosodic phrase (see chapter 3 for a discussion on this).

It is conceivable that English speakers insert a phrase boundary in order to resolve a rhythmic clash following repair strategies found in rhythm rule contexts where the insertion of an extra beat or a pause may be used in order to dissolve the stress clash caused by adjacent lexical stresses (Lieberman and Prince, 1977; Selkirk, 1984).

After introducing the experiment's participants, the following sections present the analyses conducted in order to answer these questions.

#### **4.2.3. Participants**

Sixteen (4 men, 12 women) speakers were recorded at the Western Sydney University in Australia in a quiet room. Their mean age was 27.38 (range: 19 to 36). None of the participants had known speech or reading disorders. All participants were naïve as to the purpose of the experiment. They were paid for their participation.

### 4.2.3.1. The use of Australian English speakers

Since most studies on English focus marking and on rhythmic aspects of English investigated American English speakers (e.g. Pierrehumbert and Hirschberg, 1990; Breen et al., 2010; Breen and Clifton, 2011), British English speakers (e.g. Henrich et al., 2014) or New Zealand English speaker (e.g. Calhoun, 2012) it is advisable to think about possible consequences of the speakers' variety of the study at hand. Recording Australian speakers had primarily pragmatic reasons that emanated from a cooperation with the Western Sydney University. While American English, especially via TV and media, and British English are certainly still the dominant target varieties for German Learners of English (for which the current group also serves as control group), Australian English has, according to my intuitions, joined these varieties as a target variety for more and more students, mainly due to stays abroad down under before or during their studies.

In any case, the crucial question is whether we expect any different behaviour from speakers of these three varieties in terms of alternation preferences and the prosodic marking of focus. While there are a number of studies investigating Australian intonation they mainly concentrate on a phenomenon referred to as “high rising terminals”, or “uptalk”, which describes the rising intonation at the end of statements, and has been found to be especially common in Australian English (e.g. Fletcher and Harrington, 2001; Fletcher et al., 2005; Fletcher and Loakes, 2010; Guy and Vonwiller, 1989). To the best of my knowledge, there is no research that indicates any major differences regarding the phenomena under investigation in the study at hand between Australian and American or British English. Thus, I do not expect Australian speakers to behave differently with respect to the prosodic marking of information structure and the preference for rhythmic alternation. The ToBI labelling system as introduced by Beckman and Ayers (1997) claims to cover general American, standard Australian, and southern British English, so that we also would not expect any differences in terms of pitch accent inventory.

### 4.2.4. Procedure

Twenty sentences per condition (*clash*, *no clash*, *F1*, *F2*) were distributed over 4 lists using a Latin Square Design so that each participant read only one answer per context question. The experimental sentences in each list were pseudo-randomised for each participant so that the first 3 mini-dialogues were fillers and that sentences of the same condition were not successive. The stimuli of the second experiment presented in the following section (4.3) were integrated in the current experiment, so that one list contained 20 experimental sentences of the “clash”-experiment, 20 sentences of the second



experiment which I will call “shift”-experiment, and 100 filler sentences. The context questions of each question-answer pair had been previously recorded spoken by an Australian English native speaker. The context speaker was instructed to read the questions neutral (that is without any presumed context other than the one given) as well as natural. The L1 recordings took place in a quiet room at the Western Sydney University. The mini-dialogues (both question and answer) were presented on a screen, preceded by instructions and a context story that was designed to make the question-answer pairs more plausible. Parallel to the experiment with German native speakers, the story introduced an aunt who had the whole very big family over. The family talked all day about private issues and about the family business. Since the aunt’s hearing as well as her memory was not the best any more she asked her grandchild in the evening what has been discussed all day. The aunt’s question was the pre-recorded context-question and the participants were asked to click on a symbol which triggered playing of the question on loudspeakers.

Participants were instructed to first silently read the dialogue, listen to the context question by clicking on the loudspeaker and then to produce the answer. This means they were prepared for their answer sentence. This is important to note since it has been shown that rhythm effects are more likely to be found in prepared speech, as already mentioned in 3.3.2 (cf. Tilsen, 2012). The speakers controlled the appearance of each new dialogue themselves by pressing a key on a keyboard. They were instructed to repeat their productions in case of misreadings. The instructions were also given verbally. Figure 4.1 gives an example of how the stimuli were presented. For female participants the presentation was adapted with the picture of a female grandchild (see 3.1).

This resulted in 320 produced answers. Three recordings had to be excluded from all analyses presented in the following because there were either hesitations or repetitions in them, so that 317 recordings were analysed. One item was accidentally recorded in the wrong condition (it was carrying the wrong code) so that condition *clash* has four additional cases and condition *F2* lacks four recordings.

#### **4.2.5. General remarks on the statistical analyses of this chapter**

All statistical analyses of this chapter, including the statistical analyses presented in the following subsections, were performed in R 3.5.0 (R Core Team, 2018), using the function `glmer` with the packages `lme4` (Bates et al., 2013). For data processing, and visualisation the R packages `'tidyverse'` (Wickham, 2017), `'plyr'` (Wickham, 2011), `'ggplot2'` (Wickham, 2009), and `'jtools'` (Long, 2018) were used.

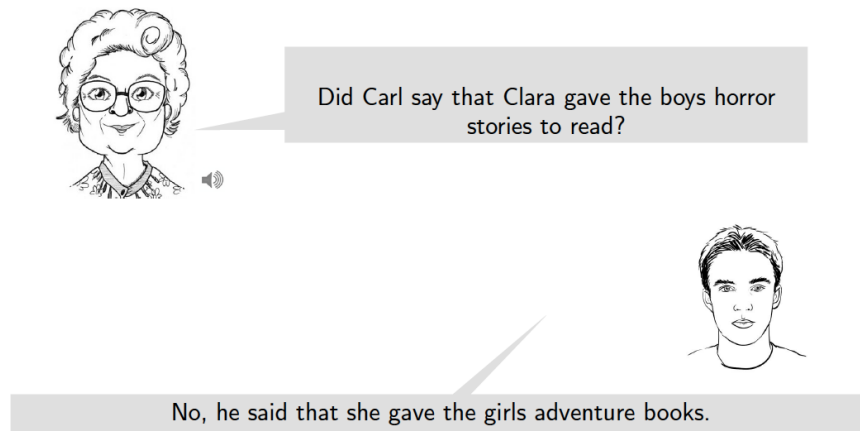


Figure 4.1.: Screenshot of the stimuli presentation. In this example the double-focus condition *no clash* is presented. Clicking on the loudspeaker played the recorded context question.

I used an alpha level of 0.05 for all statistical tests.

#### 4.2.6. Analysis of pitch accent placement

The annotation of pitch accent placement followed the same procedure described in 3.4. The aim of this annotation task was to find out whether English speakers realise both focus accents or whether one of the accents is omitted in rhythmical clash environments, that is when marking both foci prosodically would result in two directly adjacent pitch accents violating the preference for an alternation of weak and strong.

All recordings were played to three annotators using the Demo-window from praat (Boersma, 2001). One of the annotators was an American English native speaker and the other two annotators were German native speakers (one of which was myself). All annotators were prosodically trained. In order to gain a semantically unbiased judgement the answers were played without the respective context question and in randomised order. The annotators were asked to click on the word or words which sounded prominent to them from the verb onwards (see figure 4.2). Each recording was played three times in the course of the annotation experiment. The two experiments of this chapter (“clash”

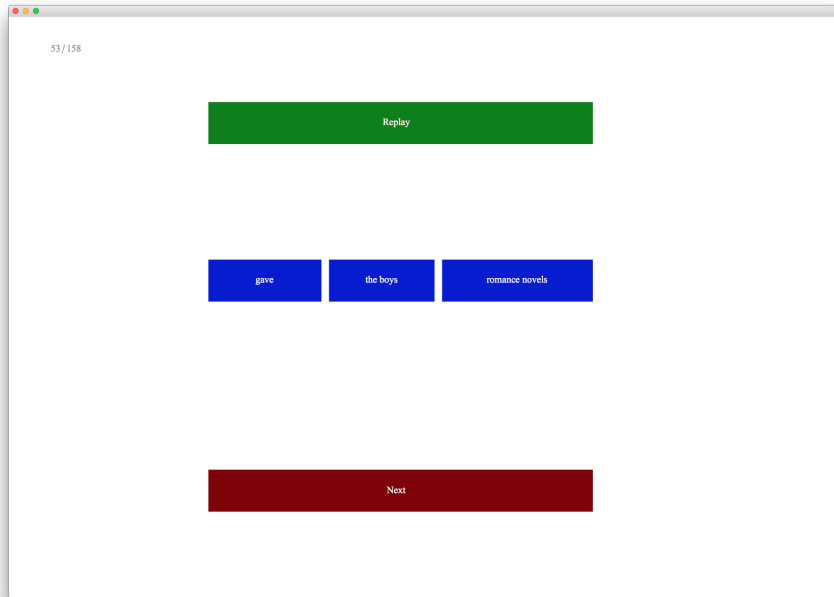


Figure 4.2.: Screenshot of the annotation task with praat’s demo window

and “shift”) were presented in one annotation task.<sup>3</sup> The task was divided into 12 blocks à 160 sentences per group. This resulted in a total of 1917 annotations, with 957 for the current “clash”-experiment. Due to the separate blocks the annotators could take breaks in between. The annotators were able to replay each recording up to ten times before a decision was required. For the analysis, I considered a constituent to be pitch accented when a prominence was perceived at least 4 times, that is by at least two annotators (see page 62 for a more elaborate explanation for choosing this threshold).

#### 4.2.6.1. Intra-annotator and inter-annotator agreement

Both intra-annotator and inter-annotator agreement were calculated in order to see to what extent the annotators agreed in judging prominences. I again applied Fleiss’ Kappa to calculate the degree of agreement between multiple raters on a scale from 0 (chance agreement) to 1 (perfect agreement) (Fleiss, 1971). Kappa was calculated on the agreement on the prominence pattern (prominences on both NPs, on NP1 or on NP2) given for each recording by one annotator compared to the other two annotations of the same recording by the same annotator (intra-annotator agreement) and to the annotations of this recording by the other annotators (inter-annotator agreement).

<sup>3</sup>The two language groups (L1 and L2, sharing the same stimuli), however, were annotated separately.

For the English L1 data, the intra-annotator agreements were  $\kappa=0.81$  for annotator one,  $\kappa=0.69$  for annotator two and  $\kappa=0.76$  for annotator three. This means that all annotators had substantial to almost perfect agreement within their own ratings (cf. Landis and Koch, 1977). The inter-annotator agreements for the prominence judgement task was  $\kappa=0.6$ , which means that annotators agreed moderately on the prominence patterns of the recordings as played to them.

### 4.2.6.2. Statistical analysis

I performed a generalised linear mixed model using the logit link function. Since I wanted to investigate the relationship between the pitch accent placement and the different conditions, the dependent variable was the number of pitch accents (binomial: one or two). As fixed factor, I included *condition* (*clash*, *no clash*, *F1*, *F2*). As random effects, I had intercepts for *subject* and *item*.<sup>4</sup> An overall effect for condition was tested for significance by performing a likelihood ratio test comparing the model including *condition* to the model without it (cf. Baayen, 2008; Baayen et al., 2008; Winter, 2013). The model with an AIC value (Akaike’s information criterion) of at least two points smaller was considered the better model (c.f. Burnham and Anderson, 2002).

### 4.2.6.3. Results

There is an overall effect of condition as determined by the likelihood ratio test ( $\chi^2(1)=34.33$ ,  $p < 0.0001$ ).

Table 4.1 gives an overview of the effects of the different levels as predicted by the final model. The intercept (first line) refers to condition *clash* to which the other conditions are compared to. The estimates here are log-odds. Figure 4.3 shows the results for the number of prominences as annotated by condition. On the x-axis we see the four experimental conditions and on the y-axis the probability for two prominences is given. Looking at the control conditions *F1* and *F2*, where the answer sentences have only one contrastive focus, we can see that single-focus sentences only rarely are produced with two prominences which means that they are mostly produced with only one pitch accent, as expected, namely in 71% of *F1* ( $\beta = -1.13$ ,  $SE = 0.36$ ,  $p < 0.01$  compared to *clash*), and in 75% of *F2* ( $\beta = -1.38$ ,  $SE = 0.38$ ,  $p < 0.001$  compared to *clash*).

In the double-focus sentences, English L1 speakers realised a pitch accent on **both** focused constituents in only 52% of the *clash* condition and in 61 % of the *no clash* con-

---

<sup>4</sup>Since a random slope-model turned out to be too complex for the data (like for the German data, see 3.4.2), I chose intercept-only models for the following analyses. When random-slopes were used, this is mentioned in the text.

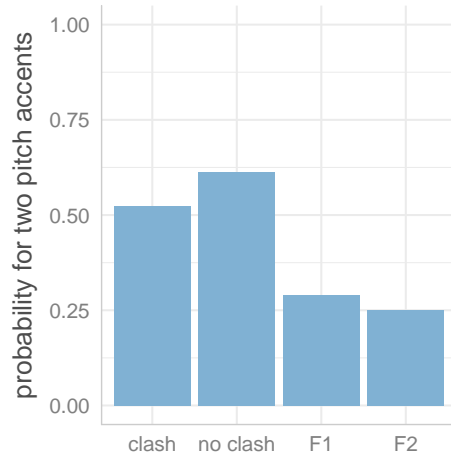


Figure 4.3.: English L1: Number of pitch accents by condition (x-axis). The probability for two pitch accents (on both foci) is on the y-axis.

dition. The difference between these two conditions does not reach significance ( $\beta=0.40$ ,  $SE=0.34$ ,  $p=0.24$ ), this means that the rhythmic manipulation did not affect pitch accent placement.

Table 4.1.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the first annotation task of the English L1 data.*

	Estimate	Std. Error	z value	p value
(Intercept)	0.1258	0.3047	0.413	0.67963
cond. no clash	0.4004	0.3415	1.173	0.24097
cond. F1	-1.1258	0.3576	-3.148	0.00164
cond. F2	-1.3792	0.3756	-3.672	0.00024

Figure 4.4 shows which of the two NPs speakers accented.<sup>5</sup> Looking again at the single-focus sentences first, we can see that speakers mostly realised a pitch accent on the first NP in *F1* and on the second NP in *F2*, leaving the respective given object mostly unaccented. I therefore assume that the participants were engaged in the task and were paying attention to the context questions.

Looking at the double-focus conditions we can see that whenever there was only one pitch accent realised, it was almost always realised on the second focus. There are a

<sup>5</sup>Note that the numbers in this graph deviate slightly from the numbers as presented in Figure 4.3. This is because Figure 4.4 comprises the mean of all ratings by the annotators, while Figure 4.3 represents the 4 out of 9 approach thus having only one rating per sentence.

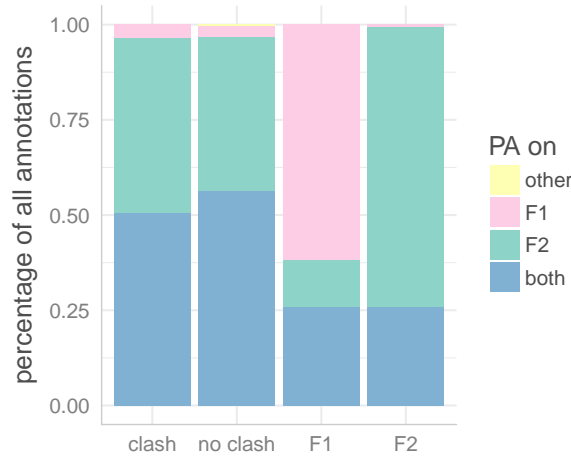


Figure 4.4.: English L1: Distribution of pitch accents by condition (x-axis) on NP1, NP2 or on both NPs.

few numbers labelled “others” - these are instances where the annotators, or one of the annotators labelled a prominence on neither of the two objects. This happened in two cases.

#### 4.2.7. Analysis of pitch accent type

From the analysis of pitch accent placement presented above, we know that speakers realised two pitch accents in just over half of the double focus sentences (52% of *clash* and 61% of *no clash* sentences). The analysis presented in this section has a closer look at these pitch accents as to their shape. The aim is to answer the question whether the type of pitch accent chosen by the speakers was affected by the rhythmic environments.

The results for the German speakers showed that speakers preferred a pattern with a rising pitch accent on the first focus in rhythmically dispreferred sentences (*clash*) and a double-peak pattern (a fall on both foci) in regular contexts (*no clash*). From the study by Wang and Féry (2018) we also know that the rise-fall realisation is an option for German speakers mostly in short sentences regardless of rhythmic make-up. And also in this thesis’ study on German, rise-falls were also frequently found in the *no clash* sentences (albeit less often than in *clash* contexts).

For English, however, focused words in double-focus contexts of previous studies were found to be always realised with falling contours - on the first as well as on the second focus (Eady et al., 1986; Liu, 2010). So far, it seems that a rising contour is not an

option to prosodically mark a focus in English. In order to see whether this is a stable finding and also holds for rhythmically dispreferred structures, the data was annotated with respect to pitch accent type.

#### 4.2.7.1. Procedure

I annotated a subset of the data comprising the two double-focus conditions *clash* and *no clash* whenever at least 4 out of 9 annotations marked both NPs as prominent (that is all cases where both foci were considered to be pitch accented, see 4.2.6). Thus, 43 sentences of condition *clash* and 51 sentences of condition *no clash* were labelled with respect to pitch accent type and phrase boundary. The results concerning the use of phrase boundaries will be presented in 4.2.9.

#### 4.2.7.2. The problem of comparing pitch accent labels of different labelling conventions

Since this data is labelled in order to compare it to similar data by different language groups (L1 German, L2 English), an important issue and complication needs to be addressed. When comparing phonological categories between two languages and when making predictions about possible difficulties for language learners, the system used to describe the intonation categories is essential. Different labelling conventions with possibly differing underlying assumptions can make the comparison problematic and in the end may come to different predictions, even within the autosegmental-metrical framework. This becomes apparent when we look at studies that differed in the choice of the intonation description.

Jilka (2000) for example, bases his analysis of German on the categories proposed by GToBI(S) (Mayer, 1995) with its five pitch accent types coming to the conclusion that German, mainly only using two (H\*L and L\*H), has less pitch accents than American English which may lead to difficulties for English learners and vice versa. In her prosodic typology, Jun (2005) compares English pitch accents as labeled within Mainstream American English-ToBI (MAE-ToBI) to German annotated with GToBI, and concludes that German has **more** distinct pitch accent types than English. One crucial source for differences within the same language is the assumption concerning leading and trailing tones or, as Grabe (1998) puts it, the ‘accentual cut’. There is some debate on whether or not the tones in German and English are only left-headed or both left- and right-headed. Left-headedness means that the tone, or contour, directly preceding the accented syllable is not part of the actual pitch accent. This is what Gussenhoven

postulates for English and Dutch (Gussenhoven, 1984, 1988, 1992, 2016) and Féry (1993) postulates for German. The Stuttgart System, on which the analysis of the German data in chapter 3 is based, builds on Féry's approach. GToBI (German ToBI, see Baumann et al. 2001) and Mainstream American Tobi (Beckman and Pierrehumbert, 1986; Beckman et al., 2005; Silverman et al., 1992) however, postulate the presence of mixed-headed pitch accents, that is the tone preceding the syllable associated with the pitch accent (called *leading* tone) and the tone following the accented syllable is assumed to come from other phenomena, such as the tonal marking of boundaries.

Using systems with different underlying assumptions makes it rather difficult to compare data. In the case of a rising-falling accent, that is a peak on the accented syllable that is preceded by a rise and followed by a fall, a system with leading tones would refer to it as a rising accent, looking at the part preceding the peak (also called an 'on-ramp' analysis), while a trailing-tone system will refer to it as a falling accent, looking at what comes after the peak ('off-ramp' analysis).<sup>6</sup> To simply map the different categories to categories of the other system is therefore not always possible, since information on leading- or trail tone respectively may be missing.<sup>7</sup>

As a complicating factor the English data at hand also serves as control data for the comparison to the realisations by German learners of English, as presented in chapter 4. When analysing learners productions, it is not at all straightforward what system is best to use in order to describe L2 data - with labels used to describe intonation events in the L1 (i.e. German) or using rather the repertoire of the target language (i.e. English), running the risk to not being able to catch realisations outside of either of the two.

Considering these complications, I decided to use the same category labels for all three language groups in order to be able to compare the realisations by L1 English speakers to the realisations by the L1 German speakers and to make predictions about the realisations of English L2 speakers. These labels therefore stand somewhat aside the mainstream labeling conventions for English mainly serving the purpose of comparability. Grabe (1998) suggests a similar inventory in comparing German and English intonation. She bases her annotation on a modified version of Gussenhoven's model for British English and Féry's model for German. In both languages she transcribes rises with  $L^*+H$  and falls with  $H^*+L$ . The difference to the annotation here is that in the study at hand also  $H^*$  is used for a high pitch accent that is not followed by a fall, but linked

---

<sup>6</sup>See for example Gussenhoven (2016) on this issue.

<sup>7</sup>As an attempt to increase the comparability of different intonation description systems for German, DIMA (Deutsche Intonation: Modellierung und Annotation) has been founded with the aim to develop a description system that can translate back to the different systems, see Kügler et al. (2015, 2019).



to the next pitch accented syllable. (See also page 88 for a discussion of H\*.) The labels are described in a way so that the reader may be able to map them to any preferred system of their own. The different categories are based on whether the accented syllable is perceived as high (H\*) or low (L\*<sup>8</sup>).

They are introduced in 3.5 and repeated here:

**H\*L** a high tonal target on the accented syllable followed by a fall on the same or on the post-accented syllable. This label also includes its downstepped version, that is when the tonal target on NP2 is high but lower than the tonal target of NP1 (!H\*L).<sup>9</sup>

**H\*** a high tonal target on the accented syllable without a fall, but here the contour stays high until the next pitch accent's high tonal target (linked leveltone).

**L\*H** a low tonal target on the accented syllable followed by a rise on the same syllable or on the postaccented syllable. This label also includes L\*, that is a rising accent whose H-tone is reached with the next pitch accent.

In cases of uncertainty the data was played to a phonetician experienced in the annotation of intonation and discussed until a decision was reached.

There were two cases in which I did not perceive any pitch accent on the second NP even though there were 4 out of 9 judgements in favour of a prominence. Since this is such a low number, I disregarded these two instances from all further graphical and statistical investigation in order to improve readability. These cases are from condition *no clash* so that 43 recordings in the *clash* condition and 49 recordings in the *no clash* condition were analysed. In the rest of the recordings, speakers always realised an H\*L on the second focus which is expected given that the pitch accent on NP2 is the last accent in a statement. I will refer to the contours using the same terminology as in the German part:

**fall-fall:** The first focus accent is realised with an H\*L accent, this means that both foci were realised with a falling contour.

**high-fall:** The first focus is realised with an H\* accent, that is with raised F0, but there is a no low trail tone, that is the perception is not that of a fall (H\* on NP1 followed by H\*L on NP2).

<sup>8</sup>I am aware that these categories may differ in fine phonetic detail across the three language groups (Atterer and Ladd (2004) for example, found that pre-nuclear rises are aligned later in German than in English.) The phonetic implementation, however, is not under investigation in this study, so that the categorical labels are sufficient for our purposes.

<sup>9</sup>See 3.5.1 for a comment on downstep.

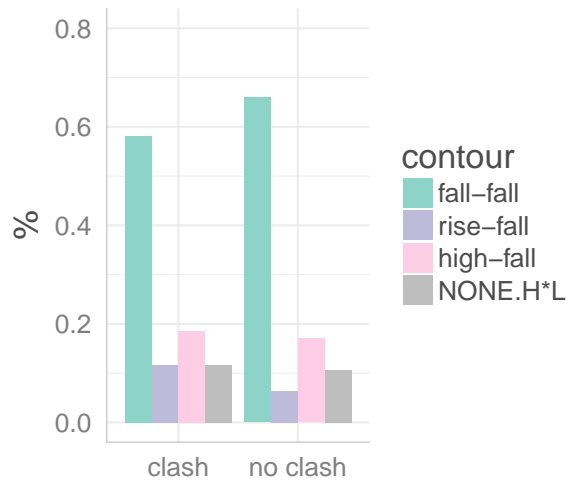


Figure 4.5.: The distribution of pitch accent types across conditions

**rise-fall:** The first focus is realised with an L\*H, that is a rising contour.

In 10 cases (5 of condition *clash* and 5 of condition *no clash*) I did not perceive a pitch accent on the first focus even though the annotators perceived a prominence at least four out of nine times. Since the number is not trivial, these cases are included in the following descriptions. I refer to these cases as NONE.H\*L (no accent on NP1 and H\*L on NP2).

#### 4.2.7.3. Statistical analysis and results

Figure 4.5 shows the distribution of contours across the two double-focus conditions. We can see that, contrary to the German data, speakers preferred the fall-fall pattern in both conditions, that is irrespective of the rhythmic environment. In order to statistically verify that there is indeed no significant difference between the conditions, I ran a generalised linear mixed model using the logit link function. I used the probability for fall-fall as dependent variable (binomial) and included condition (*clash* and *no clash*) as fixed factor. As random factors I included an intercept for item and for subject. A model comparison using a likelihood ratio test comparing this model to the null-model (without the fixed factor *condition*) revealed that the two conditions are indeed not significantly different ( $\chi^2(1)=0.57$ ,  $p=0.45$ ).

### 4.2.8. Analysis of pitch accent relation

The previous analysis showed that, contrary to what has been found for German, the rhythmic environment does not affect the choice of pitch accent type. Both focus accents are mostly realised with a fall (fall-fall pattern) across conditions. This is in line with previous studies on double-focus sentences in English (Eady et al., 1986; Liu, 2010). What we do not know yet is whether these two pitch accents are also equivalent in prominence. According to mainstream prosodic theory there is only one maximally prominent (nuclear) accent per prosodic domain (Nespor and Vogel, 1986; Selkirk, 1995; Truckenbrodt, 1995). Results for German double focus, from this thesis as well as by Wang and Féry (2018) showed that a phrase may nevertheless accommodate two equally prominent pitch accents. In order to see how English speakers realise the focus accents in terms of prominence relation, a second annotation task was conducted. The research question was: does the rhythmic environment affect the relation between two focus accents? More specifically, do speakers realise two fully fledged emphatic pitch accents or is one pitch accent decreased in clash contexts?

#### 4.2.8.1. Procedure

All sentences that were realised with two prominences (that is the same subset as for the pitch accent type annotation) were again played to the three annotators. They judged each recording three times in randomised order without the context question and blind to the condition. The two adjacent focused NPs were shown on the screen and the annotators were instructed to mark which of the words is more prominent or whether they are the same. They chose the more prominent item by clicking on the respective NP. If they considered both NPs to be equally prominent, they could click on the button “equal”. If necessary, the annotators could listen to the recordings up to ten times before making a decision. They were instructed to take two breaks during the task. This task was implemented in praat using ExperimentMFC (Boersma, 2001), see Figure 4.6. Like in the German data, I considered an item to be more prominent when it was marked as more prominent by at least five out of nine annotations (see section 3.6 for an explanation).

#### 4.2.8.2. Inter-annotator and intra-annotator agreement

Agreement between the annotators as well as between the individual annotations within each annotator were again calculated using Fleiss’ Kappa in the same way as described in 4.2.6.1. The three labels on which agreement was measured are “prominences are

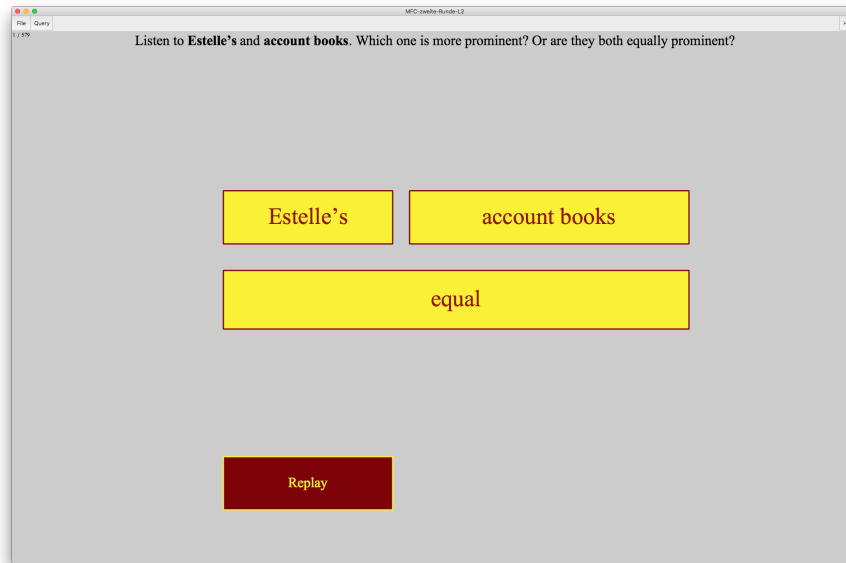


Figure 4.6.: Screenshot of the second annotation task with praat's ExperimentMFC

equally strong”, “NP1 is stronger” and “NP2 is stronger”. The inter-annotator agreement for this task was  $\kappa=0.19$ , which, on a scale from 0 to 1 means that there was only “slight agreement” between the annotators (cf. Landis and Koch, 1977). Intra-annotator agreement was  $\kappa=0.68$  for the first annotator (substantial agreement), and  $\kappa=0.50$  for annotator two and  $\kappa=0.45$  for annotator three (moderate agreement). These results suggest that the task of marking the relational prominence was more difficult than the previous task of marking where prominences are. This issue will be addressed in the discussion of these results (4.2.10).

### 4.2.8.3. Statistical analysis

In order to analyse the relationship between rhythmic environment and prominence relation of the focus constituents, I conducted a generalised linear mixed model using the logit link function with prominence (*same*, or *not same*) as the dependent variable. As fixed factors I tested condition (*clash*, *no clash*) and as random factor I included an intercept for speaker.<sup>10</sup> A main effect for the fixed factor *condition* was tested via likelihood ratio test comparing the model including condition to the null model without condition. The two conditions were considered to be significantly different if the model containing the factor *condition* had an AIC-value of at least 2 points smaller than the

---

<sup>10</sup>An additional intercept for item proved to be too complex for the data, probably due to the small data set.

null model and a p-value  $<0.05$ . The cases for which I could not perceive a particular pitch accent type on one of the NPs were not considered in the statistical analysis.

#### 4.2.8.4. Results

Figure 4.7 shows the effect of *condition* (x-axis) on the probability for two equally prominent focus accents (y-axis). In the majority of cases the two pitch accents were marked as being equally prominent, namely in 75.68% of the *clash* sentences and in 83.33% of the *no clash* sentences. Comparing the model with condition to the null model via likelihood ratio test revealed that the difference between the two conditions is not significant ( $\chi^2(1) = 0.84$ ,  $p=0.36$ ).

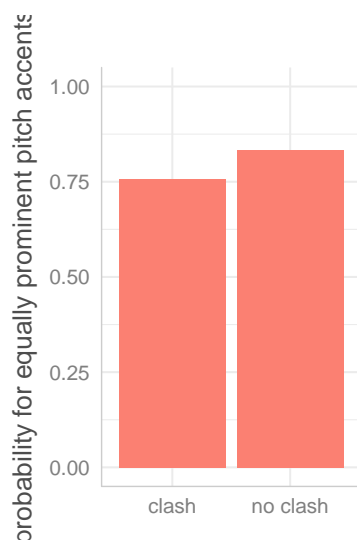


Figure 4.7.: The effect of condition (x-axis) on the probability for two equally strong pitch accents (y-axis).

Whenever the two pitch accents were not the same, the second one was mostly the more prominent one, see Figure 4.8, where the dark orange indicates the proportion of equally prominent foci, yellow a more prominent second focus and the light green part the proportion of more prominent first foci.

#### 4.2.9. Analysis of phrasing

In order to understand whether speakers used phrase boundaries in the double-focus sentences, two approaches were taken. Phrase boundaries were manually annotated and word duration as acoustic cue was measured.

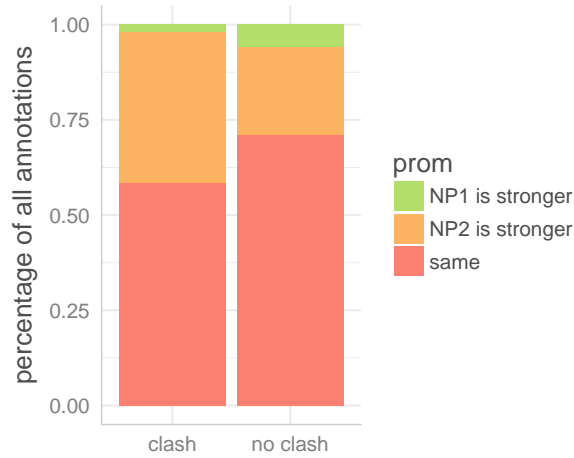


Figure 4.8.: Distribution of the strongest prominence by condition (x-axis): NP1, NP2 or both NPs are the same.

#### 4.2.9.1. Boundary annotation

In addition to pitch accent type, intermediate phrase boundaries and intonation phrase boundaries were annotated on the same subset as introduced above (all double-focus sentences that were realised with two pitch accents). The procedure and labels were the same as in the German experiment (see 3.7) – the two levels of phrase boundaries were annotated (by me) focussing on the audio impression.

In English, the same deliberations apply as in German (see section 3.7): are double-focus sentences realised within one intermediate phrase or does each focus constitute its own phrase, and are phrasing strategies in double-focus sentences guided by the rhythmic environment. As outlined before, it is usually assumed that each phrase only hosts one main pitch accent. In the case of double-focus, Ladd (2008) states that this principle can be maintained if we assume that each focus constitutes its own intermediate phrase. Whether speakers actually realise two phrases still needs to be tested experimentally. For German, it was found that speakers realised the two focus accents within two intermediate phrases in about 75%. Rhythmic preferences did not seem to affect the phrasing (see section 3.7).

**Statistical analysis** Table 4.2 shows the distribution of intermediate phrase boundaries by condition. In 67.44% of the *clash* sentences with two accents and in 63.27% of the *no clash* sentences with two accents speakers realised an intermediate phrase boundary. An intonation phrase boundary with a high boundary tone between the two foci, as reported

Table 4.2.: *Frequency of intermediate phrase boundaries by condition.*

	ip	no ip
clash	<b>67.44</b>	32.56
no clash	<b>63.27</b>	36.73

as an option for German (Wang and Féry, 2018) was not used by any of the speakers. Using a generalized linear mixed model (logit link function) with the presence of an ip (binary) as dependent variable and condition (*clash*, *no clash*) as fixed factor, and an intercept of subjects and for items as random factors it was tested whether there was a difference in the probability for an ip between the conditions. A likelihood ratio test comparing the model with condition to the null model without condition showed that the two conditions are not significantly different ( $\chi^2(1)=0.23$ ,  $p=0.63$ ).

#### 4.2.9.2. Word duration

In order to substantiate the manual phrase-boundary annotation with evidence derived from acoustic cues, word duration of the first foci in the double-focus conditions were measured and compared to the duration of the same word in the single-focus condition (*F1*). See (45), where the three respective conditions are repeated. The duration of the first focused word (in bold) is compared across conditions. It is known that an important acoustic cue for phrase boundaries in English (e.g. Turk and Shattuk-Hufnagel, 2007; White, 2002), and possibly all of the world’s languages (c.f. e.g. Vaissière, 1983), is lengthening of pre-boundary material. In *F1*, *girls* carries the focus accent and the following material is given. It is, therefore, assumed that there is no phrase break after NP1. In both *clash* and *no clash*, speakers realised an ip boundary after *girls* in the majority of cases according to the manual annotations. I therefore expect that this boundary is reflected in longer word durations compared to the same word in condition *F1*.

The utterances were automatically segmented into words (Rapp, 1995), and word duration of the focused word in NP1 was extracted by means of the speech synthesis system Festival (Festival, 2010).

(45) Did Carl say that Clara gave the boys horror stories to read?

No, he said that...

**clash:** ... she gave [the **GIRLS**]<sub>NP1</sub> [ROMance novels]<sub>NP2</sub>.

**no clash:** ... she gave [the **GIRLS**]<sub>NP1</sub> [adVENTure books]<sub>NP2</sub>.

**F1:** ... she gave [the **GIRLS**]<sub>NP1</sub> [horror stories]<sub>NP2</sub>.

**Statistical analysis** I conducted a linear mixed model with word duration as dependent variable and condition as fixed factor. As random factors I included an intercept for subject and for item. A main effect for condition was tested by means of a likelihood ratio test, as in the other analyses. I determined the effect of the different variable levels (*clash*, *no clash*, *F1*) by means of the t-values, assuming that t-values of  $> |2|$  indicate significance.

**Results** A model comparison between this model and the null model excluding condition revealed that there is a main effect for condition ( $\chi^2(1)=11.69$ ,  $p<0.01$ ). Table 4.3 gives the output of the final model. The double-focus conditions do not differ in word duration, the first focus in condition *no clash* is not significantly longer than in condition *clash* ( $\beta=0.01$ ,  $SE=0.01$ ,  $t=0.54$ ). Unexpectedly, however, the single-focus in *F1* is significantly longer than the first focus in condition *clash* ( $\beta=0.04$ ,  $SE=0.01$ ,  $t=3.17$ ). See figure 4.9 for a visualisation of the effect of condition on the duration of NP1. The three conditions (*clash*, *no clash*, *F1*) are on the x-axis and word duration (in ms) is on the y-axis.

Table 4.3.: *Estimated coefficient, Standard Error, and t-value for the linear mixed model for word duration in the English L1 data.*

	Estimate	Std. Error	t value
(Intercept)	0.385246	0.026315	14.640
cond. no clash	0.007207	0.013410	0.537
cond. F1	0.041896	0.013224	3.168

#### 4.2.10. Summary and discussion of results

The aim of this experiment was to elicit and analyse data that gives insight into how alternation preferences interact with prosodic focus marking in English. The results are meant to be compared to the results found for German in the sense of a comparative analysis. Additionally, knowing the behaviour of English L1 speakers, we can then predict and compare realisations by English L2 speakers for the same material as will be reported in 5.3.

The research questions for the English L1 group were first of all the same as for the German L1 group. They are repeated here:

1. Do alternation preferences affect pitch accent placement? More specifically, are



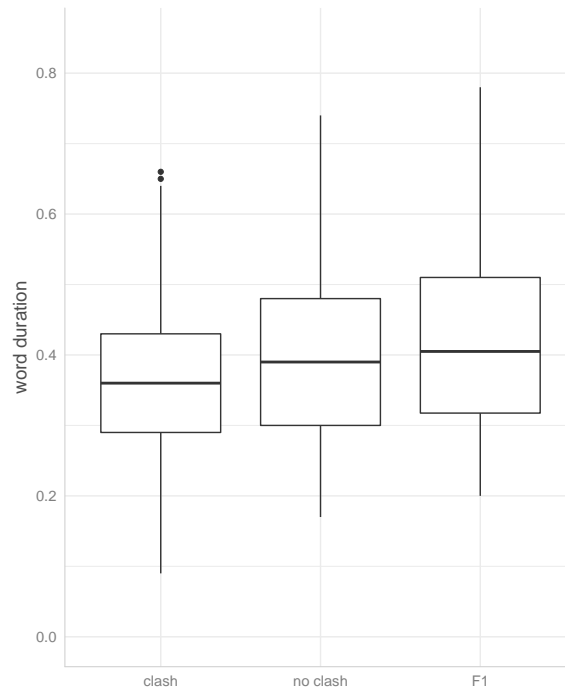


Figure 4.9.: The duration of the focused word in NP1 by condition clash, no clash and F1.

information-structurally required pitch accents omitted in rhythmic clash environments?

2. Do alternation preferences affect speakers' choice of pitch accent type used for focus marking?
3. Does the rhythmic environment affect the relation between pitch accents? More specifically, are two equivalent pitch accents more likely in rhythmically preferred sentences (*no clash*)? And is the relation between the pitch accents asymmetrical in *clash* environment, that is, one pitch accent is subordinate in prominence?
4. Does the rhythmic environment affect prosodic phrasing? More specifically, is the insertion of (a) an intermediate phrase boundary or (b) an intonation phrase boundary a strategy that speakers adopt more often in clash environments?

The following will summarise and discuss the results in light of these research questions.

### 4.2.10.1. Pitch accent placement

Since the reduction of prominence is a strategy that has been found in the context of lexical stress clash (Grabe and Warren, 1995; Horne, 1990; Vogel et al., 1995) it was hypothesised that speakers omit or reduce a pitch accent when it is directly adjacent to a second pitch accent. This was tested in the context of double-focus sentences, that is, both pitch accents are information-structurally induced to mark focus, and more specifically corrective focus in the experiment at hand.

The results did not show an effect of rhythmic manipulation, meaning speakers did not behave differently when there was a syllable between the two potential pitch accents than when there was no material in between the two potential pitch accents. This is in line with the German experiment where we did not see an effect of the rhythmic manipulation in the double-focus sentences on pitch accent placement. It was speculated there that the prosodic marking of a correction is too strong a constraint to be affected by alternation preferences, since mostly both foci were accented (in 71% of *clash* and in 79% of *no clash* sentences). What is striking in the English data, however, is the low number of realisations where both foci are marked with a pitch accent. Only 52% in the *clash* condition, and 61% in the *no clash* condition were realised with two pitch accents. This questions the explanation that the requirement to prosodically mark corrective focus is too strong to be bended by rhythmic constraints in English. If this were the case then we would expect a higher number of double-accent productions in both double-focus conditions.

Considering that the speakers mostly accented the focused NP in the single-focus conditions and left the given NP unaccented (in 71% of *F1* and in 75% of *F2*), I assume that the speakers paid attention to the context question and therefore, that the low number of double-accent productions is not task induced.<sup>11</sup> This is also supported by the fact that whenever there was only one pitch accent realised, it was almost always realised on the second focus, conforming with all-new sentence productions, while a task-induced confound would presumably result in various productions.

Explanations as to why the first pitch accent was often omitted in both *clash* and *no clash* contexts are only speculative at this point. It is conceivable that for English

---

<sup>11</sup>On the other hand, however, the number of double-accent productions in the single-focus conditions is unexpectedly high, comparable to the German results. While a pre-nuclear pitch accent on given material preceding the focus is not uncommon (c.f. e.g. Calhoun, 2010; Féry, 2011), such as in double-accent productions in the single-focus condition *F2*, we would expect post-focal  $F_0$  lowering in the single-focus condition *F1* (Eady et al., 1986). A high number of post-focal accents in this condition (about 29%) may be a side-effect caused by the reading task that was, despite all the care taken, not very natural.

speakers one unaccented syllable between the two foci is not enough in order to dissolve the clash. In English, unstressed syllables are much more reduced than in German, with the vowel typically realised as a schwa (e.g. Delattre, 1969), so that there is a difference in quality between the English and German material between the two potential focus accents. This may entail that the condition *no clash* is still rhythmically dispreferred even though the focus accents are not directly adjacent on the lexical level, but only at a higher level in the prosodic hierarchy.

Previous studies on English double-focus (Eady et al., 1986; Liu, 2010, see above) found that both foci are realised with a pitch accent, so that we cannot assume a general dispreference of marking multiple foci in a sentence. In an experiment on Mandarin Chinese, Kabagema-Bilan et al. (2011) found something similar, namely that the first focus in a double-focus context was not marked with increased  $F_0$  and duration when there was only one syllable between the first and the second focus, which points towards an effect of the rhythmic context.

The experiment presented in the section 4.3 will shed some light on this issue. There, the first focus in the double-focus conditions consists of stress-shift items (e.g. *thirteen*). This means a shift of stress to the initial syllable adds more material between the two foci. If this results in more double-accent realisations than in the present experiment this would hint at a rhythmic effect here in both conditions.

In order to get a better understanding in this matter, I listened to all double-focus recordings again where only one focus was annotated as being prominent. I found that in 39% of these cases (about evenly distributed across the two conditions) the given verb was pitch accented, as exemplified in (46).

- (46) Did Carl say that Clara gave the boys horror stories to read?  
No, he said that she GAVE the girls ROMANCE novels.

It seems like the clash was resolved in some cases by shifting the prominence from the focused word *girls* to the given verb *gave*. Shifting prominence is something English L1 speakers are used to do in rhythm rule contexts where stress is shifted to the left in order to avoid two successive stressed syllables (*thirtéen* → and *thirteen mén*). It may be the case that prominences on a higher level can also be shifted in order to avoid two directly adjacent pitch accents as has been suggested by Riester and Piontek (2015) who found instances of pitch accents in a German radio news corpus that appeared to be shifted away from the focused word to a structurally normally unaccented word, when a second pitch accent was directly adjacent.<sup>12</sup> The prominence annotation for the verbs is,

---

<sup>12</sup>See chapter 1 for more details on this findings.

however, solely based on my own impression, so that further testing is needed in order to understand whether the shift of prominence to other material outside the focus domain preceding the focus is a systematic strategy in clash or ‘near-clash’ environments.

### 4.2.10.2. Pitch accent type

Whenever two pitch accents were realised in the double-focus conditions, speakers mostly used the fall-fall pattern, that is each of the foci was realised with a fall. This is in line with previous findings by Eady and colleagues (1986), namely that the pitch accents with which English speakers in their experiment marked foci were falling in both single and double-focus sentences. For German, it was found that speakers have the option to mark the first focus in double-focus sentences with a rising pitch accent (Wang and Féry, 2018, and also this thesis 3.5). Even though the prosodic marking of single-focus is fairly similar across English and German, the two languages apparently differ when it comes to the prosodic marking of multiple foci in a sentence. While speakers of English seem to prefer to mark each focus with a fall regardless of whether another focused word follows, German speakers can choose between a high or a low tonal target on the first focus. Alternation of pitch accent types in clash environment, thus, is not available to English speakers. This may have direct consequences on the placement of pitch accents as discussed above. Not being able to adapt the pitch accent type to the rhythmic circumstances may consequently lead to more omissions of pitch accents altogether.

### 4.2.10.3. Pitch accent relation

When English L1 speakers marked each focus prosodically with a fall, both pitch accents were mostly judged to be equivalent in prominence. And this was the case in rhythmically regular (in about 83% of *no clash*) as well as rhythmically dispreferred sentences (in about 75% of *clash* sentences). This is conform with previous findings for double-focus sentences by Eady et al. and also by Liu (2010) that  $F_0$  and duration of focused words is increased in both single and double-focus sentences to about the same degree. The finding here is also different from the findings for German where the majority of first foci were realised less prominent than the following second focus, especially in clash environments.

The inter-rater agreement for the task of judging the prominence relation was rather low – with a kappa-value of 0.19 the agreement among the annotators was only “slightly” above chance and even lower than in the corresponding annotation task in the German experiment. This suggests that this task was more difficult than judging which word or

words are prominent in the sentence. An explanation for the low kappa may be that the two accents were actually not very different in prominence or the difference was so small that annotators were unsure in their judgements. It is also possible that small differences were caused by different cues which were differently valued by the three annotators. Additionally, the pitch accents almost always were of the same kind, namely falls, which may have been more difficult to judge in terms of prominence relation than pitch accents of different type as it was often the case in the German data. In the same way as already addressed in the previous chapter, it is also for English potentially problematic to assume two equally strong pitch accents as these results suggest. Given the prevalent idea that there is only one most prominent pitch accent in an intonation phrase, such an assumption may have theoretical consequences since the results suggest that in cases of double focus there does not necessarily have to be only one strongest pitch accent. More studies on double focus are needed in order to draw further conclusions.

#### 4.2.10.4. Phrasing

We only know little about how the presence of multiple foci affects the phrasing of a sentence in English. Previous studies on English double focus did not explicitly report on phrasing strategies (Eady et al., 1986; Liu, 2010). To my understanding, they imply that both foci are realised within one phrase. In the experiment at hand, it was hypothesised that English speakers may add a phrase-boundary in clash environments in order to dissolve the clash. However, the speakers realised an intermediate phrase boundary in the majority of sentences regardless of rhythmic environment, namely in about 67% of clash sentences and in about 63% of no clash sentences.

A realisation of double-focus sentences within two intonation phrases with a high boundary tone between the foci, as found in some cases by Wang and Féry (2018) for German, was not found in the English data – the English L1 speakers in the present experiment never realised the two foci within two intonation phrases.<sup>13</sup> It has to be noted again, however, that annotating phrase boundaries in these contexts with adjacent pitch accents is tricky since acoustic cues used to identify phrase breaks may be acoustically concealed by the first focus accent.

In order to back up the manual annotations for ip-boundaries, word duration was measured as an additional indicator for the presence of a boundary (cf. Kentner and Féry, 2013). I expected longer words in first focus position in the double-focus conditions where I had annotated phrase boundaries for the majority of sentences, compared to the

---

<sup>13</sup>In the German experiment parallel to this one, only one speaker used this realisation once.

single-focused word in the same position after which we do not expect to find a boundary since the rest of the phrase is given and unaccented material. However, like it was the case with the German data, the acoustic measurements did not support the boundary annotation. In fact, the focused word in the single-focus condition was significantly longer than the same word in the double-focus sentences. For the German data, I had speculated that a lengthening effect in the single-focus condition of being the last pitch accent in the utterance adds more to the focused word than the intermediate phrase in the double-focus sentences where a second emphatic pitch accent follows shortly. This second focus accent may cause some backward time pressure onto the realisation of the first focus so that effects of preboundary lengthening are decreased. This is also a possible explanation for the English data, but has to remain speculative at this point. For now I have to say that the results on phrasing are inconclusive and further testing is needed.

### 4.2.10.5. Phonological structure of the English double-focus realisations

In the following the phonological structure of the L1 English double-focus realisations as discussed above are summarised in the form of an extended metrical grid (cf. 3.8.4). The metrical grid is used to indicate prominence relations between the two foci. In order to be consistent with the autosegmental-metrical framework in which the data were analysed, the three levels given here refer to the prosodic word, the intermediate phrase and the intonation phrase (see 2.1.1 for a description of this hierarchy within the ToBI framework and page 27 for a description of the metrical grid in traditional metrical phonology.) A grid mark on the first level refers to the lexical stress, a grid mark on the second level to a prominence in the intermediate phrase (conforming to a prenuclear pitch accent here) and a third gridmark on level three indicates the strongest pitch accent in the intonation phrase (conforming to the nuclear accent). The ToBI labels below the respective syllables give the most common contour used.

There are three major realisations found for English L1 speakers. (47) depicts how English speakers mostly realised the double-focus sentences in both conditions (*clash* and *no clash*; the example here is from the condition *no clash*), namely with two equally prominent falling pitch accents with an intermediate phrase boundary between them. (48) gives the second most frequent realisation, here the condition *clash*, that is only one falling pitch accent on the second focus while the first focus is not accented. In some of these cases speakers realise a prominence on the given verb preceding the first focus as indicated by example (49). This is the third common pattern and accounts for roughly

17% of the data.

- |      |  |     |     |     |                     |
|------|--|-----|-----|-----|---------------------|
|      |  | ×   | ×   |     |                     |
|      |  | ×   | ×   |     | Intonation phrase   |
|      |  | ×   | ×   |     | intermediate phrase |
|      | ×  | ×   | ×   |     | prosodic word       |
| (47) | [( ... gave the girls) <sub>ip</sub> (adventure books) <sub>ip</sub> ] <sub>IP</sub> . |     |     |     |                     |
|      |  | H*L | -   | H*L |                     |
|      |  |     | ×   |     | Intonation phrase   |
|      |  |     | ×   |     | intermediate phrase |
|      | ×  | ×   | ×   |     | prosodic word       |
| (48) | [( ... gave the girls) (romance novels) <sub>ip</sub> ] <sub>IP</sub> .                |     |     |     |                     |
|      |  |     | H*L |     |                     |
|      |  |     | ×   |     | Intonation phrase   |
|      | ×  |     | ×   |     | intermediate phrase |
|      | ×  | ×   | ×   |     | prosodic word       |
| (49) | [( ... gave the girls) (romance novels) <sub>ip</sub> ] <sub>IP</sub> .                |     |     |     |                     |
|      |  |     | H*L |     |                     |

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

While the previous experiment investigated alternation preferences and pitch accenting for information structure, the current experiment has a closer look at rhythmic repair strategies on the lexical stress level and how this may interact with the prosodic marking of contrastive focus.

One of these repair strategies is the so-called ‘rhythm rule’. As a rhythmic repair strategy in lexical stress clash environments, the rhythm rule has been subject to a considerable amount of both theoretical considerations (e.g. Liberman and Prince, 1977; Selkirk, 1984; Hayes, 1995; Gussenhoven, 1991; Kiparsky, 1966) and experimental investigations (e.g. Kentner, 2012a; Tilsen, 2012; Breen and Clifton, 2011; Schlüter, 2005; Wagner and Fischenbeck, 2002; Grabe and Warren, 1995; Vogel et al., 1995; Bohn et al., 2011; Henrich et al., 2014). It refers to the process of shifting one of two clashing stresses onto another syllable in stress clash environments.

There are different assumptions about what constitutes a stress clash. And also different languages have different configurations of clash (see Nespors and Vogel, 1986). For English, Standard Metrical Theory defines a stress clash as two adjacent stresses on a metrical level (Selkirk, 1984; Liberman and Prince, 1977; Nespors and Vogel, 1989). More specifically, Nespors and Vogel (1986, p. 98) propose that in English clash is defined as

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

two adjacent stresses at the word level, allowing one unstressed syllable between two stresses involved in a stress clash. A number of more recent studies, however, imply that a stress clash constitutes adjacent stresses at the syllable level (e.g. Vogel et al., 1995; Henrich et al., 2014), that is when two directly-adjacent syllables both carry lexical stress. Stimuli in Vogel et al. (1995), for example, comprised short phrases (with or without context) such as for example *thirteen clients* in the clash condition, and *thirteen canoes* in the no clash condition.

In cases of stress clash, alternation of strong and weak syllables may be restored by shifting one of the clashing stresses to another syllable. In English, this is possible when the word has a secondary stress that precedes its primary stress; then stress is shifted to the left. This is the process that is referred to as the “rhythm rule” (Lieberman and Prince, 1977). For the purpose of illustration ‘thirteen men’ (or ‘thirteen women’) became quite famous serving as a prototype of cases prone to the process of restoring alternation. The adjacency of the two lexical stresses (*-téén* and *mén*) results in a dispreferred stress clash. According to metrical phonology, alternation is restored by shifting stress in *thirteen* to the initial syllable thereby avoiding a stress clash with the following stressed syllable *men* and establishing an alternation of strong and weak.<sup>14</sup> See example 50 for a representation of the rhythm rule by means of a metrical grid (see also page 27 on the use of metrical grids).

$$\begin{array}{cccccc}
 & & \times & & & \times \\
 & \times & \times & \times & & \times \\
 \times & \times & \times & \times & \times & \times \\
 \text{(50) } & \text{thir} & \text{t} & \text{e} & \text{en} & \text{men} \rightarrow \text{thir} & \text{t} & \text{e} & \text{en} & \text{men}
 \end{array}$$

Lieberman and Prince (1977) refer to this process of stress retraction as *Iambic Reversal*.

It is often debated whether stress is actually shifted, in the sense of a measurable increase of prominence in the secondary stressed syllable and a decrease of prominence in the primary stressed syllables, as predicted by the metrical account (Selkirk, 1984). In fact, previous studies on English have found that decreasing prominence of the primary stressed syllable (rather than increasing prominence on the secondary stressed syllable) is the dominant repair strategy in rhythm rule contexts, since acoustic effects on the syllable to which stress is shifted have not been found (Grabe and Warren, 1995; Horne, 1990; Vogel et al., 1995). What we perceive as a stress shift then, so the argument, results from the relative weakness of the final syllable in comparison with the initial

<sup>14</sup>Another repair strategy discussed by metrical accounts is lengthening of the stressed syllable so that a beat is added, and pause insertion (Lieberman and Prince, 1977; Selkirk, 1984; Nespors and Vogel, 1986).



one which is now perceptually stronger. This has led to the suggestion that stress shift is a primarily perceptual phenomenon (e.g. Grabe and Warren, 1995; Tomlinson et al., 2014).

It has also been found that the initial syllable often was perceived as more prominent than the final one even in sentences without stress clash on the lexical level, as in *Chinese canoes* (e.g. Vogel et al., 1995; Beckman et al., 1990). These findings are in line with a ‘deletion analysis’ (Gussenhoven, 1991) proposing that what is perceived as stress shift in rhythm rule contexts is caused by the deletion of the medial pitch accent in a sequence of three pitch accents - on the first (*thir-*), on the last (*-teen*) and on the following syllable (*men*). Also Bolinger (1958; 1965) suggests that the main protagonist in stress shift examples is a pitch accent that wants to be placed as early in a phrase as possible, while the final pitch accent in a phrase wants to be as late as possible, so that a phrase like *thirteen men* will be realised with an early pitch accent on the initial syllable and a late pitch accent on the last word (THIRteen MEN). But also a phrase without a stress clash, like *thirteen canoes* will be produced this way (THIRteen canOES) according to his account. This pitch accent view is different from the metrical account in that it sees intonational pitch accents as phonologically autonomous from word stress, while metrical theories treat pitch accents in a phrase not differently from stress patterns within words. This is reflected in how prominence relations are depicted in metrical grids where the heights of the grid marks represent relative prominences on all levels (cf. Shattuck-Hufnagel, 1995; Tilsen, 2012). Integrated accounts offer to bring both views together, the preference for alternation and regularity intuition of the metrical account and the early and late prominence tendency, suggesting that a deletion or movement of stress in the metrical grid may cause the deletion or movement of a pitch accent, or the opposite direction, a change in pitch accent location may trigger a change in lexical stress pattern (Shattuck-Hufnagel, 1995; Bolinger, 1981). That is, speakers will not only mark intonation phrases with early and late pitch accents but also place pitch accents in a way that avoids directly adjacent pitch accents (pitch accent clash).

What makes it difficult to draw comprehensive conclusions from the various experimental investigations of the phenomenon ‘stress shift’, is that studies used different sentence material with different information-structural contexts making it difficult to compare results, especially when there are no phonological analyses that give insight into how speakers had prosodically structured the utterances. This is why we still do not fully understand how prominences on the phrase level, such as pitch accents used for focus marking, and rhythmic repair strategies triggered by stress clashes on the lexical level interact. Studies often tested constructions where the first constituent is the first

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

---

accent in the phrase and the second constituent carries the nuclear accent. This raises the question of whether an asymmetry of prominences, as in “Move the beat that is less prominent” (Selkirk, 1984, p.174), is indeed a necessary requirement for the rhythm rule to operate (cf. Shattuck-Hufnagel, 1995). The experiment at hand contributes to answering that question because it looks at environments with phrase-medial adjacent focus accents on both members that are possibly equally prominent (see experiment 1).

The study presented in this chapter will thereby contribute to extend the investigated circumstances by looking at how repair strategies in clash contexts interact with the prosodic marking of information-structural categories such as corrective focus and givenness.

There are some intuitions and findings concerning corrective focus marking in rhythm rule contexts. While Gussenhoven states that a focused constituent will preserve all its accents (hence no deletion of a medial accent), he also notes that in corrective contexts, such as *She doesn't live in AriZona, she lives in CAliFORnia*, preserving both accents seems odd (1991, p.32).

According to Liberman and Prince (1977), stress shift will not take place when the word carries the nuclear accent of a phonological phrase. Beckman et al. (1990) found empirical evidence for this intuition in some instances where stress was not shifted when the target word was emphatically contrasted to some given alternative. In a double-focus sentence, however, the phonological phrase may in fact carry two equally prominent accents (see Wang and Féry 2018, who actually claim that there are two nuclear pitch accents), a context that has not been investigated with respect to stress shift.

The current experiment will therefore investigate the interplay of the rhythm rule with focus accent placement looking at different focus scenarios. It aims at answering the following questions: How does the rhythm rule operate under (corrective) focus marking? Does focus accent override the rhythm rule?

In addition to that, the design allows us to look at various other questions that are still debated. Since the marking of a corrective focus typically requires to increase prominence lending cues on the speaker side, it should become apparent whether stress is actually shifted since a mere reduction of prominence on the clashing syllable is not enough to arrive at an emphatic pitch accent. The different conditions (see below) should also give insight into what constitutes a stress clash, that is under what circumstances is stress shift triggered (if at all under focus marking).

In addition to giving insight into the status of the rhythm rule under corrective focus marking, the experiment may shed light on several questions that arise from the previous experiment:

1. L1 speakers often had produced only one pitch accent in double-focus sentences even when there was one unaccented syllable between the two potential focus accents. In this experiment, repair strategies are possible since there is a secondary stressed syllable to which prominence could be moved (see below). The speakers' behaviour in resolving clashes when stress shift is possible may give an explanation as to why speakers often omitted focus accents.
2. Results on phrasing as possible repair strategies obtained by means of annotation and duration measurements have been inconclusive so far. Since the rhythm rule is argued to only occur within one phrase (Nespor and Vogel, 1986), finding shifted stresses would suggest that the two foci were indeed realised within one phrase.

### 4.3.1. Method

The data for the current study was elicited via a reading production experiment together with the data presented in the preceding chapter (see 4.2.4).

### 4.3.2. Stimulus design

The experimental stimuli were designed to elicit adjacent pitch accents via double-focus environments as has been described in the experiments above. Example (60) presents the four conditions for one item. Capital letters indicate focus expected to be marked by pitch accent, square brackets indicate focus. The first focus consisted of a word with secondary stress on the first, and main stress on the final syllable. The second focus had lexical stress on the first syllable in the *clash* condition, resulting in a stress clash on the lexical level, and on the second syllable in the *no clash* condition, i.e. with an unaccented syllable between the two potential pitch accents. In addition to the double-focus conditions, two single focus conditions were tested (*S-Foc* and *given*). Both conditions include a stress clash on the lexical level, but in *S-Foc* (= 'single focus') the target word is focused and the following word given; while in *given*, the target word is given and the following word focused.

The first focus, that is the target word possibly undergoing stress shift, was always either an adjective (e.g. *unfair*), a *-teen* number (*fifteen*) or the first constituent in a compound (e.g. *infra-red microscopes*) with secondary stress on the first and primary stress on the final syllable.

It has been shown that more frequent words, such as *àntique bóok*, are more likely to undergo stress shift than less frequent words, such as e.g. *àrcáne sórt* (Hammond,

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

1999). This is why also these stimuli were controlled for frequency so that the mean frequencies of words and compounds in the respectively contrasting positions did not significantly differ from each other. The Frequency values were taken from the English corpora collection from Leipzig Wortschatz corpus (Quasthoff and Richter, 2005). The mean frequency of the first focus in the context question was 12.4 (i.e. “the” is  $2^{12.4}$  more frequent than position 1 in the context question); the mean frequency of the first focus in the target sentence (which is contrasted to the first position in the context question) was 13.5 ( $t = -1.3152$ ,  $df = 37.967$ ,  $p\text{-value} = 0.1963$ ).

The mean frequency of the second position in the context question was 12.7, while the mean frequency value of the correcting second focus in the **clash** sentence was 13.65 ( $t = -0.8944$ ,  $df = 37.835$ ,  $p\text{-value} = 0.3768$ ) and in the **no clash** sentence 13.1 ( $t = -0.4021$ ,  $df = 37.841$ ,  $p\text{-value} = 0.6899$ ).

(51) *Did Anna say that she met an Indian programmer?*

No, she said that ...

**clash** ... she met [a JapaNESE] [ARchitect].

**no clash** ... she met [a JapaNESE] [acCOUNtant].

**S-Foc** ... she met [a JapaNESE] programmer.

*Did Anna say that she met a Japanese programmer?*

No, she said that ...

**given** ... she met a Japanese [ARchitect].

See (52) and (53) for an example of the two double-focus conditions represented by means of metrical grids. (52) shows the metrical grid for the *clash* context, assuming two equally prominent accents in the double-focus contexts, as 52% of *clash* sentences were realised in the previous experiment suggests, while (53) represents the *no clash* condition having a gridmark on a lower level between the two accented syllables, as realised in 61% of *no clash* sentences in the previous experiment (the remaining sentences in these conditions were mainly realised without a focus accent on the first focus).

(52) ... a Japanese architect

```

      ×  ×
     ×  ×  ×
    × × ×  × × ×

```

(53) ... a Japanese accountant

```

      ×      ×
     ×  ×  ×
    × × ×  × × ×

```

#### 4.3.3. Participants

The participants of this experiment were the very same as described in section 4.2.3.

#### 4.3.4. Procedure

Twenty sentences per condition (*clash*, *no clash*, *S-Foc*, *given*) were distributed over 4 lists using a Latin Square Design so that each participant read only one answer per context question.

The stimuli from experiment 2 were elicited together with the stimuli from the previous experiment so that each list additionally contained the 20 sentences from that experiment and 100 filler sentences. The experimental sentences in each list were pseudo-randomized for each participant so that the first three mini-dialogues were fillers and that sentences of the same condition were not successive. The context questions for each question-answer pair had been previously recorded spoken by a native speaker of Australian English who was instructed to read the questions in a neutral and natural way. The mini-dialogues were presented on a screen, preceded by a context story that was designed to make the question-answer pairs more plausible. Participants were instructed to first silently read the dialogue, listen to the context question and then to produce the answer. After the dialogues, the participants read a list containing the target words in isolation interspersed with filler words with various stress patterns. Thus 320 target sentences and 320 isolated words were recorded. Three sentences were excluded due to hesitations and verbal errors so that 317 sentences were analysed.

#### 4.3.5. Stress-shift annotation

Three phonetically trained listeners (two German speakers and one American English speaker) listened to the target words (removed from context) and the words produced in isolation, and indicated whether main stress was on the first or the last syllable by clicking on the respective button (see Figure 4.10 for an example); each recording was presented twice in the course of the judgement task.

I considered stress to be shifted when at least four out of the six listener judgments marked main stress on the first syllable. In 4.5% of cases neither *shift* nor *no shift* could be determined. Given the low number, I removed these cases from the statistical analysis when 4 out of 6 judgments agreed on equal prominence.

##### 4.3.5.1. Inter- and intra-annotator agreement

Intra- and inter-annotator agreement were determined using Fleiss' Kappa (Fleiss, 1971). All annotators had substantial to almost perfect agreement within their own ratings ( $\kappa=0.86$ ,  $\kappa=0.76$ ,  $\kappa=0.74$ ), inter-annotator agreement was  $\kappa=0.49$ , which means that

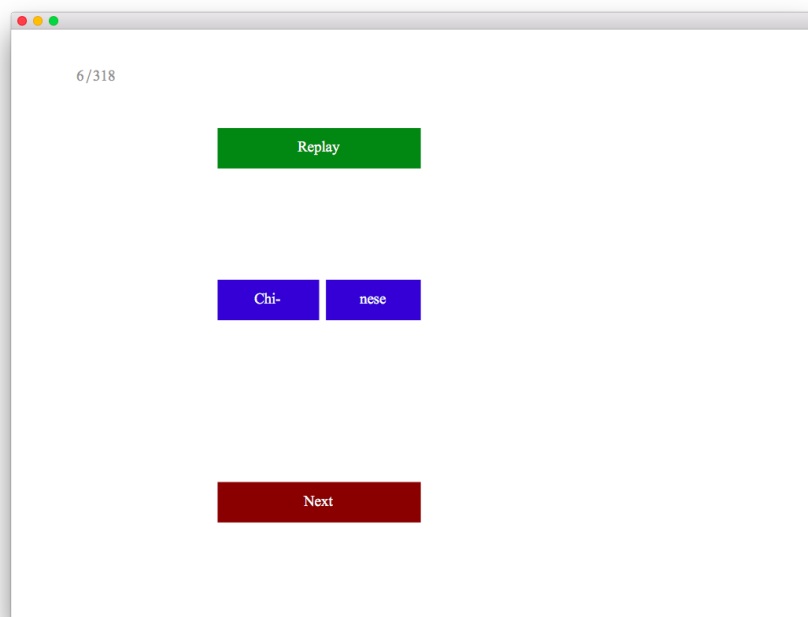


Figure 4.10.: Screenshot of the stimuli presentation for the stress shift annotation. The target word was removed from its context and presented in isolation. The annotators clicked on the box containing the stressed syllable.

annotators agreed moderately (Landis and Koch, 1977).<sup>15</sup>

### 4.3.6. Pitch-accent annotation

In order to understand the interaction of stress clash, stress shift and pitch accent placement, the three annotators judged the presence of pitch accents, or, more specifically, whether they perceived the words involved in the rhythm rule as prominent or not. All answer sentences were played without the respective context question, and in randomized order. This task was annotated together with the prominence annotation of experiment 1 (that is the previously presented experiment), this means that the sentences of both experiments were presented in the same task in randomised order. Each recording was played three times in the course of the annotation task.

Following the procedure of the previously reported experiments, I considered a constituent to be pitch accented when a prominence was perceived at least 4 times, that is by at least two annotators.

#### 4.3.6.1. Inter- and intra-annotator agreement

Intra-annotator agreement was substantial to almost perfect ( $\kappa=0.82$  for annotator one,  $\kappa=0.75$  for annotator two and  $\kappa=0.75$  for annotator three), and inter-annotator agreement was moderate ( $\kappa=0.62$ ).

### 4.3.7. Statistical analyses

All statistical analyses were performed in R 3.5.0 (R Core Team, 2018), using the function `glmer` from package `lme4` (Bates et al., 2013).

To investigate the relationship between stress shift and the five conditions (the four conditions of 4.3.2 plus the realisation of the target word in isolation), I performed a generalized linear mixed effects analysis using the logit link function, with *stress shift* as the dependent variable. I included *condition* as a fixed factor, and random intercepts for *subjects* and for *items*.

To investigate the correlation between stress shift and pitch accent placement, a second analysis was performed on a subset of the data comprising the two double-focus conditions (*clash* and *no clash*) using the presence of two pitch accents as dependent variable. I included an interaction between *condition* and *stress shift* as a fixed factor,

---

<sup>15</sup>The lower inter-annotator agreement was mainly caused by the third annotator who often differed in her perception. I ran another analysis excluding this annotator which did not yield different results in any meaningful way.

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

---

and random intercepts for *subjects* and *items*. An overall effect of condition and the interaction term respectively was tested for significance by performing a likelihood ratio test comparing the model including the fixed factor to the model without it (cf. Baayen, 2008; Baayen et al., 2008; Winter, 2013). The model with an AIC value (Akaike's information criterion) of at least two points smaller was considered the better model (c.f. Burnham and Anderson, 2002).

In the following I will first present the results of the first analysis looking at the circumstances in which stress was shifted. After that I will present the results from the second analysis looking at whether stress shift influenced pitch accent placement in double-focus sentences.

#### 4.3.8. Results

There is an overall effect of condition as determined by the likelihood ratio test ( $\chi^2(1) = 270.32$ ,  $p < 0.0001$ ). Figure 4.11 shows the probability of stress shift (y-axis) by the five different conditions (x-axis). In condition *clash*, where two focus-marking pitch accents are required on directly adjacent syllables, stress was shifted in 75% of the cases. In condition *no clash*, stress was also shifted in the majority of cases (68%), even though the two foci were separated by one syllable in this condition, that is there was no stress clash on the lexical level. The two double-focus conditions were not significantly different ( $\beta = -0.42$ ,  $SE = 0.38$ ,  $p = 0.27$ ).

Stress was shifted even more often (89%,  $\beta = 1.22$ ,  $SE = 0.48$ ,  $p < 0.05$ ) when the target word was given and the following word focused (*given*).

Stress was **not** shifted in 77% of the cases when the target word was focused and the following word given (S-Foc) ( $\beta = -2.68$ ,  $SE = 0.41$ ,  $p < 0.0001$ ), and in 84% of the cases when the word was produced in isolation ( $\beta = -3.19$ ,  $SE = 0.35$ ,  $p < 0.0001$ ). Table 4.4 gives an overview of the effects of the different levels as predicted by the final model. The intercept (first line) refers to condition *clash* to which the other conditions are compared. (The estimates are log-odds here, again.)

Looking at the pitch accenting in the double-focus conditions (Figure 4.12), there is a main effect for the interaction between stress shift and pitch accent placement as determined by the likelihood ratio test ( $\chi^2(1) = 147.59$ ,  $p < 0.05$ ). We found that while shifting stress by tendency increases the probability for the presence of two pitch accents from 69% to 91% (estimated) in condition *clash* ( $\beta = 1.54$ ,  $SE = 0.81$ ,  $p = 0.06$ ), it decreases it in condition *no clash* from 90% to 75% ( $\beta = -2.66$ ,  $SE = 1.09$ ,  $p < 0.05$ ). Table 4.5 gives the estimates for the model with the interaction, intercept is condition *clash*.



Table 4.4.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the shift annotation task.*

	Estimate	Std. Error	z value	p value
(Intercept)	1.2787	0.3585	3.566	<0.001
no clash	-0.4226	0.3839	-1.101	0.270955
S-Foc	-2.6820	0.4174	-6.425	<0.0001
given	1.2219	0.4805	2.543	<0.05
isolation	-3.2192	0.3500	-9.199	<0.0001

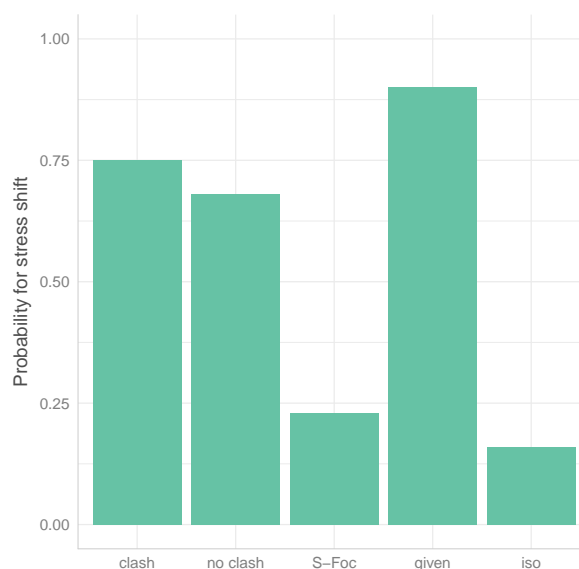


Figure 4.11.: The probability for stress shift by condition in the English L1 data

### 4.3.9. Summary and discussion of results

This experiment aimed at getting insight into how and when the rhythm rule operates in sentences with varying focus structure. Specifically, the question was whether corrective focus overrides stress shift as has been previously assumed (see above).

Example (54) repeats the conditions for ease of reference. Comparing the three conditions in which the target word was focused (*clash*, *no clash*, *S-Foc*), I found that L1 speakers hardly ever shift stress when the following word is given (*S-Foc*). The stress clash on the lexical level is not enough to trigger stress shift of the more prominent focused target word. When a second focus accent followed the target word, however, such as in the double-focus conditions *clash* and *no clash*, stress was shifted most of the time.

(54) *Did Anna say that she met an Indian programmer?*

### 4.3. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

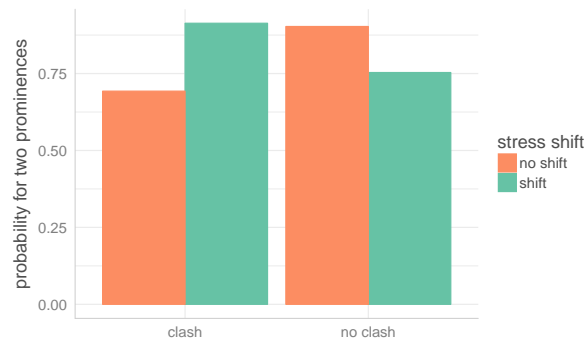


Figure 4.12.: The propability for two pitch accents by condition and stress shift in the English L1 data

Table 4.5.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the pitch accent annotation task of the English L1 data.*

	Estimate	Std. Error	z value	p value
(Intercept)	0.8134	0.7982	1.019	0.3082
cond. no clash	1.4171	0.8798	1.611	0.1072
stress shift	1.5441	0.8110	1.904	0.0569
no clash:shift	-2.6581	1.0914	-2.435	0.0149

No, she said that ...

**clash** ... she met [a JapaNESE] [ARchitect]..

**no clash** ... she met [a JapaNESE] [acCOUNtant].

**S-Foc** ... she met [a JapaNESE] programmer.

*Did Anna say that she met a Japanese programmer?*

No, she said that ...

**given** ... she met a Japanese [ARchitect].

Concerning the first research question whether a corrective focus accent overrides the rhythm rule, it can therefore be specified that it does override the rhythm rule (so stress is not shifted) when it is the last accent in the phrase (the nuclear accent), and that it does not override the rhythm rule (so stress is shifted) when it is followed by another pitch-accented syllable as in the double-focus conditions. It seems that the second prominence needs indeed to be stronger (or equally strong) than the first prominence in order to trigger stress shift, as is also the case in *given*, where speakers almost always shifted the weaker stress to the left. While in *S-Foc*, the weaker prominence in the postnuclear word

is not able to trigger the stronger prominence to shift (cf. Selkirk, 1984).

The fact that the annotators frequently perceived stress shift in the *given* condition suggests that the rhythm rule also operates in deaccented speech speaking against proposals that stress shift phenomena are caused by deleting or shifting a pitch accent (cf. Gussenhoven, 1991; Shattuck-Hufnagel, 1995). A factor possibly contributing to frequent stress shift in this condition, however, may lie in the make-up of the dialogues possibly causing a priming effect. Consider the example dialogue *Did Anna say that she met a Japanese programmer? - No, she said that she met a Japanese ARchitect.* - the speaker of the context question may have shifted the stress on, here, *Japanese* to the initial syllable due the following initial stress on, here, *programmer*, priming an equal stress pattern in the answer sentence where the same word is repeated. Insights gained from this condition may therefore be of limited value.

Having a closer look at the double-focus conditions, it was found that English L1 speakers in this study often shift stress not only when corrective focus accents are located on directly adjacent syllables (as in condition *clash*), but also when the corrective focus accents are separated by one syllable (as in condition *no clash*), similar to findings by Vogel et al. (1995) and Beckman et al. (1990). This suggests that stress shift is not only triggered by directly adjacent lexical stresses on the syllable level, as implied by for example Henrich et al. (2014) or Vogel et al. (1995), but also by prominences of a higher level in the prosodic hierarchy as suggested by Standard Metrical Phonology (Lieberman and Prince, 1977; Selkirk, 1984; Nespor and Vogel, 1986). This experiment could clarify that this includes pitch accents used for focus marking.

Since the context required an emphatic corrective focus accent the results suggest that stress clashes are not only resolved by stress or accent reduction, but that prominence is actually shifted onto the initial syllable even under focus marking and even when the clash is only present above the lexical level.

The fact that stress shift was perceived without its shift triggering context also speaks against the claim that stress shift is a mainly perceptual phenomenon. It has to be noted however, that the high intra-rater agreements but lower inter-rater agreement in determining stress shift indicates that listeners may have different strategies in determining word stress.

Another question was whether the possibility to shift the accented syllable away from the second focus accent yields in more double-accent productions, considering that the speakers in experiment 1, where such repair strategies were not available, often omitted the first focus accent. Looking at this interplay of stress shifting and pitch accenting it was shown that there is a correlation between stress shift and pitch accent placement.

#### 4.4. Conclusion of this chapter

---

While most of the double-focus sentences were in fact realised with two pitch accents (ca. 75%) the two conditions are affected differently: Concerning the *clash* condition, there are more often two prominences when stress is shifted, whereas if stress is not shifted, we find more instances where the pitch accent on the first focused word is omitted. This suggests that alternation preferences may also interact with the prosodic marking of focus by pitch accent placement such that pitch accents are not realized in order to prevent a pitch accent clash. This result, however, barely achieves significance and the proposal needs further testing. Concerning the *no clash* condition, where the adjacent pitch accents are separated by one syllable, the correlation of pitch accent placement and stress shift goes in the other direction - if stress is shifted, speakers more often omitted the first focus accent than when stress was not shifted. As to why this is the case I have no explanation so far. A more fine-grained analysis of the realised pitch accents may be needed to draw further conclusions.

The fact that speakers produced two pitch accents in over 75% of the shiftable contexts but only in about 57% in the contexts of the experiment where stress shift was not an option (experiment 1), however, does support the assumption that English L1 speakers may have omitted pitch accents due to rhythmic reasons in the previous experiment and that the one unaccented syllable between the two focus accents was not enough to dissolve the dispreferred clash.

Another question that was left open after the first experiment was whether speakers used phrase breaks between the pitch accents as a repair strategy in clash contexts, since results obtained by means of manual boundary annotations and duration measurements were inconclusive. The finding that speakers mostly shifted stress in both double-focus conditions suggests that the two focus constituents were in fact mainly realised within one phrase since the rhythm rule is said to only operate when the words participating in the rule are grouped into one prosodic phrase (Nespor and Vogel, 1986; Selkirk, 1984; Liberman and Prince, 1977; Tilsen, 2012).

#### 4.4. Conclusion of this chapter

The experiments in this chapter investigated the interplay of alternation preferences and corrective focus marking in productions by native Australian English speakers. The overall research question of this thesis is whether alternation preferences, that is the preference for an alternation of strong and weak that both German and English share, can account for some of the variation that we find in the realisation of focus. For German, I found that yes, the rhythmic context does indeed affect the marking of corrective focus

in terms of the choice of pitch accent type and the degree of prominence. Pitch accent placement, however, does not seem to be affected in German. Despite the similarities between the two languages in terms of both, rhythmic preferences and prosodic focus marking, English speakers seem to have different strategies in dealing with the conflict between rhythmic and semantic constraints. Alternating the tonal contour to a rise fall pattern (like in German) does not seem to be an option in English clash contexts. Drawing from both English experiments, however, there is evidence that the preference for alternation affects the placement of pitch accents, since focus accents are more likely to be realised when there is the option to shift the pitch accent to the initial syllable. Concerning the interplay of the rhythm rule and focus, it could be shown that the rhythm rule does also operate under corrective focus marking if the focused word is not the last accented word in the phrase. Furthermore, I found that stress is shifted even when there is no stress clash on the lexical level supporting a stress clash definition of the standard metrical account.



## 5. Alternation preferences and focus marking in L2 English

This chapter presents data collected from non-native speakers, namely German learners of English. After having established how alternation preferences and focus marking interact in L1 German and English, the experiments presented in the following are designed to investigate how L2 speakers deal with a possible conflict between pragmatic requirements, rhythmic well-formedness and optional rhythmic repair strategies. The first experiment mirrors the one conducted in German and the first of the two English experiments. The major research question for that experiment is whether alternation preferences under corrective focus marking are transferred from the L1 to the L2. In order to make predictions about possible difficulties or advantages German Learners of English may have in an alternation-focus-marking conflict in English, that is which aspects may be transferred to the L2, the results obtained so far for the two native groups will be reviewed in light of Mennen's L2 Intonation Learning theory as introduced in 2.3.2.

The second experiment presented in 5.4 repeats the shift experiment presented in the L1 English chapter (see 4.3). That is, it again looks at prosodic realisation of focus in different focus and rhythmic environments within rhythm rule contexts. This means that speakers have the possibility to shift the primary stress of the first participant in a stress clash to the initial syllable such as in *thirteen wómen*. The main research question here is: what is the status of the rhythm rule in non-native speech? Are rhythmic repair strategies required in the context of corrective focus, given that reparation is optional?

Regardless of L1 strategies that may be transferable, it is conceivable that alternation only plays a minor role in non-native speech. It has been shown that L2 speakers have a reduced ability to generate expectations (cf. RAGE hypothesis Schafer et al., 2014). Given that rhythm in native speech has a guiding function and helps to create expectations of what comes next (see 2.2.2), repair strategies may not be indicated in the absence of such expectations in non-native speech. Additionally, Tilsen (2012) found that repair strategies in stress clash contexts are more often used in read and prepared speech than

## 5.1. Comparison of German and English results in light of LILt, and predictions for L2 speakers

---

in spontaneous speech. Considering the greater cognitive load for non-native speakers, L2 speech may be comparable to less prepared speech, and therefore less prone to rhythmic adjustments. Other aspects beyond L1 transfer such as the ones just mentioned will be addressed in the discussion sections of this chapter.

Before going into the details of these two experiments, predictions regarding difficulties for learners will be formulated by means of Mennen's L2 intonation learning theory (LILt, Mennen, 2015).

Note that the term L2 is used throughout the thesis to refer to the English spoken by the advanced German learners in the study at hand. There are of course different scenarios in which a language is acquired, as a foreign language via formal instruction, a second language in the course of immigrating to the second language speaking area or as a heritage language. The language learners investigated in the study at hand all live in Germany, and mainly acquired their English skills via formal instruction and sometimes some longer or shorter visits to an English speaking country. Hence, strictly speaking, English is a foreign language to them. For ease of reference I will still refer to them as L2 speakers.

## 5.1. Comparison of German and English results in light of LILt, and predictions for L2 speakers

In order to understand learner difficulties, we need to establish how similar the L1 and the target language is in the phenomena under investigation. Analogous to phonetic similarity to which models of segmental speech learning refer when predicting the relative difficulty to acquire segments, intonational similarity needs to be established. An advantage of the study at hand is that the elicited data for both L1s is as parallel as possible to each other and to the L2 data, and therefore highly comparable which is important in order to make a valid claim about the second language (cf. Rasier and Hiligsmann, 2007).

As introduced in section 2.3.2, Mennen's (2015) L2 Intonation Learning theory (henceforth LILt) encompasses four dimensions along which language similarity can be determined within the autosegmental-metrical framework. This means that languages' intonation can be compared with respect to discrete categories. I will briefly repeat the dimensions here and will then elaborate which dimensions apply in our case.

The first dimension is the *systemic* dimension and covers the inventory and the distribution of, in our case, pitch accents in German and in English. For the comparison of



pitch accent inventories the labelling conventions play an important role, as described in section 4.2.7.2. This is why I will refer to the categories as used in this study. All pitch accent types used by the German speakers were also found in the English realisations. So that the two language groups do not differ in the systemic dimension. In her contrastive analysis on German and English intonation, Grabe (1998) concludes that German and English are indeed very similar in terms of the inventory of intonational categories. Where the languages differ, according to her findings, is the acoustic phonetic realisation of the falling pitch accent ( $H^*+L$  in her analysis) in terms of alignment (with German having a later peak than English). Others have also found that German and English differ in the alignment of the peak (Atterer and Ladd, 2004). That is the second dimension, the *realisational* dimension, more specifically, how these different pitch accent types are phonetically implemented. Fine phonetic details of the contours were not investigated and compared in this study, so that this is not the dimension in which we are interested here. If we consider the relative degree of prominence of the two pitch accents as part of the phonetic implementation, however, we did find differences between the two language groups: Whenever English speakers realised both foci with a pitch accent, these were mostly perceptually judged to be equivalent in prominence, regardless of rhythmic context. German speakers on the other hand often realised the first focus subordinate in prominence to the second focus, especially in clash environments. Since this analysis only gives information on the relation between two pitch accents and not on any acoustic details, contributions to insights into deviations in the realisational dimension are, however, limited.

The third dimension characterises differences and similarities with respect to the functions of the pitch accent types, that is the *semantic* dimension. The function of the pitch accents investigated here is to mark corrective focus. The prosodic marking of focus is very similar in the two languages, as has been established in section 2.1.3. Also the speakers of both language groups in this study reliably realise a single corrective focus (in the two single-focus conditions) in the same way, namely with a high tonal target on the pitch-accented syllable followed by a fall in pitch ( $H^*L$ ), suggesting that the semantic dimension is not at play here. It could be argued that the marking of a double-focus is a different function than single-focus marking. One difference that was found here between German and English was that the first focus may be realised with a rising pitch accent in German (this study and also Wang and Féry, 2018) while the first focus in English is preferably realised with a fall like in single-focus sentences (this study and also Eady et al. 1986).

The effect on pitch accent type in these double-focus sentences is, however, not gov-

erned by meaning, that is the function, but by the presence of a second pitch accented constituent. This is why I do not consider this difference to be a matter of the semantic dimension but rather of the frequency dimension, which is the fourth and last dimension of Mennen's LILt. The frequency dimension refers to differences or similarities languages may have in the frequency of use and distribution of the language's inventory, in our case here, pitch accent types. It is here where German and English have been found to differ, since German speakers often use rises in environments where English speakers would not.

The question is how rhythmic aspects can be integrated into a second language learning model such as LILt. From the literature we know that the two languages share the preference for rhythmic alternation (see section 2.2). The results gained so far suggest, however, that these preferences may affect different aspects of focus marking. In German, the choice of pitch accent type is affected by alternation preferences since the rise is the preferred contour when there is a pitch accent clash, that is two pitch accents on directly adjacent syllables. While in English, alternation preferences possibly affect pitch accent placement – in almost half of the double-focus sentences of experiment 1, that is where stress could not be shifted, English speakers did not accent the first focus, while German speakers marked both foci in more than 70%. German speakers, however, often reduced the first focus accent in prominence, even more often in clash environments, while English speakers, when both foci were prosodically marked, mostly realised two equally prominent pitch accents.<sup>1</sup> It is not straightforward in which dimension rhythmic preferences and their role in influencing the use and distribution of intonation categories should best be sorted. It appears that languages may differ in to what degree they are affected by rhythmic constraints and how and where these preferences are shown. Since the consequences of rhythmic interference are seen in the frequency and distribution of tonal categories – in our case this manifests in pitch accent placement and pitch accent type - it can be argued that rhythmic interference is part of the frequency dimension, as depicted above, but also plays into other dimensions, since they also affect the realisational dimension (pitch accents are reduced due to rhythmic constraints) and the semantic dimension (focus marking changes due to rhythmic constraints). To address similarities and differences between languages with respect to how they are affected by alternation preferences is important since differences in degree and effect of rhythm may be a source for difficulties for L2 intonation learners in so far as they may use the correct

---

<sup>1</sup>As discussed before, this may be a correlate of pitch accent type, assuming that rising accents are less prominent than falls per se and German speakers often used rising accents for the first focus position. I will come back to this speculation in the general discussion (6.1).

inventory for the correct functions with the correct phonetic implementations, but in the wrong context. Learners may, for example, correctly use a falling pitch accent to mark a corrective focus in English, but may use a rising pitch accent when this focus is directly adjacent to another pitch accented syllable thus producing a falsely pattern.

Table 5.1 gives an overview of the results concerning the strategies of L1 German and L1 English speakers in clash contexts under corrective focus marking, pointing out the similarities and differences.<sup>2</sup>

Table 5.1.: Overview of results concerning prevalent repair strategies in L1 German, L1 English.

language	(single) focus marking	repair strategies in clash environments
German	pitch accent with high target and post-focal compression (H*L)	pitch accent <b>placement</b> : no omissions pitch accent <b>type</b> : alternating contours preferred (L*H H*L) pitch accent <b>relation</b> : reduced prominence on 1 <sup>st</sup> focus
English	pitch accent with high target and post-focal compression (H*L)	pitch accent <b>placement</b> : omissions seem possible pitch accent <b>type</b> : no alternating contours (H*L H*L) pitch accent <b>relation</b> : no reduced prominence on 1 <sup>st</sup> focus (when realised)
	→ same	→ different

Now that we have established similarities and dissimilarities between German and English as gained from the previous experiments, what does that mean for our L2 speakers? The advantage of classifying cross-linguistic similarities and dissimilarities with LILt according to Mennen is to predict and “identify where deviations from the native form occur and whether they occur more frequently in some than in other dimensions” (2015:175). As outlined earlier, LILt shares the assumption of Flege’s Speech Learning Model (1995) that similar (but not identical) categories are more difficult to acquire than very differ-

<sup>2</sup>Note that the results on phrasing are not included here. As discussed earlier, for both language groups results from the manual boundary annotation and the word duration measurement remain inconclusive in the sense that I do not want to draw any further conclusions from there at this point.

## 5.2. General remarks on the statistical analyses of this chapter

---

ent categories. Different categories are much easier for learners to identify as such and consequently establish a new category in their interlanguage, while similar categories may not even be perceived as being not the same and therefore sorted into the already existing respective L1 category.

Considering how close the German and English intonation systems are with respect to the phenomena discussed above, ultimate attainment may be difficult for German learners of English when it comes to choosing the right contour in the right context. This means that transfer effects in the choice of pitch accent types are to be expected. More specifically, L2 speakers may transfer the use of rises in the first focus position. A second question is what role rhythmic preferences play in the L2 given that repair strategies are optional in both languages albeit apparently to different degree. While German allows the alternation of pitch accent types in clash contexts to prosodically mark two adjacent corrective foci, this does not seem to be an option in English. Do alternation preferences exhibit the same effect as in the speakers' L1 or are there differences in ranking when it comes to a conflict between the prosodic marking of information structure and the preference for alternation in the L2?

In order to investigate how L2 speakers realise double-focus sentences in different rhythmic environments, the exact same experiment was conducted with L2 speakers of English. That is material and analyses were the same as for the English L1 group. The results for the replication of the first experiment are presented in the following sections. After that I will present the results for the replication of experiment 2, that is, double-focus and single-focus sentences with potential stress-shift items.

## 5.2. General remarks on the statistical analyses of this chapter

All statistical analyses of this chapter were performed in R 3.5.0 (R Core Team, 2018), using the function `glmer` with the packages `lme4` (Bates et al., 2013). For data processing and visualisation the R packages `'tidyverse'` (Wickham, 2017), `'plyr'` (Wickham, 2011), `'ggplot2'` (Wickham, 2009), and `'jtools'` (Long, 2018) were used.

I used an alpha level of 0.05 for all statistical tests.

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

The data for the current study was elicited via a reading production experiment. It replicates the experiment conducted with L1 English speakers, which itself was designed to be maximally comparable to the German study presented in chapter 3. Thus, stimuli and analysis procedure are the same as explained in the previous chapter (see section 4.2). For ease of reference, the stimulus design is repeated in the following.

#### 5.3.1. Stimulus design

The double-focus conditions are repeated in (55) and the single-focus conditions in (56).

(55) Did Carl say that Clara gave the boys horror stories to read?

No, he said that...

**clash:** ... she gave [the GIRLS]<sub>NP1</sub> [ROMance novels]<sub>NP2</sub>.

**no clash:** ... she gave [the GIRLS]<sub>NP1</sub> [adVENTure books]<sub>NP2</sub>.

(56) Did Carl say that Clara gave the boys horror stories to read?

No, he said that...

**F1:** ... she gave [the GIRLS]<sub>NP1</sub> [horror stories]<sub>NP2</sub>.

**F2:** ... she gave [the boys]<sub>NP1</sub> [ROMance novels]<sub>NP2</sub>.

#### 5.3.2. Participants

Sixteen (2 men, 14 women) German learners of English (henceforth L2 English) participated in the experiment. Their mean age was 23.25 (range: 19-28). They were all students at the University of Stuttgart. Fourteen speakers were students of English, ten of them within a degree programme leading to secondary teacher accreditation. One speaker studied German and one Linguistics. They all indicated German as their native language. In Germany, English is typically introduced as a school subject in the first year of the secondary school (at around age 11)<sup>3</sup>, this means that all participants have been learning English as a foreign language for at least 7 years. According to their indication in the questionnaire, which they all filled out after the reading task, they were proficient to highly proficient in writing, reading, speaking and listening to English. They all found

---

<sup>3</sup>In some areas it is already introduced in primary school at age 7.

the task rather easy (1-3 on a scale from 1-7). When asked to assess their own performance in the reading task on a scale from 1 (very bad) to 7 (very good), the majority of speakers (11) rated themselves with 5-7, the remaining 5 rated themselves from 2 to 4 (even though they all found the task rather easy, see above). None of the participants had known speech or reading disorders. All participants were naïve as to the purpose of the experiment. They were paid for their participation.

#### 5.3.3. Procedure, annotations and analyses

The L2 recordings took place in a sound attenuated chamber at the University of Stuttgart. Data elicitation, annotations and statistical analyses followed the same procedure as in the L1 experiment described above and are therefore not repeated in detail here. In order to avoid effects caused by listening to Australian English as a variety of English to which German speakers may be less accustomed, the context questions were produced by an American English native speaker (while it was an Australian speaker in the experiment with Australian speakers).<sup>4</sup> Fourteen recordings were excluded due to hesitations, repetitions, or misperceptions of lexical word stress for example in names (e.g. 'Lucille instead of Lu'cille). One sentence was not recorded due to a technical problem. This leaves us with 305 recordings that were analysed.

#### 5.3.4. Analysis of pitch accent placement

The L2 data was analysed in the same way as the German L1 and the English L1 recordings. That is three annotators (the same as for the English L1 data) judged which word or words are prominent in the sentence. Each recording was played three times to each of the annotators. Again, I considered a word to be pitch accented when at least four out of nine annotations marked the word as prominent (see 3.4.1 for more details on this procedure).

##### 5.3.4.1. Intra-annotator and inter-annotator agreement

For the L2 data, the intra-annotator agreements were  $\kappa=0.8$  for annotator one,  $\kappa=0.74$  for annotator two and  $\kappa=0.78$  for annotator three. This means that all annotators agreed substantially to almost perfectly within their own ratings also for the rating of L2 recordings. The inter-annotator agreements for the prominence judgement task was  $\kappa=0.59$ ,

---

<sup>4</sup>It can be assumed that German speakers have more encounters with American English, especially via media influence, see page 102, where this has been addressed.

which means that the agreement across annotators on the prominence patterns of the recordings was moderate (cf. Landis and Koch, 1977).

### 5.3.4.2. Statistical analysis

In order to investigate the effect of focus and rhythmic environment on pitch accent placement, I performed a generalised linear mixed model using the logit link function. The number of pitch accents (one or two) was the dependent variable, as fixed factor I included condition with the levels *clash*, *no clash*, *F1* and *F2*. As random factors I included intercepts for participants and items. An overall effect for condition was tested for significance by using likelihood ratio tests comparing the model including *condition* to the model without it (cf. Baayen, 2008; Baayen et al., 2008; Winter, 2013). The model with an AIC value (Akaike's information criterion) of at least two points smaller was considered the better model (c.f. Burnham and Anderson, 2002).

### 5.3.4.3. Results

There is an overall effect of condition in the L2 data, as determined by the likelihood ratio test ( $p < 0.0001$ ). Table 5.2 gives an overview of the effects of the different levels as predicted by the final model. The intercept (first line) refers to condition *clash* to which the other conditions are compared. The estimates here are log-odds. Figure 5.1 shows the results for the number of prominences as annotated by condition. On the x-axis we see the four experimental conditions and on the y-axis the probability for two prominences (i.e. both foci were perceived as prominent).

Looking at the control conditions *F1* and *F2*, where the answer sentences have only one contrastive focus, we can see that single-focus sentences are mostly produced with only one pitch accent, although *F2* receives more often (48%,  $\beta = -2.06$ ,  $SE = 0.45$ ,  $p < 0.0001$  compared to *clash*) two prominences than *F1* (27%,  $\beta = -3.29$ ,  $SE = 0.50$ ,  $p < 0.0001$  compared to *clash*). In double-focus sentences, English L2 speakers produced a pitch accent on **both** focused constituents in most of the cases (83% in *clash* and 80% in *no clash*), the rhythmic manipulation did not show an effect on pitch accent placement here ( $\beta = -0.17$ ,  $SE = 0.46$ ,  $p = 0.70$ ).

Figure 5.2 shows which of the two noun phrases (NPs) speakers accented.<sup>5</sup> Looking at the single-focus sentences, we can see that speakers mostly realised a pitch accent

---

<sup>5</sup>Note that the numbers in this graph deviate slightly from the numbers as presented in figure 5.1. This is because figure 5.2 comprises the mean of all ratings by the annotators, while figure 3.3 represents the 4 out of 9 approach thus having only one rating per sentence.

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

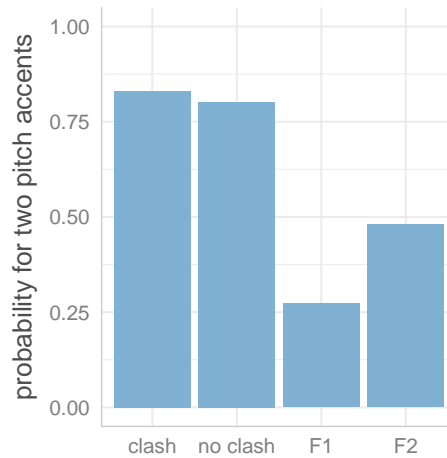


Figure 5.1.: Number of pitch accents by condition (x-axis) in L2 English. The probability for two pitch accents (on both foci) is on the y-axis.

Table 5.2.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the first annotation task of the English L2 data.*

	Estimate	Std. Error	z value	p value
(Intercept)	2.0227	0.4748	4.260	<0.0001
cond. no clash	-0.1746	0.4636	-0.377	0.706
cond. F1	-3.2937	0.4975	-6.621	<0.0001
cond. F2	-2.0562	0.4515	-4.554	<0.0001

on the first NP in *F1* and deaccented the given constituent NP2. There are quite a high number (48%) of pitch accents on the given NP1 in condition *F2* (which will be discussed below).

Looking at the double-focus conditions we can see that whenever there was only one pitch accent realised, it was almost always realised on NP2. There are two cases, labelled as “other”, in which neither NP1 nor NP2 was perceived as prominent.

#### 5.3.5. Analysis of pitch accent type

##### 5.3.5.1. Procedure

As outlined above, the L2 recordings were annotated using the same labels as in the German data and the English L1 data. In four cases I was not entirely sure as to the type of pitch accent. These cases were played to a second phonetician experienced



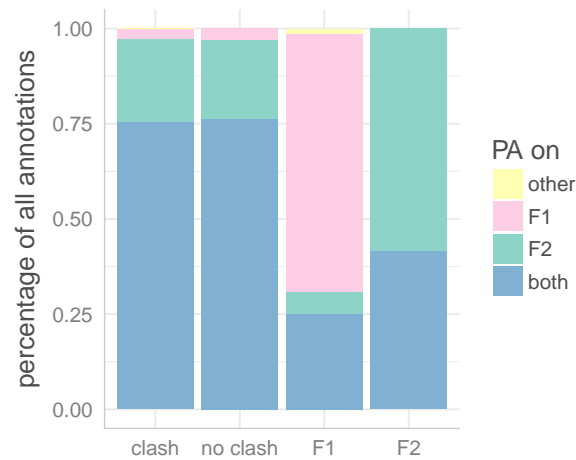


Figure 5.2.: English L2: Distribution of pitch accents by condition (x-axis) on object 1, object 2 or on both objects.

in the annotation of intonation and discussed until a decision was reached.

There was one sentence in the L2 group that was produced with a rising contour on the second object, and two cases without a pitch accent on NP2. Due to the low number, I excluded these three cases from all following graphical and statistical analyses for reasons of readability. This left 64 sentences of condition *clash* and 58 sentences in condition *no clash* that were statistically analysed.

### 5.3.5.2. Statistical analysis and results

Figure 5.3 shows the distribution of contours across the two double-focus conditions and Table 5.3 gives the proportional frequency of the pitch accent types used by the speakers per condition. We can see that, like in the English L1 data, the double-peak contour is the preferred realisation in both double-focus conditions. However, contrary to L1 English speakers, L2 speakers also frequently used the rise-fall contour, like the German L1 speakers. Similar to the German data, it looks as if condition *clash* increases the likelihood for high-fall and rise-fall productions.

In order to test statistically whether the conditions are indeed significantly different, I conducted a generalised linear mixed model using the logit link function. I used the probability for *double-peak* as dependent variable (binomial, 1 or 0) and included *condition* (*clash* and *no clash*) as fixed factor. As random factors I included an intercept for *item* and for *subject*. A model comparison using a likelihood ratio test comparing the final model and the null-model (without the fixed factor *condition*) revealed that the

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

Table 5.3.: *Frequency of pitch accent types across conditions (it reads as <pitch accent on NP1>. <pitch accent on NP2>). The preferred pitch accent type in each condition is highlighted in bold.*

	H*L.H*L	H*.H*L	L*.H.H*L	NONE.H*L
clash	<b>39.06</b>	20.31	31.25	9.38
no clash	<b>63.79</b>	6.90	27.59	1.72

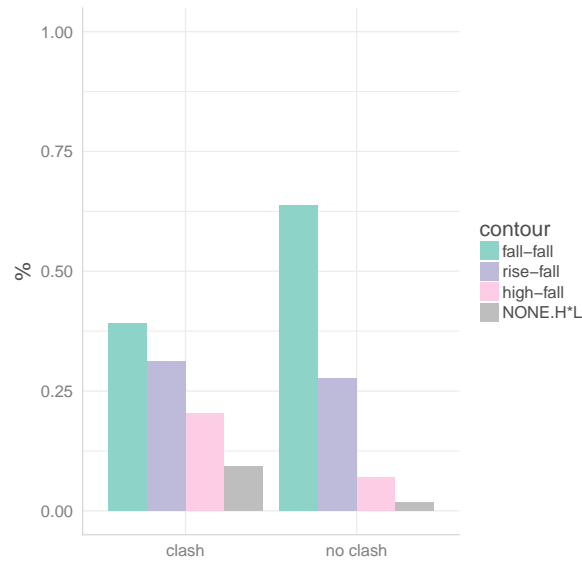


Figure 5.3.: The distribution of pitch accent types across conditions

two conditions are indeed significantly different ( $p < 0.01$ ). Table 5.4 gives an overview of that final model.

Table 5.4.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear model predicting the presence of the double-peak contour by condition. Intercept is condition clash.*

	Estimate	Std. Error	z value	p value
(Intercept)	-0.5040	0.3056	-1.649	0.09910
cond. no clash	1.1436	0.4189	2.730	<0.01

#### 5.3.6. Analysis of pitch accent relation

When English L1 speakers marked both foci with a pitch accent, the two accents were mostly equivalent in prominence, regardless of rhythmic environment. German L1 speakers, however, realised the first focus accent subordinate to the second focus pitch accent,

although the accents were more often equal in rhythmically regular contexts. English speakers, however, almost always used falls to mark the foci.<sup>6</sup> What does this mean for L2 speakers? The previous analyses revealed that L2 speakers realise most double-focus sentences with two pitch accents. They mostly chose a double-peak pattern like English L1 speakers, but more often in rhythmically regular sentences (*no clash*), like German L1 speakers. Unlike in L1 English there is also a high number of rising accents in first focus position, like in L1 German. For the relation between these pitch accents this can mean that German learners may realise a prenuclear accent on the first focus like they do in their L1, especially in clash environments.

### 5.3.6.1. Procedure

In the exact same way as in the other previously presented production experiments, all sentences that were realised with two prominences were again played to the three annotators. They judged each recording three times in randomised order without the context question and blind to the condition. The two adjacent focused NPs were shown on the screen and the annotators were instructed to mark which of the words is more prominent or whether they are the same. They chose the more prominent item by clicking on the respective NP. If they considered both NPs to be equivalent in prominence, they could click on the button “equal”. If necessary, the annotators could listen to the recordings up to ten times before making a decision. They were instructed to take two breaks during the task. This task was implemented in praat using ExperimentMFC (Boersma, 2001), see Figure 4.6 on page 114. I considered an item to be more prominent when it was marked as more prominent by at least five out of nine annotations (see 3.6 for more details on this decision).

### 5.3.6.2. Inter-annotator and intra-annotator agreement

Agreement between the annotators as well as between the individual annotations within each annotator were again calculated using Fleiss’ Kappa in the same way as described in section 4.2.6.1. The three labels on which agreement was measured are “prominences are equally strong”, “NP1 is stronger” and “NP2 is stronger”. The inter-annotator agreement for this task was  $\kappa=0.17$ , which, on a scale from 0 to 1 (0 meaning chance level) means that there was only “slight agreement” between the annotators (cf. Landis and Koch, 1977). Intra-annotator agreement was  $\kappa=0.64$  for the first annotator (substantial

---

<sup>6</sup>A possible correlation between intonation contour and prominence relation could not be tested statistically unfortunately, due to sparse data.

agreement),  $\kappa=0.28$  for annotator two (fair agreement) and  $\kappa=0.43$  for annotator three (moderate agreement). These results are similar to the kappa values for the same tasks in L1 German and English and suggest that the task of marking the relational prominence was more difficult than the previous task of marking where prominences are. I will resume to this issue in the discussion section.

#### 5.3.6.3. Statistical analysis

In order to analyse the relationship between rhythmic environment and prominence relation of the focus constituents, I conducted a generalised linear mixed model using the logit link function with prominence (*same*, or *not same*) as the dependent variable. As fixed factors I tested condition (*clash*, *no clash*), as random factors I included intercepts for subject and for item. Main effects for the fixed factor were tested via likelihood ratio test comparing the model with the factor to the model without the factor. If the model containing the factor had an AIC-value of at least 2 points smaller than the model without the factor and a p-value  $<0.5$  the factor was considered significant. I excluded the cases where I could not perceive a particular pitch accent type on one of the NPs from the statistical analysis.

#### 5.3.6.4. Results

Figure 5.4 shows the effect of condition (x-axis) on the probability for two equally prominent focus accents (y-axis). In the majority of cases the two pitch accents were marked as being equally prominent, namely in 86.21% of the *clash* sentences and in 89.47% of the *no clash* sentences. The factor *condition* did not significantly improve the model: Comparing the model with condition to the null model via likelihood ratio test revealed that there is indeed no difference between the two conditions ( $\chi^2(1)= 0.28$ ,  $p=0.89$ ). When the prominences are not the same, the one on NP2 is the stronger one (see Figure 5.5)

#### 5.3.7. Analysis of Phrasing

In order to understand whether speakers used phrase boundaries in the double-focus sentences, two approaches were taken. Phrase boundaries were manually annotated and word duration as acoustic cue was measured.

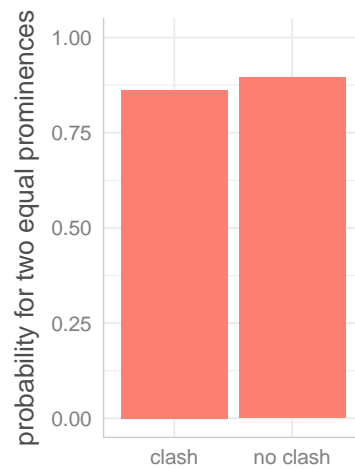


Figure 5.4.: The effect of condition (x-axis) on the probability for two equally strong pitch accents (y-axis).

### 5.3.7.1. Boundary annotation

Intermediate phrase boundaries and intonation phrase boundaries were annotated alongside pitch accent type on all double-focus recordings with two pitch accents by one annotator (myself) based on my auditory impression. The question is whether L2 speakers insert a phrase boundary in clash environments in order to create space between the two adjacent focus accents.

**Statistical analysis** A generalised linear mixed model (logit link function) with the presence of an ip (yes/no) as dependent variable and condition (*clash*, *no clash*) as fixed factor, and an intercept of subjects and for item as random factors was conducted in order to test whether there was a difference in the probability for an ip between the conditions. A main effect for condition was determined by using a likelihood ratio test comparing the model with condition to the null model without condition.

Table 5.5 shows the distribution of intermediate phrase boundaries by condition. In 73.44% of the *clash* sentences with two accents and in 54.10% of the *no clash* sentences with two accents speakers realised an intermediate phrase boundary. An intonation phrase boundary was not used by any of the speakers. The difference between the two condition was significant, condition *clash* increased the probability of an ip-boundary ( $\beta=1.19$ ,  $SE=0.42$ ,  $p<0.05$ ). Table 5.6 gives the output of the final model.

None of the speakers realised the two foci within two intonation phrases with a high boundary tone, as it was sometimes found for longer sentences in Wang and Féry's study

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

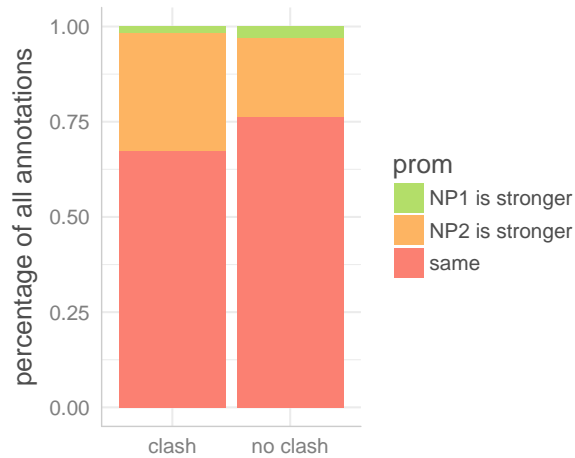


Figure 5.5.: Distribution of the strongest prominence by condition (x-axis): NP1, NP2 or both NPs are the same.

Table 5.5.: *Frequency of intermediate phrase boundaries by condition.*

	ip	no ip
clash	<b>73.44</b>	26.56
no clash	<b>54.10</b>	45.90

on German double-focus.

#### 5.3.7.2. Word duration

In the same way as in the other experiments, word duration of the first foci in the double-focus conditions were measured in order to substantiate the manual phrase-boundary annotation with evidence derived from acoustic cues.

Therefore, utterances were automatically segmented into words (Rapp, 1995), and word duration of the focused words in NP1 was extracted by means of the speech synthesis system Festival (Festival, 2010). In order to find possible acoustic effects of phrase

Table 5.6.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear model predicting the presence of an ip-boundary by condition. Intercept is condition clash.*

	Estimate	Std. Error	z value	p value
(Intercept)	1.1906	0.3869	3.077	<0.01
cond. no clash	-0.9859	0.4207	-2.343	<0.05

boundaries after the first focus in double focus conditions, duration of that first focus was compared to the duration of the focused word in the F1 condition, that is where no second focus is following and thus no boundary is expected. (See example (45) on page 117 for more details.)

**Statistical analysis** A linear mixed model was conducted with word duration as dependent variable and condition as fixed factor. As random factors I included an intercept for subject and for item. A main effect for condition was tested by means of a likelihood ratio test, as in the other analyses.

**Results** The difference between the conditions did not reach significance ( $\chi^2(1)=3.08$ ,  $p=0.2$ ) so that the results regarding word duration as an indicator for a prosodic boundary cannot reinforce the manual boundary annotation.

### 5.3.8. Summary and discussion

The previous sections investigated the interplay of alternation preferences and focus marking in L2 English. Given the general similarity between German and English but the different strategies to deal with a possible conflict between the preference for alternation and the prosodic marking of focus, it has been predicted that German learners of English will have difficulties here, transferring their L1 strategies in clash contexts which leads to deviations in L1L2's frequency dimension. Deviations were especially predicted for the marking of the first focus with a specific pitch accent type, since it is here that the two languages differ the most. (While English speakers quite robustly realised both focus accents with a high tonal target on the accented syllable, German speakers often used a rise with a low tonal target on the first of the clashing focus accents in order to arrive at a tonal alternation.)

The following section will discuss the L2 results in light of these predictions.

#### 5.3.8.1. Pitch accent placement

For a better overview and comparison of the results, Figure 5.6 repeats the prominence annotation results for all language groups. Following the sequence of presentation, 5.6a shows the L1 German group, 5.6b the L1 English group and 5.6c the results for the language learners.

Similar to both the L1 German and the L1 English group, the L2 learners were not influenced by the rhythmic manipulation in their placement of pitch accents in double-

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

---

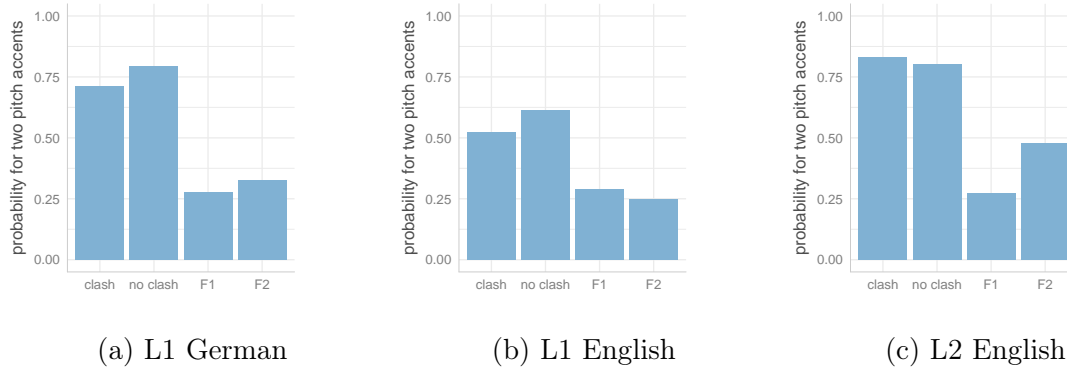


Figure 5.6.: Number of pitch accents for the three language groups (a) L1 German, (b) L1 English and (c) L2 English. The conditions are on the x-axis and the probability for two pitch accents on the y-axis.

focus contexts. The vast majority of sentences in both conditions (*clash* and *no clash*) were realised with two pitch accents, one on each focus. With over 80% of double-accent productions, the L2 speakers realised more pitch accents than the German L1 group (about 75%) and a lot more than the English L1 group (about 57%). The prosodic marking of focus seems to be a very strong constraint for L2 speakers (like following a motto “function first!”). Producing a higher number of pitch accents compared to L1 speakers has been found in several L2 studies (cf. Rasier and Hiligsmann, 2007). Gut (2009), for example, reports on two studies who found that L2 learners of English produce more pitch accents than native speakers (Juffs, 1990, for Chinese learners of English; and Grosser, 1997, for Austrian learners of English). In Gut (2000) she reports on findings that German learners of English produced as many pitch accents as the native speakers, but on different elements. This is, however, not the case here: While the learner group realises more pitch accents than the native group, these realisations are not erroneous but reflect the prosodic marking as required by the information structure. If this would coincide with an overuse of pitch accents we would expect a high number of double-accent productions also in the single-focus conditions *F1* and *F2*. So let us look again at how these sentences were realised: in the majority of sentences with focus on NP1 and given NP2, L2 speakers accented only the focused word and left the given NP2 unaccented (73%), being in line with both native language groups. The double-accent production in the other single-focus condition where NP1 is given and focus is on NP2, is, however, higher (48%) than it was in both native groups (33% for L1 German and 25% for L1 English). A prenuclear accent on given material, especially in longer stretches such as in the stimuli at hand, is not uncommon, as already outlined before. Although the even



higher number of pre-nuclear accents in this context may be attributed to a tendency for L2 learners to accentuate given information, as observed for learners of different L1s (cf. the overview of studies in Gut (2009) who cites for this effect Grosser (1982, 1997) for German learners, Ramirez Verduro (2002) for Spanish learners, Wennerstrom (1994) for Spanish, Japanese and Thai learners; Udofot (1997) for Nigerian English).<sup>7</sup>

### 5.3.8.2. Pitch accent type

Most difficulties for L2 learners were expected in the choice of pitch accent type. While the pitch accent inventories of German and English are identical in this study, they differ in frequency and distribution of the different types. In this study's experiment on native German speakers it has been found that German speakers have the option to realise the first focus in a double-focus sentence with a rising pitch accent, which they preferably chose to do in contexts of pitch accent clash. In the corresponding experiment with native Australian English speakers it was found that this is not an option for English speakers regardless of rhythmic context. Given the similarity in the other dimensions it was hypothesised that ultimate attainment for L2 speakers will be difficult in the frequency dimension since they may map the use of rises to contexts where this is not indicated in English. Additionally, it was hypothesised that they may be guided by alternation preferences in the choice of pitch accent type resulting in more rises in clash environments. Figure 5.7 shows the results for all language groups. Again, the German speakers are on the left (5.7a), the L1 English results in the middle (5.7b) and the learner results on the right (5.7c). The results revealed that L2 speakers are somewhat in between the L1 groups in terms of pitch accent type distribution. They preferred falling pitch accents (fall-fall pattern) in both double-focus conditions like the English L1 speaker, but also frequently used rises as the first focus accent, and significantly more often in clash contexts. It seems as if sometimes L2 speakers reverted to their native repair strategy in pitch accent clash contexts in order to restore tonal alternation using a pitch accent that is not licensed in English. Using the wrong contour can lead to misinterpretation on behalf of the listeners. Braun and Tagliapietra (2011), for example, found that highly proficient German learners of Dutch interpret both the double-peak pattern and the hat pattern contrastively even though only the double-peak pattern has contrastive meaning in Dutch. Future work will investigate how these alternation effects on the choice of pitch accent type will alter the intended meaning and interpretation by

---

<sup>7</sup>It has to be noted, however, that the German learners as investigated by Grosser (1982, 1997) were school children and thus rather early beginners and not directly comparable to the highly proficient speakers in the study at hand.

### 5.3. Experiment 1: Alternation preferences and corrective focus in L2 English

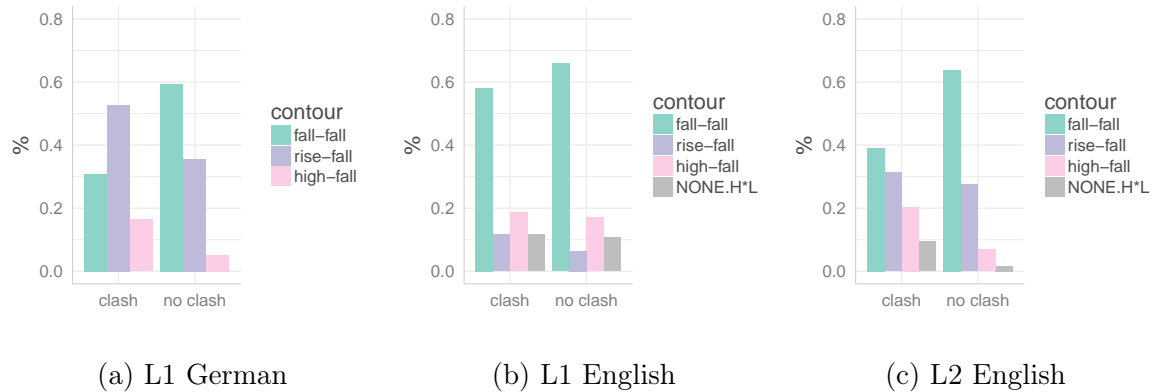


Figure 5.7.: Pitch accent types per condition as used by the different language groups (a) L1 German, b) L1 English and c) L2 English. The conditions (*clash*, *no clash*) are on the x-axis and how often the contours were used in percent are on the y-axis.

the native listener.

#### 5.3.8.3. Pitch accent relation

Figure 5.8 gives an overview of the results regarding the prominence relation between the two focus accents in the double-focus condition. While German speakers mainly realised the first focus accent as subordinate (prenuclear) to the second focus accent, the prominence of the first accent (if realised) in English was not affected by the presence of a second focus, replicating findings by Eady et al. (1986). English L2 speakers, too, reliably realised two equally prominent pitch accents on both foci, regardless of the rhythmic environment. This indicates that L2 English speakers have acquired the prominence relations of English double-focus. The fact that the two focus accents were perceived as equally prominent in over 80% of the time even though speakers frequently used rises (in over 28% of the time) suggests that the perceived prominence is not solely guided by accent type, as was speculated earlier. The L2 speakers are also the group that most reliably realises two pitch accents for both foci in the double-focus conditions (see above). More elaborate phrase boundaries may enhance the effect, see below.

Note, however, that we cannot deduce from this finding that L2 speakers realised the two foci especially prominent. We can only say that the two focus accents are perceived as equally prominent. Phonetic analyses will be needed in order to see how saliently they marked the foci compared to the environment.

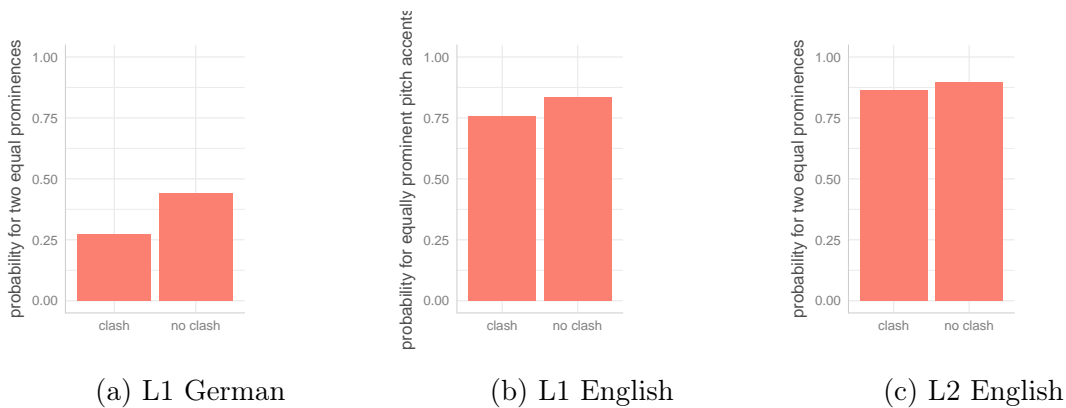


Figure 5.8.: Pitch accent relation per condition as used by the different language groups (a) L1 German, b) L1 English and c) L2 English. The conditions (*clash*, *no clash*) are on the x-axis and the probability for two equally prominent foci are on the y-axis.

#### 5.3.8.4. Phrasing

While clash contexts were not found to be different from no clash contexts in terms of boundary placement in L1 German and L1 English, L2 speakers were found to realise an intermediate phrase break significantly more often in the clash conditions – that is, when focus accents were realised on adjacent syllables. Repairing a clash situation by pause insertion or syllable lengthening is discussed in metrical phonology as a strategy in dispreferred clash contexts (cf. Selkirk, 1984; Liberman and Prince, 1977). This is why it was hypothesised that speakers may use a phrase boundary between the two adjacent focus accents. Since the L2 speakers almost always realise both focus accents and with equal prominence and mainly with the same accent type, it seems like the main strategy resolving the clash for L2 speakers is adding a phrase boundary between the focus accents. Considering that this context is particularly difficult to, on the one hand, process, and on the other hand, articulate this is not all too surprising. Additionally, it has been found that reverting to duration as a cue is a universal strategy for L2 learners (Bohn, 1995). Comparing the results to the other language groups shows, however, that intermediate phrase breaks were also annotated in the majority of L1 German and L1 English double-focus sentences (> 70% in German and > 60% in L1 English). While there was no difference between the two double-focus conditions in these two groups, L2 speakers actually realised less phrase breaks in the no clash condition (ca. 54%). This is not surprising in the sense that it is the condition where we assumed that a phrase break is not necessary given the syllable between the focus accents, but it is surprising that it is the L2 group that actually shows this effect, given that L2 speakers have shown

to realise shorter phrases than native speakers (e.g. Gut, 2009, 2000; Grosser, 1993).<sup>8</sup> Again, I have to note, however, that the annotation of phrase breaks in these contexts is difficult due to the emphatic pitch accent context and, thus, may be flawed. An effect of phrasing as found with the manual annotation could also not be validated by the word duration measurements, so that results on phrasing are, again, inconclusive.

### 5.3.8.5. Phonological structure of the L2 English double-focus realisations

Examples (57) to (59) represent the phonological structure of the most common realisations of the two double-focus conditions by the German learners of English. 57 is an example of a sentence in the *clash* condition which L2 speakers mainly realised with an intermediate phrase break between the two foci (ca. 75%). Also the majority of the *no clash* condition was realised that way, but with ca. 54% significantly less often. The remainder of the sentences in this condition were realised within one intermediate phrase as depicted in example (58).<sup>9</sup> In both the two-phrase and the one-phrase version, speakers preferably realised a falling pitch accent on both foci. Both focus accents were perceived as equally prominent in the vast majority of the double-focus sentences (in over 86% of the time).<sup>10</sup> After the fall-fall pattern as the most frequently used contour, speakers also often used a rising pitch accent on the first focus, especially in *clash* contexts, as illustrated in (59) (as found in the German data).

		×	×		
		×	×		Intonation phrase
		×	×		intermediate phrase
×		×	×		prosodic word
(57)	[(... gave the girls) <sub>ip</sub> (romance novels) <sub>ip</sub> ] <sub>IP</sub> .				
		H*L	-	H*L	
		×	×		Intonation phrase
		×	×		intermediate phrase
×		×	×		prosodic word
(58)	[( ... gave the girls adventure books) <sub>ip</sub> ] <sub>IP</sub> .				
		H*L	H*L		

---

<sup>8</sup>Gut (2009) finds, however, that non-native phrasing very much depends on the speaking style with shorter and more often incomplete phrases in free speech compared to read speech.

<sup>9</sup>Assuming the results as obtained by the manual boundary annotation, but see discussion above.

<sup>10</sup>Note, however, the low inter-annotator agreement discussed above which may discount this result to some extent.

	×	×	
	×	×	
×	×	×	

(59) [(... gave the girls)<sub>ip</sub> (romance novels)<sub>ip</sub>]<sub>IP</sub>.

L\*H - H\*L

Intonation phrase  
intermediate phrase  
prosodic word

### 5.3.8.6. Comparing the results of all language groups – overview

Table 5.7 summarises the results as obtained for all language groups and provides an overview of the comparative analysis for this experiment. It shows the different phenomena under investigation, namely pitch accent placement, pitch accent type, pitch accent relation and phrasing in double-focus environments and names the most common repair strategies used by the three language groups in clash contexts. Note that the results on phrasing are given in parentheses since further investigations are needed to substantiate these results.

Table 5.7.: Overview of results with prevalent repair strategies in L1 German, L1 English and L2 English. PA = pitch accent.

	German	L1 English	L2 English
PA placement	no omissions	omissions seem possible	no omissions
PA type	alternating contours preferred (L*H H*L)	no alternating contours (H*L H*L)	no alternating contours (H*L H*L) preferred, but alternation frequently used (L*H H*L)
PA relation	reduced prominence on 1st focus	no reduced prominence on 1 <sup>st</sup> focus accent	no reduced prominence on 1 <sup>st</sup> focus accent
Phrasing	(frequent ip breaks in both conditions)	(frequent ip breaks in both conditions)	(more phrase breaks in clash contexts)

## 5.4. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

From the previous experiment we learnt how L2 speakers mark adjacent foci in rhythmically dispreferred sentences. Experiment 2 will now have a closer look at whether, and how, L2 speakers use optional repair strategies in clash contexts under corrective

#### 5.4. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

---

focus marking. The second experiment therefore replicates experiment 2 as described in section 4.3 investigating the application of the rhythm rule under corrective focus marking by L2 speakers.

The rhythm rule, that is shifting the lexical stress to another syllable in order to avoid stress clashes, is not unfamiliar to German speakers. While in English stress shift is only possible to the left, that means stress is shifted to an earlier syllable, shift in German is possible in both directions, so it can also be moved to the right. This may happen in phrases, like *anziehen* → *den Röck anziehen*, compounds, like *Báhnhof* → *Háuptbahnhòf* and after prefixation (*síchtbar* → *únsichtbàr*), where the now secondary-stressed syllable moves to the right away from the new primary-stressed syllable (Kiparsky, 1966; Wagner and Fischenbeck, 2002; Bohn et al., 2011). While findings by Bohn et al. (2011), obtained from an ERP-study, showed that German speakers are perceptively sensitive to rhythmic irregularities in rhythm rule contexts, Wagner and Fischenbeck (2002) found that when looking at production stress shift is not a very regular phenomenon in German, suggesting that stress shift is less grammaticalised than in English. Given that German speakers are less prone to use the option compared to English speakers, we may not find stress shift in rhythm rule contexts to affect such a strong prosodic function as the marking of corrective foci. In the previous experiment I found that our L2 speakers' realisation of focus did not seem to be affected by rhythmic manipulations of the context. This suggests that optional rhythmic adjustments may play a minor role in L2 speech. Tilsen (2012) found that repair strategies in stress clash contexts are more often used in read and prepared speech than in spontaneous speech. While it was read speech that was under investigation here, L2 speech may be comparable to less prepared speech, and therefore less prone to rhythmic adjustments, considering the greater cognitive load for non-native speakers. This may additionally contribute to less frequently used rhythmic adjustment strategies.

Little is known about the role of alternation preferences in L2 so far. We know, however, that non-native speakers tend to divide their utterances into smaller phrases (Gut, 2009, 2000; Grosser, 1993). The rhythm rule, however, only applies within one phrase so that a prominence across the phrase boundary is not expected to trigger stress shift (Nespor and Vogel, 1986; Selkirk, 1984; Liberman and Prince, 1977; Tilsen, 2012). If this is so we would expect that speakers do not shift stress given that the respective stressable word would be the last word in the phrase. Since the L2 speakers in this study are assumed to be highly proficient (university students), the universal observation of shorter phrases in L2 speech may not preponderate. Therefore, it will be interesting to see whether L2 speakers of English whose L1 is German a) apply the rhythm rule and

b) in which contexts they make use of it.

Following this, the more specific research questions for this experiment are the following:

1. Are rhythmic repair strategies acquired in the context of corrective focus?
2. Are L2 speakers more likely to produce two focus accents when there is a possibility to shift the pitch accents to an earlier syllable?
3. If L2 speakers apply the rhythm rule, how do they make use of stress shift in different information-structural environments?

### 5.4.1. Method

The data for the current study was elicited via a reading production experiment together with the data presented in the preceding section (see the first experiment in this chapter presented in 5.3).

### 5.4.2. Stimulus design

The experimental stimuli were the exact same as the stimuli for the experiment 2 conducted with the English L1 speakers in Sydney (chapter 4). They are repeated here for better reference. For details on the design and for what factors the sentences were controlled, the reader may be referred to section 4.3.2. (All stimuli for all experiments are in the appendix.)

(60) *Did Anna say that she met an Indian programmer?*

No, she said that ...

**clash** ... she met [a JapaNESE] [ARchitect].

**no clash** ... she met [a JapaNESE] [acCOUNtant].

**S-Foc** ... she met [a JapaNESE] programmer.

*Did Anna say that she met a Japanese programmer?*

No, she said that ...

**given** ... she met a Japanese [ARchitect].

### 5.4.3. Participants

The participants of this experiment were the very same as described in section 4.2.3.

#### 5.4.4. Procedure and analyses

Data elicitation, annotations and statistical analyses followed the same procedure as in the L1 experiment described in the sections 4.3.4 to 4.3.7, except that the context questions were produced by an American English native speaker, as already mentioned in the previous experiment. Thus, two annotation tasks were carried out by three annotators (the same as in the previous experiment), the first task determined whether stress was shifted or not (that is, whether stress is perceived on the initial or the final syllable), and the second task determined whether both members participating in the rhythm rule are prominent. For the stress shift annotation, the target words were presented in isolation and for the pitch accent annotation the answer sentences were presented without the respective context question.

**Inter-annotator and intra-annotator agreement** Concerning the judgement of stress shift, agreement ratings were similar to the L1 data – annotators of the L2 data agreed with themselves also substantially to almost perfectly, and moderately with each other ( $\kappa=0.59$ ).

For the pitch accent annotation, intra-annotator-agreements were substantial to almost perfect ( $\kappa=0.83$ ,  $\kappa=0.85$  and  $\kappa=0.79$ ), and inter-annotator agreement was substantial ( $\kappa=0.70$ ).

**Data removed** In 4.2% of target words both syllables were rated as having equal stress. I removed these cases from the statistical analysis when at least 4 ratings agreed on this. 9 sentences were excluded due to hesitations and verbal errors. I found that among the items there are three (*home-grown*, *hotel*, *Hong Kong*) which were frequently (>50%) realized with initial stress when read in isolation. Since there is a possibility that this is due to learners' misperception of the stress pattern, I removed all sentences containing these items from the analysis. These three words accounted for 15% of the data.

#### 5.4.5. Statistical analyses

To investigate the relationship between stress shift and the five conditions, I performed a generalized linear mixed effects analysis using the logit link function, with *stress shift* as the dependent variable. I included *condition* as a fixed factor, and random intercepts for *subjects* and for *items*.



To investigate the correlation between stress shift and pitch accent placement, a second analysis was performed, again, on a subset of the data comprising the two double-focus conditions (*clash* and *no clash*) using the presence of two pitch accents as dependent variable. I included an interaction between *condition* and *stress shift* as a fixed factor, and random intercepts for *subjects* and *items*. In the following I will first present the results of the first analysis looking at the circumstances in which stress was shifted. After that I will present the results from the second analysis looking at whether stress shift influenced pitch accent placement in double-focus sentences.

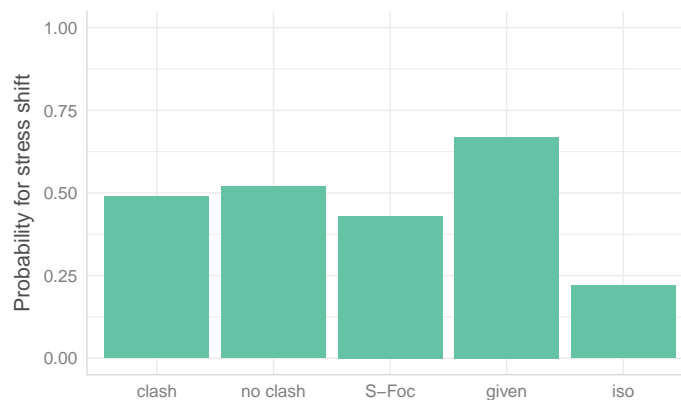
### 5.4.6. Results of stress shift annotation

There was a main effect for condition as obtained by model comparison (likelihood ratio test) comparing the model with condition to the null model ( $\chi^2(1) = 82.369$ ,  $p < 0.0001$ ). Table 5.8 gives an overview of the estimates as given by the final model.

Table 5.8.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the shift annotation task of the L2 data.*

	Estimate	Std. Error	z value	p value
(Intercept)	0.006817	0.418896	0.016	0.9870
no clash	0.185828	0.404633	0.459	0.6461
S-Foc	-0.353591	0.401771	-0.880	0.3788
given	0.910396	0.411466	2.213	0.0269
isolation	-1.661016	0.336952	-4.930	<0.0001

Figure 5.9.: The propability for stress shift by condition in the English L2 data



## 5.4. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

See Figure 5.9 for an overview of when L2 speakers shifted stress. The conditions are on the x-axis, the probability for stress shift is on the y-axis. Like the L1 group, the L2 speakers mostly did not shift stress when the target word was uttered in isolation (69%), and shifted stress in the majority of the *given* cases, where the target word is given (71%). In the double-focus conditions (*clash* and *no clash*), L2 speakers shifted stress in about 55% of the cases. The difference in metrical context between the conditions did not affect the probability of stress shift ( $\beta=0.19$ ,  $SE=0.40$ ,  $p=0.65$ ). Overall, the L2 speakers shifted stress considerably less often across the double-focus conditions compared to the L1 speakers. When the target word was focused and the following word was given (*S-Foc*), L2 speakers also frequently shifted stress (43% of the cases), which is not significantly different from the double-focus conditions ( $\beta=-0.35$ ,  $SE=0.41$ ,  $p=0.38$ ). This is unlike the English L1 speakers, who mostly did not shift stress in this condition.

### 5.4.7. Results of pitch accent annotation

There was no interaction between stress shift and pitch accenting: neither stress shift ( $\chi^2(1) = 1.3949$ ,  $p= 0.238$ , as compared to the null model) nor condition (*clash*, *no clash*,  $\chi^2(1) = 0.1171$ ,  $p= 0.732$ , obtained by model comparison) affected the pitch accenting in the double-focus conditions. L2 speakers realized both focus accents in 73% of the *clash* cases and in 80% of the *no clash* cases. Figure 5.10 shows the results of the prominence annotation. The two double-focus conditions are on the x-axis and the probability for both foci to be prominent is on the y-axis. Table 5.9 presents the estimates as predicted by the final model (without the interaction term).

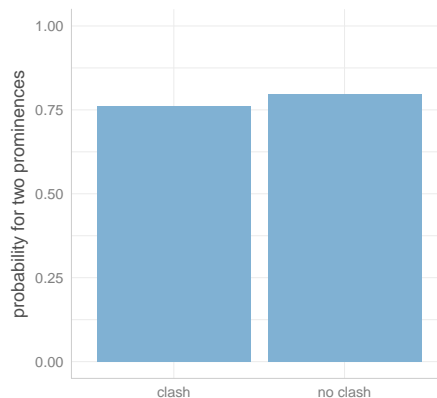


Figure 5.10.: The probability for two pitch accents by condition in the English L2 data

Table 5.9.: *Estimated coefficient, Standard Error, z-value and estimated p-value for the generalised linear mixed model for the pitch accent annotation task of the English L2 data.*

	Estimate	Std. Error	z value	p value
(Intercept)	0.8134	0.7982	1.019	0.3082
cond. no clash	1.4171	0.8798	1.611	0.1072
stress shift	1.5441	0.8110	1.904	0.0569
no clash:shift	-2.6581	1.0914	-2.435	0.0149

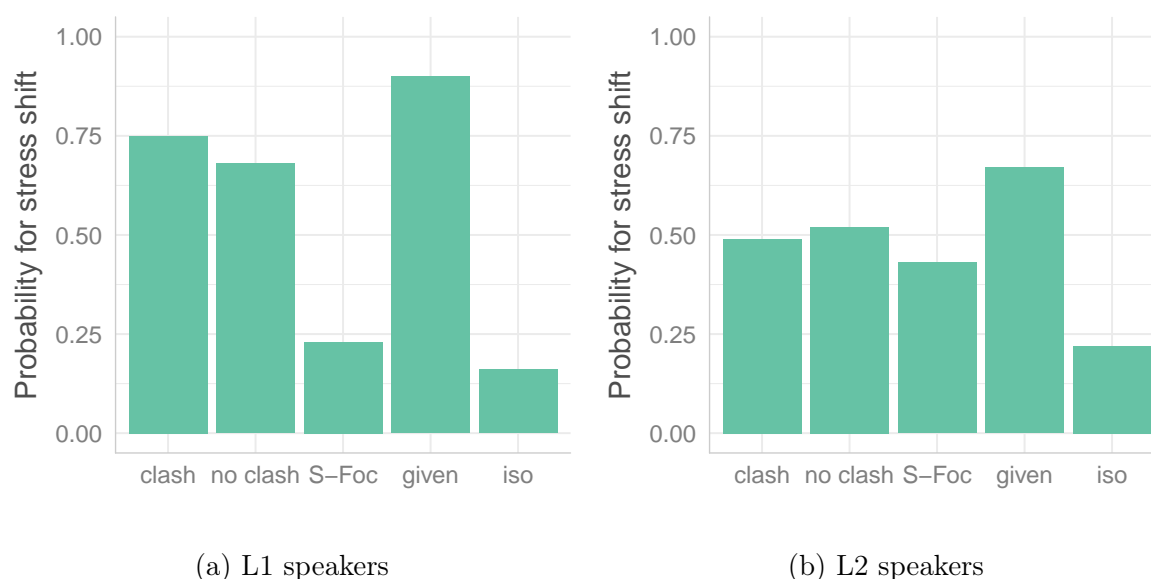


Figure 5.11.: Stress shift results for the L2 speakers (b) in comparison to the results obtained for the L1 speakers (a). The conditions (*clash*, *no clash*, *S-Foc*, *given* and the word spoken in *isolation*) are on the x-axis and the probability for stress to be shifted on the y-axis.

#### 5.4.8. Summary and discussion of results

The research question for the second experiment with German learners of English was whether, and in which environments, L2 speakers apply the rhythm rule under corrective focus marking. What has been found is that L2 speakers do indeed make use of stress shift, so they must have acquired the optional rule. However, they apply it less often than L1 speakers, as hypothesised above. Figure 5.11 shows the results of the two language groups again, here in direct comparison. Following the order of presentation, 5.11a presents the probability of stress shift by condition for L1 English speakers and 5.11b the same for the English learners. Here we can see the difference in how frequently stress shift as a stress clash repair strategy is used by the two groups. While the L1

#### 5.4. Experiment 2: Alternation and corrective focus marking in Rhythm Rule contexts

---

speakers had shifted stress most of the time in the two double-focus sentences (*clash* and *no clash*), the L2 group applied stress shift only in a bit over half of the cases.<sup>11</sup> The different rhythmic environment between the two double-focus conditions did not affect shifting of stress in either group so that the L2 results support the conclusion that stress shift is also triggered by adjacent prominences above the lexical level, since lexical stresses are only adjacent in the *clash* condition. A reason for the overall less application of the rhythm rule may be that stress shift as a rhythmic repair mechanism is not a very regularly used strategy in L1 German (Wagner and Fischenbeck, 2002), so that missing shifts could be attributed to L1 transfer. Another reason for differences in rhythm rule application may be a difference in phrasing compared to L1 speakers, since L2 speakers have been found to produce shorter intonation phrases (Gut, 2009, 2000; Grosser, 1993). A phrase break between the two foci would lead to less shifting since the target word would be the last accent in the phrase, and the rhythm rule only operates within a prosodic phrase (Lieberman and Prince, 1977; Nespor and Vogel, 1986; Selkirk, 1984). Note, however, that the stress shift data was not annotated as to ToBI labels, including boundaries. In experiment 1, however, intermediate phrase boundary annotation suggested that L2 speakers used overall less phrase breaks across the conditions than both L1 groups. Given the comparable stimuli between the two experiments and the high proficiency of the speakers this suggests that different phrasing strategies may not be the cause for the different results regarding the use of stress shift. That is, assuming less stress shift application resulting from shorter phrases is not plausible.

Another explanation for less stress shifting by L2 speakers may lie in the nature of non-native speech: Tilsen (2012) found that repair strategies in stress clash contexts are more often used when speech was prepared than in spontaneous speech. Considering the greater cognitive load for non-native speakers, L2 speech may be comparable to less prepared speech, and therefore less prone to rhythmic adjustments.

Concerning stress shift and pitch accent, I did not find a correlation between stress shift and pitch accent placement, L2 speakers realized both foci with a pitch accent in more than 75% of the time regardless of metrical environment or whether stress was shifted or not. Like in the previous experiment, the number of two focus accents is particularly high in the L2 group. Marking a corrective focus with a pitch accent seems to be a very strong constraint for German learners of English that is not easily toppled by rhythmic well-formedness ambitions.<sup>12</sup>

---

<sup>11</sup>An explorative look into speaker-specific results did not reveal any sub-group tendencies so that I cannot say that there are some speakers who always shift stress and some who never do so. The variability across items is greater than across speakers.

<sup>12</sup>It has also been shown that there is a tendency for L2 speakers, even highly proficient ones, to

An additional difference between the L1 and L2 speakers in this study is that L2 speakers frequently shifted stress when the following word was given and deaccented (condition *S-Foc*: *Indian programmer? No, JApanese programmer!*). In fact, stress was shifted in this condition equally frequent as in the double-focus conditions. While L1 speakers reliably did not shift stress in this case, the corrective focus did not override stress shift in the learner group, even though it was the last accent of the phrase (since the following material is given). It seems like the L2 speakers have not acquired the necessity of a weak-strong asymmetry of prominences for the rhythm rule to apply and overgeneralise to contexts in which this constraint is not given (since the shift-triggering word is given and thus less prominent than the word whose stress shift it triggers). Overgeneralisation of accent rules by L2 speakers has been found by various studies (e.g. Rasier and Hiligsmann, 2009; Archibald, 1997)<sup>13</sup>, although in these studies the overgeneralisation is attributed to L1 transfer, which in our case here is not plausible (since in German stress shift in this context would not be expected). Non-native speakers, even proficient learners and regardless of the L1, have also been found to differentiate less between accented and unstressed syllables and tend to overarticulate in comparison to native speakers (cf. Barry, 2012b). If this was the case in the study at hand, the given word may have triggered stress shift due to its unreduced form. Acoustic analyses are needed to look into that possibility.

When the target word is given and the following word is focused (condition *given*), L2 speakers reliably shift stress just as the L1 speakers did. It has to be noted, however, that L2 speakers in the previous experiment frequently accented given material (in 48% of the sentences in this condition, see 5.3.8.1 for a discussion of this finding), so that it is conceivable that the given target word in experiment 2 was also potentially pitch accented. Pitch accents on these conditions, however, were not annotated, so that a more detailed analysis of the realization of the given target word is needed in order to shed light on this issue. As already mentioned for the L1 data where stress in this condition was also mainly shifted, it has to be noted that these results may be influenced by a possible priming effect resulting from the make up of the context question (e.g. *Did Anna say that she met a Japanese programmer? - No, she said that she met a Japanese ARchitect.*). The speaker of the context question may have shifted the stress on, here, *Japanese* to the initial syllable due the following initial stress on, here, *programmer*, thus

---

generally realise more pitch accents (cf. Rasier and Hiligsmann, 2007) which may additionally play into the findings here, adding to the number of pitch accents used.

<sup>13</sup>See also Rasier and Hiligsmann (2007), who cites Hiligsmann and Rasier (2002); Ramirez Verdugo (2002); Rasier (2003, 2006) for similar findings.

priming early stress in the answer sentence.

## 5.5. Conclusion of this chapter

The two experiments in this chapter investigated realisations by German learners of English and how these speakers dealt with possible conflicts between a preference for alternation and the prosodic marking of focus. Comparing the realisation of the respective L1 groups (that is L1 English speakers and L1 German speakers) deviations were expected in the frequency dimension of Mennen's L2 intonation learning theory. The phenomena has been attributed to the frequency dimension given that German and English share an inventory in terms of intonation categories (systemic dimension), and both typically realise corrective focus with a high tonal target on the accented syllable (semantic category). Where they have been found to differ, however, is the strategy to deal with alternation constraints under corrective focus: in German, a low tonal target for the first focus in double focus sentences seems to be licensed in order to adhere to a tonal alternation while for English speakers adapting the pitch accent type does not seem to be allowed but instead the first focus accent was frequently omitted adhering to an alternation of strong and weak. These different strategies lead to more frequent rises for German speakers and less frequent pitch accents for English speakers – a characteristic of the frequency dimension playing well into the semantic dimension in affecting the prosodic marking of information structural categories such as corrective focus.<sup>14</sup>

In the first experiment L2 speakers showed a general “loyalty” to the semantic function of intonation in reliably realising prominent foci in double-focus sentences. Tonal alternation was restored sometimes by using a rise-fall pattern, especially for directly adjacent accented syllables, but to a lesser extent than by the German L1 group. The rhythmic environment did not seem to be an equally driving factor in L2 realisations compared to the L1 groups. This was also shown with the second experiment which investigated the interplay of corrective focus marking and the rhythm rule, or more specifically whether lexical stress was shifted to an earlier syllable under corrective focus. While L2 speakers indeed frequently applied the rhythm rule, they did that overall less often than the L1 speakers and overgeneralised it to contexts where L1 speakers reliably did not shift stress.

To conclude, results for the L2 speakers of this study, thus, suggests that highly proficient German learners of English show no difficulties in the systemic and the semantic

---

<sup>14</sup>See 5.1 for a more elaborate explanation of LILT's dimensions and how they apply to this study.

dimension, but that the frequency dimension, or more specifically, strategies in the quest for alternation (as part of the frequency dimension) are more difficult to fully acquire and are ranked lower or are acquired at a later stage than both the systemic and the semantic dimension.





## 6. General conclusions, contributions and outlook

This thesis set out to investigate the interplay between alternation preferences and the prosodic marking of focus in German, English and German learners of English. With German and English, two well-investigated stress-timed languages were chosen which share a number of similarities: They both have been found to prefer an alternation of strong and weak in production, perception and processing of sentences. They also both use pitch accenting to indicate focus structure, and they have been shown to do so by similar means. It has been found for both languages, however, that there is variation in the mapping of pitch accents and information-structure categories, and that we cannot assume a strict one-to-one relationship. Against this backdrop, this study investigated whether the preference for rhythmic alternation can account for variation in the prosodic marking of focus. I found that, despite their similarity, the two languages in fact rely on different strategies in situations in which alternation and focus marking are working in opposite directions.

These differences may play a role in second language acquisition. Prominent models of second language learning, such as Flege's (1995) SLM, and Mennen's (2015) LILt assume that learners will have more difficulties in similar but not identical categories. This study set out to investigate in which dimension alternation preferences operate in an L2, and whether different strategies are transferred or attained. The following section will summarise the main findings of this undertaking.

### 6.1. Discussion of the main findings and general outlook

**German and English** In both languages, the rhythmic manipulation in the experimental material does not seem to affect the prosodic marking of focus – there was no difference between the two double-focus conditions. However, the results with respect

to pitch accent placement nevertheless differ between the two language groups. While the German speakers realised both focus accents by means of prominences in about 75% of cases (71% in clash and 79% in no clash sentences), the English speakers realised the two prominences in only about 56% of cases (in 51% of the clash and in 62 % of the no clash sentences). I had assumed that this comparably lower number may be a consequence of the rhythmic environment after all, and no-clash sentences do not have enough material between the two foci in order to be rhythmically preferred. The higher degree of vowel reduction, and the frequent use of stress shift have been addressed as factors which might contribute to this difference between the two languages. The results obtained from the second experiment with English L1 speakers on rhythm rule contexts indeed point in the direction of a rhythmic effect, since more focus accents were realised in both double-focus conditions when stress could be shifted. A question to pursue in future work is how much material is necessary for English speakers to reliably realise adjacent focus accents, and for stress shift not to occur. Considering influences of speech rate on matters of rhythmic timing (e.g. Dellwo and Wagner, 2003) it may be worth having a closer look at alternation repair strategies under focus marking in different speech rates and styles.

It has to be noted that in all studies the raw numbers show a difference between the clash condition and the no-clash condition, in so far as pitch accents are more often omitted in clash contexts. This is what was also shown in the pilot study for the respective double-focus constructions involving two adjacent NPs. Given that pitch-accent omissions on rhythmic grounds have been found before (e.g. Riester and Piontek, 2015; Kentner, 2012a), this raises the question whether pitch accent placement may indeed be affected in prosodically less marked structures or less emphatic environments in German. It remains to be investigated in which contexts strong focus-to-accent requirements are overridden by rhythmic preferences.

However, the main difference in strategy between the two language groups manifests itself in the choice of pitch accent type in clash environments. German speakers often revert to a clash reduction in terms of melodic alternation – by realising alternating pitch accent targets. Consequently, in a sequence of two focus accents they may realise the first one with a rising accent, that is with a low tonal target on the accented syllable (L\*). This was not found for English speakers who almost always realised both foci with a high starred accent. English and German have been found to differ in the frequency dimension before. Mennen et al. (2012), for example, found a significantly higher number of L\* accents in German than in English, while English speakers used H\* accents more often than German speakers. This difference in frequency may be, possibly among other

things, a consequence of the current finding that German speakers prefer an alternation of low and high, particularly in clash environments, and in general when the pitch accents are close together, as shown by our corpus study presented in section 2.2.3.2.

The finding that German speakers frequently use low tonal targets for corrective focus marking – considering that corrective focus is on the far end of a “scale of emphasis” – is all the more surprising given the results by Scuffil (1982), who in a comparative study of German and English intonation found that for German, “low pitch inhibits perceptual prominence in a way that it does not for English” (p.264). English speakers, however, almost exclusively used high tonal targets to prosodically mark the corrective contrasts. Scuffil’s finding is in line with the results obtained from the relational prominence judgement task in so far as the first focus accent (which was often L\*H) was predominantly found to be less prominent than the second one in the German, but not in the English experiment. It was speculated before that the asymmetry of prominences in German is in fact a correlate of pitch accent type with rising pitch accents perceived as less prominent than falls. The results obtained for the L2 speakers, however, does not support this assumption: the L2 group used different pitch accent types (also frequently rises albeit to a lesser extent than the L1 group) but still realises equally prominent foci most of the time. This additional finding suggests that the difference between the L1 groups is not a confound of the different types of pitch accents used. The prominence relation results are to be treated with caution, however, given the low inter-rater agreements in all language groups.

The finding that pitch accent type can be adapted according to alternation preferences in German is another example of the “many to many mapping between prosodic form and discourse function” (Roettger et al., 2019), in other words, specific communicative functions that prosody has, such as signalling information structure, may be realised in several ways. This means that one discourse function may be realised with various accent types, or one accent type used for various discourse functions. This thesis contributes to the identification of factors that bring speakers to choose one form over the other. Perception studies are needed in order to understand how listeners perceive and interpret these variants in different contexts.

Finally, it remains to be noted that we do find rhythmic readjustment also in focus contexts, that is, when the adjustable unit is focus marked, contrary to what has been previously claimed (Grabe and Warren, 1995; Hayes, 1995; Kentner, 2012b). It seems as if rhythmic readjustment or repair strategies on focus accents are not blocked because this focus carries the main sentence accent but because in single-focus sentences this is the last prominence in the phrase. The presence of a second sentence accent such as

in double-focus sentences, however, generates adjustments with respect to pitch accent type and degree of prominence in German, and with respect to pitch accent placement and stress shift in English. With respect to English stress shift, evidence presented in this thesis argues against a deletion account and suggests an actual shift of prominence in the form of a focus accent. This thesis contributes to clarifying the contexts and requirements of rhythmic repair strategies and extends the investigated contexts of the rhythm rule.

**L1 and L2** As often found in L2 research, the L2 speakers are somewhat between the two L1 groups with respect to their prosodic realisation of focus in clash environments. Deviations from the target norm were found with respect to two aspects. First, while the L2 speakers realised the majority of double-focus sentences with a fall-fall pattern (H\*L H\*L), like the L1 English speakers, they also frequently used a rising L\*H accent on the first focus, especially in clash-contexts, like the L1 German group, albeit to a lesser degree. The use of an L\*H on the first of two foci, however, does not seem to be acceptable in English. The contour here is caused by the transfer of native repair strategies in clash environments, which leads to misproductions in the L2. Secondly, the L2 speakers use optional repair strategies of the target language to a lesser degree than L1 English speakers – while L2 speakers regularly shift stress in rhythm rule contexts, they do so, on the one hand, less often than the native speakers, and on the other hand, often apply stress shift in more contexts than licensed in L1 English. Further research is needed in order to investigate how these productions are perceived by native speakers. In the first case, transfer effects lead to erroneous productions. In the second case, however, not shifting the stress in rhythm rule contexts may not be considered erroneous given that this repair strategy is optional. This may be expressed in different degrees of acceptability.

Overall, the results obtained from the L2 speakers in this study suggest that alternation preferences play only a minor role in their L2 – while alternation preferences did affect focus marking in the L1 groups, the effects were sliced down in L2 speech. Requirements within the semantic dimension, which in our case manifests in the prosodic marking of focus by means of a pitch accent, on the other hand, were reliably met, even to a greater extent than speakers in the L1 groups. These findings suggest that aspects in the frequency dimension, which in our case are consequences emanating from the preference for alternation, are more difficult to fully acquire and are ranked lower or are acquired at a later stage than both the systemic and the semantic dimension. In his investigation of second language learners, Neufeld (1988) found that speakers often

fail to produce faultless speech even though they are able to perceive erroneous productions. Because of this finding he argues that adult learners tend to “deprioritise linguistic levels in production that do not contribute directly to meaning” (Neufeld, 1988, 1997, as cited in Leather, 1999, p. 7). Even though he focused on the segmental level in his investigation, this may be applicable to intonational categories. As already mentioned, it should be interesting to see whether this priority scale is mirrored in foreign accent ratings. With respect to the relatively young L2 intonation learning model LILt, this thesis contributes to testing and promoting this model in L2 research. More studies are needed in order to “identify where deviations from the native form occur and whether they occur more frequently in some than in other dimensions” (Mennen, 2015, p.175).

## 6.2. Summary

To conclude I would like to revisit and answer the research questions outlined in 2.4.

1. How do alternation preferences affect the realisation of corrective focus in German?

The results presented in chapter 3 show that the rhythmic context affects the choice of pitch accent type and the degree of prominence. In rhythmic clash contexts, German speakers prefer an alternation of low and high tonal targets. Pitch accent placement, however, does not seem to be affected.

2. How do alternation preferences affect the realisation of corrective focus in English? And what is the status of the rhythm rule under corrective focus marking in English?

The results presented in chapter 4 show that focus accents are more likely to be realised when there is the option to shift the pitch accent to an earlier syllable away from the clash. Contrary to results found for German, an alternation of pitch accent types does not seem to be an option in English. Contrary to previous assumptions, stress shift is possible under corrective focus marking when it is not the last focus in the phrase.

3. Are alternation preferences transferred from L1 German to L2 English? And are rhythmic repair strategies acquired in the context of corrective focus?

The results presented in chapter 5 suggest that the option to use a rising pitch accent type on a corrective focus in clash environments is transferred from German, resulting in more rises in clash environments (but less than in their L1). The requirement to mark corrective focus by pitch accent placement, however, is not

### 6.3. Open questions for future research

---

affected by the rhythmic context – L2 speakers almost always realised both focus accents irrespective of whether the location of pitch accent can be shifted or not. Stress shift as a strategy in rhythm rule contexts is acquired but overgeneralised to unlicensed contexts. These findings suggest that L2 speakers are less guided by alternation preferences than native speakers of both languages.

With respect to the overall research questions, these findings lead me to conclude that

- the preference for alternation can influence the prosodic marking of focus and is a source for variation in the realisation of information-structure categories. The rhythmic context should therefore be taken into account when assigning semantic to phonological categories.
- even though German and English share the preference for alternation, it affects prosodic focus marking differently.
- strategies emanating from alternation preferences are transferred and can cause misproductions in an L2. L2 speech itself, however, is less driven by rhythmic factors.

### 6.3. Open questions for future research

**Alternation preferences and focus marking in silent reading** As outlined in section 2.2.3, it has been found by various studies that alternation preferences play a role even in silent reading (e.g. Breen and Clifton, 2011; Rothermich et al., 2012; Magne et al., 2010; Kentner, 2012a). It is also known that adding new information to the discourse model elicits an electrophysiological response in EEG studies in the form of an N400 (Baumann and Schumacher, 2012) – an ERP component often associated with semantic integration. It remains to be seen how focus and alternation preferences interact in silent reading. The study by Henrich et al. (2015) showed that irregularities elicit weaker responses when the attention is shifted towards the preceding word by means of inducing a narrow focus. The target words themselves were, however, not investigated under narrow focus marking.

First results are obtained from a study by Schaufler and Augurzky (2016) which investigates effects of alternation preferences and the interplay with focus marking in silent reading. We conducted an ERP study using the same sentence design as in the study presented in this thesis. We found that both information structure and rhythmic

context affect the processing of double-focus sentences. The former was reflected in an N400 for the second unexpected focus, as has also been found by Baumann and Schumacher (2012). The latter, that is, the rhythmic effect, showed in a difference between the clash and the no clash condition. What was unexpected, however, was the direction of the effect: the potential pitch accent clash in the clash condition led to an N400 reduction compared to the no clash condition. Two possible explanations were given: First, the second focus accent on NP2 may not have been fully realised since adjacent pitch accents are dispreferred. On the prosodic level the discourse model is therefore more minimal since it involves only one pitch accent in the target sentence, which reduces processing costs. Second, the expectation of a minimal discourse (i.e. only one focus) triggers specific lexical predictions, namely the given NP2 from the context question, which in our design always had initial stress (e.g. *Schülerinnen*, see page 55). The similarity in form with the then encountered NP2, which was also initially stressed, contrary to the no clash condition, may have led to a reduction of the N400, thus indicating fewer processing costs for the irregular context compared to the regular one. In a follow up study, we used the same double-focus material but without the potential pitch accent clash (Augurzky and Schauffler, 2018). Instead, we manipulated the prosodic similarity between the NP2s and the expected given NP (same stress pattern, or different stress pattern). The difference between the double-focus conditions was not replicated, arguing against the form-based similarity explanation and suggesting that it was indeed the rhythmic manipulation that led to the differences in processing costs. It seems as if readers' expectancies in multiple focus environments are not triggered by a preference for minimal discourse alone, but may also be modulated by suprasegmental demands, even in silent reading. More research is needed in order to understand how alternation preferences affect the incremental processing of multiple focus and why or when the violation of alternation preferences may be beneficial in these cases.

**Perception of rhythmic repair strategies under focus marking** As mentioned before, perception experiments are needed in order to fully understand the role of alternation preferences in the structure of discourse. Looking at the German production results, the question arises as to whether the various double-focus realisations are equally accepted in their given contexts. Specifically, how is the rising contour on the first of two subsequent foci, and also a reduced prominence on a corrective focus perceived? Is it interpreted as contrastive? In their perception study on pitch accent types and contrastiveness interpretation, Kügler and Gollrad (2015) found that it is particularly the height of the peak that contributes to the interpretation of contrastiveness. Many

of the German contrastive-focus realisations, however, either have a low target on the pitch accented syllable, or a reduced prominence compared to the adjacent focus – does this form affect the interpretation of the function, or is it accepted or even expected by the listeners in this context as allotonic variation? With respect to the English results, speakers frequently omitted or shifted the first of two directly adjacent focus accents. The same question also applies here – do these omissions and/or shifts affect listeners’ interpretation? Concerning the methodology, both acceptability judgements as well as more indirect measures such as visual world paradigms are conceivable.

The question of perception of course applies especially to L2 speech. We have seen some transfer effects of German alternation strategies (rising contours) and some overuse of English strategies (stress shift in single-focus constituents). How do L1 listeners or interlocutors perceive these realisations, and what is their contribution to foreign accent? And the other way round: how will English learners of German realise focus accents under clash conditions and how will their realisations be interpreted by German native speakers?

With respect to learners, it is important to see how they understand the use of intonational categories within particular contexts rather than in isolation. This is particularly true for L2 speakers of similar languages like German and English, which have been shown to differ in the frequency dimension. In a perception study on intonation contours with advanced German learners of English, Puga et al. (2017) found that while learners performed native-like in some aspects, they did significantly worse in more marked contexts.

Insights from production and perception on alternation preferences may also further speech synthesis systems, which so far often lack the variety of natural speech, and make synthesised speech more natural.

**“In vitro and in situ”** The design of a highly controlled experiment such as the ones in this thesis allows the researcher to neatly single out individual effects. However, this fine control also posts a number of limitations, first and foremost the lack of naturalness. Even though speakers were asked to prepare their sentences before reading them in an attempt to approximate more natural utterances, the speakers’ realisations may not be representative of spontaneous speech. Future work should investigate alternation preferences in discourse outside the lab (by e.g. using corpora, as e.g. Schlüter, 2005; Wagner, 2010; Schauffler et al., 2015a). Other limitations may lie within the object under investigation itself. Considering that speakers are prone to avoid clash structures it may be difficult to see tonal adjustment on adjacent accents “in the field”. In a corpus



investigation on English similar to the one by Schauffler et al. (2015a) in German, it would be interesting to see, however, whether tonal alternation is found at all in a systematic way in English speech.

**Speaker-specific strategies** Among all the other aspects and questions that are still left to be posed there is one I would like to address last but not least, and that is the issue of speaker-specific strategies and realisations. It has been pointed out various times throughout this thesis that we cannot reliably map prosodic form to function in a one-to-one way, because there is variation in how specific functions are prosodically realised and the functions that are realised with one prosodic tool. This thesis identified alternation preferences as one of the factors which contribute to this variation. However, the importance of speaker-specific strategies and styles may play an important role here and not all speakers – and listeners – may be equally driven by the preference for alternation. Recent work addresses the importance of speaker- (and listener-)specific behaviour in the mapping of prosodic form onto discourse function (e.g. Grice et al., 2017; Cangemi et al., 2015; Roettger et al., 2019). With respect to rhythmic alternation, it may be the case that some speakers may have more “rhythmical talent” (Wagner, 2008) and may be more dedicated to rhythmic-wellformedness than others. “Rhythmic training” may also play a role as studies with musicians (Geiser et al., 2010) or participants trained in poetry-reading (Wagner, 2010) suggest. Geiser et al. (2010), for example found that musicians are more sensitive to regular patterns of strong and weak beats (as shown in a mismatch negativity paradigm). According to Johansson (2008), music training also facilitates pitch processing (in both music and language) and furthers phonological awareness.<sup>1</sup> More research on speaker-specific and listener-specific strategies and factors will add to the picture of the role of alternation preference.

Musicians or not – “rhythm is a dancer”, the dance may just not always be the same.

---

<sup>1</sup>Marques et al. (2007), for example, found that musicians reacted stronger to pitch violations than non-musicians.



# Bibliography

- Abercrombie, D. (1965). *Studies in phonetics and linguistics*. Oxford University Press, London.
- Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press, Edinburgh.
- Anttila, A., Adams, M., and Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes*, 25(7-9):946–981.
- Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, 10:215–240.
- Archibald, J. (1997). The acquisition of second language phrasal stress: a pilot study. In Hannahs, S. and Young-Scholten, M., editors, *Focus on phonological acquisition*, pages 263–289. Oxford University Press, Amsterdam/Philadelphia.
- Atterer, M. and Ladd, D. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics*, 32(2):177–197.
- Auer, P. and Uhmann, S. (1988). Silben- und akzentzählende Sprachen: Literaturüberblick und Diskussion. *Zeitschrift für Sprachwissenschaft*, 7(2):214–259.
- Augurzky, P. (2008). Prosodic balance constrains argument structure interpretation in German. Poster presented at AMLaP 2008, Cambridge.
- Augurzky, P. and Schaufliker, N. (2018). Why two is not always better than one – an ERP study on minimality-based and prosodic predictions in German discourse processing. Poster presentation at AMLaP 2018, Berlin.
- Baayen, H. (2008). *Analyzing Linguistic Data*. Cambridge University Press.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.

- Bard, E. G. and Aylett, M. P. (1999). The dissociation of deaccenting, givenness, and syntactic role in spontaneous speech. In *Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco)*, volume 3, pages 1753–1756.
- Baroni, M., Bernardini, S., Ferraresi, A., and Zanchetta, E. (2009). The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226.
- Barry, W. (2012a). Learner corpora in second language prosody research and teaching. In Trouvain, J. and Gut, U., editors, *Non-Native Prosody. Phonetic Description and Teaching Practice*, pages 145–167. Mouton de Gruyter, Berlin, New York.
- Barry, W. (2012b). Rhythm as an L2 problem: How prosodic is it? In Trouvain, J. and Gut, U., editors, *Non-Native Prosody. Phonetic Description and Teaching Practice*, pages 97–120. Mouton de Gruyter, Berlin, New York.
- Bates, D., Maechler, M., and Bolker, B. (2013). *lme4: Linear mixed-effects models using Eigen and syntax*. R package version 0.999999-2.
- Baumann, S. (2006). *The Intonation of Givenness – Evidence from German*. Ph.D. thesis, Saarland University.
- Baumann, S., Becker, J., Grice, M., and Mücke, D. (2007). Tonal and Articulatory Marking of Focus in German. In *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany.
- Baumann, S. and Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, 38(10):1636 – 1657. Prosody and Pragmatics.
- Baumann, S., Grice, M., and Benzmüller, R. (2001). GToBI - a phonological system for the transcription of German intonation. In Puppel, S. and Demenko, G., editors, *Prosody 2000. Speech Recognition and Synthesis*, pages 21–28. Poznan: Adam Mickiewicz University, Faculty of Modern Languages and Literature.
- Baumann, S. and Riester, A. (2012). *Referential and lexical givenness: Semantic, prosodic and cognitive aspects*, pages 119–162. De Gruyter Mouton, Berlin, Boston.
- Baumann, S. and Röhr, C. T. (2015). The perceptual prominence of pitch accent types in German. In *Proceedings of the ICPHS Glasgow*.

- Baumann, S. and Schumacher, P. B. (2012). (De-)accentuation and the processing of information status: Evidence from event-related brain potentials. *Language and Speech*.
- Baumann, S. and Winter, B. (2018). What makes a word prominent? Predicting untrained German listeners' perceptual judgments. *Journal of Phonetics*, 70:20 – 38.
- Beckman, M. and Ayers, G. (1997). Guidelines for ToBI transcription.
- Beckman, M. and Hirschberg, J. (1999). The ToBI annotation conventions. [http://www.ling.ohio-state.edu/~tobi/ame\\_tobi/annotation\\_conventions.html](http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html).
- Beckman, M. and Pierrehumbert, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook*, 3:255–310.
- Beckman, M. E., Hirschberg, J., and Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In Jun, S.-A., editor, *Prosodic typology: The phonology of intonation and phrasing*. Oxford University Press, Oxford.
- Beckman, M. E., Swora, M. G., Rauschenberg, J., and de Jong, K. (1990). Stress shift, stress clash, and polysyllabic shortening in a prosodically annotated discourse. In *Proceedings of the 1990 International Conference on Spoken Language Processing*, pages 5–8, Kobe.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In Strange, W., editor, *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pages 171–232. York Press, Timonium, Maryland.
- Best, C. T. and Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn, O.-S. and Munro, M. J., editors, *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, pages 13–34. John Benjamins Publishing, Amsterdam.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10):341–345.
- Bohn, K., Wiese, R., and Dohmas, U. (2011). The status of the rhythm rule within and across word boundaries. In *Proceedings of the 17th International Conference of the phonetic sciences 2001, Hong Kong*, pages 332–335.
- Bohn, O.-S. (1995). Cross-language speech perception in adults. first language transfer doesn't tell it all. In Strange, W., editor, *Speech perception and linguistic experience. Issues in cross-language research*, pages 279–304. York Press, Baltimore.

- Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, 14:109–149.
- Bolinger, D. L. (1965). *Forms of English: accent, morpheme, order*. Harvard University Press, Cambridge.
- Bolinger, D. L. (1981). *Two kinds of vowels, two kinds of rhythm*. Indiana University Linguistics Club, Bloomington, Indiana.
- Boula de Mareüil, P. and Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent. *Phonetica*, 63:247–267.
- Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, 49(4):451–493.
- Braun, B. and Ladd, D. R. (2003). Prosodic correlates of contrastive and non-contrastive themes in German. In *8th European Conference on Speech Communication and Technology*, pages 789–792, Geneva, Switzerland.
- Braun, B. and Tagliapietra, L. (2011). On-line interpretation of intonational meaning in L2. *Language, Cognition and Neuroscience*, 26:224–235.
- Breen, M. and Clifton, C. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of memory and language*, 64(2):153–170.
- Breen, M., Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Language and cognitive processes*.
- Brown, G. (1983). Prosodic structure and the given/new distinction. In Cutler, A. and Ladd, D. R., editors, *Prosody: Models and Measurements*, pages 67–77. Springer, New York.
- Büiring, D. (1997). *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. Routledge Studies in German Linguistics. Routledge, London and New York.
- Büiring, D. (2003). On D-trees, beans, and B-accent. *Linguistics and Philosophy*, 26:511–545.
- Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2 edition.

- Burroni, F. and Tilsen, S. (2020). Prominence clash does not induce rhythmic adjustments in Italian. In *Proceedings of the 10th International Conference on Speech Prosody*, Tokyo, Japan.
- Calhoun, S. (2010). The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86:1–42.
- Calhoun, S. (2012). The theme/rheme distinction: Accent type or relative prominence? *Journal of Phonetics*, 40(2):329 – 349.
- Cangemi, F., Krüger, M., and Grice, M. (2015). Listener-specific perception of speaker-specific production in intonation. In Fuchs, S., Pape, D., Petrone, C., and Perrier, P., editors, *Individual Differences in Speech Production and Perception, volume 3*. Peter Lang.
- Carter, P. (2005). Quantifying rhythmic differences between Spanish, English, and Hispanic English. In Gess, R. S. and Rubin, E. J., editors, *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th linguistic symposium on Romance languages*, volume 272, pages 63–75. John Benjamins, Amsterdam, Philadelphia.
- Chafe, Wallace, L. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. N., editor, *Subject and Topic*, pages 27–55. New York: Academic Press.
- Chen, A. (2012). *Prosodic investigation on information structure*, pages 251–286. Mouton DeGruyter.
- Chen, S. and Fon, J. (2008). The peak alignment of prenuclear and nuclear accents among advanced L2 English learners. In *Proceedings of Speech Prosody (SP-2008)*, pages 643–645, Campinas, Brazil.
- Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. Harper & Row, New York.
- Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press, 32 East 57th Street, New York, N.Y. 10022, 1 edition.
- Dauer, R. M. (1983). Stress-timing and syllable timing re-analysed. *Journal of Phonetics*, 11:51–62.

- Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. volume 7, pages 295 – 326. De Gruyter Mouton, Berlin, Boston.
- Dellwo, V., Fourcin, A., and Abberton, E. (2007). Rhythmical classification based on voice parameters. In *Proceedings of ICPPhS 2007*, pages 1129–1132, Saarbrücken, Germany.
- Dellwo, V. and Wagner, P. (2003). Relations between language rhythm and speech rate. In *Proceedings of ICPPhS 2003*, Barcelona.
- Dogil, G. (2003). Understanding prosody. In Rickheit, G., Herrmann, T., and Deutsch, W., editors, *Psycholinguistik—Ein internationales Handbuch / Psycholinguistics—An International Handbook*, pages 544–565. de Gruyter, Berlin.
- Eady, S. and Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, 80 2:402–15.
- Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., and Lotts, D. W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, 29(3):233–251. PMID: 3695759.
- Eckart, K., Riester, A., and Schweitzer, K. (2012). A discourse information radio news database for linguistic analysis. In Chiarcos, C., Nordhoff, S., and Hellmann, S., editors, *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*.
- Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, 27(2):315–330.
- Féry, C. (1993). *German Intonational Patterns (Linguistische Arbeiten)*. Max Niemeyer Verlag.
- Féry, C. (2007). The fallacy of invariant phonological correlates of information structural notions. In Féry, C., Fanselow, G., and Krifka, M., editors, *The Notions of Information Structure*.
- Féry, C. (2013). Focus as prosodic alignment. *Natural Language and Linguistic Theory*, 31:683–734.



- 
- Féry, C. (2017). *Intonation and Prosodic Structure*. Cambridge University Press, Cambridge; New York.
- Féry, C. and Krifka, M. (2008). Information structure. Notional distinctions, ways of expression. In van Sterkenburg, P., editor, *Unity and diversity of languages*, pages 123–136. John Benjamins.
- Féry, C. and Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*.
- Festival (2010). Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. IMS German Festival home page. [www.ims.uni-stuttgart.de/phonetik/synthesis](http://www.ims.uni-stuttgart.de/phonetik/synthesis).
- Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In Strange, W., editor, *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, pages 233–277. York Press, Timonium, Maryland.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–382.
- Fletcher, J., Grabe, E., and Warren, P. (2005). Intonation variation in four dialects of English: the high rising tune. In Jun, S.-A., editor, *Prosodic Typology – The Phonology of Intonation and Phrasing*, pages 390–409. Oxford University Press, Oxford.
- Fletcher, J. and Harrington, J. (2001). High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, 58:215–229.
- Fletcher, J. and Loakes, D. (2010). Interpreting rising intonation in Australian English. In *Proceedings of Speech Prosody 2010, Chicago*.
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2):285–319.
- Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In Fujimura, O., editor, *Vocal Physiology: Voice Production, Mechanisms and Functions*, pages 347–355. Raven, New York.
- Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua*, 121(13):1906 – 1922. New insights into the Prosody-Syntax interface: Focus, phrasing, language evolution.

- Féry, C. and Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language*, 82(1):131–150.
- Gamer, M., Lemon, J., and Singh, I. F. P. (2012). *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.
- Gast, V. (2010). Contrastive topics and distributed foci as instances of sub-informativity: A comparison of English and German. In Breul, C. and Göbbel, E., editors, *Comparative and Contrastive Studies of Information Structure*, pages 15–50. John Benjamins, Amsterdam.
- Geiser, E., Sandmann, P., Jäncke, L., and Meyer, M. (2010). Refinement of metre perception–training increases hierarchical metre processing. *The European journal of neuroscience*, 32(11):1979–1985.
- Geiser, E., Schmidt, C., Jancke, L., and Meyer, M. (2006). Neural correlates of rhythm processing in speech perception. In *Proceedings of Speech Prosody 2006, Dresden*.
- Giegerich, H. (1985). *Metrical phonology and phonological structure: German and English*. CambridgeUniversityPress, Cambridge.
- Goldschneider, J. and DeKeyser, R. (2005). Explaining the “Natural Order of L2 Morpheme Acquisition” in English: A meta-analysis of multiple determinants. *Language Learning*, 55:27 – 77.
- Grabe, E. (1998). *Comparative Intonational Phonology: English and German*. Ponsen en Looien. Ph.D. thesis.
- Grabe, E. and Low, E. L. (2002). Duration variability in speech and the rhythm class hypothesis. In Warner, N. and Gussenhoven, C., editors, *Papers in Laboratory Phonology 7*, pages 515–546. Mouton de Gruyter, Berlin.
- Grabe, E. and Warren, P. (1995). Stress shift: do speakers do it or do listeners use it? In Connell, B. and Arvaniti, A., editors, *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence*, pages 95–110. Cambridge University Press, Cambridge.
- Grice, M., Reyelt, M., Benz Müller, R., Mayer, J., and Batliner, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. In *Proceedings of ICSLP*, pages 1716–1719.

- Grice, M., Ritter, S., Niemann, H., and Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, 64:90–107.
- Grosser, W. (1993). Aspects of intonational L2 acquisition. In Kettemann, B. and Wieden, W., editors, *Current Issues in European Second Language Acquisition Research*, pages 81–94. Gunter Narr, Tübingen.
- Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. *Second Language Speech: Structure and Process*, pages 211–228.
- Grüter, T., Rohde, H., and Schafer, A. J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In *Proceedings of BUCLD 38*. Cascadilla Press.
- Guenther, F. H. (2015). The neural control of speech: From computational modeling to neural prosthesis. In *Proceedings of the 18th International Conference of the phonetic sciences 2015, Glasgow*.
- Gussenhoven, C. (1983). Stress shift and the nucleus. *Linguistics*, 21:303–339.
- Gussenhoven, C. (1983a). Focus, mode and the nucleus. *Journal of Linguistics*, 19:377–417.
- Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
- Gussenhoven, C. (1988). Adequacy in intonation analysis. In van der Hulst, H. and Smith, N., editors, *Autosegmental studies on pitch accent*. Foris, Dordrecht.
- Gussenhoven, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology*, 8:1–35.
- Gussenhoven, C. (1992). Sentence accents and argument structure. In Roca, I. M., editor, *Thematic structure: its role in Grammar*. Foris, Berlin.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.
- Gussenhoven, C. (2008). Types of focus in English. In Lee C., Gordon M., B. D., editor, *Topic and Focus. Studies in Linguistics and Philosophy, vol. 82*, pages 83–100. Springer, Dordrecht.

- Gussenhoven, C. (2016). Analysis of intonation: the case of MAE\_ToBI. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 7(1):10.
- Gut, G. (2003a). Prosody in second language speech production: The role of native language. *Fremdsprachen Lehren und Lernen*, 32:133–152.
- Gut, U. (2000). *Bilingual acquisition of intonation*. Niemeyer, Tübingen.
- Gut, U. (2003b). Non-native speech rhythm in German. In *Proceedings of ICPHS 2003*, pages 2437–2440, Barcelona.
- Gut, U. (2009). *Non-native Speech. A corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Peter Lang GmbH, Frankfurt am Main.
- Gut, U. and Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition*, 36:283–302.
- Guy, G. and Vonwiller, J. (1989). The high rise tones in Australian English. In Collins, P. and Blair, D., editors, *Australian English*, pages 21–34. University of Queensland Press, St Lucia.
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, 3:37–81 (part 1), 199–244 (part 2).
- Hammond, M. (1999). Lexical frequency and rhythm. In Darnell, M., Moravcsik, E., Newmeyer, F., Noonan, M., and Wheatley, K., editors, *Functionalism and Formalism in Linguistics, Vol. 1*, pages 329–358. John Benjamins.
- Hartmann, K. and Zimmermann, M. (2009). Morphological focus marking in Gùrùntùm (West Chadic). *Lingua*, 119(9):1340–1365.
- Hayes, B. (1984). The phonology of rhythm in English. 15(1):33–74.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. The University of Chicago Press, Chicago.
- Henrich, K., Alter, K., Wiese, R., and Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An {ERP} study on English compounds. *Brain and Language*, 136:19 – 30.

- Henrich, K., Wiese, R., and Domahs, U. (2015). How information structure influences the processing of rhythmic irregularities: ERP evidence from German phrases. *Neuropsychologia*, 75:431 – 440.
- Hiligsmann, P. and Rasier, L. (2002). De zinsaccentuering in de tussentaal van franstalige leerders van het nederlands. *n/f*, 1:37–53.
- Hobbs, J. R. (1990). The Pierrehumbert-Hirschberg theory of intonational meaning made simple: comments on Pierrehumbert and Hirschberg. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communications*, pages 313–323. MIT Press, Cambridge, MA.
- Horne, M. (1990). Empirical evidence for a deletion formulation of the rhythm rule in English. *Linguistics*, 28:959–981.
- Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, 121(13):1870 – 1889. New insights into the Prosody-Syntax interface: Focus, phrasing, language evolution.
- Ito, K., Speer, S., and Beckman, M. (2004). Informational status and pitch accent distribution in spontaneous dialogues in English. In *Proceedings of Speech Prosody 2004 (Nara, Japan)*.
- Jilka, M. (2000). *The Contribution of Intonation to the Perception of Foreign Accent*. Doctoral Dissertation, Arbeiten des Instituts für Maschinelle Sprachverarbeitung (AIMS), Vol.6(3), University of Stuttgart.
- Johansson, B. (2008). Language and music: What do they have in common and how do they differ? A neuroscientific approach. *European Review*, 16:413–427.
- Juffs, A. (1990). Tone, syllable structure and interlanguage phonology: Chinese learners' stress errors. *International Review of Applied Linguistics in Language Teaching*, 21:99–115.
- Jun, S.-A. (2005). Prosodic typology. In Jun, S.-A., editor, *Prosodic Typology – The Phonology of Intonation and Phrasing*, pages 430–458. Oxford University Press, Oxford.
- Jun, S.-A. (2012). Prosodic typology revisited: Adding macro-rhythm. In *Proceedings of the 6th International Conference on Speech Prosody*.

- Jun, S.-A. (2014). *Prosodic typology: By prominence type, word prosody, and macro-rhythm*, pages 520–539.
- Kabagema-Bilan, E., Lòpez-Jimènez, B., and Truckenbrodt, H. (2011). Multiple focus in Mandarin Chinese. *Lingua*, 121(13):1890–1913.
- Kanerva, J. M. (1990). *Focus and Phrasing in Chichewa Phonology*. Garland, New York.
- Kelly, M. H. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, 30(2):107–137.
- Kelly, M. H. and Bock, J. K. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3):389–403.
- Kentner, G. (2012a). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition*, 123(1):1–20.
- Kentner, G. (2012b). *Linguistic rhythm and sentence comprehension in reading*. Doctoral dissertation, Goethe-Universität Frankfurt.
- Kentner, G. and Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review*, 30(2):277–311.
- Kimbell, A. E. and Cole, J. (2014). Avoidance of stress clash in perception of conventional American English. In *Proceedings of Speech Prosody 7, 2014, Dublin*.
- Kingdon, R. (1958). *The Groundwork of English Intonation*. Longman, London.
- Kiparsky, P. (1966). Über den deutschen Akzent. *Studia Grammatica VII: Untersuchungen über Akzent und Intonation im Deutschen*.
- Kohler, K. J. (2009). Rhythm in speech and language. *Phonetica*, 66:29–45.
- Krifka, M. (2007). Basic notions of information structure. In Féry, C., Fanselow, G., and Krifka, M., editors, *Working papers of the SFB632. Interdisciplinary Studies on Information Structure 6*, pages 13–55. Universitätsverlag Potsdam, Potsdam.
- Kügler, F. and Calhoun, S. (2020). Prosodic encoding of information structure: A typological perspective. In Gussenhoven, C. and Chen, A., editors, *The Oxford Handbook of Language Prosody*. OUP, Oxford.
- Kügler, F. and Gollrad, A. (2015). Production and perception of contrast: The case of the rise-fall contour in German. *Frontiers in Psychology*.

- Kuhl, P. and Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect”. In Strange, W., editor, *Cross-Language Speech Perception*, pages 121–154, Baltimore. York Press.
- Kügler, F., Baumann, S., Andreeva, B., Braun, B., Grice, M., Neitsch, J., Niebuhr, O., Peters, J., Röhr, C., Schweitzer, A., and Wagner, P. (2019). Annotation of German intonation: DIMA compared with other annotation systems. In *Proceedings of the 19th International Conference of the phonetic sciences 2019, Melbourne*, pages 1297–1301.
- Kügler, F. and Féry, C. (2017). Postfocal downstep in German. *Language and Speech*, 60:260–288.
- Kügler, F., Smoliboeki, B., Arnold, D., Braun, B., Baumann, S., Grice, M., Jannedy, S., Michalsky, J., Niebuhr, O., Peters, J., Ritter, S. B., Röhr, C., Schweitzer, A., Schweitzer, K., and Wagner, P. (2015). DIMA – annotation guidelines for German intonation. In *Proceedings of the 18th International Conference of the phonetic sciences 2015, Glasgow*.
- Ladd, D. (2008). *Intonational Phonology*. Cambridge University Press, Cambridge, UK, 2nd edition.
- Ladd, D. R. (1983). Phonological features of intonational peaks. *Language*, 59:721–759.
- Ladd, R. (1996). *Intonational phonology*. Cambridge University Press.
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–174.
- Leather, J. (1999). *Second-language speech research: An introduction*, pages 1–58. Blackwell, Oxford.
- Li, A. and Post, B. (2014). L2 acquisition of prosodic properties of speech rhythm. *Studies in Second Language Acquisition*, 36:223–255.
- Lieberman, M. Y. and Pierrehumbert, J. (1984). Intonational invariants under changes in pitch range and length. In Aronoff, M. and Oehrle, R. T., editors, *Language Sound Structure*, pages 157–233. MIT Press, Cambridge, MA.
- Lieberman, M. Y. and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8:249–336.

- Liu, F. (2010). Single vs. double focus in English statements and yes/no questions. In *Proceedings of the Fifth International Conference on Speech Prosody*, Chicago.
- Long, J. A. (2018). *jtools: Analysis and Presentation of Social Scientific Data*. R package version 1.0.0.
- Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterisations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech*, 43:377–401.
- Magne, C., Gordon, R., and Midha, S. (2010). Influence of metrical expectancy on reading words: An ERP study. In *Proceedings of the Fifth International Conference on Speech Prosody*, pages 1–4, Chicago.
- Marques, C., Moreno, S., Castro, S. a., and Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of cognitive neuroscience*, 19:1453–63.
- Mayer, J. (1995). Transcribing German intonation – the Stuttgart system. Technical report, Universität Stuttgart.
- Mayer, J. (1997). Intonation und Bedeutung. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung (Univ. Stuttgart)*, AIMS, 3(4).
- Mennen, I. (1998). Can language learners ever acquire the intonation of a second language? In *Proceedings of STiLL - Speech Technology in Language Learning, Marholmen (Sweden)*, pages 17–20.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32:543–563.
- Mennen, I. (2015). Beyond segments: towards an L2 intonation learning theory (LILt). In DelaisRoussaire, E., Avanzi, M., and Herment, S., editors, *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*, pages 7–27. Springer-Verlag, Berlin, Heidelberg.
- Mennen, I., Schaeffler, F., and Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, 131.
- Möbius, B. (1995). Components of a quantitative model of German intonation. In *Proceedings of the 13th International Congress of Phonetic Sciences (Stockholm)*, volume 2, pages 108–115.



- 
- Möhler, G. (2001). Improvements of the PaIntE model for  $F_0$  parametrization. Technical report, Institute of Natural Language Processing, University of Stuttgart. Draft version.
- Möhler, G. and Conkie, A. (1998). Parametric modeling of intonation using vector quantization. In *Proceedings of the Third International Workshop on Speech Synthesis (Jenolan Caves, Australia)*, pages 311–316.
- Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17(1):17–34.
- Nanette Veilleux, S. S.-H. and Brugos, A. (2006). Transcribing prosodic structure of spoken utterances with ToBI. <https://ocw.mit.edu>.
- Nava, E. and Zubizarreta, M. L. (2009). Order of L2 acquisition of prosodic prominence patterns: Evidence from L1 Spanish / L2 English speech. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*, pages 175–187, Somerville, MA. Cascadilla Proceedings Project.
- Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Foris, Dordrecht, Holland.
- Nespor, M. and Vogel, I. (1989). On clashes and lapses. *Phonology*, 6:69–116.
- Nooteboom, S. G. and Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information: an experiment. *Journal of the Acoustical Society of America*, 82:1512–1524.
- O’Brien, M. and Gut, U. (2011). Phonological and phonetic realisation of different types of focus in L2 speech. In Magdalena Wrembel, Malgorzata Kul, K. D.-K., editor, *Achievements and perspectives in SLA of speech: New Sounds 2010 Volume 1*, page 205–216. Peter Lang.
- Ordin, M. and Polyanskaya, L. (2015). Perception of speech rhythm in second language: the case of rhythmically similar L1 and L2. *Frontiers in Psychology*.
- Ordin, M., Polyanskaya, L., and Ulbrich, C. (2011). Acquisition of timing patterns in second language. In *Proceedings of Interspeech 2011*, pages 1129–1132.
- Özçelik, O. (2016). The prosodic acquisition path hypothesis: Towards explaining variability in L2 acquisition of phonology. *Glossa: a Journal of General Linguistics*, 1(1):1–48.

- O'Connor, J. D. and Arnold, G. F. (1973). *Intonation of Colloquial English*. Longman, London.
- Palmer, H. E. (1922). *English Intonation with Systematic Exercises*. Heffer, Cambridge.
- Pierrehumbert, J. and Beckman, M. (1988). *Japanese Tone Structure*. MIT Press, Cambridge, MA.
- Pierrehumbert, J. and Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*, pages 271–311. MIT Press, Cambridge, MA.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, Massachusetts Institute of Technology.
- Pike, K. L. (1945). *The intonation of American English*. University of Michigan Press, Ann Arbor.
- Pitt, M. A. and Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, 16:564–573.
- Pointon, G. (1980). Is Spanish really syllable-timed? *Journal of Phonetics*, 8:293–304.
- Polyanskaya, L. (2015). *Contribution of prosodic timing patterns into perceived foreign accent*. PhD thesis, Bielefeld: Universitätsbibliothek Bielefeld.
- Puga, K., Fuchs, R., Setter, J., and Mok, P. (2017). The perception of English intonation patterns by German L2 speakers of English. In *Proceedings of Interspeech 2017, Stockholm, Sweden*, pages 3241–3245.
- Quasthoff, U. and Richter, M. (2005). Projekt Deutscher Wortschatz. *Babylonia*, 3:33–35.
- Quené, H. and Port, R. F. (2003). Rhythmical factors in stress shift. *IULC Working Papers*.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramirez Verdugo, D. (2002). Non-native interlanguage intonation systems: a study based on a computerized corpus of Spanish learners of English. *ICAME Journal*, 36:115–132.

- Ramus, F., Dupoux, E., and Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of ICPHS 2003*, pages 337–342, Barcelona.
- Ramus, F., Nespors, M., and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73:265–292.
- Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models – An aligner for German. In *Proc. of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry" (Russia)*.
- Rasier, L. (2003). Le système accentuel de l’interlangue d’apprenants francophones du néerlandais. *Actes du colloque international 'Interfaces prosodiques/Prosodic interfaces'*, pages 79–84.
- Rasier, L. (2006). *Prosodie en vreemdetaalverwerving. Accentdistributie in het Frans en Nederlands als vreemde taal*. PhD thesis, UCLouvain.
- Rasier, L. and Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. 28:41–66.
- Rasier, L. and Hiligsmann, P. (2009). Exploring the L1-L2 relationship in the L2 acquisition of prosody. In *Online proceedings of first and second languages: Exploring the relationship in pedagogy-related contexts*, pages 1–4, Oxford, United Kingdom.
- Riester, A. and Baumann, S. (2013). Focus triggers and focus types from a corpus perspective. *Dialogue and Discourse*, 4(2):215–248.
- Riester, A., Brunetti, L., and De Kuthy, K. (2018). Annotation guidelines for questions under discussion and information structure. In Evangelia Adamou, K. H. and Vanhove, M., editors, *Information Structure in Lesser-Described Languages: Studies in Syntax and Prosody*, pages 403–443. Benjamins, Amsterdam.
- Riester, A. and Piontek, J. (2015). Anarchy in the NP. When new nouns get deaccented and given nouns don’t. *Lingua*.
- Rietveld, A. C. M. and Gussenhoven, C. (1985). On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13(3-4):299–308.
- Roach, P. (1982). On the distinction between “stress-timed” and “syllable-timed” languages. In Crystal, D., editor, *Linguistic Controversies*, pages 73–79. Arnold.

- Roettger, T., Mahrt, T., and Cole, J. (2019). Mapping prosody onto meaning – the case of information structure in american english. *Language, Cognition and Neuroscience*.
- Rooth, M. (1985). *Association with Focus*. PhD thesis, University of Massachusetts, Amherst, MA.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.
- Rooth, M. (1995). Focus. In Lappin, S., editor, *Handbook of Contemporary Semantic Theory*, pages 271–297. Blackwell, Oxford, UK.
- Rothermich, K., Schmidt-Kassow, M., and Kotz, S. A. (2012). Rhythm’s gonna get you: Regular meter facilitates semantic sentence processing. *Neuropsychologia*, 50(2):232–244.
- Saito, K. (2018). Advanced second language segmental and suprasegmental acquisition. In Paul Malovrh, A. B., editor, *The handbook of advanced proficiency in second language acquisition*. Wiley Blackwell.
- Schafer, A., Grüter, T., and Rohde, H. (2014). The role of discourse-level expectations in non-native speakers’ referential choices. In *Boston University Conference on Language Development*, Boston.
- Schauffler, N. and Augurzky, P. (2016). Rhythm influences the processing of double-focus sentences: ERP evidence from German. Poster presentation at AMLaP 2016, Bilbao.
- Schauffler, N., Augurzky, P., Schweitzer, K., and Lewandowski, N. (2012). Effects of prosodic balance in different elicitation techniques. Poster presented at AMLaP 2012, Riva del Garda.
- Schauffler, N., Schweitzer, A., Schweitzer, K., and Augurzky, P. (2015a). Avoiding melodic clashes in pitch accent production: a corpus study. In Leemann, A., Kolly, M.-J., Dellwo, V., and Schmid, S., editors, *Trends in Phonetics and Phonology in German-speaking Europe.*, page 125. Peter Lang.
- Schauffler, N., Turco, G., and Augurzky, P. (2015b). Producing multiple contrastive accents in German: rhythmic and syntactic factors. In *Proceedings of the ICPPhS Glasgow*.

- Schiering, R. (2007). The phonological basis of linguistic rhythm: Cross-linguistic data and diachronic interpretation. *Language Typology and Universals*, 60:337–359.
- Schlüter, J. (2005). *Rhythmic grammar: The influence of rhythm on grammatical variation and change in English*. Mouton de Gruyter, Berlin.
- Schmerling, S. (1976). *Aspects of English Sentence Stress*. University of Texas Press, Austin, Texas.
- Schwarzschild, R. (1999). GIVENness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics*, 7:141–177.
- Schweitzer, K., Eckart, K., Dogil, G., and Augurzky, P. (2012). Prosodic balance. Poster presentation at the 13th conference on laboratory phonology (LabPhon 13).
- Schweitzer, K., Riester, A., Walsh, M., and Dogil, G. (2009). Pitch Accents and Information Status in a German Radio News Corpus. In *Proceedings of Interspeech 2009*, pages 877–880, Brighton, UK.
- Scuffil, M. (1982). *Experiments in comparative intonation. A case-study of English and German*.
- Selkirk, E. (1984). *Prosody and Syntax: The Relation Between Sound and Structure*. MIT Press, Cambridge.
- Selkirk, E. (1995). Sentence prosody: Intonation, stress, and phrasing. In Goldsmith, J., editor, *Handbook of Phonological Theory*, pages 550–569. Blackwell, Oxford.
- Shattuck-Hufnagel, S. (1995). The importance of phonological transcription in empirical approaches to “stress shift” versus “early accent”: comments on Grabe and Warren, and Vogel, Bunnell and Hoskins. In Connell, B. and Arvaniti, A., editors, *Papers in Laboratory Phonology IV. Phonology and Phonetic Evidence*, pages 128–140. Cambridge University Press, Cambridge.
- Shih, S., Grafmiller, J., Futrell, R., and Bresnan, J. (2015). Rhythm’s role in genitive construction choice in spoken English. In Vogel, R. and van de Vijver, R., editors, *Rhythm in Cognition and Grammar*, pages 207–234. De Gruyter.
- Silverman, K., Backman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., and Hirschberg, J. (1992). ToBI: A standard for Labeling English Prosody. In *Proceedings of the 1992 International Conference on Spoken Language Processing*, volume 2, pages 867–870, Banff, Canada.

- Sweet, H. (1876). Words, logic, and grammar. *Transactions of the Philological Society*.
- Terken, J. and Hirschberg, J. (1994). Deaccentuation of words representing ‘given’ information: effects of persistence of grammatical function and surface position. *Language and Speech*, 37:125–145.
- Tilsen, S. (2011). Metrical regularity facilitates speech planning and production. *Laboratory Phonology 2*, pages 185–218.
- Tilsen, S. (2012). Utterance preparation and stress clash: Planning prosodic alternations. In Fuchs, S., Perrier, P., Weirich, M., and Pape, D., editors, *Speech production and perception: planning and dynamics*, pages 119–156. Peter Lang Verlag, Frankfurt am Main.
- Tomlinson, J. M. J., Liu, Q., and Fox Tree, J. E. (2014). The perceptual nature of stress shifts. *Language, Cognition and Neuroscience*, 29(9):1046–1058.
- Trofimovich, P. and Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. 28:1–30.
- Truckenbrodt, H. (1995). *Phonological phrases – their relation to syntax, focus, and prominence*. PhD thesis, Massachusetts Institute of Technology.
- Truckenbrodt, H. (2007). The syntax phonology interface. In de Lacy, P., editor, *The Cambridge handbook of phonology*, pages 435–456. Cambridge University Press, Cambridge.
- Turk, A. and Shattuk-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, 35:445–472.
- Uhmann, S. (1991). *Fokusphonologie. Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie*. Niemeyer, Tübingen.
- Vaissière, J. (1983). Language-independent prosodic features. In Cutler, A. and Ladd, D. R., editors, *Prosody: Models and measurements*. Springer, Berlin.
- van Maastricht, L., Krahmer, E., and Swerts, M. (2015). Prominence patterns in a second language: Intonational transfer from Dutch to Spanish and vice versa. *Language Learning*, 66(1):124–158.

- van Maastricht, L., Krahmer, E., Swerts, M., and Prieto, P. (2019). Learning direction matters: A study on L2 rhythm acquisition by Dutch learners of Spanish and Spanish learners of Dutch. *Studies in Second Language Acquisition*, 41(1):87–121.
- Vogel, I., Bunnell, H. T., and Hoskins, S. (1995). The phonology and phonetics of the rhythm rule. In Connell, B. and Arvaniti, A., editors, *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, Papers in Laboratory Phonology, pages 111–127. Cambridge University Press.
- Vogel, R., van de Vijver, R., Kotz, S., Kutscher, A., and Wagner, P. (2015). Function words in rhythmic optimisation. In Vogel, R. and van de Vijver, R., editors, *Rhythm in Cognition and Grammar*, pages 255–276. De Gruyter.
- Wagner, M. and Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25:905–945.
- Wagner, P. (1999). The synthesis of German contrastive focus. In *Proceedings of the 14th International Congress of Phonetic Sciences (San Francisco)*, volume 2, pages 1529–1532.
- Wagner, P. (2005). Great expectation - introspective vs. perceptual prominence ratings and their acoustic correlates. In *Proceedings of Interspeech 2005*, pages 2381–2384.
- Wagner, P. (2008). *The rhythm of language and speech: Constraining Factors, Models, Metrics and Applications*. online publication.
- Wagner, P. (2010). Two sides of the same coin? Investigating iambic and trochaic timing and prominence in German poetry. In *Proceedings of Speech Prosody 2010*.
- Wagner, P. and Fischenbeck, E. (2002). Stress perception and production in German stress clash environments. In *Proceedings of Speech Prosody 2002*.
- Wang, B. and Féry, C. (2018). Prosody of dual-focus in German: Interaction between focus and phrasing. *Language and Speech*, 61(2):303–333.
- Watson, D. G., Tanenhaus, M. K., and Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H\* vs. L+H\*. *Cognitive Science*, 32(7):1232–1244.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: a study of non-native speakers. *Applied Linguistics*, 15:399–420.

- White, L. and Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35:501–522.
- White, L. S. (2002). *English speech timing: A domain and locus approach*. PhD dissertation, University of Edinburgh.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wickham, H. (2011). The split-apply-combine strategy for data analysis. *Journal of Statistical Software*, 40(1):1–29.
- Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. R package version 1.2.1.
- Wieden, W. (1993). Aspects of acquisitional stages. In Kettemann, B. and Wieden, W., editors, *Current Issues in European Second Language Acquisition research*, pages 125–135. Gunter Narr Verlag, Tübingen.
- Willems, N. (1983). *English Intonation from a Dutch Point of View*. Doctoral Dissertation, University of Utrecht.
- Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv: 1308.5499. [<http://arxiv.org/pdf/1308.5499.pdf>].
- Wunderlich, D. (1988). *Der Ton macht die Melodie - zur Phonologie der Intonation des Deutschen*, pages 1–40. Niemeyer, Tübingen.
- Xu, Y. (2004). Understanding tone from the perspective of production and perception. *Language and Linguistics*, 5:757–797.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, pages 220–251.
- Yule, G. (1980). Intonation and givenness in spoken discourse. *Studies in Language*, pages 271–286.
- Zerbian, S. (2006). *Expression of Information Structure in the Bantu Language Northern Sotho*. PhD dissertation, Humboldt University of Berlin.



- Zerbian, S. (2015). Markedness considerations in L2 prosodic focus and givenness marking. In Delais-Roussaire, E., A. M. and Herment, S., editors, *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*, pages 171–188. Springer-Verlag, Berlin, Heidelberg.
- Zimmerer, F., Andreeva, B., Jügler, J., and Möbius, B. (2015). Comparison of pitch profiles of German and French speakers speaking French and German. In *Proceedings of ICPHS 2015*, Glasgow.
- Zimmermann, M. and Onea, E. (2011). Focus marking and focus interpretation. *Lingua*, 121:1651–1670.



# A. Materials for the L1 German experiment (chapter 3)

## A.1. Context story

*Am Samstag ist die gesamte Verwandtschaft bei Tante Ilse zu Besuch. Es wird viel beredet, über Privates und über das gemeinsame Familienunternehmen. Da Tante Ilse recht schwerhörig ist und auch das Gedächtnis nicht mehr so gut funktioniert, fragt sie am Abend ihre Enkelin nach den Neuigkeiten der vielen Besucher und worüber alles geredet wurde.*

## A.2. Experimental stimuli

### Dialogues in the *clash* condition

1. Hat Vera gesagt, dass Timo die Gläser Sängerinnen gegeben hat?  
Nein, sie hat gesagt, dass Timo das Kostüm Schwimmerinnen gegeben hat.
2. Hat Stefan gesagt, dass Wolfgang den Schlüssel Mieterinnen gegeben hat?  
Nein, er hat gesagt, dass Wolfgang das Papier Anwältinnen gegeben hat.
3. Hat Lisbeth gesagt, dass Margot den Metzger Inderinnen empfohlen hat?  
Nein, sie hat gesagt, dass Margot den Friseur Ungarinnen empfohlen hat.
4. Hat Heike gesagt, dass Moritz die Spende Staatsanwälten gegeben hat?  
Nein, er hat gesagt, dass Moritz den Verweis Aufsichtsräten gegeben hat.
5. Hat Lena gesagt, dass Robert die Puppe Golferinnen gegeben hat?  
Nein, sie hat gesagt, dass Robert den Ballon Schneiderinnen gegeben hat.
6. Hat Timo gesagt, dass Judith die Gelder Bäuerinnen gegeben hat?  
Nein, er hat gesagt, dass Judith den Pokal Reiterinnen gegeben hat.

## A.2. Experimental stimuli

---

7. Hat Jonas gesagt, dass Eva die Nachricht Bauarbeitern gegeben hat?  
Nein, er hat gesagt, dass Eva die Gebühr Hundehaltern gegeben hat.
8. Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?  
Nein, sie hat gesagt, dass Tobi das Klavier Lehrerinnen gegeben hat.
9. Hat Erwin gesagt, dass Alex die Kisten Bäckerinnen gegeben hat?  
Nein, er hat gesagt, dass Alex das Regal Strickerinnen gegeben hat.
10. Hat Heike gesagt, dass Anna das Werkzeug Vorgesetzten gegeben hat?  
Nein, sie hat gesagt, dass Anna das Porträt Niederländern gegeben hat.
11. Hat Dieter gesagt, dass Stefan das Flugblatt Richterinnen gegeben hat?  
Nein, er hat gesagt, dass Stefan das Gewehr Mieterinnen gegeben hat.
12. Hat Lisa gesagt, dass Gitte die Flaschen Krankenpflegern gegeben hat?  
Nein, sie hat gesagt, dass Gitte den Beton Altenpflegern gegeben hat.
13. Hat Kati gesagt, dass Holger die Orgel Sammlerinnen gegeben hat?  
Nein, sie hat gesagt, dass Holger die Skulptur Händlerinnen gegeben hat.
14. Hat Stefan gesagt, dass Kerstin die Stangen Kellnerinnen gegeben hat?  
Nein, er hat gesagt, dass Kerstin das Gestell Bauarbeitern gegeben hat.
15. Hat Daniel gesagt, dass Betti den Schneider Mitarbeitern empfohlen hat?  
Nein, er hat gesagt, dass Betti den Verein Jugendlichen empfohlen hat.
16. Hat Birgit gesagt, dass Lena die Führung Lehrpersonen überlassen hat?  
Nein, sie hat gesagt, dass Lena den Verkauf Rechtsanwälten überlassen hat.
17. Hat Moritz gesagt, dass Markus den Roller Flugbegleitern empfohlen hat?  
Nein, er hat gesagt, dass Markus den Chauffeur Abgesandten empfohlen hat.
18. Hat Gregor gesagt, dass Karin den Hocker Zahntechnikern empfohlen hat?  
Nein, er hat gesagt, dass Karin das Skalpell Frauenärzten empfohlen hat.
19. Hat Uta gesagt, dass Stefan das Alphorn Filmemachern gegeben hat?  
Nein, sie hat gesagt, dass Stefan das Geweih Österreichern gegeben hat.
20. Hat Vera gesagt, dass Cora die Möbel Näherinnen gegeben hat?  
Nein, sie hat gesagt, dass Cora das Geschirr Malerinnen gegeben hat.

### Dialogues in the *no clash* condition

1. Hat Vera gesagt, dass Timo die Gläser Sängerinnen gegeben hat?  
Nein, sie hat gesagt, dass Timo das Kostüm Athletinnen gegeben hat.
2. Hat Stefan gesagt, dass Wolfgang den Schlüssel Mieterinnen gegeben hat?  
Nein, er hat gesagt, dass Wolfgang das Papier Geschäftsfrauen gegeben hat.
3. Hat Lisbeth gesagt, dass Margot den Metzger Inderinnen empfohlen hat?  
Nein, sie hat gesagt, dass Margot den Friseur Bulgarinnen empfohlen hat.
4. Hat Heike gesagt, dass Moritz die Spende Staatsanwälten gegeben hat?  
Nein, er hat gesagt, dass Moritz den Verweis Verteidigern gegeben hat.
5. Hat Lena gesagt, dass Robert die Puppe Golferinnen gegeben hat?  
Nein, sie hat gesagt, dass Robert den Ballon Floristinnen gegeben hat.
6. Hat Timo gesagt, dass Judith die Gelder Bäuerinnen gegeben hat?  
Nein, er hat gesagt, dass Judith den Pokal Autorinnen gegeben hat.
7. Hat Jonas gesagt, dass Eva die Nachricht Bauarbeitern gegeben hat?  
Nein, er hat gesagt, dass Eva die Gebühr Australiern gegeben hat.
8. Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?  
Nein, sie hat gesagt, dass Tobi das Klavier Studentinnen gegeben hat.
9. Hat Erwin gesagt, dass Alex die Kisten Bäckerinnen gegeben hat?  
Nein, er hat gesagt, dass Alex das Regal Dozentinnen gegeben hat.
10. Hat Heike gesagt, dass Anna das Werkzeug Vorgesetzten gegeben hat?  
Nein, sie hat gesagt, dass Anna das Porträt Realschülern gegeben hat.
11. Hat Dieter gesagt, dass Stefan das Flugblatt Richterinnen gegeben hat?  
Nein, er hat gesagt, dass Stefan das Gewehr Juristinnen gegeben hat.
12. Hat Lisa gesagt, dass Gitte die Flaschen Krankenpflegern gegeben hat?  
Nein, sie hat gesagt, dass Gitte den Beton Elektrikern gegeben hat.
13. Hat Kati gesagt, dass Holger die Orgel Sammlerinnen gegeben hat?  
Nein, sie hat gesagt, dass Holger die Skulptur Flötistinnen gegeben hat.
14. Hat Stefan gesagt, dass Kerstin die Stangen Kellnerinnen gegeben hat?  
Nein, er hat gesagt, dass Kerstin das Gestell Mechanikern gegeben hat.

## A.2. Experimental stimuli

---

15. Hat Daniel gesagt, dass Betti den Schneider Mitarbeitern empfohlen hat?  
Nein, er hat gesagt, dass Betti den Verein Beschäftigten empfohlen hat.
16. Hat Birgit gesagt, dass Lena die Führung Lehrpersonen überlassen hat?  
Nein, sie hat gesagt, dass Lena den Verkauf Gesellschaftern überlassen hat.
17. Hat Moritz gesagt, dass Markus den Roller Flugbegleitern empfohlen hat?  
Nein, er hat gesagt, dass Markus den Chauffeur Herausgebern empfohlen hat.
18. Hat Gregor gesagt, dass Karin den Hocker Zahntechnikern empfohlen hat?  
Nein, er hat gesagt, dass Karin das Skalpell Akustikern empfohlen hat.
19. Hat Uta gesagt, dass Stefan das Alphorn Filmemachern gegeben hat?  
Nein, sie hat gesagt, dass Stefan das Geweih Historikern gegeben hat.
20. Hat Vera gesagt, dass Cora die Möbel Näherinnen gegeben hat?  
Nein, sie hat gesagt, dass Cora das Geschirr Französinnen gegeben hat.

### **Dialogues in the *F1* condition**

1. Hat Vera gesagt, dass Timo die Gläser Sängerinnen gegeben hat?  
Nein, sie hat gesagt, dass Timo das Kostüm Sängerinnen gegeben hat.
2. Hat Stefan gesagt, dass Wolfgang den Schlüssel Mieterinnen gegeben hat?  
Nein, er hat gesagt, dass Wolfgang das Papier Mieterinnen gegeben hat.
3. Hat Lisbeth gesagt, dass Margot den Metzger Inderinnen empfohlen hat?  
Nein, sie hat gesagt, dass Margot den Friseur Inderinnen empfohlen hat.
4. Hat Heike gesagt, dass Moritz die Spende Staatsanwälten gegeben hat?  
Nein, er hat gesagt, dass Moritz den Verweis Staatsanwälten gegeben hat.
5. Hat Lena gesagt, dass Robert die Puppe Golferinnen gegeben hat?  
Nein, sie hat gesagt, dass Robert den Ballon Golferinnen gegeben hat.
6. Hat Timo gesagt, dass Judith die Gelder Bäuerinnen gegeben hat?  
Nein, er hat gesagt, dass Judith den Pokal Bäuerinnen gegeben hat.
7. Hat Jonas gesagt, dass Eva die Nachricht Bauarbeitern gegeben hat?  
Nein, er hat gesagt, dass Eva die Gebühr Bauarbeitern gegeben hat.

8. Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?  
Nein, sie hat gesagt, dass Tobi das Klavier Schülerinnen gegeben hat.
9. Hat Erwin gesagt, dass Alex die Kisten Bäckerinnen gegeben hat?  
Nein, er hat gesagt, dass Alex das Regal Bäckerinnen gegeben hat.
10. Hat Heike gesagt, dass Anna das Werkzeug Vorgesetzten gegeben hat?  
Nein, sie hat gesagt, dass Anna das Porträt Vorgesetzten gegeben hat.
11. Hat Dieter gesagt, dass Stefan das Flugblatt Richterinnen gegeben hat?  
Nein, er hat gesagt, dass Stefan das Gewehr Richterinnen gegeben hat.
12. Hat Lisa gesagt, dass Gitte die Flaschen Krankenpflegern gegeben hat?  
Nein, sie hat gesagt, dass Gitte den Beton Krankenpflegern gegeben hat.
13. Hat Kati gesagt, dass Holger die Orgel Sammlerinnen gegeben hat?  
Nein, sie hat gesagt, dass Holger die Skulptur Sammlerinnen gegeben hat.
14. Hat Stefan gesagt, dass Kerstin die Stangen Kellnerinnen gegeben hat?  
Nein, er hat gesagt, dass Kerstin das Gestell Kellnerinnen gegeben hat.
15. Hat Daniel gesagt, dass Betti den Schneider Mitarbeitern empfohlen hat?  
Nein, er hat gesagt, dass Betti den Verein Mitarbeitern empfohlen hat.
16. Hat Birgit gesagt, dass Lena die Führung Lehrpersonen überlassen hat?  
Nein, sie hat gesagt, dass Lena den Verkauf Lehrpersonen überlassen hat.
17. Hat Moritz gesagt, dass Markus den Roller Flugbegleitern empfohlen hat?  
Nein, er hat gesagt, dass Markus den Chauffeur Flugbegleitern empfohlen hat.
18. Hat Gregor gesagt, dass Karin den Hocker Zahntechnikern empfohlen hat?  
Nein, er hat gesagt, dass Karin das Skalpell Zahntechnikern empfohlen hat.
19. Hat Uta gesagt, dass Stefan das Alphorn Filmemachern gegeben hat?  
Nein, sie hat gesagt, dass Stefan das Geweih Filmemachern gegeben hat.
20. Hat Vera gesagt, dass Cora die Möbel Näherinnen gegeben hat?  
Nein, sie hat gesagt, dass Cora das Geschirr Näherinnen gegeben hat.

### Dialogues in the *F2* condition

1. Hat Vera gesagt, dass Timo die Gläser Sängerinnen gegeben hat?  
Nein, sie hat gesagt, dass Timo die Gläser Schwimmerinnen gegeben hat.
2. Hat Stefan gesagt, dass Wolfgang den Schlüssel Mieterinnen gegeben hat?  
Nein, er hat gesagt, dass Wolfgang den Schlüssel Anwältinnen gegeben hat.
3. Hat Lisbeth gesagt, dass Margot den Metzger Inderinnen empfohlen hat?  
Nein, sie hat gesagt, dass Margot den Metzger Ungarinnen empfohlen hat.
4. Hat Heike gesagt, dass Moritz die Spende Staatsanwälten gegeben hat?  
Nein, er hat gesagt, dass Moritz die Spende Aufsichtsräten gegeben hat.
5. Hat Lena gesagt, dass Robert die Puppe Golferinnen gegeben hat?  
Nein, sie hat gesagt, dass Robert die Puppe Schneiderinnen gegeben hat.
6. Hat Timo gesagt, dass Judith die Gelder Bäuerinnen gegeben hat?  
Nein, er hat gesagt, dass Judith die Gelder Reiterinnen gegeben hat.
7. Hat Jonas gesagt, dass Eva die Nachricht Bauarbeitern gegeben hat?  
Nein, er hat gesagt, dass Eva die Nachricht Hundehaltern gegeben hat.
8. Hat Melli gesagt, dass Tobi das Schlagzeug Schülerinnen gegeben hat?  
Nein, sie hat gesagt, dass Tobi das Schlagzeug Lehrerinnen gegeben hat.
9. Hat Erwin gesagt, dass Alex die Kisten Bäckerinnen gegeben hat?  
Nein, er hat gesagt, dass Alex die Kisten Strickerinnen gegeben hat.
10. Hat Heike gesagt, dass Anna das Werkzeug Vorgesetzten gegeben hat?  
Nein, sie hat gesagt, dass Anna das Werkzeug Niederländern gegeben hat.
11. Hat Dieter gesagt, dass Stefan das Flugblatt Richterinnen gegeben hat?  
Nein, er hat gesagt, dass Stefan das Flugblatt Mieterinnen gegeben hat.
12. Hat Lisa gesagt, dass Gitte die Flaschen Krankenpflegern gegeben hat?  
Nein, sie hat gesagt, dass Gitte die Flaschen Altenpflegern gegeben hat.
13. Hat Kati gesagt, dass Holger die Orgel Sammlerinnen gegeben hat?  
Nein, sie hat gesagt, dass Holger die Orgel Händlerinnen gegeben hat.
14. Hat Stefan gesagt, dass Kerstin die Stangen Kellnerinnen gegeben hat?  
Nein, er hat gesagt, dass Kerstin die Stangen Bauarbeitern gegeben hat.



15. Hat Daniel gesagt, dass Betti den Schneider Mitarbeitern empfohlen hat?  
Nein, er hat gesagt, dass Betti den Schneider Jugendlichen empfohlen hat.
16. Hat Birgit gesagt, dass Lena die Führung Lehrpersonen überlassen hat?  
Nein, sie hat gesagt, dass Lena die Führung Rechtsanwälten überlassen hat.
17. Hat Moritz gesagt, dass Markus den Roller Flugbegleitern empfohlen hat?  
Nein, er hat gesagt, dass Markus den Roller Abgesandten empfohlen hat.
18. Hat Gregor gesagt, dass Karin den Hocker Zahntechnikern empfohlen hat?  
Nein, er hat gesagt, dass Karin den Hocker Frauenärzten empfohlen hat.
19. Hat Uta gesagt, dass Stefan das Alphorn Filmemachern gegeben hat?  
Nein, sie hat gesagt, dass Stefan das Alphorn Österreichern gegeben hat.
20. Hat Vera gesagt, dass Cora die Möbel Näherinnen gegeben hat?  
Nein, sie hat gesagt, dass Cora die Möbel Malerinnen gegeben hat.

### A.3. Filler items

1. Hat Marion erzählt, dass Nora den Kranz selbst gemacht hat?  
Ja, und dass Phillip ihr dabei geholfen hat.
2. Hat Ursula erzählt, dass Johann ihr beim Einkaufen geholfen hat?  
Nein, sie hat erzählt, dass Johann ihr das Auto Gewaschen hat.
3. Hat Simon gesagt, dass er über Weihnachten in Urlaub fährt?  
Nein, er hat gesagt, dass er nach Weihnachten In Urlaub fährt.
4. Hat Charlotte gesagt, dass Marianne sie nicht mehr besuchen kommt?  
Ja, anscheinend ist Marianne nicht mehr gut zu Fuß.
5. Hat Boris gesagt, wann er mit dem Studium fertig sein wird?  
Nein, er spricht nicht gerne darüber.
6. Hat Antje gesagt, dass Frank das Untergeschoss vermietet hat?  
Nein, sie hat gesagt, dass Frank das Ferienhaus Verkauft hat.
7. Hat Gregor erzählt, dass er Eva einen Heiratsantrag gemacht hat?  
Nein, er hat erzählt, dass er vorhat Eva einen Antrag zu machen.

### A.3. Filler items

---

8. Hat Maren gesagt, dass sie einen Kindergartenplatz bekommen haben?  
Nein, sie hat gesagt, dass sie erst nächsten Jahr Einen Kindergartenplatz bekommen.
9. Hat Peter von seiner neuen Wohnung erzählt?  
Nein, Peter war doch gar nicht da.
10. Hat Konrad gesagt, warum Sybille nicht mitgekommen ist?  
Ja, sie musste heute wohl länger arbeiten.
11. Hat Sabine gesagt, dass sie das Essen nicht mochte?  
Nein, sie hat gesagt, dass sie das Essen sehr gut fand.
12. Hat Sarah gesagt, dass Thomas den Kuchen gebacken hat?  
Nein, sie hat gesagt, dass Thomas den Kuchen Gegessen hat.
13. Hat Marion erzählt, dass Jens seine Arbeit verloren hat?  
Ja, aber er wollte dort ohnehin nicht mehr arbeiten.
14. Hat Frank von seinem Urlaub in Italien erzählt?  
Ja, er hat erzählt, dass es ihm dort gut gefallen hat.
15. Hat Ann gesagt, dass Jonas das Schlafzimmer grün gestrichen hat?  
Nein, sie hat gesagt, dass Jonas ein grünes Sofa Gekauft hat.
16. Hat Liz gesagt, dass Peter den neuen Fernseher gekauft hat?  
Ja, und sie meint er sei viel zu teuer gewesen.
17. Hat Olaf gesagt, dass er eine neue Freundin hat?  
Ja, er hat sie bei der Arbeit kennengelernt.
18. Hat Bernd gesagt, dass Nora überlegt zurück nach Deutschland zu kommen?  
Nein, sie wird noch mindestens zwei Jahre In Wien bleiben.
19. Hat Susi gesagt, ob es Ruth wieder besser geht?  
Ja, es gab in letzter Zeit wohl nur gute Nachrichten.
20. Hat Max erzählt, dass seine Mutter jetzt Theater spielt?  
Ja, und sie singe auch im Chor.
21. Hat Kim gesagt, dass Julian ein neues Auto hat?  
Ja, einen Kombi mit Platz für die ganze Familie.

22. Hat Maria erzählt, dass der kleine Max nicht gut hören kann?  
Nein, sie hat gesagt, dass er wahrscheinlich Eine Brille braucht.
23. Hat Joseph gesagt, dass er Morgen nochmal kommt?  
Ja, er kommt nochmal um dir bei der Umstellung Des Schlafzimmers zu helfen.
24. Hat Sybille gesagt, dass Jens mit Italienern zusammen wohnt?  
Nein, sie hat erzählt, dass er bei Spanierinnen wohnt.
25. Hat Marion die Pflanzen im Wintergarten gegossen?  
Ja, sie hat sich um alle Pflanzen gekümmert.
26. Haben die Kinder alle Geschenke mitgenommen?  
Ja, sie haben sich sehr gefreut und alles eingepackt.
27. Hat Markus die Gartenstühle wieder zurück gestellt?  
Nein, er musste früher gehen, aber ich mach das Nachher.
28. Hat Florian gesagt, dass Frank und er die Filiale in Bonn übernehmen?  
Nein, sie wollen wohl lediglich Peter unter die Arme Greifen.
29. Hat Kerstin erzählt, wie es Marie in der neuen Schule ergeht?  
Ja, sie hat gesagt, dass Marie sehr gerne in die neue Schule geht.
30. Hat Gerhard die alten Zeitungen mitgenommen?  
Ja, er könne noch etwas damit anfangen, meinte er.
31. Waren Franzi und Olaf mit dem alten VW-Bus da?  
Nein, den mussten sie letztes Jahr schon verschrotten.
32. Hat Vera gesagt, dass Julian eine neue Stelle hat?  
Ja, deshalb konnte er heute auch nicht dabei sein.
33. Hat Jens gesagt, dass Miriam Nachwuchs erwartet?  
Nein, da hat er von einer Freundin erzählt.
34. Hat Clemens gesagt, dass Norbert das restliche Essen mitgenommen hat?  
Nein, er hat gesagt, dass Norbert das restliche Essen Mitgenommen hätte, aber es war ja nichts übrig.
35. Hat Frieder gesagt, wie es Hans mit seinen Augen geht?  
Ja, er hat erzählt, dass Hans' Augen seit der Operation Wieder sehr gut sind.

### A.3. Filler items

---

36. Hat Uli gesagt, dass Margot jetzt in einem Kulturzentrum arbeitet?  
Ja, genau, sie ist jetzt immer vormittags dort.
37. Hat Stefan gesagt, dass Wolfgang die Geschäftsübergabe bereut?  
Nein, er meinte, dass Wolfgang froh ist, dass jetzt alles Abgewickelt ist.
38. Hat Gitte gesagt, ob es Lisa wieder besser geht?  
Ja, Lisa steht wieder voll im Berufsleben.
39. Hat Stefan gesagt, dass sie das alte Lager ausgeräumt haben?  
Er hat gesagt, dass sie erst angefangen haben und Es noch viel Arbeit sei.
40. Hat Eva gesagt, dass sie noch bei der Stadtverwaltung arbeitet?  
Ja, genau, sie hat jetzt einen unbefristeten Vertrag.

## B. Materials for the English experiments (chapters 4 and 5)

### B.1. Context story

*On Saturday, the whole family is visiting Aunt Isabelle. There's a lot of conversation about private issues and family business. Since Aunt Isabelle is quite hard of hearing and her memory isn't so good anymore, she asks her granddaughter about what the visitors were talking about.*

### B.2. Experimental stimuli for the experiments in 4.2 and 5.3

#### Dialogues in the *clash* condition

1. Did Freddy say that little Johnny fed the cats raspberries?  
No, he said that he fed the mice apricots.
2. Did Lucas say that he recommended Donna folding-chairs for her camping trip?  
No, he said that he recommended Lucille safety pins.
3. Did John say that Judy made his dad orange marmalade?  
No, he said that she made his kids elderberry juice.
4. Did your mum say that she lent Russel overalls for a costume party?  
No, she said that she lent Maureen eye shadow.
5. Did Carter say that he showed Dylan animal pounds in Corsica?  
No, he said that he showed Maxine falconry birds.
6. Did Eva say that she cooked Lenny vegetable stew?  
No, she said that she cooked Denise mulligan stew.

## B.2. Experimental stimuli for the experiments in 4.2 and 5.3

---

7. Did Ryan say that he lost Rory's master key and she got into trouble?  
No, he said that he lost Estelle's access code.
8. Did Logan say that he showed his friend jellyfish populations in Africa?  
No, he said that he showed his son elephant populations.
9. Did Miss Jane say that she brought the park wallabies?  
No, she said that she brought the zoo nightingales.
10. Did Richard say that he offered the court sentence reductions?  
No, he said that he offered the firm amnesty clauses.
11. Did Maggie say that George brought his mum flowers this morning?  
No, she said that he brought his wife jewelry.
12. Did Nora say that Henry promised his aide chocolate cake for the workshop?  
No, she said that he promised his boss ice cream cake.
13. Did Taylor say that he engraved Riley's fountain pen?  
No, he said that he engraved Renee's hunting knives.
14. Did Ivy say that Boris offered the church swimming courses?  
No, she said that he offered the school holiday programs.
15. Did Rosy say that she showed Freddy weasels in the park?  
No, she said that she showed Patrice wildcats in the park.
16. Did Carl say that Clara gave the boys horror stories to read?  
No, he said that she gave the girls romance novels.
17. Did Ginger say that she sold Micah's food processor?  
No, she said that she sold Jeanette's washing machine.
18. Did Joan say that she gave the coach fitness equipment?  
No, she said that she gave the dean answering machines.
19. Did Mike say that he showed Parker mountain equipment for their Tibet trip?  
No, he said that he showed Nicole hospital locations.
20. Did Eva say that Horace gave the cop evidence bags?  
No, she said that he gave the judge risk assessments.

**Dialogues in the *no clash* condition**

1. Did Freddy say that little Johnny fed the cats raspberries?  
No, he said that he fed the mice zucchini.
2. Did Lucas say that he recommended Donna folding-chairs for her camping trip?  
No, he said that he recommended Lucille repair kits.
3. Did John say that Judy made his dad orange marmalade?  
No, he said that she made his kids Spaghetti sauce.
4. Did your mum say that she lent Russel overalls for a costume party?  
No, she said that she lent Maureen suspenders.
5. Did Carter say that he showed Dylan animal pounds in Corsica?  
No, he said that he showed Maxine alpaca farms.
6. Did Eva say that she cooked Lenny vegetable stew?  
No, she said that she cooked Denise alfalfa sprouts.
7. Did Ryan say that he lost Rory's master key and she got into trouble?  
No, he said that he lost Estelle's account books.
8. Did Logan say that he showed his friend jellyfish populations in Africa?  
No, he said that he showed his son chimpanzee populations.
9. Did Miss Jane say that she brought the park wallabies?  
No, she said that she brought the zoo flamingos.
10. Did Richard say that he offered the court sentence reductions?  
No, he said that he offered the firm exemption clauses.
11. Did Maggie say that George brought his mum flowers this morning?  
No, she said that he brought his wife champagne.
12. Did Nora say that Henry promised his aide chocolate cake for the workshop?  
No, she said that he promised his boss lasagna.
13. Did Taylor say that he engraved Riley's fountain pen?  
No, he said that he engraved Renee's engagement ring.
14. Did Ivy say that Boris offered the church swimming courses?  
No, she said that she offered the school vacation programs.

## B.2. Experimental stimuli for the experiments in 4.2 and 5.3

---

15. Did Rosy say that she showed Freddy weasels in the park?  
No, she said that she showed Patrice raccoons in the park.
16. Did Carl say that Clara gave the boys horror stories to read?  
No, he said that she gave the girls adventure books.
17. Did Ginger say that she sold Micah's food processor?  
No, she said that she sold Jeanette's harmonica.
18. Did Joan say that she gave the coach fitness equipment?  
No, she said that she gave the dean assessment reports.
19. Did Mike say that he showed Parker mountain equipment for their Tibet trip?  
No, he said that he showed Nicole emergency numbers.
20. Did Eva say that Horace gave the cop evidence bags?  
No, she said that he gave the judge insurance claims.

### **Dialogues in the *F1* condition**

1. Did Freddy say that little Johnny fed the cats raspberries?  
No, he said that he fed the mice raspberries.
2. Did Lucas say that he recommended Donna folding-chairs for her camping trip?  
No, he said that he recommended Lucille folding chairs.
3. Did John say that Judy made his dad orange marmalade?  
No, he said that she made his kids orange marmalade.
4. Did your mum say that she lent Russel overalls for a costume party?  
No, she said that she lent Maureen overalls.
5. Did Carter say that he showed Dylan animal pounds in Corsica?  
No, he said that he showed Maxine animal pounds.
6. Did Eva say that she cooked Lenny vegetable stew?  
No, she said that she cooked Denise vegetable stew.
7. Did Ryan say that he lost Rory's master key and she got into trouble?  
No, he said that he lost Estelle's master key.



8. Did Logan say that he showed his friend jellyfish populations in Africa?  
No, he said that he showed his son jellyfish populations.
9. Did Miss Jane say that she brought the park wallabies?  
No, she said that she brought the zoo wallabies.
10. Did Richard say that he offered the court sentence reductions?  
No, he said that he offered the firm sentence reduction.
11. Did Maggie say that George brought his mum flowers this morning?  
No, she said that he brought his wife flowers.
12. Did Nora say that Henry promised his aide chocolate cake for the workshop?  
No, she said that he promised his boss chocolate cake.
13. Did Taylor say that he engraved Riley's fountain pen?  
No, he said that he engraved Renee's fountain pen.
14. Did Ivy say that Boris offered the church swimming courses?  
No, she said that he offered the school swimming courses.
15. Did Rosy say that she showed Freddy weasels in the park?  
No, she said that she showed Patrice weasels in the park.
16. Did Carl say that Clara gave the boys horror stories to read?  
No, he said that she gave the girls horror stories.
17. Did Ginger say that she sold Micah's food processor?  
No, she said that she sold Jeanette's food processor.
18. Did Joan say that she gave the coach fitness equipment?  
No, she said that she gave the dean fitness equipment.
19. Did Mike say that he showed Parker mountain equipment for their Tibet trip?  
No, he said that he showed Nicole mountain equipment.
20. Did Eva say that Horace gave the cop evidence bags?  
No, she said that he gave the judge evidence bags.

### Dialogues in the *F2* condition

1. Did Freddy say that little Johnny fed the cats raspberries?  
No, he said that he fed the cats apricots.
2. Did Lucas say that he recommended Donna folding-chairs for her camping trip?  
No, he said that he recommended Donna safety pins.
3. Did John say that Judy made his dad orange marmalade?  
No, he said that she made his dad elderberry juice.
4. Did your mum say that she lent Russel overalls for a costume party?  
No, she said that she lent Russel eye shadow.
5. Did Carter say that he showed Dylan animal pounds in Corsica?  
No, he said that he showed Dylan falconry birds.
6. Did Eva say that she cooked Lenny vegetable stew?  
No, she said that she cooked Lenny mulligan stew.
7. Did Ryan say that he lost Rory's master key and she got into trouble?  
No, he said that he lost Rory's access code.
8. Did Logan say that he showed his friend jellyfish populations in Africa?  
No, he said that he showed his friend elephant populations.
9. Did Miss Jane say that she brought the park wallabies?  
No, she said that she brought the zoo nightingales.
10. Did Richard say that he offered the court sentence reductions?  
No, he said that he offered the court amnesty clauses.
11. Did Maggie say that George brought his mum flowers this morning?  
No, she said that he brought his mum jewelry.
12. Did Nora say that Henry promised his aide chocolate cake for the workshop?  
No, she said that he promised his aide ice cream cake.
13. Did Taylor say that he engraved Riley's fountain pen?  
No, he said that he brought Riley's hunting knife.
14. Did Ivy say that Boris offered the church swimming courses?  
No, she said that he offered the church holiday programs.

15. Did Rosy say that she showed Freddy weasels in the park?  
No, she said that she showed Freddy wildcats in the park.
16. Did Carl say that Clara gave the boys horror stories to read?  
No, he said that she gave the boys romance novels.
17. Did Ginger say that she sold Micah's food processor?  
No, she said that she sold Micah's washing machine.
18. Did Joan say that she gave the coach fitness equipment?  
No, she said that she gave the coach answering machines.
19. Did Mike say that he showed Parker mountain equipment for their Tibet trip?  
No, he said that he showed Parker hospital locations.
20. Did Eva say that Horace gave the cop evidence bags?  
No, she said that he gave the cop risk assessments.

### **B.3. Experimental stimuli for the experiments in 4.3 and 5.4**

#### **Dialogues in the *clash* condition**

1. Did Angela say that she saw a newborn zebra in the zoo?  
No, she said that she saw a full-grown rhino.
2. Did Pete say that he enjoyed the Jamaican photographs?  
No, he said that he enjoyed the Nepalese scenery.
3. Did Conny say that she broke the video player?  
No, she said that she broke the CD changer.
4. Did Jenny say that she collects Indian violins?  
No, she said that she collects Hong Kong cellos.
5. Did Portia say that she witnessed a serene protest?  
No, she said that she witnessed a full-blown riot.
6. Did Jimmy talk about a German restaurant?  
No, he talked about a Chinese enterprise.

### B.3. Experimental stimuli for the experiments in 4.3 and 5.4

---

7. Did Daniel say that he ordered twelve vodka bottles?  
No, he said that he ordered fifteen whiskey bottles.
8. Did Alan say that he received a Kenyan package?  
No, he said that he received a Maltese letter.
9. Did Anna say that she met an Indian programmer?  
No, she said that she met a Japanese architect.
10. Did Molly say that she works as a senior researcher?  
No, she said that she works as a trainee lecturer.
11. Did Bob say that he visited an African theatre?  
No, he said that he visited a Lebanese restaurant.
12. Did Andrew say that his wife is a NATO analyst?  
No, he said that she is a UN emissary.
13. Did Laura say that she uses UV lasers at work?  
No, she said that she uses infrared microscopes.
14. Did Ellen say that she works with national senators?  
No, she said that she works with overseas actresses.
15. Did Anne say that Paul invited her to do some morning exercise?  
No, she said that he invited her to an afternoon meeting.
16. Did Francis say that he fears illegal processes in his new job?  
No, he said that he fears unfair judgements.
17. Did Benni praise the Belgian strawberries?  
No, he praised the home-grown celery.
18. Did Betty say that she bought a wooden megaphone?  
No, she said that she bought a bamboo vibraphone.
19. Did Amy say that she is an expert in royal couples?  
No, she said that she is an expert in TV actors.
20. Did Jensen say that his daughter is a hostel owner?  
No, he said that she is a hotel manager.

**Dialogues in the *no clash* condition**

1. Did Angela say that she saw a newborn zebra in the zoo?  
No, she said that she saw a full-grown giraffe.
2. Did Pete say that he enjoyed the Jamaican photographs?  
No, he said that he enjoyed the Nepalese geography.
3. Did Conny say that she broke the video player?  
No, she said that she broke the CD recorder.
4. Did Jenny say that she collects Indian violins?  
No, she said that she collects Hong Kong violas.
5. Did Portia say that she witnessed a serene protest?  
No, she said that she witnessed a full-blown assault.
6. Did Jimmy talk about a German restaurant?  
No, he talked about a Chinese musician.
7. Did Daniel say that he ordered twelve vodka bottles?  
No, he said that he ordered fifteen champagne bottles.
8. Did Alan say that he received a Kenyan package?  
No, he said that he received a Maltese response.
9. Did Anna say that she met an Indian programmer?  
No, she said that she met a Japanese accountant.
10. Did Molly say that she works as a senior researcher?  
No, she said that she works as a trainee mechanic.
11. Did Bob say that he visited an African theatre?  
No, he said that he visited a Lebanese museum.
12. Did Andrew say that his wife is a NATO analyst?  
No, he said that she is a UN ambassador.
13. Did Laura say that she uses UV lasers at work?  
No, she said that she uses infrared photography.
14. Did Ellen say that she works with national senators?  
No, she said that she works with overseas attorneys.

### B.3. Experimental stimuli for the experiments in 4.3 and 5.4

---

15. Did Anne say that Paul invited her to do some morning exercise?  
No, she said that he invited her to an afternoon event.
16. Did Francis say that he fears illegal processes in his new job?  
No, he said that he fears unfair requirements.
17. Did Benni praise the Belgian strawberries?  
No, he praised the home-grown asparagus.
18. Did Betty say that she bought a wooden megaphone?  
No, she said that she bought a bamboo marimba.
19. Did Amy say that she is an expert in royal couples?  
No, she said that she is an expert in TV celebrities.
20. Did Jensen say that his daughter is a hostel owner?  
No, he said that she is a hotel employee.

### **Dialogues in the *S-Foc* condition**

1. Did Angela say that she saw a newborn zebra in the zoo?  
No, she said that she saw a full-grown zebra.
2. Did Pete say that he enjoyed the Jamaican photographs?  
No, he said that he enjoyed the Nepalese photographs.
3. Did Conny say that she broke the video player?  
No, she said that she broke the CD recorder.
4. Did Jenny say that she collects Indian violins?  
No, she said that she collects Hong Kong violins.
5. Did Portia say that she witnessed a serene protest?  
No, she said that she witnessed a full-blown protest.
6. Did Jimmy talk about a German restaurant?  
No, he talked about a Chinese restaurant.
7. Did Daniel say that he ordered twelve vodka bottles?  
No, he said that he ordered fifteen vodka bottles.

8. Did Alan say that he received a Kenyan package?  
No, he said that he received a Maltese package.
9. Did Anna say that she met an Indian programmer?  
No, she said that she met a Japanese programmer.
10. Did Molly say that she works as a senior researcher?  
No, she said that she works as a trainee researcher.
11. Did Bob say that he visited an African theatre?  
No, he said that he visited a Lebanese theatre.
12. Did Andrew say that his wife is a NATO analyst?  
No, he said that she is a UN analyst.
13. Did Laura say that she uses UV lasers at work?  
No, she said that she uses infrared lasers.
14. Did Ellen say that she works with national senators?  
No, she said that she works with overseas senators.
15. Did Anne say that Paul invited her to do some morning exercise?  
No, she said that he invited her to an afternoon exercise.
16. Did Francis say that he fears illegal processes in his new job?  
No, he said that he fears unfair processes.
17. Did Benni praise the Belgian strawberries?  
No, he praised the home-grown strawberries.
18. Did Betty say that she bought a wooden megaphone?  
No, she said that she bought a bamboo megaphone.
19. Did Amy say that she is an expert in royal couples?  
No, she said that she is an expert in TV couples.
20. Did Jensen say that his daughter is a hostel owner?  
No, he said that she is a hotel owner.

**Dialogues in the *given* condition**

1. Did Angela say that she saw a full-grown zebra in the zoo?  
No, she said that she saw a full-grown rhino.
2. Did Pete say that he enjoyed the Nepalese photographs?  
No, he said that he enjoyed the Nepalese scenery.
3. Did Conny say that she broke the CD player?  
No, she said that she broke the CD changer.
4. Did Jenny say that she collects Hong Kong violins?  
No, she said that she collects Hong Kong cellos.
5. Did Portia say that she witnessed a full-blown protest?  
No, she said that she witnessed a full-blown riot.
6. Did Jimmy talk about a Chinese restaurant?  
No, he talked about a Chinese enterprise.
7. Did Daniel say that he ordered fifteen vodka bottles?  
No, he said that he ordered fifteen whiskey bottles.
8. Did Alan say that he received a Maltese package?  
No, he said that he received a Maltese letter.
9. Did Anna say that she met an Japanese programmer?  
No, she said that she met a Japanese architect.
10. Did Molly say that she works as a trainee researcher?  
No, she said that she works as a trainee lecturer.
11. Did Bob say that he visited a Lebanese theatre?  
No, he said that he visited a Lebanese restaurant.
12. Did Andrew say that his wife is a UN analyst?  
No, he said that she is a UN emissary.
13. Did Laura say that she uses infrared lasers at work?  
No, she said that she uses infrared microscopes.
14. Did Ellen say that she works with overseas senators?  
No, she said that she works with overseas actresses.



15. Did Anne say that Paul invited her to an afternoon exercise?  
No, she said that he invited her to an afternoon meeting.
16. Did Francis say that he fears unfair processes in his new job?  
No, he said that he fears unfair judgements.
17. Did Benni praise the home-grown strawberries?  
No, he praised the home-grown celery.
18. Did Betty say that she bought a bamboo megaphone?  
No, she said that she bought a bamboo vibraphone.
19. Did Amy say that she is an expert in TV couples?  
No, she said that she is an expert in TV actors.
20. Did Jensen say that his daughter is a hotel owner?  
No, he said that she is a hotel manager.

#### **B.4. Filler items**

1. Did Jonathan say that Jenny bought a new TV?  
No, he said that Jenny bought a new toaster.
2. Did Gretchen say that she bought a blue hat?  
No, she said that she found a blue hat in the park.
3. Did Paula say that Henry met a movie director?  
No, she said that he would like to meet one.
4. Did George say that Isabel failed her driving test?  
No, he said that she failed her medical exam.
5. Did Berta say that she works as a children's nurse?  
No, she said that she takes care of old people.
6. Did Daniel say that his wife owns a restaurant?  
No, he said that his wife manages a restaurant.
7. Did Andy say that he used a new recipe for the stew?  
Yes, he said that he got the recipe from a colleague.

#### B.4. Filler items

---

8. Did Anna say that she reads old newspapers?  
Yes, she found them in her attic.
9. Did Denise say that she made pasta salad?  
Yes, she used her grandmother's recipe.
10. Did Ella and Carl get engaged?  
Yes, and they want you to come to the wedding in June.
11. Did Ellen say that John quit university?  
Yes, he wants to travel the world now.
12. Did Judy say that she went to the cinema yesterday?  
Yes, but she didn't like the movie.
13. Did Kristen say that Ben adopted a puppy?  
No, she said that he adopted a little cat.
14. Did Helga say that she set fire to her kitchen?  
No, but she forgot the pie and had to clean the whole oven.
15. Did Boris say that his daughter moved to China?  
No, he said that she moved to Canada.
16. Did Elsa say that Jason had a car accident?  
Yes, he was picked up by the ambulance but everything is ok.
17. Did Adrian say that his brother is a news reporter?  
No, he said that his brother reports the weather in the radio.
18. Did Harry say that he wants to go on a cruise?  
Yes, he's been saving up for months.
19. Did Nora say that her husband is in the military?  
Yes, he's currently deployed in Afghanistan.
20. Did Tommy say that his cat ran away?  
No, he said that his goldfish died.
21. Did Cathy say that her mum lost her job?  
No, she said that her mom switched departments.

22. Did Justin say that he wrote his memoirs?  
Yes, but he is still looking for a publisher.
23. Did Tom say that he went to the new theatre?  
No, he said that he wanted to but there were no tickets left.
24. Did Josh say that his sister is pregnant?  
Yes, and he said that she is hoping for twins.
25. Did Nicole say that she works in a bakery?  
No, she said that she works in a gym.
26. Did Marilyn say that Joseph needs a haircut?  
No, she said that Joseph is going bald.
27. Did Patricia say that she wrote a poem?  
No, she said that she wrote the lyrics for Ben's songs.
28. Did Mona say that she bought a vegan cake?  
Yes, and everyone seemed to like it.
29. Did Bruno say that he began to play the saxophone?  
No, he said that he began to play the drums.
30. Did Adam say that he lost his phone?  
Yes, he's been looking everywhere.
31. Did Eric say that he attended the church meeting last week?  
Yes, he said that the church meeting went well.
32. Did Sarah say that Peter failed the final exam?  
Yes, unfortunately Peter had a headache.
33. Are Robert and Pete going to travel to Africa in autumn?  
Yes, they have already booked their tickets.
34. Did Sophie say that she started her violin lessons?  
Yes, she said that she loves her violin teacher.
35. Did Jonathan say why his girlfriend wasn't there?  
Yes, he said that Janie had broken up with him.

#### B.4. Filler items

---

36. Did Ethan invite the neighbours over for dinner?  
Yes, but the Andersons weren't at home.
37. Did Josie bake an apple cake?  
Yes, the cake was very delicious.
38. Did Kathy say that she is still working on her honours thesis?  
Yes, she is going to hand it in next week.
39. Did Peter say what he wants to do for his birthday?  
Yes, he said that he would like to go to a Mexican restaurant.
40. Did Mateo say that he is going to get married?  
Yes, he said that he is excited about the wedding.
41. Did Daniel say that he had knee surgery?  
Yes, he had to walk on crutches but it is getting better.
42. Did Anne say that she broke her arm?  
No, she said that she only bruised it, fortunately.
43. Did William say that he likes his new sweater?  
Yes, he said that the sweater is cozy.
44. Did Eleanor say that she wants to sell the house?  
Yes, the rent is too high.
45. Did Anthony say that he bought a new car?  
No, he is going to lease a new car.
46. Did Chris explain why he quit his studies?  
No, he didn't want to talk about his studies.
47. Did Peter complain about his birthday present?  
No, he really seems to like his new pajamas.
48. Did everybody watch the soccer game last Friday?  
Yes, except Jonas who had to work late last Friday.
49. Did Peter say that he's planning to travel to Europe in autumn?  
No, he's planning to travel to Africa next year.

50. Did Linda say that she is allergic to peanuts?  
No, she said that she is allergic to dust.
51. Did Nora say that her cousin is a doctor?  
No, she said that her uncle runs a drug store.
52. Did Caroline say that she is taking a bus to school?  
No, she always rides her bike to school.
53. Did Peter say that he finally got his driver's license?  
No, he said that he failed the driving test again.
54. Did Anne say that she doesn't like her new teachers?  
No, she said that her classmates are very nice.
55. Did Claus say that he finished his race thing in time?  
No, unfortunately he didn't qualify for the Marathon.
56. Did Julia say that she went to the Rolling Stones concert?  
No, she said that she wanted to but the tickets were sold out.
57. Did Andrea say that she will attend the family camp this year?  
No, she said that she wants to work all summer.
58. Did Laney say that she wanted to go to the park?  
No, she wanted to go to the public pool.
59. Did Louis say that he will take the old cupboard?  
Yes, he said that it fits well into his new flat.
60. Did the dog steal the cookies?  
No, the children ate them all.
61. Did Daisy say that her computer crashed?  
No, she said that she had some problems with her printer.
62. Did Alec say that he wants to rent a car?  
Yes, he's tired of taking the bus to work.
63. Did Sara talk about the new furniture?  
Yes, she's very fond of the new closet.

#### B.4. Filler items

---

64. Shelby say that she likes her living room?  
No, she said that she plans to paint the walls afresh.
65. Did Anthea say that she wants a divorce?  
No, she just applied for marriage counselling.
66. Did Suzy say that her brother lost his job?  
No, she said that her brother was relocated.
67. Did Calvin say that he built a table?  
Yes, he's been planning it for years and finished it yesterday.
68. Did Ismael make it to this party on time?  
No, there's been a delay at the train station.
69. Did Marcus say that he finished his doctorate?  
Yes, the date for the defense is in April.
70. Did Madita say that her new boyfriend plays the flute?  
Yes, she said that he even plays in an orchestra.
71. Did Jill say that she wants to eat vegan?  
Yes, and she is very good at cooking vegan meals now.
72. Did Theodore say that his wife is in Las Vegas?  
Yes, she is there with her friend's Bachelorette party.
73. Did Beatrice like the decoration?  
Yes, she was amazed by the piñata.
74. Did Martin bring his CD player?  
Yes, he has good taste in music.
75. Did Katherine say that she liked the appetizer?  
No, but she always complains about everything.
76. Did Charles say that she gave John yellow books?  
No, she said that she gave him concert tickets.
77. Did Debbie say that she found the stolen boxes?  
No, she said that she found old cameras.

78. Alex say that she left Kerry at home?  
No, she said that she left the little dog in the garden.
79. Did Sandra explain why Ben didn't go shopping?  
Yes, she said that he was busy at work.
80. Did Peter say that he broke Simon's oven?  
No, he said that he broke Simon's kettle.
81. Did Kate say that Franklin bought the apple pies?  
No, Franklin wasn't even here today!
82. Did Robin say that she played golf with Victor?  
No, she said that she played tennis with Paul.
83. Did Kristen say that she visited her mother?  
No, she said her mother could already leave the hospital.
84. Did Winston say that he met Rufus last night?  
No, he said that he met Rufus last week.
85. Did Adrian say that he decided to go to Spain?  
Yes, but he decided to go to France first.
86. Did Nancy say that she sold her bike?  
Yes, she said that she wanted a new one.
87. Did Emily say that she had her hair cut?  
Yes, she said that she didn't like the braids anymore.
88. Did Melinda say that she enjoyed her holidays in Malta?  
No, she said that she could not go this year.
89. Did Pamela say that she gained some weight?  
No, she said that she lost some weight.
90. Did Rebecca say that she stopped smoking?  
Yes, she said that she finally did it.
91. Did Richard say that he was finished with his project?  
No, he said that he could not finish it on time.

#### B.4. Filler items

---

92. Did Scarlett say that she met her ex-boyfriend?  
Yes, she met him accidentally in the café.
93. Did Molly say that she got flowers from Robert?  
No, she said that she got flowers from Steve.
94. Did Charles buy the magazines for me?  
Yes, he put them on the little table in the living room.
95. Did Charles water the plants in the garden?  
No, he forgot to but I will do that later.
96. Did the Abrahams like the food?  
Yes, they all said it was delicious.
97. Did Mary say that her exams were too difficult?  
Yes, she complained that many students failed almost all of the exams.
98. Did Mike say that the police caught the murderer?  
No, but he said that they had a suspect.
99. Did Oliver say that they discovered a new planet?  
No, he said that he read that they found a new star.
100. Did Winston say that he is happy with the stock market these days?  
Yes, he said that it stabilized during the last years.