Alina Gregori
Matrikelnummer: 6689544
Studiengang: Master Linguistics

geb. am 18.02.2000 in Hanau

Co-speech Gestures, Information Structure and Prosody: A Corpus Study on Prominence Peak Alignment

Abschlussarbeit zur Erlangung des akademischen Grades
Master of Arts (M.A.)
in
Linguistics

an der Johann Wolfgang Goethe-Universität,
Frankfurt am Main
angefertigt am Institut für Linguistik
des Fachbereichs 10

vorgelegt von
Alina Gregori

Erstgutachter: Prof. Dr. Frank Kügler

Zweitgutachterin: M.A. Corinna Langer

Abgabe der Arbeit am: 29.08.2022

Table of Contents

List of Figures

List of Tables

Abstract

This thesis investigates how prosodic prominence and non-referential gestures align under the mediation of Information Structure (IS) in German. Pitch accents are ordered on a scale from low to high prominence dependent on their f0 height. The investigated gestural component in this study is the apex, which is the most prominent point of a gestural stroke. Two parameters of IS, information status and focus, are taken into account as possible mediators of the synchronisation between pitch accents and non-referential gestures. The study addresses the questions whether a) the occurrence and b) temporal alignment of prosody and non-referential gestures are sensitive to IS in German spontaneous speech.

The data are taken from a German speech corpus containing task oriented spontaneous speech. The factors „pitch accents", „apices", „information status" and „focus" are annotated and afterwards the occurrence and alignment of gestures in relation to pitch accents and IS categories are extracted for evaluation in two separate analyses.

The distribution analysis of the annotated corpus showed a tendency of non-referential gestures to not appear on speech material coded for IS, which might be a result from their non-referential nature. However on apices accompanying IS referents, prominence influences a) the frequency of gesture occurrence on the referent and b) the likelihood of alignment of the apex with each pitch accent type. As a side effect, the presence of apices increases the accuracy of alignment of pitch accents and information status along the pitch accent prominence scale. A temporal synchronisation analysis showed that apices and pitch accents tend to align with one another, preferring a small distance between them. In general this is in accordance with established phonology-gesture synchrony rules, since apices precede rather than follow their nearest pitch accent. IS facilitates this synchronisation, resulting less deviation between the two factors and and a more accurate consideration of the phonological synchrony rule. The effect of the Information Structural types is not sorted straightforwardly along prosodic prominence.

To conclude, the results suggest that gestures and pitch accents tend to align following different synchronisation principles. This synchronisation is influenced by IS, though not fully straightforward. These basic insights on the interaction of intonation, gestures and IS open the field for in depth exploration of the role of prominence in gestural research and the discourse function of non-referential gestures.

Zusammenfassung auf Deutsch

Diese Abschlussarbeit befasst sich mit der Synchronisation von Intonation und nicht-referierenden Handgesten in der deutschen Sprache. Tonhöhenakzente haben im Deutschen funktionale Effekte, die die Eindeutigkeit eines Diskurses unterstützen und Informationen hervorheben können. Die Akzente werden als Hoch- und Tiefpunkte der Grundfrequenz eines Sprechers identifiziert und können anhand ihrer Höhe mit unterschiedlicher Prominenz ihrer Konstituente assoziiert werden. Dabei ist ein Tiefpunkt der Frequenz ein Anzeichen für geringe Prominenz und steigende Frequenz kennzeichnet steigende Prominenz. Gesten werden unterteilt in referierende Gesten, welche eine semantische Verbindung zum gesprochenen Inhalt haben, und nicht-referierende Gesten, welche diese semantische Verbindung nicht haben, aber zur Entwicklung und Steuerung einer Unterhaltung beitragen. In dieser Arbeit werden nicht-referierende Gesten betrachtet, um die Interaktion von Gesten und Tonhöhenakzenten näher zu untersuchen, da diese eine enge Verbindung zu Intonation haben. Gesten lassen sich in kleinere Komponenten unterteilen, deren einzig obligatorische der „Stroke“ ist, welcher die eigentliche Geste darstellt (andere Bewegungen sind Vor- oder Nachbereitung des Strokes). Jeder Stroke hat einen „Apex“, den prominentesten Punkt eines Strokes. Dort stoppt die Bewegung und endet oder wechselt die Richtung. Der Apex ist die untersuchte Konstituente der nicht-referierenden Gesten in dieser Arbeit. Informationsstruktur (IS) ist eine pragmatische Komponente von lexikalischen Diskursreferenten, die deren Prominenz bestimmt. Der Informationsstatus eines Referenten gibt an, ob dieser bereits im Diskurs erwähnt wurde, also wie neu der Referent ist. In dieser Arbeit werden drei Level von Informationsstatus unterschieden: „given“, „accessible“ und „new“. Fokus kennzeichnet die Konstituente in einer Intonationsphrase, welche die größte Informativität besitzt, oder welches unter Ausschluss von Alternativen ausgewählt wird. In dieser Arbeit wird zwischen „new-information“ Fokus und „contrastive“ Fokus unterschieden, aber auch der Unterschied zwischen fokussierten und nicht fokussierten Konstituenten allgemein berücksichtigt.

Die Fragestellungen, die in dieser Arbeit adressiert werden, lauten: Beeinflusst Informationsstruktur das Vorkommen von nicht-referierenden Gesten? Ist prosodische Akzentuierung mit nicht-referierenden Gesten zeitlich synchronisiert, und wie beeinflusst Informationsstruktur die Synchronisation? Für die Untersuchung wurde ein Korpus deutscher Spontansprache verwendet, dessen Dialoge auf einer Wegbeschreibungsaufgabe für die Teilnehmenden

basieren. Das verwendete „SaGA“ Korpus besteht aus 280 Minuten Material und enthält Audio- sowie Videoaufnahmen der Unterhaltungen und Gestentyp-Annotationen. Für diese Arbeit wurden Apexe von nicht-referierenden Gesten, Informationsstatus, Fokus und Tonhöhenakzente der IS-Referenten zusätzlich annotiert. Anschließend wurden die Vorkommen und Positionen der Apexe in Relation zu den Tonhöhenakzenten und die Angaben zur IS extrahiert und in einer Distributions- sowie einer temporalen Analyse untersucht.

Insgesamt wurden deutlich weniger nicht-referierende Gesten gefunden als prosodische Akzente und diese Gesten kamen weniger als in der Hälfte der Fälle innerhalb eines IS-Referenten vor. Wenn Gesten während IS-Referenten produziert wurden, ließ sich ein Einfluss von linguistischer Prominenz auf a) die Häufigkeit des Vorkommens von Apexen sowie b) das gemeinsame Vorkommen von Apexen und Tonhöhenakzenten verschiedener Prominenz feststellen. Die Präsenz von Gesten verbesserte außerdem die Annäherung von Tonhöhenakzenten und Informationsstruktur gemessen an ihrer Prominenz. Die zeitliche Analyse ergab, dass Tonhöhenakzente und Apexe dazu tendieren, nah beieinander produziert zu werden. Dabei erschien der Apex im Durchschnitt kurz vor dem entsprechenden Tonhöhenakzent, was mit anderen Untersuchungen der wissenschaftlichen Literatur übereinstimmt. Informationsstruktur verstärkte nicht nur diese Tendenz, sondern verbesserte auch allgemein die Synchronisation von Apexen und Akzenten, indem sich die Abweichung der beiden Faktoren mit prominenteren Leveln der Informationsstruktur verringerte.

Das geringe Vorkommen der Gesten auf IS-Referenten könnte durch ihre Funktion im Diskurs und die fehlende semantische Komponente bedingt sein. Dennoch synchronisieren Tonhöhenakzente und Apexe miteinander. Insgesamt ist aufgefallen, dass die Prominenz nicht linear ansteigend als Faktor zu bewerten war, da zum einen viele weniger prominente Referenten akzentuiert wurden, was möglicherweise auf die Aufgabe der Unterhaltungen zurückzuführen ist. Zum anderen ordneten sich „accessible“ Referenten auf der Prominenzskala nicht zwischen „given“ und „new“ ein, sondern mit einer niedrigeren Prominenz.

Abschließend lässt sich feststellen, dass sowohl Tonhöhenakzente und Apexe von nicht-referierenden Gesten miteinander synchronisieren, als auch, dass Prominenz in Form von Informationsstruktur einen positiven Einfluss auf diese Relation hat, auch wenn Apexe IS-Referenten vermeiden. Diese Arbeit trägt dazu bei, die Rolle von Prominenz in der Gestenforschung sowie die Aufgaben nicht-referierender Gesten im Diskurs zu beleuchten.

1. Introduction

Visual communication, specifically co-speech gestures, and its interface with spoken language have developed to be a popular topic of linguistic research in the last decades. Important work on this are approaches like KENDON 1980, who claimed that gestures and speech are strongly connected and who discussed the contribution of gestures to an utterance. MCNEILL 1992 focused on the categorisation of co-speech gestures and their synchronisation with speech. MCNEILL 1992 further claimed that gestures and speech are two modalities of the same framework. Many studies concerning co-speech gestures have investigated their semantic and pragmatic functions, how and which gestures contribute to the interpretation of speech and the development of a conversation. Since co-speech gestures also express discourse functions, it has been found that prosody and gestures complement each other semantically and functionally (MCNEILL 1992, LOEHR 2012, IM & BAUMANN 2020 among others). This thesis focusses on the interface of phonology and gesture and aims to investigate parallels between prosodic and gestural prominence and the extent of this link for German spontaneous speech. In addition, a potential influence of Information Structure as a factor of prominence on the interface of the two modalities is investigated.

Gestures are divided into referential gestures and non-referential gestures (MCNEILL 1992, KENDON 2004). Non-referential gestures are characterised by fulfilling discourse functions like highlighting information or directing the discourse and by missing direct lexical/semantic connection to the utterance. Those gestures display an abstract rhythmic motion parallel to sentence intonation (MCNEILL 1992). The amplitude, peak and timing of the non-referential gestures are able to guide a conversation and are therefore interesting factors to investigate on their interface. The investigated gestural component is the apex of a gestural stroke, representing the most prominent point of a gesture (ROHRER et al. 2020). Referential gestures comprise iconic and deictic gestures, which have the purpose to support the content and interpretation of an utterance by connecting to a referent semantically, illustrating its characteristics (MCNEILL 1992, COCHET & VAUCLAIR 2014). Those gestures have a semantic component and do not primarily contribute to the structural information of a discourse.

This thesis contributes to the young field of gestural research by investigating specifically non-referential gestures instead of the more typically considered referential gestures. The focus on non-referential gestures makes sense because of their discourse directing functions which are more similar to the features of pitch accents than semantically connected gestures. For this reason, the correlation between pitch accents and apices, the most prominent parts of intonation and gestures respectively, is investigated. This makes an analysis of the temporal synchronisation between the two modalities interesting, especially under the consideration of IS categories. German spontaneous speech dialogues are the investigated speech material in this thesis, which not only allows for new insights on the gestural interface with the German language, but also with a new type of speech. Until now, many gestural studies have investigated a) engaging rehearsed speech and b) monologues, since these elicit the use of gestures. Spontaneous speech dialogues allow for a better investigation of the discourse function of gestures in addition to providing a natural speech setting.

Phonologically, structure and prominence in the German language are indicated by fundamental frequency contours and pitch accents (highest and lowest f0 points). Pitch accents are present on words that have a lexical interpretation and are heads of certain prosodic domains. The type of pitch accent can display the degree of prominence a prosodic constituent has (FÉRY & KÜGLER 2008, BAUMANN & RÖHR 2015). It is assumed that a higher f0 height goes along with greater prominence. Relevant measures to investigate phonological prominence are the pitch accent type and the temporal position of its peak.

Another relevant factor for this study is Information Structure (IS), which is expressed by linguistic tools like prosody. IS is important across linguistic domains, also interacting with semantics and syntax. One parameter of IS is the ‚information status‘ of a lexical referent. Such a constituent can be labeled either „new“, „accessible“ or „given“ depending on whether it has been mentioned in the previous discourse or not (KRIFKA 2008). A „new“ constituent did not appear in the discourse before. A „given“ constituent was explicitly mentioned within the last few sentences of the discourse. An „accessible“ constituent was not explicitly mentioned previously, but is retrievable by the context, events and world knowledge of the interlocutors. In terms of phonology, it is assumed that new referents are the most prominent ones, followed by accessible referents, while given constituents are less prominent. The ‚newness‘ is a central factor of information status.

The second relevant parameter of IS is ‚focus‘. KRIFKA 2008 defines ‚focus‘ as the presence of alternatives, meaning that a constituent that receives focus was chosen over other potential constituents for that context. The most common uses of focus result from an interlocutor indicating new information, confirmation, parallelism to the context or correction. ‚Contrast‘ is a subtype of focus (contrastive focus), which overtly displays alternatives and rules them out. ‚Focus‘ is relevant for this investigation, because it is also known to attract phonological prominence (PIERREHUMBERT & HIRSCHBERG 1990): Focused constituents often are the most prominent constituents, the background is usually less prominent. Therefore, focus might contribute or interfere to the analogue behaviour of phonological and gestural prominence.

The interaction of co-speech gestures, phonological prominence and IS is the central topic of investigation in this thesis. This raises two research questions that are addressed for the German language in an analysis of spontaneous speech and gestural data:

1. Does Information Structure in its parameters information status and focus influence the occurrence of non-referential co-speech gestures in spontaneous German speech?
2. Is pitch accentuation temporally aligned with non-referential gesture apices in German and does Information Structure influence this alignment?

With regard to the first research question, the hypothesis is that non-referential gestures do occur more often during the articulation of focused or „new“ constituents in German, than with less phonologically prominent constituents (in line with IM & BAUMANN 2020). The hypothesis for the second research question is that alignment of non-referential gestures and phonological prominence marking can be observed (cf. LOEHR 2012), and that this alignment can be interfered by the IS characteristics, with higher prominence increasing the precision.

In order to test the hypotheses, a corpus study is conducted. The corpus data is then analysed for the occurrence of non-referential gestures on IS categories, in a distribution analysis and for the temporal alignment of pitch accents and apices in a temporal synchronisation analysis.

The thesis is organised as follows: In chapter 2, the theoretical background for each of the relevant factors is given. In section 2.1 co-speech gestures are discussed, in section 2.2 phonological prominence marking is introduced and section 2.3 concerns Information Structure. Section 2.4 provides an overview over relevant prosody-gesture link studies. In chapter 3, the corpus and its data are presented and all annotation conventions are introduced. The statistical procedure is explained as well. In chapter 4, the results of the distribution analysis are presented and in chapter 5, the temporal synchronisation analysis is described. Chapter 6 discusses and interprets the results of both analyses and gives a perspective for further research. Chapter 7 concludes this thesis.

2. Background

This chapter provides the relevant theoretical background and empirical studies for the following corpus study. In the first section 2.1, co-speech gestures are introduced, starting with the different types of gestures, followed by general composition of gestures and the approach of McNeill 1992 to speech-gesture synchrony. In the second section 2.2, basic concepts of intonation and prosodic prominence are presented. Afterwards in section 2.3, the different parameters of IS and their correlation prosodic prominence are introduced. Finally in section 2.4, known interactions of the three analysed factors are presented, summarising two relevant studies on the gesture-prosody and IS interface.

2.1 Co-speech Gestures

Co-speech hand gestures, which are accompanying spoken language can be used to express structural and semantic information in addition to or as a complement of speech. McNeill 1992 explains that gestures and speech combined are needed to receive the complete intention of the speaker: „If we were to look only at the gesture or the speech, we would have an in-complete picture of the speaker's memory and mental representation of the scene.“ (McNeill 1992, p. 13). While the whole body (used for imitation, plus the face and eyebrows) can be used as a tool to provide information going beyond what is expressed by speech, in this thesis, exclusively hand gestures are investigated. Most frequently, gestures are produced by one dominant hand in a conversation (cf. Lücking et al. 2010). Less frequently, either both hands form a gesture together, or the weak hand alone is used to produce a gesture. In the following paragraphs, the characteristics of hand gestures are introduced.

2.1.1 Types of Gestures

Gestures are known to play an important role in providing additional information in discourse. As Kendon 2004 (among others) states, interlocutors may use gestures to: „refer to something by pointing at it, […] show what something looks like, to indicate its size or its shape, to suggest a form, object or process by which an abstract idea is illustrated, or they may show, through visible bodily actions, that they are asking a question, making a plea, proposing an hypothesis, doubting the word of another, denying something or indicating agreement about it, and many other things“ (Kendon 2004, p. 1).

This statement by KENDON demonstrates the gesture distinction that is most relevant for this thesis. In the first part of the quote, he describes the information and usage of **referential gestures**. Referential gestures have a direct semantic connection to the content that is expressed in an utterance by supporting and highlighting the information that is spoken. This usually concerns lexical items and describes their characteristics more concretely, by mimicking their shape, size or position for example. This group of gestures includes deictic gestures, which are mainly used in a pointing fashion to „direct a recipient's attention towards a specific referent in the proximal or distal environment" (COCHET & VAUCLAIR 2014, p. 3), but also includes pointing into an empty space contributing to an abstract narrative. Figure 1 displays a typical deictic gesture.

Figure 1: *Example of a deictic gesture, screenshots from the investigated corpus. The speaker points to the left while saying „links" (left).*

Another group of referential gestures are iconic gestures, which display obvious occurrences of iconicity in communication. They are very diverse in their appearance as they adapt to the characteristics of the item they describe. In other words, MCNEILL 1992 defines them as „bear a close formal relationship to the semantic content of speech" (MCNEILL 1992, p.12). MCNEILL defines a subcategory of iconic gestures, so-called metaphoric gestures, which are also closely related to the element they accompany, but they rather describe it abstractly than describing a concrete object. Figure 2 displays an iconic gesture.

Figure 2: *Example of an iconic gesture, screenshots from the investigated corpus. The speaker paints a circle with two fingers of each hand while talking about a round location, saying „kreisförmig".*

In the second part of the quote from KENDON 2004, he displays the functions of **non-referential gestures**. These gestures are not related to the interpretation of produced words (MCNEILL & LEVY 1982), rather they support the sentence structure and provide structural information. Therefore, non-referential gestures are visually different to referential gestures in that they are more abstract and not by themselves assignable to a certain object or event. They accompany speech and putatively have a similar function to sentence prosody, which will be explored in this thesis. Non-referential gestures are abstract hand movements often forming short lines or curves not related to the content of the utterance but displaying discourse-relevant features (VILÀ-GIMÉNEZ & PRIETO 2021, SHATTUCK-HUFNAGEL et al. 2016, LOEHR 2012). Non-referential gestures are often referred to as beat gestures (e.g. IM & BAUMANN 2020), having a rhythmic component contributing to the sentence structure often characterised as „the hand moves along with the rhythmic pulsation of speech“ (MCNEILL 1992, p. 15). He explains that while they are the least prominent, abstract and rather short gestures, they make an important complement to sentence prosody. In Figure 3, a non-referential gesture is illustrated.

Figure 3: *Example of a non-referential gesture, screenshots from the investigated corpus. The speaker has both hands in a base position and includes only a short flick up of all fingers along the speech rhythm, while no connection to a referent can be found. The speaker says „parkähnliche“ (being/looking similar to a park) with an accent on „park“, where also the flick of the hands is produced.*

All gestures primarily having prosodic functions and no semantic component are classified as non-referential gestures and are under investigation in this thesis. The corpus that is used for this analysis, distinguishes between „discourse“ and „beat“ gestures in the non-referential category. Their distinction is explained in the introduction of the corpus in section 3.1. Not being relevant for this thesis, the distinction is not considered for this analysis. A group of gestures that are not referential and do not really concern specific sentence structure but rather control the discourse direction and condition are cohesive gestures. While not referring to a certain object or event, they also show iconic aspects with regard to the direction of the conversation (cf. MCNEILL 1992) and are not relevant to this thesis. Figure 4 gives an overview over the categorisation of the introduced gesture types.

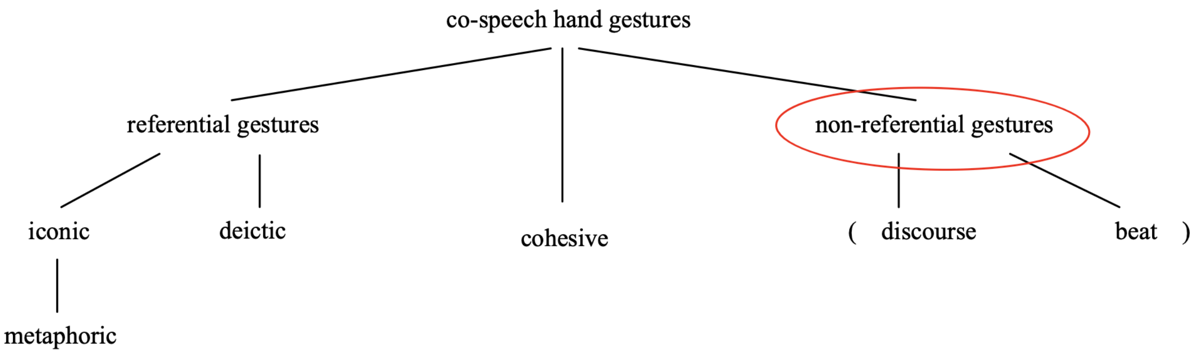


Figure 4: *Categorisation of the gesture types according to MCNEILL 1992 and LÜCKING et al. 2010. The red circle highlights the category that is investigated in this thesis.*

In this thesis, only non-referential gestures will be investigated since their relation to prosody is of interest and the semantic component does not play any role. The gestures' structure will be introduced in the following section 2.1.2.

2.1.2 Components of a Gesture

In his first famous work on gestures, KENDON 1980 suggested a hierarchical structure of gestures that consists of four constituents which are built up by the smaller constituents and are able to express a different amount of information. The apex, the item of interest in this thesis is not part of this hierarchy, but is a point within the smallest constituent of KENDON's hierarchy and is introduced directly after. KENDON's smallest constituent is the gestural **stroke.** The stroke is described as „contains the "action" of the gesture, and as such, it is the only obligatory phase of a gesture. It also often contains the kinematic peak of movement velocity, moving much faster than phases that occur just before or after a stroke." (ROHRER et al. 2020, p. 20) meaning that the stroke displays the core movement of the gesture. The next constituent is the **gestural phase** (**g-phase**) and refers to every movement or position that is not identified as the stroke. This includes the alternation of the hand between a resting position and the beginning of the stroke. Any movement before the stroke is called ‚preparation' and movement after the stroke is called ‚recovery'. Another gestural phase is the ‚hold' (before or after the stroke), where no movement is involved, but the hand remains in its position. A combination of a stroke and one or multiple g-phases is called a **gestural phrase** (**g-phrase**). These describe more completed movements from a default position to another default position and form one gesture. The last and biggest constituent of the gesture hierarchy by KENDON is the **gestural unit** (**g-unit**). G-units describe sequences of (multiple) g-phrases which start and end with the hands in a clear resting position. Over all, g-units can in principle solely consist of one stroke, but usually they contain multiple strokes which are connected by g-phases and organised into g-phrases.

While these domain can at least partly coincide with prosodic constituents (LOEHR 2012), the examined element here is the **apex** of the gesture, or more precisely of the gestural stroke. The apex is the peak of the stroke and is rather a temporal point than a movement or interval and thus suitable for a comparison to pitch accents, and for the investigation in this thesis. It is also often a turning point in the gestural movement and describes the most „extended" point of the gesture, commonly being identifiable by showing low velocity (ROHRER et al. 2020). When the term „gestures" is mentioned in this thesis, it usually refers to gesture phrases, unless defined otherwise. Figure 5 illustrates the composition of a gesture as described above.

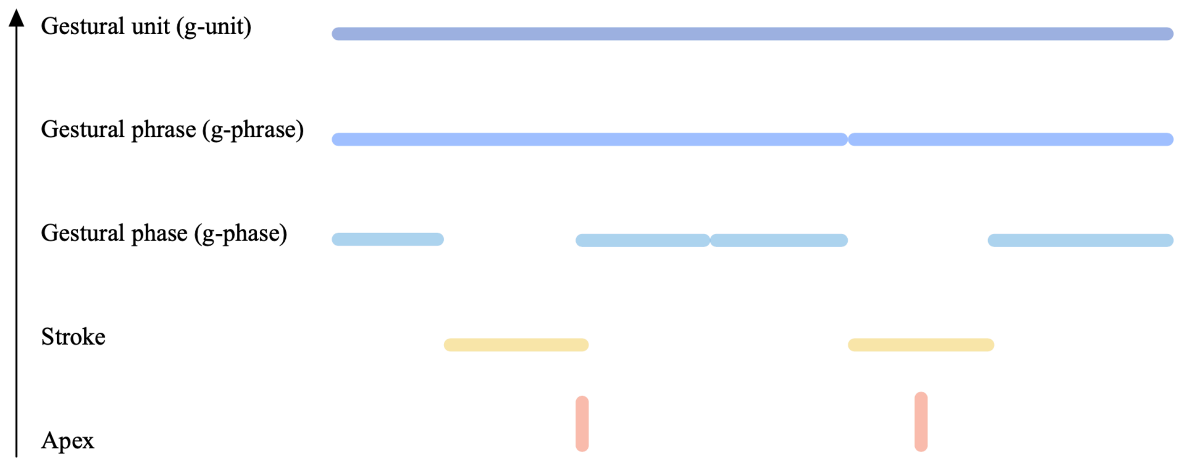


Figure 5: *Hierarchical illustration of the components of a gesture from smallest (bottom) to biggest (top) with an exemplary structure, apart from the apex in accordance with KENDON 1980. All blue components are not obligatory, the horizontal lines indicate time intervals, the vertical lines describe points in time.*

2.1.3 McNeill's Synchrony Rules

Apart from introducing the basics of gestures and known principles, MCNEILL 1992 makes statements about the synchrony of co-speech gestures and speech by assuming that gestures and speech are two aspects of the same system. McNeill first names crucial differences between the two modalities „speech“ and „gesture“:

- Global versus synthetic structure: language is synthetic, gesture is global
- (Non)combinatoriness: language is combinatorial, gestures are not
- the standard of form: languages have many rules and conventions, gestures are only often similar
- and the duality of patterning: one underlying characteristic of language according to Hockett & Hockett 1960, gestures only have one aspect, thus missing arbitrariness

Nevertheless, the parallels, complementariness and synchrony make speech and gesture two modalities in the same framework. McNeill 1992 provides five reasons confirming their relation:

1. Gestures are only used together with speech, rarely alone, which can be observed when not only considering the speaker in the discourse but also the listener. Communicative hand movements occurring without speech are normally signs or pantomime. However, gestures can sometimes occur in a speech break when the utterance is not completely finished.
2. As already mentioned, gestures and speech complement each other and express similar semantic narratives and pragmatic functions (and phonological functions as well, as the remainder of this thesis will explore). A gesture does not express the opposite to speech material.
3. McNeill mentions the temporal synchrony of gesture and speech, which has afterwards been explored a lot in the literature (e.g. Loehr 2012, Leonard & Cummins 2009, 2011) and is further investigated in this thesis. This alignment is a strong indication for one single underlying system for both modalities.
4. Another observation in favour of one common system is that gesture and speech are acquired together by children. This not only concerns a similar time of development but also the ability to learn simple parts first and later on build up on more complex structures which is observed in both modalities (Goodluck 1991, Hoff 2013).
5. The last reason McNeill brings up is connected to language impairment. Aphasia, which results from damage of the speech centres in the brain, affects the ability to speak but also the ability to gesture (Benson 1985). When the Broca areal is damaged, patients are able

to express semantically what they want to say but are troubled with grammaticality and fluency of speech (Prather et al. 1992). Similarly, these patients use referential gestures but are rarely able to use non-referential gestures. The other way around, a damaged Wernicke areal results in intact fluent speech but difficulties to express meaning that makes sense (Blumstein et al. 1982). In connection to this, patients use non-referential gestures but are troubled to use iconic and deictic gestures. Because of these five commonalities, McNeill 1992 assumes gesture and speech to be two modalities of the same underlying system and proceeds to develop rules that substantiate the temporal and interpretational synchrony between the two.

McNeill proposes three synchrony rules regarding the interface of co-speech gestures and speech in different linguistic domains. The three **Synchrony Rules** deal with phonology, semantics and pragmatics. He regards the stroke of the gesture as the part that synchronises with language in all aspects. With the phonological synchrony rule, McNeill 1992 refers to a realisation of Kendon 1980, formulating the following rule:

(1) „the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech“ (McNeill 1992, p. 26)

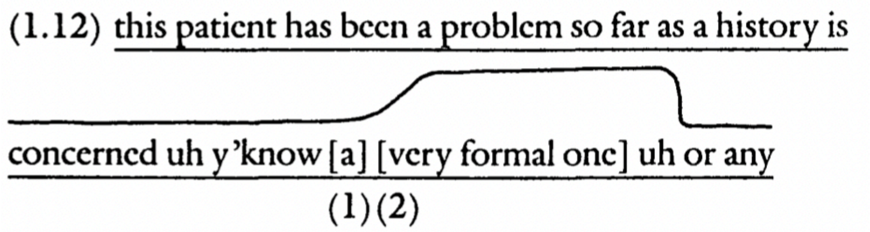


Figure 6: *Illustration of the phonological synchrony rule, taken from McNeill 1992, p. 27. (1) indicates the position of the stroke (umbrella hand moves sharply down); (2) indicates a post-stroke hold with the umbrella hand held statically. The line describes the intonation contour briefly.*

He therefore states that the gestural stroke and nuclear pitch accent synchronise. An example is shown in Figure 6. Loehr 2012 afterwards found out that bigger gestural and prosodic constituents also tend to synchronise to a smaller extend.

This constitutes to the semantic synchrony rule:

(2) „if gestures and speech co-occur they must cover the same idea unit“ (McNeill 1992, p. 27)

An ‚idea unit‘ describes the complementation of the meanings of the gesture and speech. The semantic synchrony is only established if speech and gesture not only synchronise temporally, but also express the same meaning. Having the same meaning or a complementary/specifying meaning complies to the semantic synchrony rule. Completely unrelated gestures and speech are not very common in spontaneous speech, even though it would be possible to produce them. McNeill also mentions potential difficulties for the semantic synchrony rule, namely pauses, multiple gestures and gestures exceeding one clause, but he accounts for their fulfilment of the semantic synchrony rule and explains how these rather contribute to the complementation of the modalities.

The pragmatic synchrony rule does not have any exceptions, according to McNeill:

(3) „if gestures and speech co-occur they perform the same pragmatic functions“ (McNeill 1992, p. 29)

McNeill explains that this leads to speakers only being able to employ one pragmatic function at once and he mentions that this pragmatic function is independent of a semantic one, making it possible to have additional semantic meaning expressed.

McNeill 1992 uses these three rules to define the synchrony between the two modalities gesture and speech and indicates their common origin from the same framework by this. The synchrony rules are a plausible model for the linguistic and gestural alignment and complementation in communication. In addition, in the frame of the semantic synchrony rule, the author states that the synchronisation of the two modalities does not imply a one-to-one mapping or moreover a restriction in size, but that one gesture is able to persist throughout multiple clauses as well as multiple gestures can be implemented into one clause. Still, the utterance sticks to the synchrony rules.

2.2 Intonation

As explained in section 2.1, the use of co-speech gestures goes hand-in-hand with spoken language items and can be seen as an indicator for prominence across linguistic domains. Similarly, intonation, through pitch accents in German, is an indicator for speech prominence (Ladd 2008). Being the crucial domain for prosodic prominence, „Intonation [...] refers to the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way.“ (Ladd 2008, p. 4). The Autosegmental-Metrical Theory of intonation is adapted here which consists of two parts - the autosegmental part (according to Goldsmith 1990) and the metrical part (proposed by Liberman & Prince 1977). The underlying assumption to autosegmental phonology is the existence of „two or more parallel tiers of phonological segments“ (GOLDSMITH 1990, p. 9). This proposes different phonological layers, each containing specific types of segments (e.g. tones or sounds) which are associated with each other. Segments of different tiers do not have to be associated in a one-to-one fashion or associated at all. Further, GOLDSMITH puts up the convention that connections of different autosegmental tiers are not allowed to cross each other, meaning that association-directionality is required. The previous insights to autosegmental phonology indicate that the tiers (and especially tones and sounds) are independent of each other (cf. GENZEL & KÜGLER 2011). The metrical part of autosegmental-metrical theory fundamentally concerns word-level phonology on the basis of the **Prosodic Hierarchy**, which was introduced by LIBERMAN & PRINCE 1977. The constituents of the Prosodic Hierarchy from smallest to largest are (according to FÉRY 2017): Mora < Syllable < Foot < prosodic Word < phonological Phrase < intonation Phrase < Utterance. Each domain contains at least one constituent of the next smaller domain (SELKIRK 1981). Metricality concerns prosodic domains up to the size of the prosodic word, stating that every foot contains one metrically strong (stressed) syllable. It determines the way accents are distributed rhythmically with the influence of syllable weight (cf. UHMANN 1991). In the end, this principle is applicable to and responsible for the nuclear pitch accent of a sentence. The following section 2.2.1 introduces the most important tonal feature for prominence in German: pitch accents.

2.2.1 Pitch Accents

Languages of the world use phonological tools to fulfil different purposes. In addition to other phonologically distinct units like articulatory sounds and bigger prosodic constituents forming the prosodic hierarchy, tone and stress are used in speech to express structural, but also lexical information. Primarily, tones vary in their height and contour. German as an intonation language (GUSSENHOVEN 2004, JUN 2005, FÉRY 2017 among others) does not use tones as a lexical tool, but only post-lexically. This means that tones do not influence the lexical interpretation of an item, in principle they are semantically interchangeable. Therefore, the same lexical item can receive different pitch accents in different sentences depending on the structural context. Tones convey meaning on the sentence-level, thus on higher prosodic constituents, commonly intonation phrases. These tones have the purpose of highlighting (prominent) information (LADD 1996). In languages like Mandarin Chinese (a tone language, see FÉRY 2017), the tone chosen with a produced item decides the meaning that is conveyed. In this thesis, the term „pitch“ is used to describe the perceptual, phonetic characteristics of produced sounds, and „tone“ is the corresponding phonological term suited to describe patterns, processes and contours.

In German, tones are mainly used as so called pitch accents, which are defined by LADD 1996 as „a local feature of a pitch contour – usually but not invariably a pitch change, and often involving a local maximum or minimum – which signals that the syllable with which it is associated is prominent in the utterance.“ (LADD 1996, p. 45f). LOEHR 2012 defines pitch accents as „single or clustered tones associated with a stressed syllable“ (LOEHR 2012, p. 72), meaning that they are distributed over stressed syllables (which is „an abstract lexical property of individual syllables“ LADD 2008 p. 49, after BOLINGER 1958) and mark the most prominent constituent of each level of the prosodic hierarchy. This results in one **nuclear** pitch accent per intonation phrase. A pitch accent preceding the nuclear accent is called „prenuclear“ and is phonetically weaker than the nuclear one. The „postnuclear“ part of a sentence contains all items following the nuclear accent, and even if this part contains stressed syllables, they don’t receive pitch accents by themselves. They follow the contour of the nuclear accent or the general intonation pattern of the sentence. Nuclear accents in German tend to appear near the right edge of an intonation phrase (following the Nuclear Stress Rule (NSR); CHOMSKY & HALLE 1968, LIBERMAN & PRINCE 1977). However, not only the nuclear pitch accent conveys to prominence, but other pitch accents in the same sentence as well.

Pitch accents themselves can occur in different types, depending on the height of their acoustic **fundamental frequency f0** (measured in Hz). While f0 is the most important acoustic feature to determine the type of pitch accent, acoustic measurements like intensity and duration of the accented sound adapt to pitch accents in their level of prominence. In general, pitch accents can be divided into high (H) and low (L) f0 accents (PIERREHUMBERT 1980). Naturally, high accents have a higher f0 than low accents, but this distinction cannot be delimited by absolute values, since f0 is speaker dependent, thus the differentiation of the two tones is relative to each other and to the intonation contour. Pitch accents are marked by an asterisk (T*, T is a variable for either high or low tones) behind the according H or L tone in annotation and notation. A higher pitch accent goes along with higher prominence (BAUMANN

& Röhr 2015), as will be elaborated in following sections. However, high and low tones not only serve the purpose of marking prominent items, they also contribute to the constituent structure and sentence structure of an utterance. Integrating further structural tones, the sentence melody/rhythm is built by pitch accents and boundary tones and its course indicates structural sentence information, which is not lexical (Pierrehumbert 1980, Grice et al. 2000). Generally, German declarative clauses have a decreasing intonation contour, meaning that the pitch accents have a lower f0 the further the sentence has already progressed (Pierrehumbert 1980). However, when a speaker wants to show that they have not finished talking, an intonation phrase tends to end with a higher tone, as do certain types of questions. Pitch accents and intonation contours also express sentence-level information like focus (e.g. Kügler & Calhoun 2020). Crucially, the final direction of the intonation contour is not connected to any pitch accent but indicates structural information.

In addition to sentence contours, pitch accents can be more complex and thus form a contour on a smaller constituent (Pierrehumbert 1980). These contours consist of a combination of high and low accents. A pitch contour changes height, creating rising (LH), falling (HL), fall-rise (HLH) or rise-fall (LHL) contours. Figure 7 shows examples for the relevant pitch accents introduced in this paragraph.

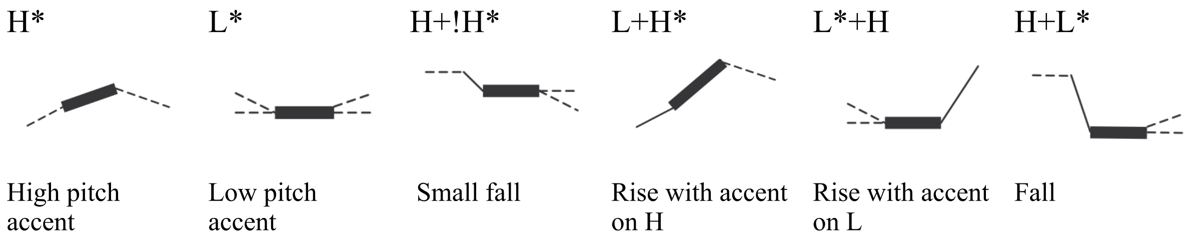


Figure 7: *Pitch accent contours and descriptions, taken from Grice & Baumann 2002, the black squares signal the (height of the) pitch accent itself, the solid lines stand for the rise or fall belonging to the pitch accent and the dotted lines indicate, how the f0 can precede or follow the accent, often displaying a higher and a lower option.*

One more important pitch accent that needs to be introduced is the downstepped high pitch accent !H* (cf. Pierrehumbert 1980, Grabe 1998, Connell 2001, Truckenbrodt 2002, 2004, Genzel & Kügler 2011). This accent is classified as a high pitch accent regarding f0, but the pitch is remarkably lower than the previous H* of the speaker in the same intonation phrase. Often, this is a result from the sentence prosody as a whole, since over the course of an utterance, the pitch usually declines. This means that the first pitch accent of an utterance receives the highest F0 value, the second one is lower. If there is another pitch accent, there is a high probability that this one is another considerable step lower than the second one. It is possible to rise the f0 height to the original height again, this often goes along with increased prominence or an intonation phrase boundary (LADD 1986, FÉRY & TRUCKENBRODT 2005, FÉRY 2010; it is no downstepped accent then). Downsteps are not obligatory to be used in a sentence at all.

Boundary tones have been mentioned as being additional tones contributing to the formation of prosodic sentence information. Opposed to pitch accents, boundary tones are not decided by syllable stress and prominence and not influenced by focus or other intonation manipulating factors. Boundary tones are characterised by high or low pitch as well (H or L) and marked by a „%“ in annotation (GToBI, GRICE et al. 2005). They occur at the edge of bigger prosodic constituents, mainly intonation phrases and also phonological phrases and their purpose is to mark those edges. This often fundamentally forms the sentence prosody and thereby gives insight into the sentence type (LADD 1996, 2008). Boundary tones are predestined to contribute to complex pitch contours or to extend existing ones. However, these tones are not directly connected to stress and prominence.

2.2.2 Prosodic Prominence

The term „linguistic prominence“ was defined by KEMBER et al. 2021 as „the expression of informal weight within utterances“ (KEMBER et al, 2021, p. 413). Intuitively, prominence applies to items that are very present, eye/attention catching for any reason in a conversation. Many different factors can lead to an item being prominent, newness, informativeness or markedness being some of them. Thus, semantic and syntactic factors, but also phonological ones can make a constituent prosodically prominent.

The underlying statement regarding prosodic prominence used in this thesis is that different pitch accents are associated with different levels of prominence (BAUMANN & RÖHR 2015). While this is a probabilistic correlation (not forced in production), a strong tendency in favour of this suggestion has been observed (KÜGLER & CALHOUN 2020). Pitch accents (f0 height), duration and intensity of a sound lead to its perceived prominence. This indicates a higher prominence of H* accents compared to L* accents. The complex pitch accents are ordered on a prominence scale around those accents. The least prominent items are those that do not receive an accent, the L* accents are the next prominent ones. After that, pitch accents arrange themselves in a rising pitch fashion with increasing prominence, the rising accent L+H* displaying the most prominent items. The resulting pitch accent prominence scale is illustrated in Figure 8 (according to Baumann et al. 2006 and Im & Baumann 2020):

Unaccented L* !H* H*+L H* L+H* →

Figure 8: *Prominence Hierarchy of pitch accents. From lowest perceived prominence (left) to highest perceived prominence (right).*

Two further assumptions have to be mentioned. First, structures that show a smaller change in their contour, like a low rise, high rise and a high fall, can easily be integrated into this scale according to the height of their pitch. A low rise is placed between the L* accent and the downstepped H, a high rise higher than H*. However, those contours are not listed in Figure 8, since they are rarely decisive for prominence, and are often not distinguished from the remaining accents in perception (Cole & Steffman 2022), as well as not associated with specific IS properties. H*+L, the early peak accent however, is integrated into the scale since this accent was investigated in prominence studies (cf. Baumann & Grice 2006) with regard to IS. There it was classified as being less prominent than a H* accent, putatively because of the fall at the end.

2.3 Information Structure

In this section, the basic terms and concepts of Information Structure (IS) that are of relevance for this thesis will be introduced. IS specifies interlocutors' information about referents in the discourse, concerning lexical referring projections (Chafe 1976, Rooth 1985, 1992, Schwarzschild 1999, Féry et al. 2007, Krifka 2008). One central concept needed to be introduced in order to explain IS is the so-called Common Ground (CG, Karttunen & Peters 1975, Stalnaker 1978, Clark 1992, 1996) of two interlocutors, which is part of every discourse. The CG is a cognitive conversational tool that contains all the information shared between the interlocutors. The most important aspects contained in the CG are the informations added throughout the conversation. However, world knowledge is a part of the CG as well, being information that is not received during the conversation, but having been acquired by the interlocutors before. The term „common ground" is defined by Clark as: „[…] a type of shared information. […] the sum of their mutual knowledge, their mutual beliefs, and mutual suppositions." (Clark 1992, p. 3). Krifka 2008 extends the idea of the CG by assuming that not only lexical information is stored in the CG (this is what Krifka calls CG-content, being identified as „truth-conditional information"), but in addition information concerning the direction or interests in a conversation (CG-management). This distinction is beyond the scope of this thesis. Therefore, both information from the CG-content as well as the CG-management are regarded as being part of a general CG.

Information status on the one hand and **focus** on the other hand are both relevant parameters of IS, and they are able to mediate pitch accents depending on their status or focus type. It is important to realise that IS is a concept present across linguistic domains. The execution, notions and influence can differ partly, especially between syntax and phonology. In this thesis the phonological interface of IS with gestures is investigated.

2.3.1 Information Status

The information status of a referent is determined by its close context. The assigned label of a referent depends on whether it was already mentioned in the discourse, or not. In this thesis, three levels of information status are distinguished (following KRIFKA 2008): An item can either be **new**, **accessible** or **given**. Newness increases from „given“ over „accessible“ to „new“, while givenness is highest on „given“ and lowest on „new“ constituents. The information status value of an item in the CG can change between its occurrences, as a result of the context, of time passed since its last occurrence and of lexical or referential characteristics. The information status of an explicit lexical referent can be determined based on the parameters as follows (according to KRIFKA 2008):

New: A new referent is defined as not having been mentioned in the previous discourse, thus being introduced by its appearance and being new to the hearer. A new referent is not part of the conversations’ CG before its occurrence, but is added to the CG by its appearance. New referents create the base for information status. When they are mentioned another time, they will receive the labels „accessible“ or „given“.

Accessible: This label is assigned to referents that have not been explicitly mentioned in the close preceding discourse, but are retrievable through different possibilities. An item is accessible when it is active in the CG for a certain reason. Referents introduced to the conversation more than a few sentences ago, or referents that concern the overall topic of the conversation are examples for typical accessible referents. If interlocutors are talking about a *bus drive* for example, the *bus* and the *road* become accessible referents throughout the complete conversation, as does the *bus drive* itself. Another factor making a referent accessible in the CG context is world knowledge, e.g.: Proper names known to all interlocutors are labelled as „accessible“. In addition to the CG, semantic relations can lead to the label „accessible“ for a referent. Talking about a specific item or scenario activates further entries in a persons mental lexicon, which are semantically related, working comparable to priming (MEYER & SCHVANEFELDT 1971). If a conversations' topic is a *kitchen* for instance, tools like *fork*, *spoon* or *plate*, and fitting electric devices like a *fridge* or an *oven* become „accessible“ referents, as do further referents that are associated with the word *kitchen*. This illustrates the role of context for information status and for conversation in general.

Additionally, pronouns are assumed to be „accessible“ referents in this thesis (following GÖTZE et al. 2007; other approaches label them as „given“, e.g. BAUMANN p.c.). Pronouns replace their substitutes which predefines the need for a pronouns substitute to be known to the interlocutors or at least being made accessible by the speaker. Personal pronouns in the first and second person are presumed to be accessible since they refer to the participants of the conversation, thus automatically being clear, identifiable and part of the CG.

Given: Given referents are those that have explicitly been mentioned in the previous discourse. While this sounds like a simple rule, several questions arise when considering real spontaneous discourse data. First, the time that has passed is an important factor to consider when deciding on the givenness of a referent. When a referent was mentioned too long ago in the discourse, it requires too much activation cost to receive the „given“ label. After how much time a referent looses its „given“ status cannot be clearly defined. A length of five clausal constituents was suggested in linguistic literature: Most guidelines assume that a referent is not „given“ anymore, when it lies more than five clausal constituents back in the discourse. However, which constituents exactly are chosen as a reference for givenness differs between existing models. While often a „sentence“ is used as the determining constituent (e.g. GÖTZE et al. 2007), others, like RIESTER & BAUMANN 2017, choose the intonation phrase as a fitting constituent. Another factor related to the activation cost question is the decision whether the clauses of both (all) interlocutors together contribute to the five sentence boundary or not. This can influence the status of the referent. For this thesis, the intonation phrases of both interlocutors are taken as the clausal units determining the information status of the referent. When a referent looses its status as „given“, it usually becomes „accessible“ for its next appearance.

As already mentioned, a constituent that is assigned its information status is called *referent*. However, not only the reference, but also lexical appearance can make an argument „given". BAUMANN & RIESTER 2013 and RIESTER & BAUMANN 2017 introduce a distinction between lexical and referential givenness. Referentially given items have been introduced to the CG in the close discourse. The terms used to describe this referent may differ between its mentions, but still lead to the label „given" (e.g. *Notfalltreppen* ‚emergency stairs' and *Treppen* ‚stairs' afterwards). On the lexical level, a constituent is given, when the exact word has explicitly been mentioned in the previous discourse. In this case it is possible that a constituent receives the status „given" because it is lexically given, even though the referent may change between the two occurrences (*Hund* and *Hund* ‚dog' when two different dogs are meant). While this distinction is important and can be extended to the whole annotation of information status, providing a more in-depth differentiation, it is not relevant for the research question of this thesis.. Therefore, both referentially and lexically given items will be viewed as „given" in this thesis.

2.3.2 Focus

The second IS parameter that is investigated with regard to its influence on prosody-gesture alignment is focus. Focus often stands in a direct relation with prosodic prominence, as pitch accents are one linguistic way of marking focus (FÉRY & KÜGLER 2008), apart from syntactic movement. In the literature, focus is often specified as serving the purpose of highlighting a constituent and in addition, informativeness and newness are mentioned as factors to attract focus (HALLIDAY 1967, JACKENDOFF 1972, SGALL et al. 1986). However, KRIFKA 2008 claims that these factors are not sufficient to fully determine the concept of focus. He proposes the following definition for focus: „Focus indicates the presence of alternatives that are relevant for the interpretation of linguistic expressions." (KRIFKA 2008, p. 247). In other words, focus expresses that the focused constituent was consciously chosen over other plausible referents. This attracts prominence, which becomes even more obvious when the alternatives are overtly present in the context as well, creating contrastive focus.

The function of focus serves a broader purpose. Highlighting, informativeness and newness do not have to be neglected as attractors of focus. More extensively, it is a tool for confirmation, correction, parallels, delimitation and the presence of implicit or explicit questions.

The concept of focus is best illustrated in language examples by looking at question-answer pairs, because different types of focus become visible then. The question provides the most direct context for a target utterance in which the focus is examined. It is assumed that information that is not given in the question (and therefore inquired by the question, often ‚replaced' by a wh-particle), is in focus in the answer and thus commonly receives the nuclear pitch accent of the sentence. Therefore, the amount and kind of information given in the question determine the size (and boundaries) of the focus constituent. The following example (4) shows a typical illustration of focus in German. ‚A' displays the question providing the context, ‚B' is the answer and contains the focus constituent marked by brackets and an index F. The primarily stressed syllable is marked by capital letters.

(4) A: Was hat Peter gebacken? ‚*What did Peter bake?*'

B: Peter hat [einen KUchen]$_F$ gebacken. ‚*Peter has baked [a cake]$_F$.*'

In (4), *einen Kuchen* is in focus since it is the constituent that is new in B opposed to A. *Ku*, the stressed syllable of the focus receives the pitch (sentence) accent. In this case, however, one cannot perceptionally distinguish between sentence B standing alone with *Kuchen* in focus and a sentence being uttered without a specific constituent in focus (for example all-new sentences, see KRIFKA 2008). There is a distinction between **narrow focus** and **broad focus**. Narrow focus is shown in example (4). The size of the constituent in focus can increase, when less information is given in the question or context. A question like in example (5) results in a broad focus constituent since all information in B is new. Sentences without a specific goal of focusing a constituent and sentences not preceded by a focus-initiating question or context are typically sentences with a broad focus.

(5) A: Was ist passiert? ‚*What happened?*'

B: [Peter hat einen KUchen gebacken]$_F$. ‚*[Peter has baked a cake]$_F$.*'

Importantly, here the nuclear accent is also placed on *KU*, based on established phonological rules preferring nuclear accents on the right of a phrase (Nuclear Stress Rule; CHOMSKY & HALLE 1968, LIBERMAN & PRINCE 1977). Thus, while the focus constituents differ in examples (4) and (5), their difference is not perceivable but only visible formally. If another constituent is in narrow focus (ergo not given in the question), a probabilistic shift of the nuclear pitch accent can be observed. This works with all lexical constituents of the sentence as examples (6) and (7) show in addition to example (4):

(6) A: Wer hat einen Kuchen gebacken? ‚*Who has baked a cake?*‘

B: [PEter]$_F$ hat einen Kuchen gebacken. ‚*[Peter]$_F$ has baked a cake.*‘

(7) A: Was hat Peter mit dem Kuchen gemacht? ‚What did Peter do with the cake?‘

B: Peter hat den Kuchen [geBACken]$_F$. ‚*Peter has [baked]$_F$ the cake.*‘

In example (6), the subject is focused and therefore, the nuclear accent is already produced on the first syllable of the sentence. In example (7), the focus is on the verb which is in sentence-final position and the sentence accent is shifted to the right. Both of the examples display narrow focus on those constituents that are inquired by the question. There are also intermediate steps in the size of focus constituents, as shown in example (8), and it is not unitedly determined which constituent size is the boundary between narrow and broad focus (cf. LADD 1980 and LAMBRECHT 1994).

(8) A: Was hat Peter gemacht? ‚*What did Peter do?*‘

B: Peter hat [einen KUchen gebacken]$_F$. ‚*Peter has [baked a cake]$_F$.*‘

The focus constituent in example (8) is determined exactly as the ones in the examples (4) - (7). The nuclear accent is also attracted by the focus and distributed within the phrase according to the phonological principles.

This is the system of focus applicable for German, which is relatively flexible in pitch accentuation. It is not unambiguously applicable to other languages. The highlighting of focus in German, that attracts accentuation and therefore prominence, is suitable for investigating temporal prominence alignment with other prominence markers since the focus and the accent have a high mapping onto each other, even though it is probabilistic.

The attraction of the pitch accent as a prominence marker becomes even more visible when **contrastive focus** is examined. Contrastive focus describes the overt presence of alternatives in the context, which are ruled out by the production and accenting of the focused constituent in the target sentence. This constellation often receives even higher prominence than a non-contrastive focus (PIERREHUMBERT & HIRSCHBERG 1990, BARTELS & KINGSTON 1994, KATZ & SELKIRK 2011). A contrastive focus construction is given in example (9).

(9) A: Corinna hat einen Kuchen gebacken. ‚*Corinna has baked a cake.*‘

B: Nein, [PEter]$_{CF}$ hat einen Kuchen gebacken. ‚*No, [Peter]$_{CF}$ has baked a cake.*‘

Often the purpose of contrastive focus is correction. In (9), interlocutor A does not ask a question, but makes a statement, which is then neglected and corrected by B. However, contrastive focus can also be used in the context of question. In example (9), A can also be reformulated as a question.

In prosody, every item that is not part of the focus constituent belongs to the background of the sentence. It is the less informative part of the sentence, often having appeared in the previous discourse and contributing to the grammaticality and formality of an utterance. The background is usually prosodically less prominent than the focus (FÉRY & KÜGLER 2008, BAUMANN & RÖHR 2015, IM et al. 2018, KÜGLER & CALHOUN 2020), even though prenuclear constituents are able to carry a secondary pitch accent. In syntax, topic often complements focus, being informative, new information but less important and therefore less prominent than focus.

2.3.3 Prominence and IS types

It was observed in research (originally by PIERREHUMBERT & HIRSCHBERG 1990, adapted by FÉRY & KÜGLER 2008, BAUMANN & RIESTER 2013, CANGEMI & GRICE 2016, IM et al. 2018, KÜGLER & CALHOUN 2020, IM & BAUMANN 2020 and more) that not only IS parameters and prosodic prominence are correlated, but that categorisations of both factors can be mapped onto each other. This means that certain pitch accents (types) have a high probability to be used with specific information status types or focus constituents. Concretely, IM & BAUMANN 2020 propose a mapping on four different levels in both factors. Their assumed pitch accent types correlating with information status are unaccented/low accented (L*) words, downstepped accents (!H*), high pitch accents (H*) and rising contours (L+H*), the latter ones denoting increased prominence. The four information status categories mapped on that are „Given", „Bridging", „Unused" and „New". „Bridging" and „Unused" items are summarised as „Accessible" in this thesis (introduced in section 2.3.1). It was introduced in section 2.3.1 as well, that the categories increase in Newness from „Given" to „New" which is a central factor for the increase of prominence. Thus, as a result, the following mapping of pitch accent types and information status is proposed (Figure 9, according to IM & BAUMANN 2020):

Unaccented, L*	!H*	H*	L+H*
↕	↕	↕	↕
Given	Bridging	Unused	New

Figure 9: *Mapping of pitch accent types and information status categories according to IM & BAUMANN 2020. Newness and prominence increase from left to right.*

An adaptation of this mapping to the information status categories by KRIFKA 2008 is assumed in this thesis, „given" being associated with unaccented and L* accents, „accessible" with !H* and H*+L and „new" with H* and L+H*, see Figure 10.

Unaccented, L*	!H*, H*+L,	H*, L+H*
↕	↕	↕
Given	Accessible	New

Figure 10: *Mapping of pitch accent types and information status categories as assumed in this thesis. Newness and prominence increase from left to right.*

For focus, a similar assumption can be met. Since it was observed that focus is associated with prominence, a distribution of the same pitch accent types applies to focus in this thesis. Unfocused material tends to associate with unaccented words and L*, regular focus with !H*, H*+L and H* and contrastive focus with H* and L+H*. Since this relation is probabilistic, it cannot be excluded that the categories receive all accent types, however the distribution of pitch accents mentioned above is most probable. It is thus assumed that one purpose of prosodic prominence is to encode IS probabilistically (FÉRY & KÜGLER 2008, BAUMANN & RÖHR 2015).

2.4 Interaction of the Factors

In this section, correlations between the presented factors „co-speech gestures", „prosodic prominence" and „IS", reported in scientific literature, are introduced. Two relevant articles by IM & BAUMANN 2020 and LOEHR 2012 presented important studies on the gesture-interface research. Both studies investigated gestural synchronisation in the English language and differ from the study in this thesis in the type of speech that is available. IM & BAUMANN 2020 looked at engaging rehearsed speech, which is not completely natural but rather planned, and focused on gesture occurrence distribution on different contexts. LOEHR 2012 investigated temporal synchronisation on different prosodic and gestural levels on complete natural speech without any pre-specified discourse topic.

2.4.1 Gesture Distribution and Occurrence according to Im & Baumann 2020

The study by Im & Baumann 2020 is a recent one. They investigated the three factors included in this thesis on engaging speech. More precisely, the authors investigate the varying correlation of pitch accent types and gesture occurrence with regard to the type of IS that is encoded. They find that gesture occurrence depends on the produced pitch accent and on IS. They use a difference between referential and lexical information in the correlation.

Im & Baumann introduce a prominence hierarchy for pitch accents, which is built up the following: L* < !H* < H* < L+H* (L* being least prominent). They state that the information status categories are organised on an increasing information status hierarchy scale: Given < Bridging < Unused < New and can be mapped one-to-one to the pitch accent types. In addition, they investigate three further types of IS categories: referential, lexical and contrastive information. The authors' goal is to combine previous insights in a three-way interaction as earlier studies found correlations either between co-speech gestures and prosodic prominence (in addition to Loehr 2012, e.g. McClave 1994 or Jannedy & Mendoza-Denton 2005) or co-speech gestures and IS (Loehr 2004 and Bergmann & Kopp 2006). In addition, the influence of prosodic prominence had not been examined before, prosodically more often temporal alignment was investigated. Im & Baumann 2020 analysed a TED Talk speech by annotating gestures, prosodic prominence and IS and counted the occurrence of gestures with items of various combinations of prosodic prominence and information status.

Their results showed that the occurrence of gestures did significantly correlate with prosodic prominence in a probabilistic relation: more prominent items tended to be accompanied by gestures more often than items with less prominent accents. An even finer division was possible, according to the prominence hierarchy. Only about 60% of the pitch accented items were accompanied by gestures at all. Similarly, occurrence of gestures significantly correlated with the information status hierarchy in referential, lexical and also contrastive information. Gestures occurred more often with new or accessible information than with given information, however they did not occur more often with referentially new than accessible information. Gestures also occurred more often with contrastive items than with non-contrastive item. Overall, information status items were only accompanied by gestures in less than 50% of the cases, contrastive words in 67%. The interaction of prosodic prominence and pitch accent types with information status was significant. When no IS information is expressed, the items often did not receive any accent or gesture. With IS information, words were mostly accompanied by pitch accents or accents with gestures (Figure 11, from IM & BAUMANN 2020, p. 693).

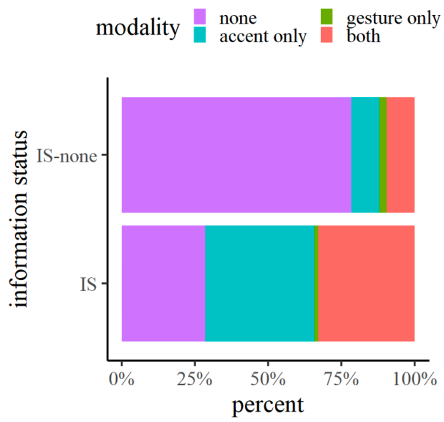


Figure 11: *Occurrences of pitch accents and gestures and their combination in presence or absence of information status. From* IM & BAUMANN *2020 p. 693.*

IM & BAUMANN assume that their observation of a correlation between gesture occurrence and prosodic prominence is a result from both modalities being subject to the „specific intention of a speaker“ (IM & BAUMANN 2020, p. 692), as their relation is probabilistic, and prosodic prominence is a more frequent tool of this intention marking than gestures are. Another observation the authors make is a slightly different behaviour between gestures and pitch accents which might lead to mismatches between gestures and prosody: while pitch accents are avoided to be adjacent to each other (NESPOR & VOGEL 1989, SHATTUCK-HUFNAGEL et al. 1994), gestures tend to be carried over across words. For information status, even if the results were less consistent, IM & BAUMANN suggest that newness facilitates the occurrence of gestures.

While the authors did not investigate the temporal alignment of gestures and speech, they examined their probabilistic parallelism and determined a co-occurrence of gestures and pitch accents dependent on IS information. IM & BAUMANN 2020 found increasing occurrence of gestures according to the prominence hierarchy as well as, in a majority, the information status hierarchy. An interaction of the three factors was also observed, which provides useful information for this thesis on spontaneous rather than rehearsed speech.

2.4.2 Temporal Synchronisation according to Loehr 2012

LOEHR 2012 provides evidence for the temporal synchronisation of prosodic constituents with gesture sequences. He states that both are used as tools for similar pragmatic functions to form a complete pragmatic expression. LOEHR determines three differently sized prosodic constituents suitable for synchronisation: pitch accents, intermediate phrases and intonational phrases. He assumes four gestural constituents: the apex of a gestural stroke, gesture phases, gesture phrases and gesture units. As the author explains, the units of both modalities are arranged on a hierarchical scale from smallest to largest and are chosen as plausible domains for synchronisation. LOEHR 2012 wants to get insight into the concrete alignment of gesture and intonation beyond the previous findings (e.g. ROTH 2002) that gestures and stressed syllables align.

In order to test the synchrony, LOEHR recorded conversations of two to three interlocutors talking about any not pre-specified topic, annotated the video and audio files and examined the modalities' temporal synchronisation observationally and statistically. He observed that a general temporal alignment of intonational and gestural elements can be confirmed, as tones occurred heaped near gestures. The structural alignment can particularly be observed between two pairs of the hierarchical modalities: as already suggested in the literature (LOEHR 2004, LEONARD & CUMMINS 2009, 2011, SHATTUCK-HUFNAGEL et al. 2007 among others), the apices of gestural strokes consistently aligned with pitch accents with only little deviation. In addition, a looser alignment between gestural phrases and intermediate phonological phrases was found, which contains the observation that the alignment concerns intermediate phrases rather than intonational phrases. LOEHR assumes that this might be a result from the amount of information packaged by the respective constituents in accordance with PIERREHUMBERT & HIRSCHBERG 1990. Figure 12 shows the histograms of deviations between apices and pitch accents in 12a) and between g-phrases and intermediate phrases in 12b).

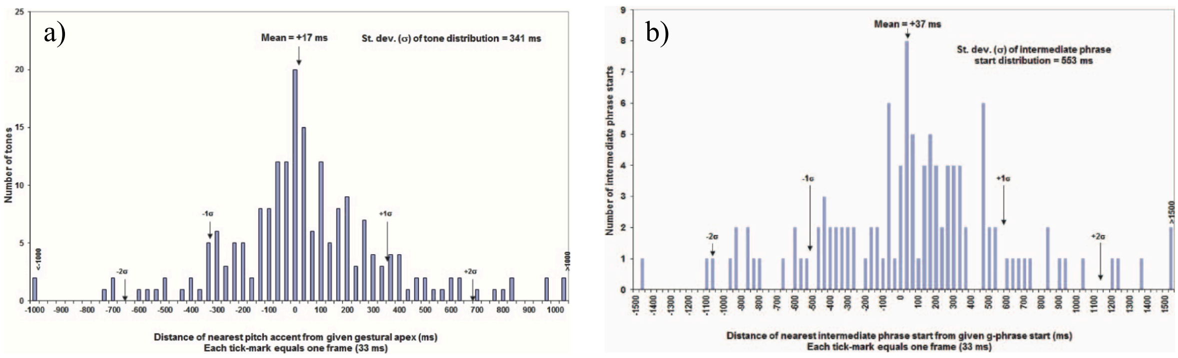


Figure 12: *Histograms of prosodic and gestural domain synchronisation in LOEHR 2012. Left: Synchronisation between pitch accents and gestural apices which have a mean deviation of 0 ms (p. 81), rigtht: Synchronisation of intermediate phrases and g-phrases with a mean deviation of +17 ms (p. 83).*

In addition, LOEHR observed that the presence of one of the intonational elements does not require the presence of a gesture. Finally, if the temporal alignment was slightly deferred, the gestural units tended to precede the phonological counterpart in accordance with MCNEILL's phonological synchrony rule. With regard to pragmatic statements of both intonation and gesture, the investigation showed a general agreement between the two modalities expressing sentence (in)completeness, IS, emphasis or contrast.

LOEHR 2012 not only underlines gestural research with empirical natural speech data, provides evidence for the alignment of gestural apices and pitch accents and introduces the alignment of bigger constituents, he also provides useful explanations for the annotation and statistical procedure of temporal gesture-speech alignment.

3. The Corpus Study - Material and Methods

A corpus study was conducted to investigate non-referential gesture occurrences and the temporal synchronisation of gesture and speech under the influence of IS. The study addresses two **research questions**:

1. Does Information Structure in its parameters information status and focus influence the occurrence of non-referential co-speech gestures in spontaneous German speech?
2. Is pitch accentuation temporally aligned with non-referential gesture apices in German and does Information Structure influence this alignment?

The **hypotheses** for both research questions result from and are underlined by previous findings from research on this synchronisation in English, which is assumed to behave similar to German in many intonational aspects. Regarding the first research question, the hypothesis is that non-referential gestures occur more often during articulation of focused or „new" constituents in German than with background or „given" information in line with the findings by IM & BAUMANN 2020 who put a strong focus on the investigation of IS facilitating or blocking the use of gestures. For research question 2, the hypothesis is formulated based on findings by SHATTUCK-HUFNAGEL et al. 2007, LOEHR 2012 and many others, saying that pitch accentuation and the apices of gestures align temporally in German. In addition, it is assumed that IS influences this alignment, in a way that more prominent IS referents increase the accuracy of accent-apex alignment.

In this chapter of the thesis, the analysed corpus (3.1), stimuli (3.2), annotation systems (3.3) and statistical methodology are presented (3.4).

3.1 The SaGA Corpus

The corpus analysed in this study is the Bielefeld **Speech and Gesture Alignment** (SaGA) **Corpus**, which was created by ANDREAS LÜCKING and colleagues from Bielefeld University and introduced in [LÜCKING et al. 2010]. The corpus contains German data of controlled natural speech conversations and works with a virtual reality environment. Twenty-five conversations between two interlocutors, of which one speaker had the main talking part, were recorded and prepared for analysis. In order to make the data comparable between subjects and control the content of each discourse, the authors worked with a task for the participants and provided the same stimulus for all conversations. For this purpose, they created a virtual reality (VR) environment, concretely a town, in which a bus ride took place along specific stops. For each conversation, one participant was asked to ride through the town with the bus and describe the surrounding of all five bus stops (sculpture, city hall, church square, chapel, fountain) and the way in between them. The conversation partner such that the partner should be able afterwards to find their way from the start to the last stop on their own. Figure 13 shows exemplary cutouts from a video of the VR town.

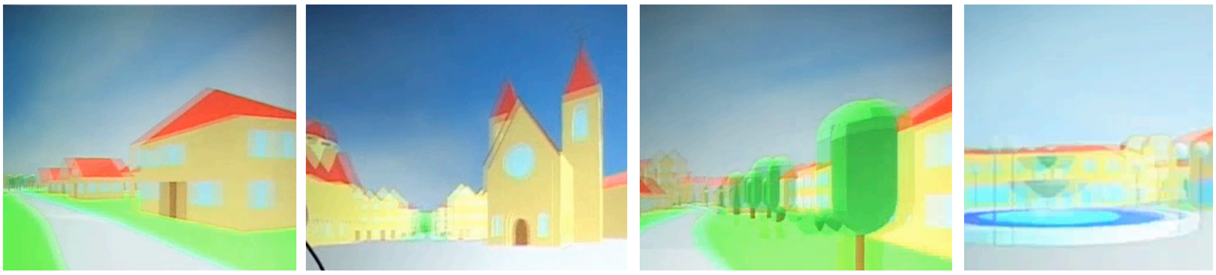


Figure 13: *VR town environment as the base stimulus for the recording of the corpus, screenshots from the bus ride video. From left to right: basic houses and the street, church square (right church), houses and trees, fountain / final station.*

The collectors of the corpus chose this direction-giving and sight-description task based on findings by Alibali 2005, who claims that „speakers gesture more when they talk about spatial topics than when they talk about abstract or verbal ones“ (Alibali 2005, p. 313). In addition, the VR town description task leads to a corpus of spontaneous speech conversations while specifying a conversation topic and therefore controlling the discourse to some extend.

All conversations were recorded in three different camera perspectives providing two videos showing each interlocutor alone once and a total view showing both interlocutors together in their conversation. In addition, the participants describing the bus ride wore electronics to measure their hand and body movements precisely and VR glasses to be able to navigate and accurately describe the VR environment. On the glasses-screen, eye movements were tracked. According to Lücking et al. 2010, throughout the 25 conversations, 280 minutes of audio and video material were collected, containing 39.435 words in total. Accompanying those almost 40.000 words, around 6.000 gestures were produced, meaning that if gestures and words would align in a one-to-one fashion, approximately 15,2% of words would be accompanied by a gesture. However, skimming through the data does not show such a one-to-one mapping, instead, gestures can last for multiple words (and one word can contain more than one gesture). The authors report almost 5.000 referential gestures and around 1.000 non-referential gestures. Lücking et al. annotated the corpus for gestures systematically and in depth using the gesture annotation software ELAN (ELAN 2021). It includes lexical annotation on the level of the word. All annotations were provided for both interlocutors, the describer and the listener, and for both hands, depending on which hand produces the gesture (two-handed gestures are indicated on both tiers). The start and end of each gesture phrase and the interval in between were marked. Gesture types (following McNeill 1992) were annotated, differentiating between „iconic“ and „deictic“ gestures for the referential category and „discourse“ and „beat“ gestures in the non-referential category. Referential gestures were delineated as suggested in McNeill 1992, „beat“ gestures in the SaGA corpus annotation showed repetition of one and the same stroke within one gesture, while „discourse" gestures were annotated when there was one distinct movement, showing more diverse motions across the gestures of this type. If gestures displayed multiple functions, this was also indicated by creating intermediate types like „iconic-deictic“ and also crossing referentiality like „iconic-beat“. In addition, „moves“ were indicated, signalling hand (or body) movements not contributing to the discourse. The gesture phases and strokes of all referential gestures were identified. Apart from the stroke, this included preparation and recovery of the hands as well as holds. For non-referential gestures, these phases were not annotated, since for previous investigations of the corpus, non-referential gestures were not considered (Lücking, p.c.).

Apart from gesture types and components, the „gesture perspective“ indicated, which interlocutor uses the gesture in the discourse. The „gesture practice“ described referential gestures themselves more in detail, specifying the movement or whether the gesture points or draws something.

These informations were annotated for all conversations, but the SaGA corpus provides more detailed information in seven conversations. In these more detailed annotations, many of these annotation categories concern the physical characteristics of the gesture. This includes the position and direction of the palms, wrists or back of hands, as well as the hand shape, the position of the hands relative to each other and the extent of the movement. Non-physical information includes semantic and discourse functions the gestures convey. Linguistically, lemmas and word categories were indicated on these conversations. The apices of strokes were indicated as well in these conversations. Phonological prominence and syllable nuclei were annotated, but not in connection to pitch accents, which is done for this study (see section 3.3.2).

3.2 Data

From the 25 discourses, all complete lexical referring speech material (NPs, PPs and pronouns) which is accompanied by a non-referential gesture to some extent and/or is coded for IS is in principle suitable for investigating the research questions of this thesis. For the study all conversations were selected, that contained audio and video material of the conversation as well as gesture type annotation. 18 conversations fulfilled this criteria, leading to 204 minutes of discourse material. The longest of the 18 available conversations had a duration of 19:26 minutes (C5), the shortest was 5:18 minutes long (C16). The average length of one conversation is 11:33 minutes. 730 of the 1.000 non-referential gestures in the corpus were found in the included material. The 657 „discourse" gestures and 73 „beat" gestures were annotated with 775 stroke apices in total (since the gesture-stroke relation is not 1-to-1). For gesture occurrence, all referential items providing IS information are considered and investigated for the amount of pitch accent and/or gesture occurrences. For temporal synchronisation investigation, only items showing and containing non-referential gestures are examined. Gestures and utterances of both conversation partners were considered.

The included data provide 5.024 data points in total, considering each interval in time that contains a pitch accent, a non-referential gesture and an IS referent, or at least one of the three.

Out of the 36 participants (18 dialogues), 20 participants were males, 16 females. However, looking deeper into the talking parts and tasks of the participants, 13 route describers where male and only five female. In four of five conversations led by a female, the listener was female as well. In total, seven listeners were male, eleven female. No age information was provided by the annotated corpus. All participants but one seem to be young adults.

3.3 Annotations

In this section, all annotation systems that were used in this study are introduced. The data of the SaGA corpus are annotated with regard to the factors „non-referential gestures", „prosodic prominence (pitch accents)" and „Information Structure" (information status and focus) following the presented systems. Gesture annotation already provided in the SaGA corpus from LÜCKING et al. 2010 was used in this study, but was not sufficient to answer the research questions. For additional gesture annotation, the MultiModal MultiDimensional system

(„M3D" system; ROHRER et al. 2020) is used. For intonation annotation, the GToBI system (GRICE et al. 2005) was used. For IS, the GÖTZE et al. 2007 Core annotation system was applied. The gesture annotation software ELAN (ELAN 2021) was used to collect the complete annotated information; see Figure 14 for an example of the ELAN interface with all annotations.

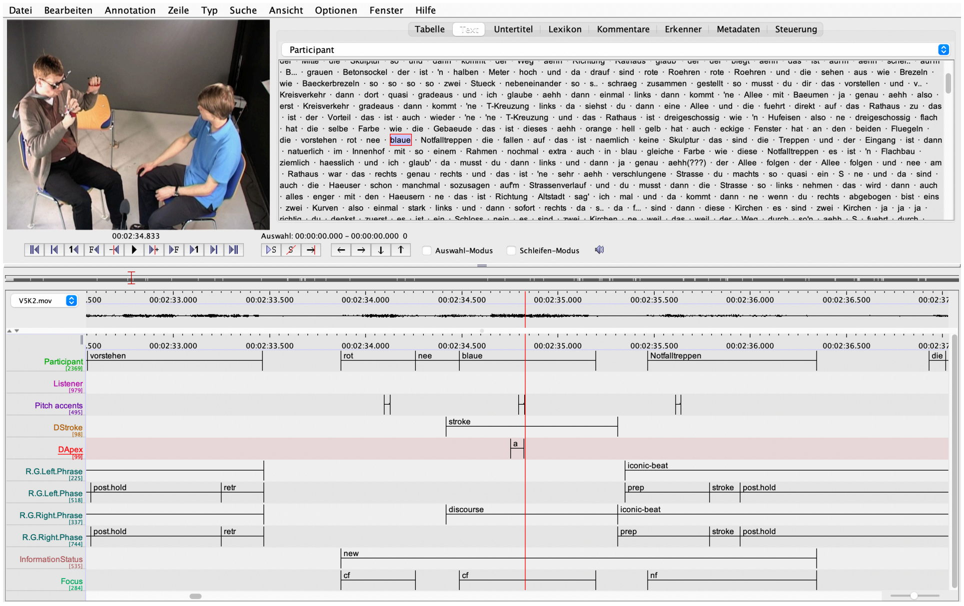


Figure 14: *ELAN annotation software surface. Top left: video file(s), top right: navigation through the file, bottom: annotations split into different tiers, containing word, gestural, intonational and IS annotations.*

3.3.1 Gesture Annotation

The annotation of non-referential gestures is introduced in this section. Gesture type annotation is adapted from the SaGA corpus in a modified way, combining „beat" and „discourse" gestures together to the type „non-referential" gestures. Since for non-referential gestures, strokes and apices were not annotated in the SaGA corpus, the annotation of these components was done for this study following the M3D annotation guidelines (ROHRER et al. 2020). During annotation, the stroke is identified a) as the central and most important movement of the g-phrase and b) by delimitation to the remaining gesture phases. A pre- or post-stroke hold can indicate the end of a stroke, while also preparation and recovery are identifiable by a non-prominent hand movement from or to the resting position respectively, often being less complex and slower than the stroke.

Dependent on whether the hand movement direction is changed during the gesture, the apex of the stroke describes the point in time, where the movement is extended the most (unidirectional stroke) or the turning point of the hand (bidirectional & multidirectional strokes). In other words, the apex is „the kinetic goal of the stroke“ (LOEHR 2007, p. 190), a point where no hand motion can be observed. To find this point in annotation, the M3D model proposes to identify the decrease of movement or velocity in the stroke, being visible through a clearer, less blurry image in the video. Since ELAN only allows time intervals to be annotated, the M3D model uses a two frame time interval, which has the length of 66 ms. The smallest possible interval in ELAN is 10ms - but intervals below 30ms are not useful in visual research, since one video frame in this analysis complies to 33ms. The right boundary of that interval refers to the apex of the stroke. Figure 15 illustrates how gesture annotation was performed in ELAN. The tiers „DStroke“ and „DApex“ were annotated specifically for this study, the remaining tiers were already contained in the corpus.

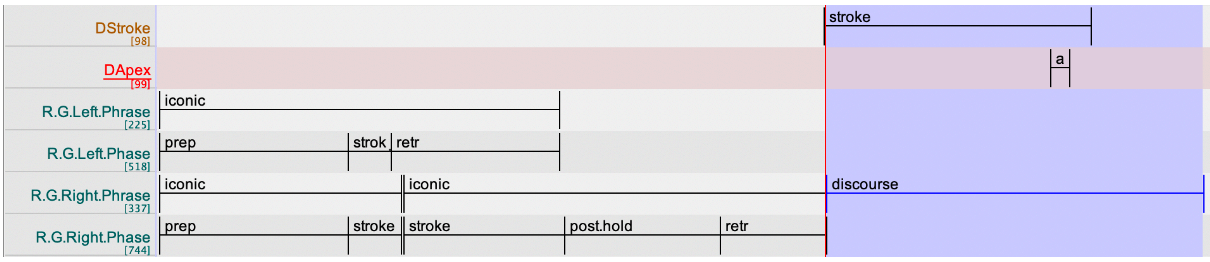


Figure 15: *Gestural annotation tiers in ELAN. Informations about gesture types, g-phrases, g-phases, strokes and apices of non-referential gestures.*

3.3.2 Prosodic Prominence Annotation

Concerning prosodic prominence, pitch accents have to be indicated in the data, since those are the tones that mark prominence on speech material. In this study, the GToBI system (originally by REYELT et al. 1996, GRICE et al. 1996, up-to-date version by GRICE et al. 2005) is used for intonational annotation. It is chosen over other renomated intonation annotation systems (DIMA, Deutsche Intonation: Modellierung und Annotation; KÜGLER et al. 2015, 2019, 2022 or KIM, Kiel Intonation Model; NIEBUHR 2019) because of its applicability to German and use of the autosegmental-metrical framework. The annotation of intonation is done with the computer software Praat (BOERSMA & WEENINK 2022).

GToBI is a variation of the ToBI (Tone and Break Indices) system, developed by SILVERMAN et al. 1992 a,b, to accurately and consistently label intonation in German. The GToBI system does not only allow to mark pitch accents and pitch accent contours, but also phrase accents and boundary tones. All tones receive a binary value of their F0-height: H (high) or L (low), relative to the mean fundamental frequency of the speaker, as H and L tones mark local maxima or minima of pitch. Pitch accents are indicated by an asterisk after the tone T: T*, pitch contour include tones directly before or after the pitch accent, added by a plus: T*+T, T+T*. Phrase accents are represented by a minus after the tone, T-, and apart from marking a phrase boundary, they can form an intonation phrase boundary in combination with boundary tones (indicated by a percent sign): T%. A sentence boundary is annotated as follows in GToBI: T-T%. Downstep is marked by an exclamation mark before the tone, !T* and upstep by a circumflex: ^T. Table 1 gives an overview over all tones that are distinguished in the GToBI labelling system and a short description of the pitch behaviour.

Pitch accents	**Description**	**Boundary Tones**	**Description**
H*	Local pitch maximum	L-L%	Low terminal Fall
L+H*	Rise on stressed syllable, late accented high tone	L-H%	Small terminal Rise
L*+H	Rise on stressed syllable, early accented low tone	H-L%	Leveled boundary
L*	Local pitch minimum	H-H%	High terminal Rise
H+L*	Fall before the low accent		
H*+L	Early peak before a Fall		
!H*	Downstepped High Accent; lower than preceding f0 peak in the same Intonation Phrase		
H+!H*	Smaller fall		

Table 1: *Pitch accents and boundary tones according to GToBI, including a description of each tone.*

As has already risen from the previous description, different types of boundaries are distinguished in the GToBI annotation. B1 boundaries are word boundaries, which usually do not receive any prosodic boundary marking. B2 boundaries are irregular boundaries, resulting from hesitation or correction. B3 are phrase boundaries, which were introduced before and usually delimit prosodic phrases. B4 boundaries are the biggest boundaries, characterising intonation phrases and elicit the biggest pause to another (intonation) phrase, compared to all other boundary types.

The GToBI annotation system is applied for pitch accents in this study. While H*+L and !H* were not counted to the basic pitch accents in GRICE et al. 2005, they depict important accents in the present analysis, since they are notions of pitch accent types. H+!H* does not play a role in this analysis. Pitch accents were annotated on the noun of the information status phrase, since the noun is the main referring part of the phrase. Possible additional accents on determiners or adjectives were not annotated. For focused constituents, the most prominent item of the phrase was annotated. If information status and focus phrases overlapped but started or ended on different points (e.g. *[geradeaus (den Weg)*$_I$ *entlang.]*$_F$ ‚straight ahead along the path‘), the focus domain could receive a second accent in addition to the information status NP (pitch accent on *geradeaus* for focus and *Weg* for information status in the example). Figure 15 shows an example of pitch accent annotation on IS referents in Praat.

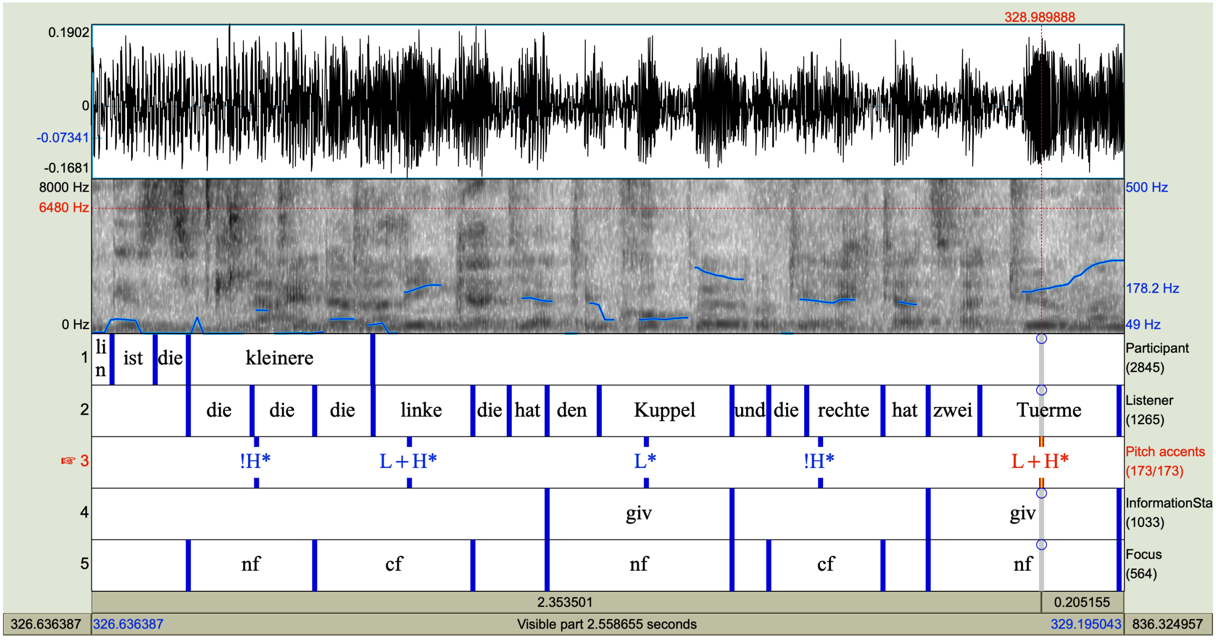


Figure 16: *Pitch accent annotation in Praat according to GToBI. Top shows oscillogram and spectrogram, bottom shows annotation tiers containing lexical word annotation, pitch accents and IS parameters.*

Pitch accents are the tones responsible for perceptive prominence, boundary tones and phrase accents will not be annotated for this study. Their purpose is to form sentence intonation. Since this is not relevant for the research questions of this thesis, they are not annotated for this study. To introduce the full annotation system of GToBI, all tones were listed here.

3.3.3 Information Structure Annotation

Two parameters of IS are investigated in this study, information status and focus. For annotation, the Potsdam annotation guidelines by Götze et al. 2007 were used. This system provides a core annotation scheme and an extended one. For this study, the Core annotation scheme is sufficient, but the extended labels are briefly introduced as well. The authors provide annotation schemes for information status, focus and topic. In addition to a detailed description of all labels they propose, Götze et al. provide many concrete examples to their labelling guideline and a roadmap of yes/no-questions helping to decide which label a referent receives according to their model. Depending on the answer to those questions a label can already be chosen or one is forwarded to the next question (see Appendix A).

Information Status: For information status, in the core scheme, three labels are possible: „given", „accessible" and „new". Referring lexical items (NPs, PPs and their pronouns, not expletives like *there*, *it*, *on the other hand*, etc.) receive a label dependent on their retrievability and accessibility in the discourse, concretely, whether they have an antecedent in the previous discourse or not. Referents receive the label „given" (giv), if the „expression has an explicitly mentioned antecedent in the previous discourse" (GÖTZE et al. 2007, p. 153), specifically in the last five sentences. An „accessible" (acc) referent has „not been mentioned, but is accessible via some kind of relation to a referent in the previous discourse, in the situative context, or the assumed world knowledge of the hearer, or a combination thereof." (GÖTZE et al. 2007, p. 156). Thus the referent has not been mentioned explicitly in the close preceding discourse. The retrievability often has an origin in semantic relations or world knowledge. Items that are not considered „given" anymore, are labeled as „accessible". Pronouns are accessible as well since they always need an antecedent in the discourse. „New" (new) referents are „new to the hearer and to the discourse" (GÖTZE et al. 2007, p. 160), they have not been introduced to the CG before, but enter the discourse by their appearance.

In the extended annotation scheme, the „given" and „accessible" status receive sub-labels depending on the following parameters. „Given" is split into „giv-active", which is accurate when the explicit referent has been mentioned within the last or current sentence, and „giv-inactive" where the referent has been mentioned in the close preceding discourse (five intonation phrase), but not the current or last sentence. „Accessible" can be split into four sub-labels. The first sub-label, „acc-sit" is assigned when the referent is accessible through the discourse situation itself. Pronouns belong to this category among others. When given or accessible referents are referred to as a group, this group receives the label „acc-aggr". Referents are labelled „acc-inf", if they are retrievable from the discourse by some kind of bridging relation such as a part-whole relation, set relation or attributing relation. The last sub-label GÖTZE et al. assume is „acc-gen" which implies that the hearer can access the target referent from their world knowledge.

Focus: With regard to Focus, two different labels are available in the core annotation scheme. The authors distinguish between „new-information focus“ (nf) and „contrastive focus“ (cf). Broad vs. narrow focus is not distinguished. According to the authors definition, focus concerns the most relevant information in a particular context, which is selected from possible alternatives. In addition, the focus domain can be of different size, depending on how much information is new or relevant, and multiple foci inside one „sentence“ (intonation phrase in this study) are possible. This especially concerns the appearance of new-information focus and contrastive focus in one sentence or contrastive focus contrasting multiple information samples. Focus is claimed to be present in non-declarative clauses like interrogatives or imperatives as well. „New-information focus“ is the part of the utterance that „provides new and missing information which serves to develop the discourse“ (GÖTZE et al. 2007, p. 172). This type of focus often contains those items that are the answer to explicit or implicit questions and do not overtly rule out other alternatives. This is the case for „contrastive focus“ items, which „evokes the notion of contrast to (an element of) another utterance“ (GÖTZE et al. 2007, p. 178). This often results in multiple contrastive foci in one sentence. GÖTZE et al. 2007 also introduce the labelling of possibly present focus operators like „only“, which is indicated by the notion „+op“.

In the extended annotation scheme, seven sub-labels are available, two for the nf-label and five for the cf-label. In new-information focus it can be distinguished, whether the focused constituent is the answer to an explicit question/has been explicitly requested by another interlocutor or not. Explicitly requested items are labelled „nf-sol“. On the other hand, referents receive the label „nf-unsol“, that have only implicitly been requested (if they were requested at all) or serve to develop the discourse. It often introduces new referents and also applies to predicates and quantificational determiners. In terms of contrastive focus, more distinctions are drawn. The first cf-type is „cf-repl“, indicating correction of something previously uttered in the (close) discourse. Another type shows the overt presence of two or more alternatives, which are both labelled as contrastive, and then the choice between those alternatives, which is then labelled as „cf-sel“. The next label, „cf-part“ is used when a previously mentioned referent that describes a group is split up. The new parts of the group that are specified by further lexical information receive the „cf-part“ label. The definition of the label „cf-impl“ is directly adopted from GÖTZE et al., containing a more complex relation.

They state that „cf-impl“ „implies that the requested information holds true not for the information provided explicitly in the answer but for other alternatives that are accessible in the context.“ (GÖTZE et al. 2007, p. 181). The last sub-label of focus is „cf-ver“, and it is used to enhance the truth value of a sentence, making the whole clause the labelled domain, but the prosodic prominence concentrates strongly on one constituent which has putatively been challenged in the previous discourse.

GÖTZE et al. 2007 provide a systematic and easy to use annotation system for IS, which offers a clear distinction in its core version and the possibility to specify each component further within their label. For this study, the core categories of the model are applied. The labelling of lexical referents concerns the referential as well as lexical accessibility of items in this discourse in this study, as a further distinction goes beyond the scope of this thesis (but cf. BAUMANN & RIESTER 2013, RIESTER & BAUMANN 2017). The close preceding discourse being defined as the „last five sentences“ for the „given“ label is a convention rather than an empirical preset. This part is not adopted from GÖTZE et al. 2007 but following BAUMANN (p.c.), who suggests to use five intonation phrases for defining the label „given“, since those are easier to determine and delimit. In addition, it is not pre-specified whether those phrases are counted for one or all interlocutors, for this study intonation phrases of both interlocutors are counted as contributing to this pool for the labelling of „given“.

While pronouns are included in the information status labelling system by GÖTZE et al. 2007, they are not annotated for the analysis of this study. Pronouns are no lexical constituents that are typically connected to IS and they rarely carry pitch accents (SELKIRK 1996, TRUCKENBRODT 2007, KÜGLER 2018), if not in focus (KRATZER & SELKIRK 2007, TRUCKENBRODT 2006, ZERBIAN & BÖTTCHER 2019). This makes them irrelevant for the present analysis.

3.4 Methodology and Statistical Analysis

All annotations are collected together in ELAN and afterwards extracted as a tab-separated file containing information about all time intervals which were annotated on each tier. Apart from the content of the interval, the start and end time of the interval, as well as the duration of each interval are contained in that file. In the first part of the analysis, the synchronisation of pitch accents and non-referential gestures is investigated on the word level by a

distribution analysis. Words that provide information status or focus information are considered. In the second part of the analysis, the exact point in time of appearance of the gesture apices and the deviation of pitch accents from that point are investigated by a **temporal synchronisation analysis**. IS is taken into account. Descriptive and inferential statistics are performed using R (R CORE TEAM 2013). The complete R-Script written to perform the statistical analysis is provided in Appendix B. Summarising the previous section 3.3, the factors and their levels presented in Table 2 are considered:

Pitch Accents	**Non-referential Gestures**	**Information Structure**	
		Information Status	**Focus**
No Accent	No Gesture	None	None
L*	Stroke overlaps with IS	Given	New-information
!H*	Apex, not on IS	Accessible	Contrastive
H+L*	Apex on IS	New	
H*+L			
H*			
L*+H			
L+H*			

Table 2: *Investigated factors and their factor levels.*

For the distribution analysis, the two rise levels (L+H* and L*+H) and two fall levels (H*+L and H+L*) have been merged together to the levels L+H* containing the rising accents and H*+L containing the falling accents. For this part of the analysis, the exact location of the accent is not important, as whole words are the domain of synchronisation, and the merged categories agree with the pitch accent prominence scale assuming pitch accent types rather than all pitch accents to be relevant (section 2.3.3, BAUMANN et al. 2006, IM & BAUMANN 2020). All pitch accents listed in Table 2 are considered for the temporal synchronisation analysis, because here, the exact position of the pitch accent is the subject of investigation, in its alignment with gestural apices.

For the distribution analysis, occurrences of gestural apices and overlapping of strokes (but not coinciding with the apex) with IS material (and as well if they occurred during speech breaks or non-IS speech material) are counted. In addition, occurrences and types of pitch accents are counted on IS material. Initially, total occurrences and distributions of all factors are shown for each of their levels. Stacked-bar plots are used to illustrate occurrences of a variable A dependent on the distribution of another variable B, indicating the percentage of each level of variable A for each level of variable B (e.g. Gesture distribution given the different pitch accent types). In order to examine the interaction of all three factors, further comparisons are drawn for two variables A and B given a certain value on variable C, specifically:

- Gesture
    - apex produced (on IS referents)
- Information Status
    - given
    - accessible
    - new
- Focus
    - Focus (new-information focus & contrastive)
    - No Focus

The significance of the results is tested using Pearson's Chi-Square-tests, since all variables are of categorical nature. A level of significance of $\alpha = 0{,}05$ is chosen. In addition, Pearson's residuals are calculated to get more specific information about the significant combinations. They are illustrated using correlation plots.

For the temporal synchronisation analysis, all apices of non-referential gestures are investigated. Their point in time of production is compared to the time of production of the nearest pitch accent, not taking semantic relations into account. Since ELAN only allows time intervals to be annotated, the right boundaries of the respective intervals are counted as the production point of pitch accents and apices. The time interval of pitch accents was set to 30 ms during the importation from Praat to ELAN. The time interval of the gestural apex was 70 ms (ca. two frames), ending with the point of lowest velocity. The distance of the nearest pitch accent and gesture points are calculated regarding their time of production in seconds and milliseconds by subtracting the time of the apex from the time of the pitch accent. The time stamps for the boundaries are received by data extraction from ELAN. For each gestural apex, the nearest pitch accent is chosen (not the other way round, since there were more pitch accents than gestures), looking at the deviation of that pitch accent from the apex. The apex therefore forms a zero-point. Pitch accents preceding the apex receive a negative distance value, pitch accents following the apex receive a positive one. Histograms are plotted to illustrate the deviations. For this purpose, the distances between the modalities are grouped into categories of frames and distances of 33 ms are grouped together. They are summarised under the median of that frame, thus 16,5 seconds off both „ends“ of the frame. All data points with a deviation of three seconds or more are grouped together and excluded from the analysis, since these words have a clear semantic displacement to speech. The standard deviation and mean deviation are calculated. It is then examined whether the synchronisation changes under the different aspects of IS. For this purpose, it was indicated in the data, whether the apex appeared on a word coded for IS. The coding is compared to the coding of the distribution analysis and it is assured that the labelling matched. Corresponding subsets of the apices are then analysed for a possible change in degree of synchronisation with the same procedure. Using qq-plots, the data are visually tested for normal distribution, before proceeding with inferential statistics. In the recent years, several studies (RASCH & GUIARD 2004, PAGANO 2010, WILCOX 2012) have proposed that an unpaired t-test (which is the appropriate statistical test for this type of continuous data and analysis) is robust also for not normally distributed data. Following this reasoning, an unpaired one sample t-test is performed to test statistical significance of the results of the general temporal synchronisation of apices and pitch accents. The chosen level of significance is $\alpha = 0{,}05$. T-tests are calculated for the general synchronisation and synchronisation of the different subsets distinguished by IS. In these tests, the mean values of the data are compared for their deviation from a mean of 0 ms. To compare the deviation differences between the different factors, box plots are drawn and their significance is tested using a one factorial ANOVA (information status) and a two-sample t-test (focus).

4. Distribution Analysis

In this chapter, the results of the investigation on synchrony of gestures and prosodic prominence on the word level are presented. Material that conveys IS information on the parameters of information status and focus is considered. First, the general distribution of all factor levels and their link to one other factor at a time are presented in section 4.1. In section 4.2, observations on interactions of all three factors are illustrated. In section 4.3, the individual and combined occurrences of pitch accents and apices are explored under the influence of IS.

4.1 General Distribution

730 non-referential gestures were extracted from the corpus, and 775 apices were found on them. 306 (39,5%) of these apices aligned with IS referents, while 469 (60,5%) did not. They appeared either on words not coded for IS or in a speech break (e.g. hesitation or turn taking,…). In 243 cases (51,8% of the non-aligning apices; 31,4% of all apices), the apex of a gesture was very close to an IS referent so that while the apex itself did not coincide with the referent, the gestural stroke overlapped. An overlap was counted when the stroke and the noun of the IS constituent overlapped for more than 100 ms. Figure 17 shows gesture distribution in the whole corpus (17a) as well as apex and stroke distribution across the gesture occurrences (17b).

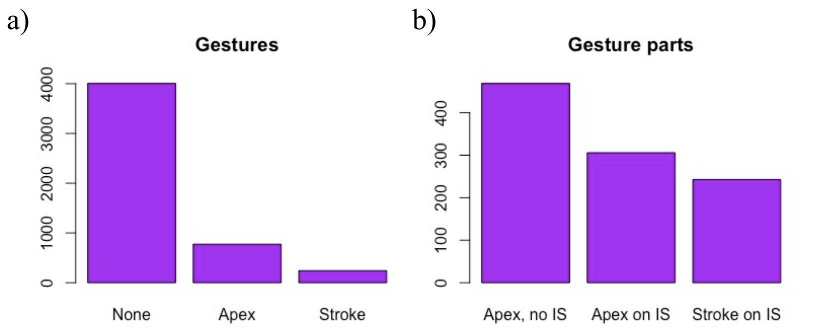


Figure 17: *Distribution of non-referential gestures. Left: Presence of gestures, right: Distribution of gesture apices dependent on IS presence. The Y-axes show the occurrence frequencies of each category.*

In total, 4.394 pitch accents were produced on referring expressions coded for IS. Across the pitch accent types, H* accents were produced most often (1.579 occurrences, 35,9%), H*+L accents least often (274 occurrences, 6,2%). The distribution of pitch accents is listed in Figure 18 and Table 3.

3.939 referring NPs were found in the corpus that were coded for information status and 2.773 phrases were in focus (Figure 20a). In 2.161 cases (54,9% of all information status phrases and 77,9% of all focused phrases) information status and focus coincided, meaning that 1.778

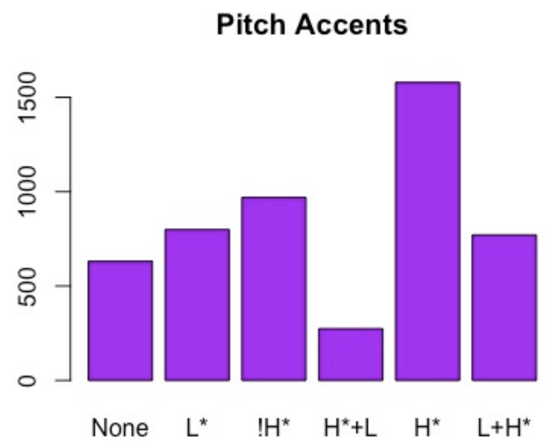


Figure 18: *Distribution of pitch accents in pitch accent types, the order of the accent types follows the Prominence Scale.*

Pitch accent	Occurrences
No accent	631
L*	799
!H*	970
H*+L	274
H*	1.579
L+H*	771

Table 3: *Distribution of pitch accents on IS across the corpus, the order of the accent types follows the Prominence Scale.*

information status referents were not in focus (45,1%) and 612 focused constituents did not contain a lexical phrase coded for information status (22,1%). In total, 4.551 phrases carried at least one type of information structural information.

For information status, there was a similar amount of „given" (1.540, 39,1%) and „new" (1.446, 36,7%) referents, and less „accessible" referents (953, 24,1%), see Figure 19. The majority of focus phrases were „new-information focus" (2.345, 84,6%), the remaining 428 (15,4%) were „contrastive" focus (Figure 20b).

Figure 19: *Distribution of information status levels.*

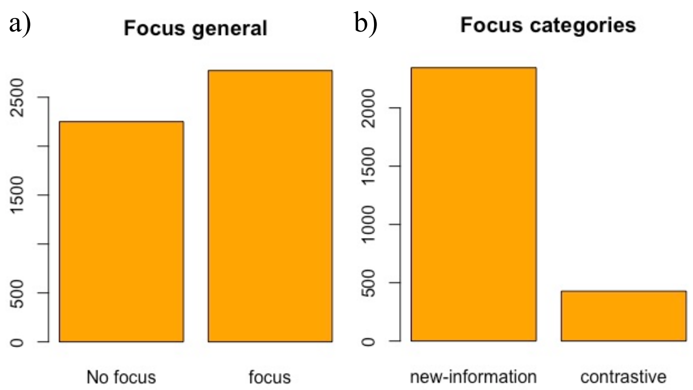


Figure 20: *Distribution of focus. Left: Focus presence and absence, right: focus categories.*

4.1.1 Pitch Accents and Information Structure

Pitch accents and IS correlated in a way that newer or more informative words are more prominent and receive higher pitch accents.

Regarding pitch accents and information status, the relative distribution of L*, !H*, H*L and H* seems very similar across all levels of information status (Figure 21). Unaccented words were found almost exclusively in the „accessible“ status, while more L+H* accents were found on the „new“ referents compared to the remaining two levels. Table 4 gives the absolute occurrences of pitch accents for each information status level.

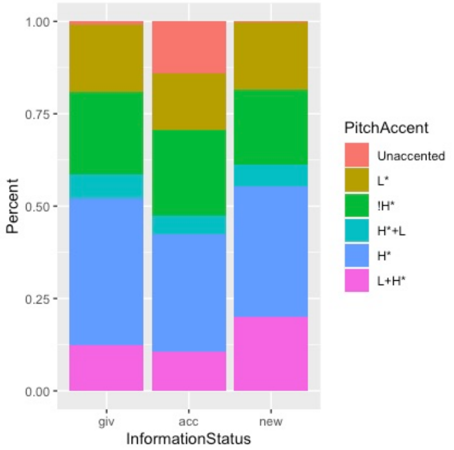


Figure 21: *Relative distribution of pitch accents on different information status levels (X-axis). Y-axis: percentage of pitch accent occurrences on total referents of that category.*

	None	Given	Accessible	New
Unaccented	481	11	134	5
L*	107	282	146	264
!H*	114	345	220	291
H*+L	42	100	48	84
H*	152	610	303	514
L+H*	189	192	102	288

Table 4: *Distribution of pitch accents on different levels of information status.*

A Pearson’s Chi-Square test revealed that these results are significant (X-squared = 408,24, df = 10, p < 0,001). Calculating Pearson’s residuals and a correlation plot (the complete residual-values for all calculations are provided in Appendix C) showed a strong positive correlation between the levels „accessible“ and „unaccented“ (res-value = 16,2) and only weak correlations for all remaining levels (cf. the correlation plot in Appendix D). As positive correlations were calculated for „new + L+H*“ (res-value = 5,09), „accessible + !H*“ (res-value = 0,89) and „given + H*“ (res-value = 2,21), for illustration purposes, the significance calculation was repeated, excluding all unaccented words. This Chi-Square test was still significant (X-squared = 47,161, df = 8, p < 0,001) and calculating the residuals revealed the same positive correlations mentioned above in detail, see the correlation plot in Figure 22.

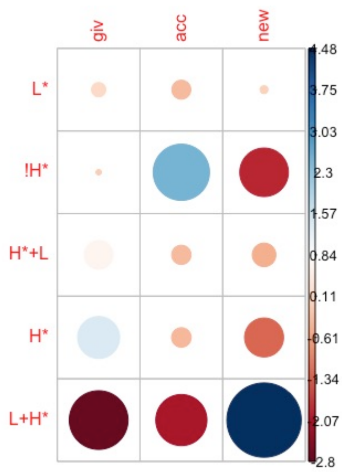


Figure 22: *Correlation plot on the correlation of pitch accents and information status (without unaccented). The size of the dot and the intensity of the color describe the strength of the correlation. Blue color: positive correlation, red color: negative correlation.*

While in this plot those correlations are weak compared to the data set including the unaccented words, it can be observed that the strongest positive correlation is between the „new“ and „L+H*“ levels (res-value = 4,48). The residuals for „accessible“ and „!H*“ were 2,57 and 1,42 for „given“ and „H*“. Negative correlations were found for „L+H*“ with „given“ (res-value = 2,79) and „accessible“ (res-value = -2,12) as well as „!H*“ with „new“ referents (res-value = -1,91) and „H*“ with „new“ (res-value = -1,23).

The occurrence of pitch accents on focused vs. non-focused material shows a similar relative frequency as information status: L*, !H*, H*+L and H* occurred comparably often on focused and non-focused constituents (Figure 23). Almost all unaccented material was not in focus (only 18 focus phrases were unaccented), the majority of rising L+H* accents was found on focused phrases (558 L+H* on focus (72,4%), 213 no focus (27,6%)). These results are significant (Chi-Square test: X-squared = 853,48, df = 5, p < 0,001). The correlation plot (Figure 24) reveals mirrored correlations for the „focus" vs. „no-focus“ conditions: the strongest correlations in both directions are found on unaccented material. „Focus“ has a strong negative correlation (res-value = -17,64) with unaccented words, „no-focus“ has a strong positive correlation (res-value = 19,58) with unaccented words. All remaining correlations for „focus“ are positive, for „no-focus“ they are negative. „L+H*“ shows the next strongest correlations with both focus conditions (res-value = 6,42 for Focus; res-value = -7,12 for „no-focus“), followed by „H*+L“ and „L*“, the weakest correlations are „H*“ and „!H*“.

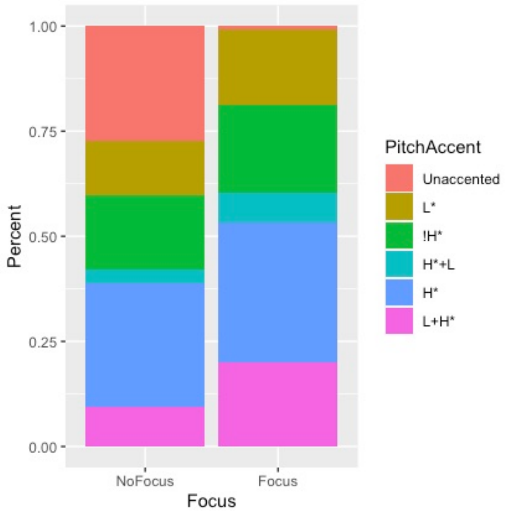


Figure 23: *Relative frequencies of pitch accents on focus presence.*

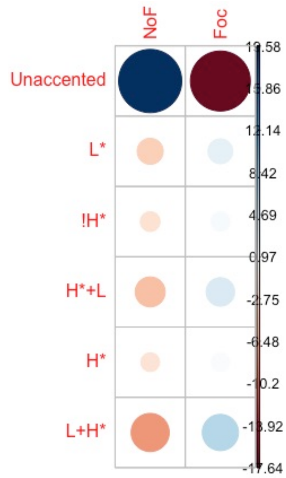


Figure 24: *Correlation plot on pitch accents and focus presence.*

Taking a deeper look into the pitch accent distribution on the different types of focus, „new-information“ focus and „contrastive“ focus, shows that the relative distribution of unaccented, L*, !H* and H*+L is similar. Relatively more H* accents are produced on contrastive focus (37,1% of all contrastive focus accents) than new-information focus (24,6% of all new-information focus accents), and more L+H* accents are produced on new-information (21,7%) than contrastive focus (11,4%), see Figure 25. Pearson’s Chi-Square test revealed that these results are significant (X-squared = 26,77, df = 5, p < 0,001), with a negative correlation between „contrastive" focus and „L+H*“ (res-value = -4,0) as well as „contrastive“ focus and the unaccented category (res-value = -1,13). All remaining correlations were positive. The strongest positive correlations are between „new-information“ focus and „L+H*“ (res-value = 1,71) and „!H*“ (res-value = 1,51) & „H*“ (res-value = 1,47) in the „contrastive“ focus condition, see Figure 26.

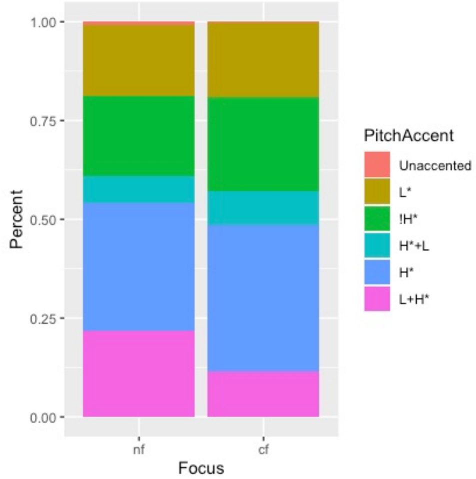


Figure 25: *Relative frequencies of pitch accents on focus categories.*

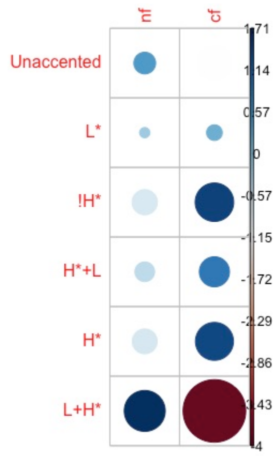


Figure 26: *Correlation plot on pitch accents and focus categories.*

4.1.2 Gestures and Information Structure

The results of gesture occurrence in connection to IS are illustrated in this section. As presented in the beginning of chapter 4, the occurrence of gestures in general (775 apices, 17%) and of gestures on IS material (306 apices, 6,7%) is a minority compared to the amount of IS referents in the corpus (4.551). The distribution of gesture occurrence on information status is given in Figure 27. The relative occurrence of gesture apices aligning with different information status levels increases with increasing newness of the referent: non-IS referents had the least apex alignment, „given“ and „accessible“ referents had more aligned apices, and „new“ referents showed the most apex-IS referent alignment. The overlap of strokes (but not the apex) with referents was similar for non-IS referents, „given“ and „new“, but higher for „accessible“ referents. Obviously, all apexes appearing on non-IS material appeared on non-IS referents.

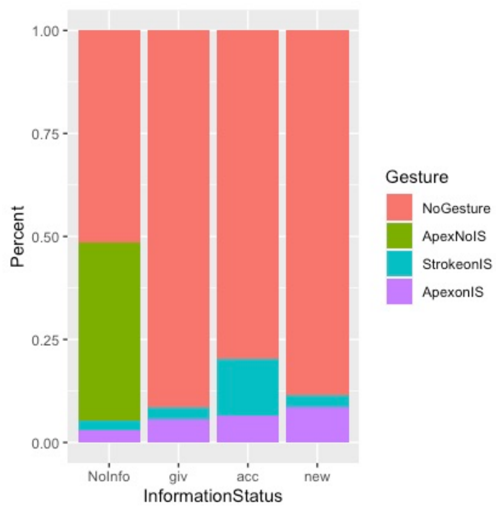


Figure 27: *Relative frequencies of gestures and parts on different levels of information status. Purple part of the bar describes the accurate alignment of apices and information status.*

Calculating a Pearson's Chi-Square test on these results indicated their significance (X-squared = 2078,3, df = 9, p < 0,001). Pearson's residuals and a correlation plot (Figure 28) revealed that the strongest positive correlation was between „ApexNoIS“ and „NoInfo“ (res-value = 36,54) and resulting negative correlations for „ApexNoIS“ with all information status levels (res-values: given = -11,9, accessible = -9,43, new = -11,6). Another strong positive correlation was observed between „StrokeonIS“ and „accessible“ (res-value = 12,5). „NoGesture“ showed positive correlations with „given“ (res-value = 5,11) and „new“ (res-value = 3,79). „ApexonIS“ had a negative correlation with „NoInfo“ (res-value = -4,1) and a positive correlation with „new“ (res-value = 3,83).

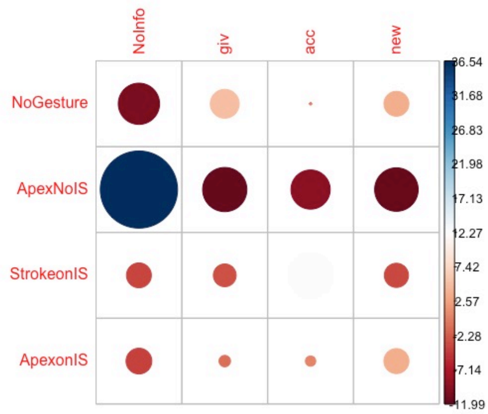


Figure 28: *Correlation plot on gestures and information status levels.*

The distribution of gestures on focus, „new-information“ focus and „contrastive“ focus shows similar results, and both differ from non-focus material. The relative occurrence of apices on focused phrases was higher (for new-information and contrastive: 203 occurrences) than on non-focus referents (103 occurrences), while the overlap of strokes (but displaced apices) was smaller for focused than for non-focused constituents. All apices on non-IS referents did appear on non-focus phrases. This results in 572 apices produced on non-focused material - adding up apices on non-focus referents, which are coded for information status and apices that occur without any IS referent. Figure 29 illustrates these results. Considering all gesture material on IS (thus apices and strokes) in total, non-focus constituents are accompanied more often by gestures than by focused constituents.

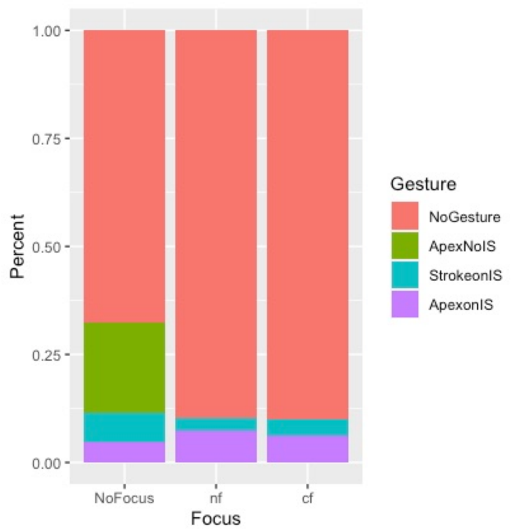


Figure 29: *Relative frequencies of gestures and parts in presence and categories of focus.*

The results are significant (Chi-Square test: X-squared = 710,34, df = 6, p < 0,001). Residuals (see the correlation plot in Figure 30) revealed a strong positive correlation between „ApexNoIS“ and „NoFocus“ (res-value = 17,86) and a positive correlation of „StrokeonIS“

with „NoFocus“ (res-value = 4,71). „ApexNoIS“ had negative correlations with „new-information“ focus (res-value = -14,8) and „contrastive“ focus (res-value = -6,32). Further positive correlations were found for „ApexonIS“ and „new-information“ focus (res-value = 2,78) and „NoGesture“ with „new-information“ focus (res-value = 5,32) and „contrastive“ focus (res-value = 2,37).

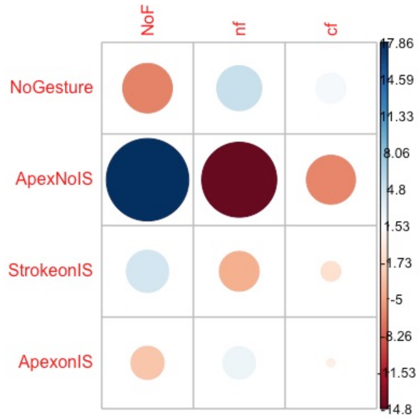


Figure 30: *Correlation plot on gestures and parts in presence and categories of focus.*

4.1.3 Pitch Accents and Gestures

The core comparison of this thesis, the synchronisation of pitch accents and gestures, is presented in this section. Pitch accents (4.394, 87,5% of all data points) occurred more often than gestures (775 apices, 15,4% of all data points). The distribution of the different gesture options accompanying pitch accent types is given in Table 5. It can be observed that all pitch accent types were most often not accompanied by a gesture and when no gesture was produced, the referent was accented in the majority of cases. Stroke overlapping (therefore a displaced apex) was most often found with unaccented words or H* accents. The apex was most often not accompanied by an accent, or by a H*. All pitch accent types aligned more often with an apex than overlapped only with the stroke.

	Unaccented	L*	!H*	H*+L	H*	L+H*	Total Gestures
No Gesture	29	733	877	242	1.422	703	4.006
Stroke overlaps	111	20	30	12	51	19	243
Apex	491 (469 no IS)	46	63	20	106	49	775
Total accents	631	799	970	274	1.579	771	

Table 5: *Distribution of gestures and gesture parts accompanying different pitch accents.*

Figure 31 illustrates the relative frequency of pitch accents on the different gesture parts.

A Pearson’s Chi-Square test showed that these results are significant (X-squared = 2577,5, df = 10, p < 0,001). Calculating Pearson’s residuals indicated a strong positive correlation between an apex and a missing accent (res-value = 39,9). Another positive correlation can be found between overlapping strokes and unaccented referents (res-value = 14,57), while the detected correlation between „NoGesture" and unaccented words is strongly negative (res-value = -21,14). Weaker negative correlations are found between „Apex" and „Stroke" and all pitch accent types, weak positive correlations between „NoGesture" and all pitch accent types (residuals in Appendix C). Figure 32 illustrates these correlations.

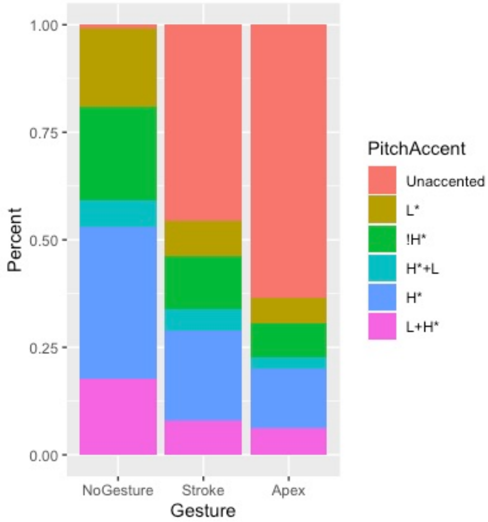


Figure 31: *Relative frequencies of pitch accents on gestures.*

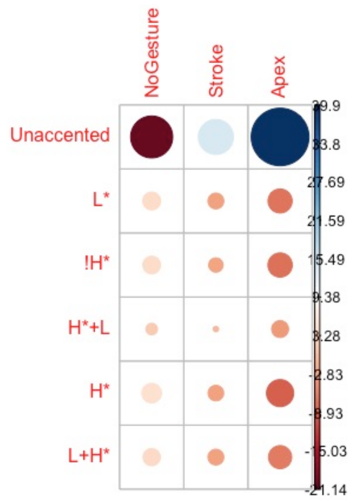


Figure 32: *Correlation plot on pitch accents and gestures.*

4.1.4 Interim Summary

The results of this section show significant correlations of different levels of each factor. For pitch accents and information status, the analysis showed that unaccented words tend to be „accessible" and „new" referents are often accompanied by L+H* accents, while „given" referents avoid L+H* accents. The results for pitch accents and focus suggest that non-focused material is rather unaccented than focused constituents, while focus phrases combine with a pitch accent type, thus show a mirrored distribution. The relevant distinction between new-information focus and contrastive focus is that contrastive focus rarely aligns with L+H*, but otherwise the pitch accent distribution is similar.

Dependencies of non-referential gestures and information status are that stroke overlaps are most frequent with „accessible" referents, and the apex tended to align best with „new" referents of all information status levels. However, „new" and „given" also tended to not align with a non-referential gesture at all. Non-focused constituents aligned best with no-IS apices or with overlapping strokes. Focus correlated positively with the avoidance of gestures, but the alignment of apices on IS referents was best on new-information focus. In general, it was significant that it was avoided for gestures to align with any IS referent in the majority of apex occurrences.

Regarding correlations of gestures and pitch accents, it appears that apices align with unaccented words most often, as well as overlapping strokes. Similarly, „NoGesture“ had a positive correlation with all pitch accent types. This shows in general a similar tendency of gestures as with IS, that non-referential gestures tend to not align with pitch accents.

4.2 Three-way Interaction of the Variable Distribution

After the distribution of every factor individually and the interaction of two factors each have been explored, in this section, three-way interactions are investigated. In order to do so, for one factor C, the distribution analysis of the two remaining factors A and B was done following the procedure in section 4.1 for each level of factor C. The distributions are then compared to each other. First, the data is reviewed with the prerequisite that an apex is produced, looking at the interactions of pitch accents and IS. Afterwards, both IS factors are used as underlying components and pitch and gestures are compared under different IS levels.

4.2.1 Gestures

First, the data are evaluated with the prerequisite that an apex is produced. The distribution of pitch accents accompanying apices given in Figures 33a and 33b can be compared to the pitch accent distribution across all data in Figure 18. In Figure 33a, unaccented words represent the majority of referents accompanied by an apex, which include non-IS referents. However, through the different pitch accent types, H* accents are the most frequent, followed by !H*, L* and L+H* in this order. In Figure 33b, the distribution across all data with an apex on IS material, resembles the same distribution of the pitch accent types, the only identifiable difference lies in the relative decrease of unaccented words (14,4% across all data vs. 7,2% with apex on IS).

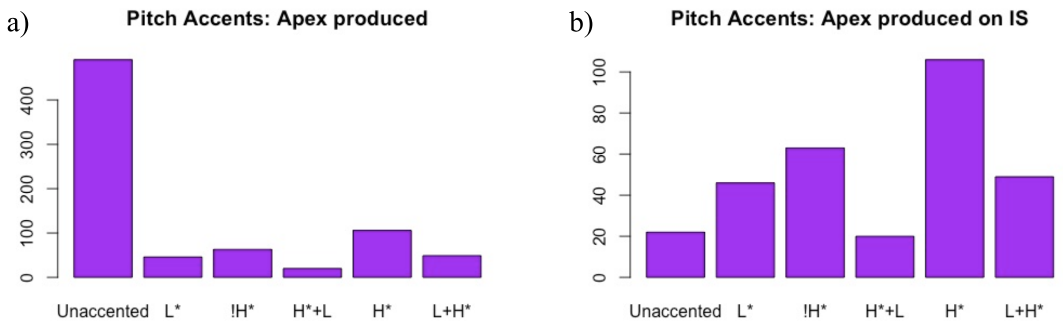


Figure 33: *Distribution of pitch accents under the prerequisite that an apex is produced. Left: apices in total, right: apex on IS referents.*

With regard to the distribution of information status levels on words with an apex, it appears that the relative distribution of „given" and „new" shifts. From a slightly higher amount of „given" referents in the general distribution (39,1% „given" and 36,7% „new", recall Figure 19 and repeated in 34a), to a higher frequency of „new" referents in the apex prerequisite conditions (28,1% „given" and 40,5% „new", Figure 34b).

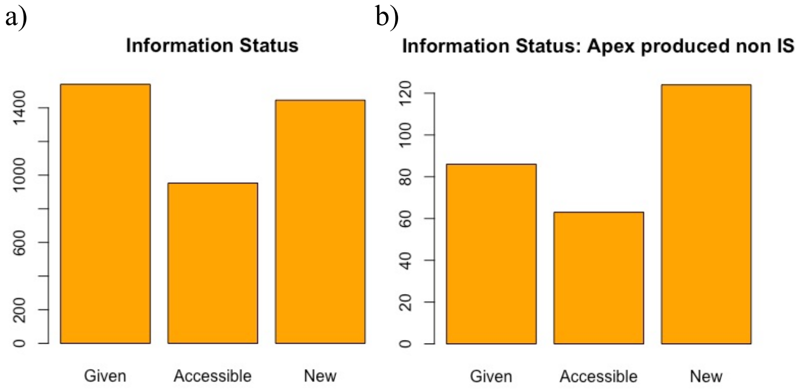


Figure 34: *Distribution of information status levels under the prerequisite that an apex is produced. Left: overall distribution, right: apex on IS referents.*

The relative frequencies of focus between „new-information" focus and „contrastive" focus did not change in relation to the presence of an apex: the distribution shown in Figure 20b does not change. What changes is the distribution in the general distinction between Focus vs. non-Focus material, see Figure 35. While in the general focus distribution (Figure 35a), there was a small majority of focus (55,2%) over No Focus (44,8%), in the condition where an apex is produced generally, non-focused constituents were more frequent (73,8%) than focused referents (26,2%), see Figure 35b. In turn, as visible in Figure 35c, when only considering the apex occurrences on IS referents, focused phrases (66,3%) predominate non-focused material (33,7%), more than in the general focus distribution.

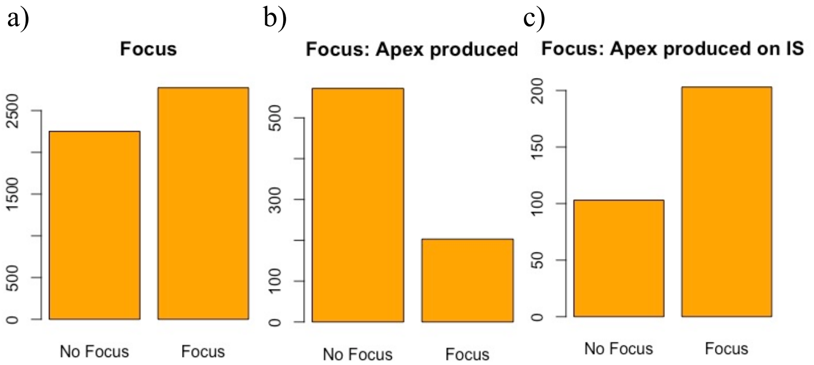


Figure 35: *Distribution of focus given that an apex is produced. Left: overall distribution, center: focus with apices in general, right: apex on IS referents.*

Regarding the comparisons of two factors given that an apex is produced, pitch accents are compared with information status and focus. A difference between the general distribution of pitch accents and the distribution under apex production (on the information status levels) is observed. In the latter case, the production of L+H* increases on „accessible" referents, leading to their relative frequency being arranged between the L+H* accents on „given" and

„new“ (see Figure 36). In addition, the amount of L* accents increases on „given“ referents, and „accessible“ referents are more often unaccented than „given“ and „new“. A Chi-Square test indicated that these results are significant (X-squared = 62,708, df = 10, p < 0,001) with residuals (Figure 37) showing positive correlations between „accessible“ referents and „Unaccented“ (res-value = 6,18) as well as L* with „given“ (res-value = 1,97). Negative correlations were between „Unaccented“ and „given“ (res-value = -2,25) and „new“ (res-value = -2,53). Pitch accent distribution on focused vs. non-focused phrases with an apex shows an approximation of the distribution on focus presence and absence. Less non-focused constituents are unaccented and less focused constituents have L+H* accents, as illustrated in Figure 38. Calculating a Chi-Square test shows that these results are significant (X-squared = 38,738, df = 5, p < 0,001). Residuals revealed correlations of „unaccented“ with both categories: positive with „no-focus“ (res-value = 4,63) and negative with „focus“ (res-value = -3,3). L* and H*+L have positive correlations with „focus“ (res-value = 0,81 & 1,02) and negative correlations with „no-focus“ (res-values = -1,14 & -1,43), see Figure 39.

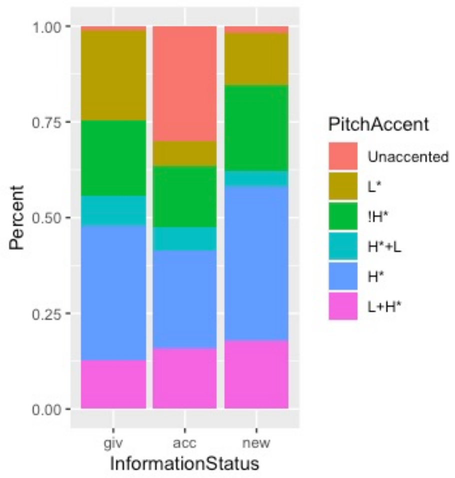


Figure 36: *Relative frequencies of pitch accents on information status when an apex is produced on the referent.*

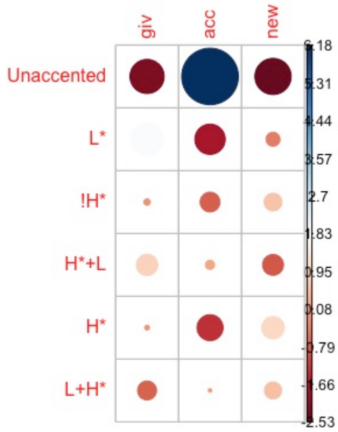


Figure 37: *Correlation plot on pitch accents and information status - apex produced on the referent.*

A Chi-Square test was also calculated for the pitch accent distribution on the two focus categories (new-information & contrastive) with a produced apex, but results were not significant (X-squared = 3,3, df = 5, p = 0,653).

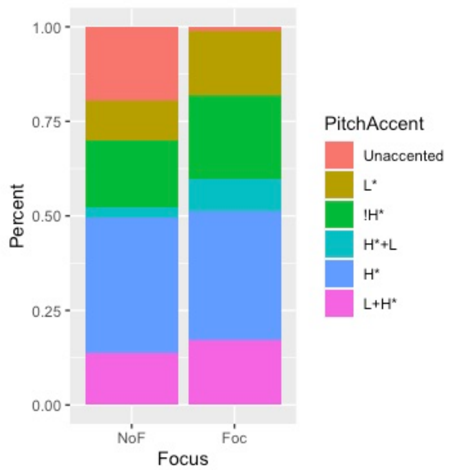


Figure 38: *Relative distribution of pitch accents and focus presence when an apex is produced on the referent.*

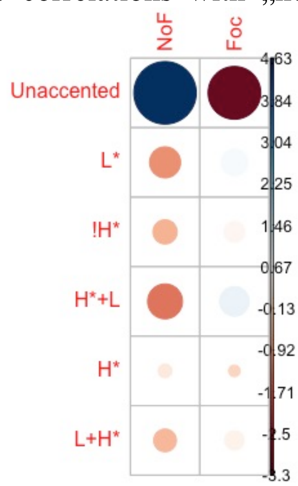


Figure 39: *Correlation plot of pitch accents and focus when an apex is produced on the referent.*

4.2.2 Information Status

This section explores the variation of distribution of pitch accents and gestures under the different levels of information status. The distributions partly illustrate observations that were made in the previous sections more in detail. Figure 40 shows the differences in distribution of pitch accents in general (Figure 18) and given a specific information status.

On the „given“ and „new“ level (Figures 40a & 40c respectively), less unaccented words are found compared to the general distribution. On the „given“ and „accessible“ level (Figures 40a & 40b), less L+H* accents are present, and more L+H* accents are found on the „new“ level (40c). More !H* accents are found on the „accessible“ level than on the two remaining levels (40b).

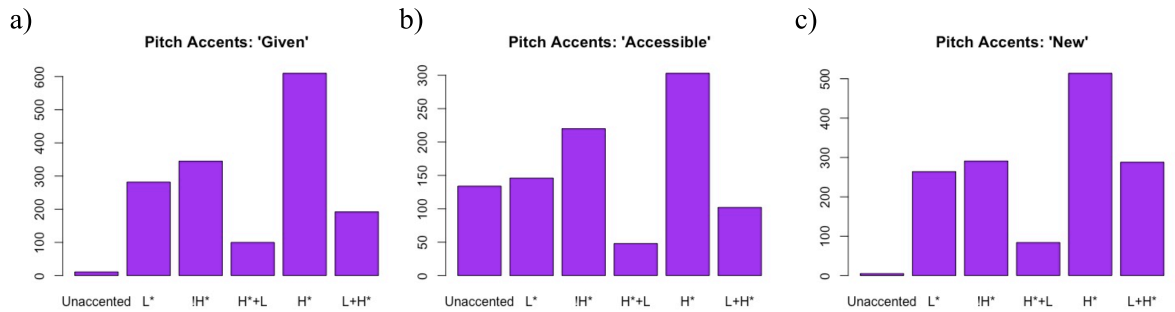


Figure 40: *Distribution of pitch accents when a specific information status is assigned. Left: „given“ referents, center: „accessible“ referents, right: „new“ referents.*

The distribution of gesture components (apices and strokes) that accompanied IS referents showed a strong variation. While in the general distribution, the amount of apices and strokes on IS referents was similar (306 apices / 55,7%, 243 strokes; see Figure 41a), the distribution differed more strongly for each information status level. With „given“ referents (Figure 41b), more apices than strokes aligned (86 apices / 64,6%, 47 strokes). For „new“ referents, the percentage of apices aligning was even higher than for „given“ referents (124 apices / 75,6%, 40 strokes; Figure 41d). „Accessible“ referents had a lot more overlap with strokes than alignment with apices (63 apices / 32,4%, 131 strokes; Figure 41c).

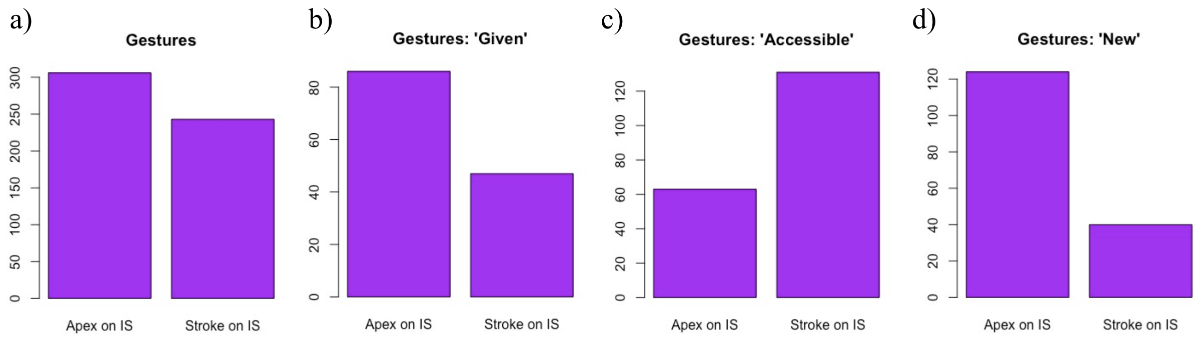


Figure 41: *Distribution of gestures under the prerequisite that a specific information status is assigned. Left: overall distribution, center left: „given“ referents, center right: „accessible“ referents, right: „new“ referents.*

Comparing the relations of pitch accents and gestures under each information status level illustrates the variation in the prosody gesture alignment in relation to prominence. The amount of aligning apices and overlapping strokes on each pitch accent type varies for the different information status levels. Figure 42 shows this variation.

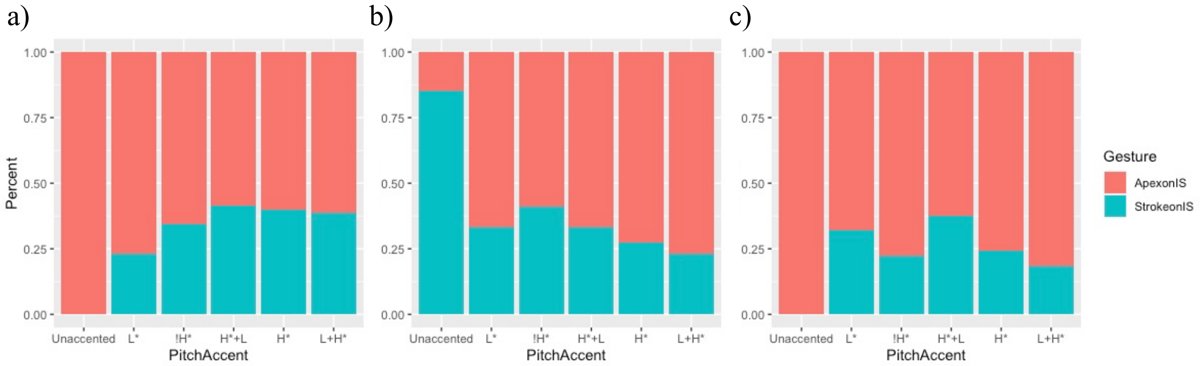


Figure 42: *Relative frequencies of pitch accents and gestures when a specific information status is assigned. Left: „given" referents, center: „accessible" referents, right: „new" referents.*

On „given" referents as in 42a, apices align best with unaccented words or L* accents (red part of the bars) and all accents higher on the Prominence Scale align worse. Regarding the distribution on „accessible" referents (Figure 42b), apices only rarely align with unaccented words and apex alignment increased (with the exception of the step from L* to !H*) with every more prominent accent. The distribution on „new" referents (Figure 42c) is similar to „accessible" referents: The apex alignment increases with more prominent accents, apart from the observation that unaccented words always align with an apex and never only with the stroke, and the alignment decreases on the step from !H* to H*+L.

Chi-Square tests for the distributions on each information status level indicate that the results for „given" and „new" are not significant („given": X-squared = 3,05, df = 5, p = 0,69 & „new": X-squared = 2,77, df = 5, p = 0,73). The results for „accessible" referents are significant (X-squared = 58,66, df = 5, p < 0,001) and showed a negative correlation between „unaccented" and apices (res-value = -3,57) and a positive correlation between „unaccented" and overlapping stroke (res-value = 2,48). All correlations of apices and pitch accent types were weak positive and all correlations of stroke and pitch accent types were weak negative.

4.2.3 Focus

The impact of focus on the prosody gesture alignment is explored in this section. Since the previous sections showed no big difference between new-information focus and contrastive focus, in this section the distinction between focus presence and absence is reviewed.

The main difference in the pitch accent distribution between the general distribution (631 unaccented / 12,6%, 771 L+H* / 15,3%; see Figure 18) and accents on focus and non-focused phrases is that non-focused constituents contain almost all unaccented words and fewer L+H* pitch accents (612 unaccented / 27,2%, 213 L+H* / 9,5%; see Figure 43a), while focused material is barely unaccented and shows relatively more L+H* accents (19 unaccented / 0,7%, 558 L+H* / 20,1%; see Figure 43b) than the general distribution.

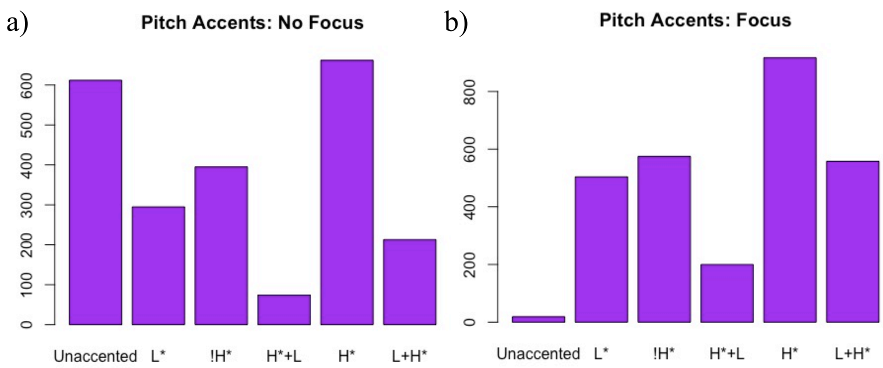


Figure 43: *Distribution of pitch accents in absence and presence of focus. Left: unfocused phrase, right: focus phrase.*

The distribution of apices and overlapping strokes also differs under the focus conditions. In the general distribution, a similar amount of apices align as apices where slightly displaced leading to overlapping strokes (306 apices / 55,7%, 243 strokes; see Figure 44a). On non-focused constituents, more often strokes overlapped than apices aligned (103 apices / 39,5%, 158 strokes; see Figure 44b). In the focused phrases, more apices align than strokes overlap (203 apices / 70,5%, 85 strokes; see Figure 44c) and the difference is bigger than in the general condition.

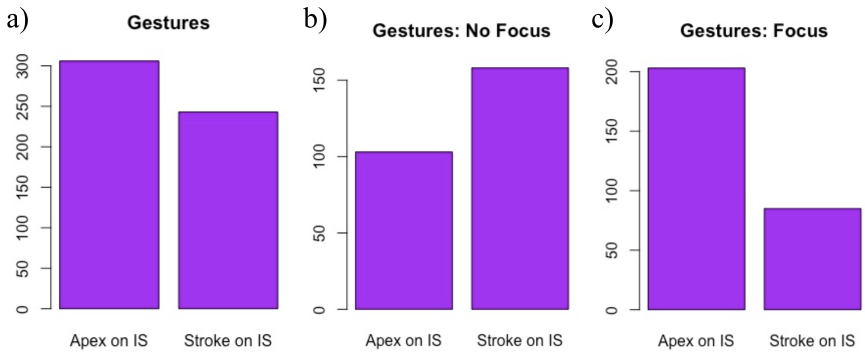


Figure 44: *Distribution of gestures in absence and presence of focus. Left: overall distribution, center: unfocused phrase, right: focus phrase.*

Comparing the relative frequencies of pitch accents and gestures on focused vs. non-focused constituents shows a rather stable alignment of apices and overlapping of strokes for all pitch accents types in the focused condition (Figure 45b). The apex alignment increases slightly the more prominent an accent is, but the difference is very small. A Chi-Square test indicated that these results are not significant (X-squared = 0,24, df = 5, p = 0,99). The distribution of pitch accents and gestures on non-focused referents shows a different picture. The alignment of apices and gestures differed for every pitch accent type. Apices aligned the least frequent with unaccented words and H*+L accents, more often with !H* and H* accents and best with L* and L+H* accents, see Figure 45a. Calculating a Chi-Square test showed that these results are significant (X-squared = 66,124, df = 5, p < 0,001), with the strongest correlations for unaccented words. A negative correlation holds between apices and „unaccented“ (res-value = -4,37) and a positive correlation between strokes and „unaccented“ (res-value = 3,53). Apart from H*+L (where no correlation existed), all pitch accent types showed weak positive correlations with apices and weak negative correlations with overlapping strokes.

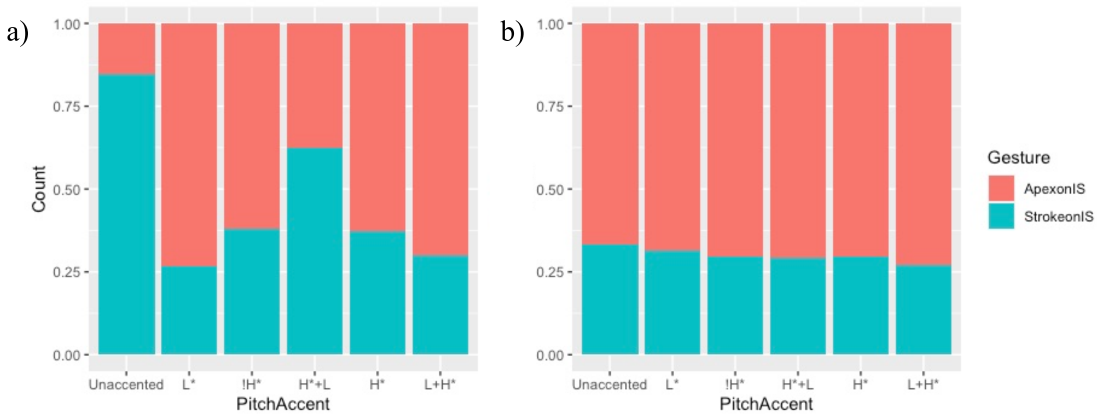


Figure 45: *Relative frequencies of pitch accents and gestures in absence and presence of focus. Left: unfocused phrase, right: focus phrase.*

4.3 Presence and Absence of Accents and Apices

The previous sections have given insight into the alignment of pitch accents and gesture parts on the lexical word level and the impact that IS, specifically information status and focus, have on this alignment. In general, IS showed an influence on pitch accents according to their assigned prominence, especially for L+H* accents being more frequent with more „new“ and „focused“ material and L* accents and unaccented are more frequent with „given“ or „accessible“ material and „non-focused“ constituents. The distribution of gesture parts (aligned apices or displaced apices leading to overlapping strokes) was also found to be influenced by IS in the way that apex alignment was strongest with „new“ and „focused“ referents. „Given“ referents also showed more apex alignment than stroke overlaps, but with a smaller difference. „Non-focused“ constituents and „accessible“ referents primarily showed stroke overlaps. IS also influenced the alignment of pitch accent types and gestures, but only significantly for the „accessible“ information status level, in a way that apices aligned better with increasing prominence of the pitch accents. Non-focused constituents also significantly influenced the pitch accent-gesture alignment, with a more mixed picture. The apex alignment of focused material was very stable.

Overall, it can be observed that the use of gestures was less frequent than the use of pitch accents, such that in the majority of cases pitch accents marked IS referents, but no non-referential gestures were used. When gestures were used, in less than half of the cases, they aligned with pitch accents let alone IS referents. The total amount of data points for each factor are given in Table 6 and the numbers of each possible combination of the occurrence of pitch accents or apices are provided in Table 7. Overlapping strokes are not included in this part of the analysis, since the indication of an overlapping stroke means that a) the associated apex does not align with the referent and b) an annotated stroke means that there is an apex.

Apices	Pitch accents	Information Status	Focus
775	4.394	3.939	2.773

Table 6: *Occurrences of each factor in the corpus.*

Information status is assembled from 1.540 „given“ referents, 953 „accessible“ referents and 1.446 „new“ referents.

pitch accent, apex	pitch accent, No apex	No pitch accent, apex	No pitch accent, No apex
284	4.109	491	140

Table 7: *Occurrences of pitch accents and apices on their own, together and absence of both.*

IS phrases that were not accompanied by either a pitch accent or an apex were the least frequent phrases. As mentioned, fewer apices occurred in the presence of a pitch accent (36,6%) than without a pitch accent (63,4%), thus on unaccented IS referents, non-IS referents or during speech breaks. Most frequently, when a pitch accent was used on an IS referent, it was not accompanied by a gestural apex (93,5%).

Regarding the changes under the different levels of IS, it appears that the presence of both pitch accents and apices is most frequent in the most prominent categories, while apices without an accent or none of the two factors most likely appeared with less prominent material. For information status it is visible that the alignment of pitch accents and apices increases from „no information status“ (33, 3%) to „given“ (85, 5,5%) and „accessible“ (44, 4,6%) to „new“ (122, 8,4%). For „new“, almost all remaining referents were accompanied by a pitch accent, but no gesture (1.319, 91,2%). Two referents had an apex but no gesture, and 3 referents had none of both. Most of the apices without pitch accents were found on „no information status“ referents (469, 43,2%), only one on „given“ and 19 (2%) on „accessible“. Most constituents bearing no apex and no gesture were found on „no information status“ (12, 1,1%) or „accessible“ (115, 12,1%) referents. These results are illustrated in Figure 46a. Focus shows a similar picture in that focus constituents have more aligned apices and pitch accents (201, 7,2%) than non-focused phrases (83, 3,7%). The majority of apices occurring without a pitch accent were produced on non-focused constituents (489, 21,7% of all non-focused material & 99,6% of all non-IS apices), as were phrases that are not accompanied by a pitch accent or an apex (123, 5,5%), shown in Figure 46b.

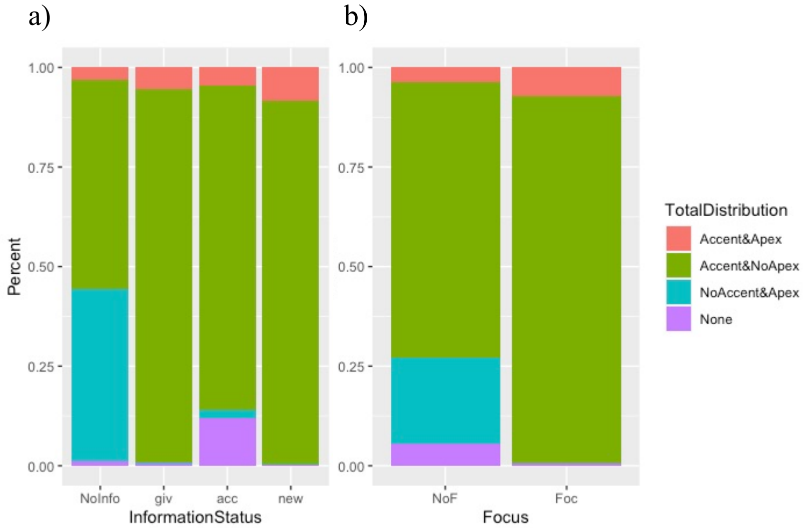


Figure 46: *Occurrences of pitch accents, gestures and their joint occurrence. Left: under specified information status levels, right: under focus absence and presence.*

The significance of these results was calculated using Pearson’s Chi-Square tests. The results for information status are significant (X-squared = 2142,9, df = 9, p < 0,001) and residuals were calculated (correlation plot in

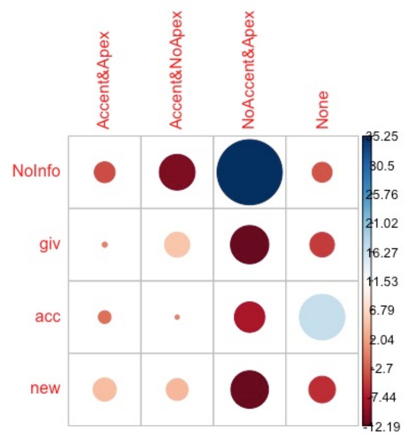


Figure 47: *Correlation plot of the occurrences of pitch accents and gestures with different information status levels.*

Figure 47). The strongest positive correlation was found between „NoAccent & Apex“ and „No information status“ (res-values = 35,25) and between „accessible“ and „NoAccent & NoApex“ (res-value = 17,16). A weaker positive correlation was found between „new“ referents and „Accent & Apex“ (res-value = 4,45) as well as positive correlations for „Accent & NoApex“ with „given“ and „new“ (res-values = 5,2 & 4). Negative correlations were found for „NoAccent & Apex“ for all information status categories (res-values: „given“ = -12,19, „accessible“ = -7,68, „new“ = -11,72), as well as „Accent & NoApex“ and „no information status“ (res-value = -10,62).

The results for Focus were significant as well (X-squared = 808,72, df = 3, p < 0,001) and correlations were explored by calculating residuals (correlation plot in Figure 48). The strongest correlations were found in the „NoAccent & Apex“ category, which was positive with „no-focus“ phrases (res-value = 18,14) and negative with „focus“ phrases (res-value = -16,34). „Focus“ had positive correlations with both categories containing pitch accents (res-values = 3,53 & 5,99) and a negative correlation with „NoAccent & NoApex“ (res-value = -6,86). „No-focus“ shows mirrored correlations, negative with the pitch accent categories (res-values = -3,92 & -6,64) and a positive correlation with referents without pitch accents and apices (res-value = 7,61).

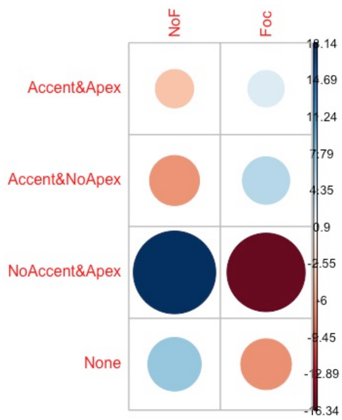


Figure 48: *Correlation plot of the occurrences of pitch accents and gestures with focus presence or absence.*

While it was found that the presence of non-referential gestures was not very high in general, and even smaller on IS referents and accompanying a pitch accent on the same word, the concrete temporal alignment of the intersection of these events is subject to the investigation in the following chapter.

5. Temporal Synchronisation Analysis

5.1 General Synchronisation

In the 775 data points available for the temporal synchronisation analysis, five pitch accents aligned exactly with an apex (0,6%). 422 pitch accents (54,5%) followed their nearest apex, 348 preceded it (44,9%). 28 pitch accents, 3,6%, occurred within one frame (+/- 16,5 ms apart from 0) of the apex. 518 pitch accents preceded or followed the apex within one second /30 frames (66,8%). The distance between the two modalities ranged from the pitch accent preceding the apex 12,2 sec to the pitch accent following the apex 7,9 seconds afterwards. In total, 12 accents were more than 5000 ms apart from the apex, 9 preceding the apex, 3 following and 44 accents had a distance of more than 3000 ms to the apex.

Figure 49 shows the deviation of pitch accents from the apices in a histogram in milliseconds, the apex being produced at 0 ms. The curve has its peak at 0 ms, and the frequencies of each deviation drop with greater distance value. Data points with a deviation of more than 3000 ms were not included in the diagram as well as in the following analysis. This led to the deletion of 44 datapoints. This is justified by the observation that there is no example in the dataset where a word is 3000 ms long. Even though the semantic connection is not considered in this analysis, a distance of 3000 ms or more makes it impossible in this dataset that the apex and the accent are produced on the same word (longest single word had a duration of 2.169 ms and a word interval was 270 ms on average). It is even unlikely that they occur in the same phrase. The resulting dataset for this analysis consists of 731 apices. In the whole temporal synchronisation analysis, negative values show the deviation of pitch accents preceding the apex, positive values show the deviation for pitch accents following the apex.

The mean deviation of this general distribution is μ =

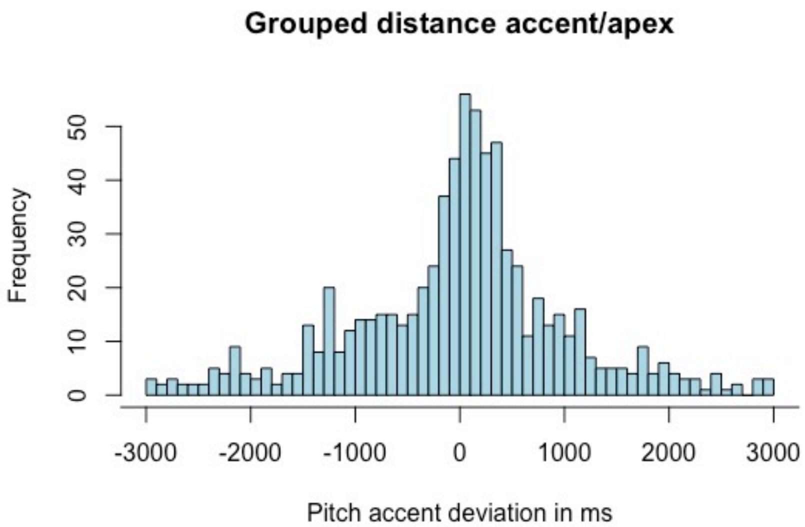


Figure 49: *Histogram on the overall deviation between apices and their nearest pitch accents. Deviations are grouped into deviation of one frame. X-axis: Pitch accent deviation (negative values mean that the pitch accent precedes the apex), Y-axis: frequency of occurrence of the deviation in the corpus in ms.*

38,63 ms, thus on average the nearest pitch accent appears 38,63 ms behind the apex. The standard deviation is σ = 385 ms. The graph in Figure 49 shows a distribution that is not shifted to the left or right but centered. The histogram in Figure 49, as well as the qq-plot that was plotted for these data in Figure 50 show an approximation to a normal distribution in the first theoretical quantiles. In the second and third quantiles, the distribution diverges from a normal distribution in both the positive and the negative direction.

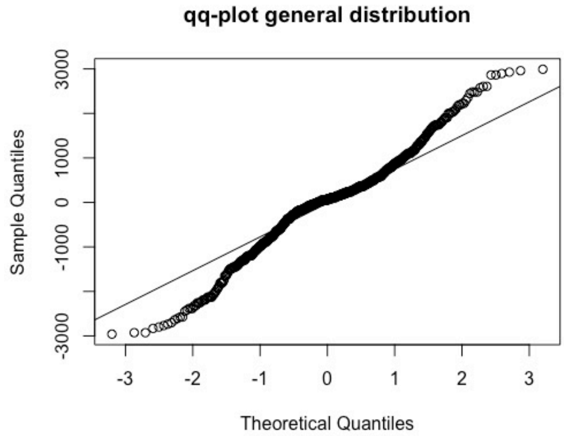


Figure 50: *qq-plot for the general deviation of pitch accents from apices to visually test the normal distribution of the data set. The line indicates a normal distribution, the dots are the datapoints from this data set in ascending order.*

A one sample t-test was significant (t = 2,71, df = 730, p < 0,01) for the general distance between apices of non-referential gestures and pitch accents in this spontaneous speech dataset. In the following sections, the variation of the deviation under the different IS parameters is explored.

5.2 Information Status

As has been explored in the previous chapter, 306 out of 775 apices were produced on a referent coded for information status. After deletion of the 44 datapoints with more than 3000 ms deviation, 272 apices on IS material remain. From these 272 referents, 123 (45,2%) were labelled as „new", 63 (23,2%) as „accessible" and 86 (31,6%) were labelled as „given" referents.

Given: Starting with „given" referents, it can be observed that a majority of accents are found within 500 ms before or behind their connected apex. 59 accents (68,6%) followed the apex, 26 (30,2%) preceded it. One apex and pitch accent aligned perfectly. Four accents (4,7%) had a distance to the apex of 720 to 820 ms and all of them followed the apex. The

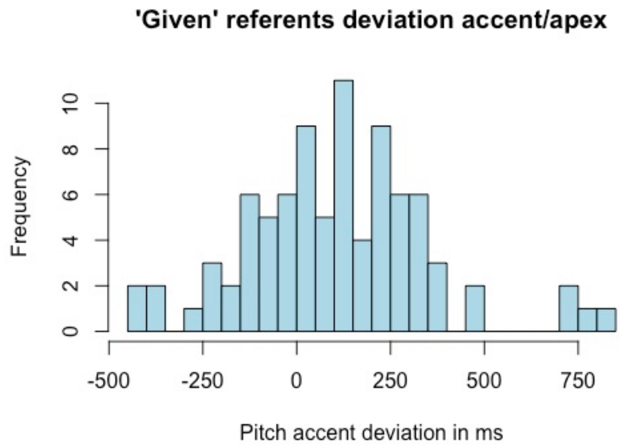


Figure 51: *Histogram on the deviation between apices and their nearest pitch accents on „given" referents in ms.*

histogram in Figure 51 shows a tendency of pitch accents to appear behind apices on given referents, also indicated by the mean deviation of $\mu_g = 103,93$ ms and standard deviation of $\sigma_g = 239$ ms. The mean deviation in this subset is higher than in the general distribution, the standard deviation is smaller. The qq-plot (Figure 52) shows an approximation to a normal distribution that shows few divergence, mainly in the positive direction of the second theoretical quantile. The one sample t-test indicates that these results are significant ($t = 4,02$, $df = 85$, $p < 0,01$).

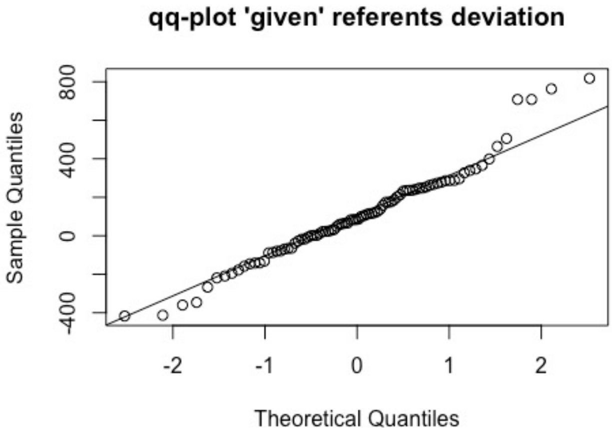


Figure 52: *qq-plot for the deviation of pitch accents from apices on „given" referents to visually test the normal distribution of the data set.*

Accessible: Regarding apices on „accessible" referents, the span of pitch accent distance to the apex is bigger than in the „given" condition, ranging between -2750 and 2880 ms. 17 accents (27%) have a deviation of more than 500 ms. More accents followed the apex (38, 60,3%) than preceded it (23, 36,5%). Two apices and pitch accents aligned completely. The histogram (Figure 53) shows a peak of distance slightly behind the apex, but no monotone slope is identifiable. The average deviation is $\mu_a = -19,07$ ms, thus a pitch accent precedes the apex on average and the standard deviation is $\sigma_a = 375,91$ ms. The qq-plot in Figure 54 shows divergence from normal distribution outside the first theoretical quantiles, especially in the negative direction. A one sample t-test indicated that these results are not significant ($t = -0,403$, $df = 62$, $p = 0,6886$).

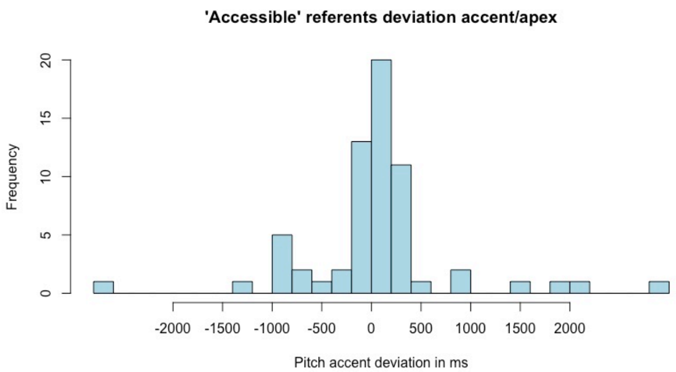


Figure 53: *Histogram on the deviation between apices and their nearest pitch accents on „accessible" referents in ms.*

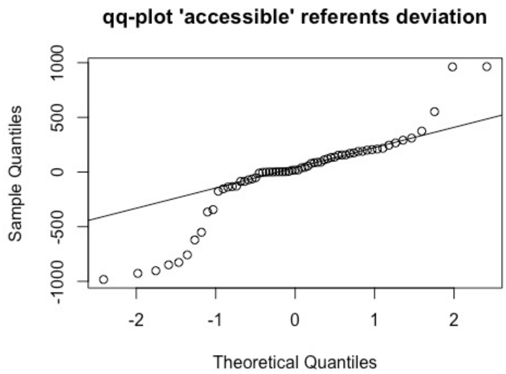


Figure 54: *qq-plot for the deviation of pitch accents from apices on „accessible" referents to visually test the normal distribution of the data set.*

New: Apices on „new“ referents had pitch accents preceding and following the apex within one second, between -950 and 800 ms, of which ten accents (8,1%) had a deviation of more than 500 ms. Most frequently, deviations were between -250 and 500 ms with more pitch accents following the apex (76 accents, 61,8%) than preceding it (46 accents, 37,4%) and the peak is around the 0 ms central point. One apex and pitch accent aligned perfectly. The mean deviation for this subset was μ_n = 70,2439 ms and the standard deviation amounted σ_n = 259,59 ms. The histogram for apices on „new“ referents (Figure 55) shows this distribution. A qq-plot (Figure 56) illustrates a high approximation to a normal distribution with only little divergence in the second theoretical quantiles. A one sample t-test indicated that the results for the „new“ subset are significant ($t = 3{,}001$, $p < 0{,}01$).

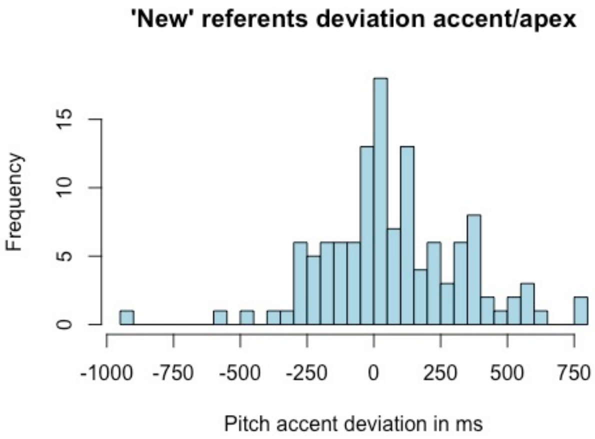


Figure 55: *Histogram on the deviation between apices and their nearest pitch accents on „new“ referents in ms.*

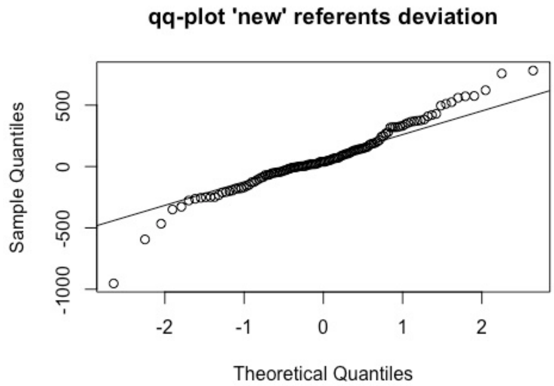


Figure 56: *qq-plot for the deviation of pitch accents from apices on „new“ referents to visually test the normal distribution of the data set.*

No information status: The deviation of constituents that are not coded for information status is explored as well. 459 data points are subject to this analysis after deleting all apices whose nearest accent was more than 3000 ms away. It can be observed that fewer pitch accents align very closely with the apex than are displaced before or behind the apex. The distribution (histogram in Figure 57) has one peak around 500 ms (behind the apex) and forms a plateau of similarly frequent occurring deviations around -1200 ms to 200 ms.

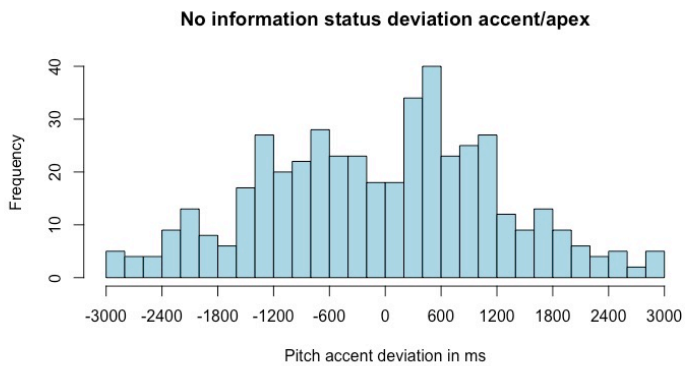


Figure 57: *Histogram on the deviation between apices and their nearest pitch accents on referents without information status in ms.*

This subset of data also contains the

most extreme deviations (-3000 and 3000 ms) and has almost the same amount of accents preceding (226 accents, 49,2%) and following the apex (232 accents, 50,4%). One apex and pitch accent aligned perfectly. 343 accents (74,7%) showed a deviation of more than 500 ms. The mean deviation was μ_{noi} = 25,84 ms, the standard deviation σ_{noi} = 432 ms. The histogram (Figure 57) as well as the qq-plot in Figure 58 do not suggest an approximation to a normal distribution across all theoretical quantiles. Almost all values diverge from the line representing a normal distribution. The performed one sample t-test indicates that these results are not significant (t = 1,28, p = 0,2008).

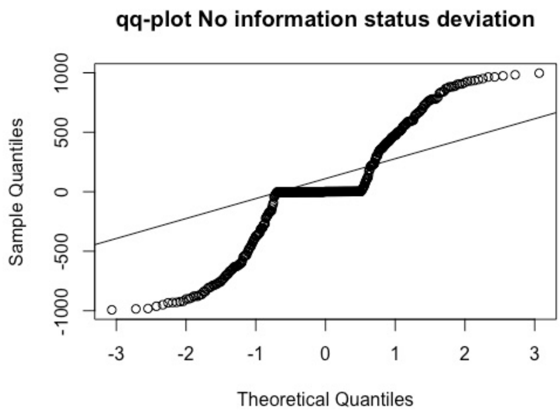


Figure 58: *qq-plot for the deviation of pitch accents from apices on referents without information status to visually test the normal distribution of the data set.*

Summary: To sum up this section, it appears that constituents coded for information status show a reduced deviation of pitch accents and prefer apices and pitch accents with a small distance. An overview of mean and standard deviations and peak frequencies of pitch accent deviations is given in Table 8.

	Given	Accessible	New	No Information Status	General Distribution
Mean	103,93 ms	-19,07 ms	70,2439 ms	25,84 ms	38,63 ms
Distance from general mean	65,3 ms	-57,7 ms	31,61 ms	-12,79 ms	0 ms
Standard Deviation	239 ms	375,91 ms	259,59 ms	432 ms	385 ms
Peak	100 - 150 ms	0 - 100 ms	0 ms	- 600 & 1000 ms	0 ms
p-value t-test	< 0,01	0,6886	< 0,01	0,2008	< 0,01

Table 8: *Summary of means and deviations under different levels of information status.*

Non-information status referents not only show a greater range of deviation, they also have a distribution of deviation frequencies that peaks circa half a second before and behind the apex. Apices on „given“ referents show the smallest deviation of pitch accents, followed by apices on „new“ referents and lastly apices on „accessible“ referents, whose deviation is still smaller than for non-information status apices. Apices on „given“ referents also show the biggest mean deviation shift of all apices, pitch accent being placed about 100 ms behind the apex. Apices on „new“ referents and non-information status apices show a small shift of pitch accents behind the apex. Apices on „accessible“ referents were the only apices showing a tendency of pitch accents preceding the apex considering their mean deviation. In „given“ referents, the mean is shifted the most compared to the general mean, followed by „accessible“, the „new“ mean shifts less and „No info“ has almost the same mean deviation. „Accessible“ and „No info“ referents show a stronger divergence from a normal distribution of the data than „given“ and „new“ referents do.

The box plots shown in Figure 59 illustrate and complement to these observations. Most obviously, the „No info“ category differs from the levels of information status in showing the highest amount of data points being outside the 95% quantiles (whiskers) of the data set. Overall, the means of all categories are similar to each other, close to 0 ms, with the „accessible“ median being highest, but with the 95% quantiles being more towards the negative (accents preceding) values. The „No info“ mean is the lowest. „Given“ referents show the smallest variance of data points, with the 95% quantiles being closest to each other. A one-factorial ANOVA was conducted exploratively keeping in mind that these data are not normally distributed. It indicated that these results are not significant ($F(3, 727) = 1,747$, $p = 0,156$).

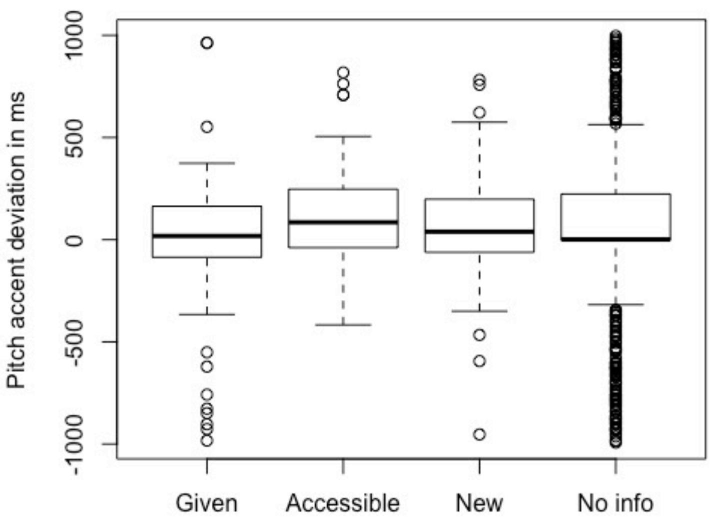


Figure 59: *Boxplots of the differing deviations of pitch accents to apices under the different levels of information status.*

5.3 Focus

Since the distribution analysis did not reveal a big difference between the two focus types „new-information“ focus and „contrastive“ focus, in this section the deviation difference is analysed distinguishing only focused constituents from non-focused constituents. Out of the 731 apices, 202 apices (27,6%) were produced during a focused constituent. 529 apices

(72,4%) were produced on non-focused material, consisting of referents only coded for information status, or non-IS words or speech breaks.

Focus: Apices on focused constituents aligned with pitch accents with a deviation between -500 and 500 ms in the majority of cases. 10 datapoints (4,9%) showed a pitch accent following the apex between 500 and 800 ms later and one pitch accent each (0,49%) preceded or followed the apex with a delay of more than one second. The majority of accents followed the apex (132, 65,3%) 65 accents preceded the apex (32,2%). All five perfectly aligning apex-accent pairs were produced on focused constituents. The histogram of this dataset (Figure 60) has its peak slightly behind the 0 ms mark and deviation frequencies drop with greater values.

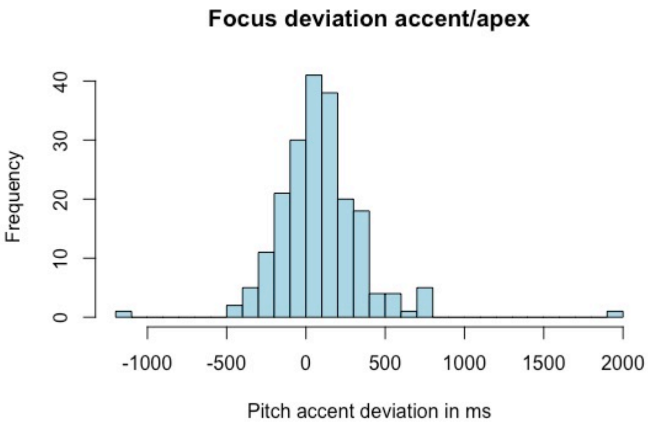


Figure 60: *Histogram on the deviation between apices and their nearest pitch accents on focus phrases in ms.*

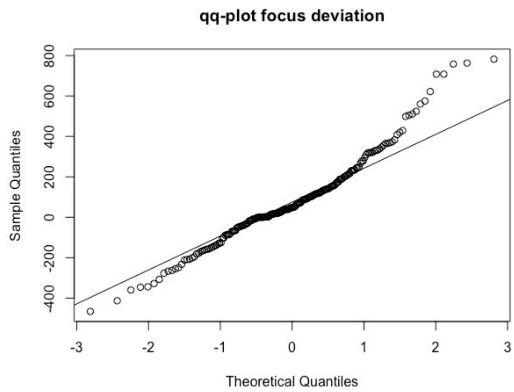


Figure 61: *qq-plot for the deviation of pitch accents from apices on focus phrases to visually test the normal distribution of the data set.*

The mean deviation of pitch accents to apices on focused constituents is μ_F = 80,766 ms and the standard deviation is σ_F = 220,2 ms. The qq-plot (Figure 61) shows an approximation to a normal distribution with accurate data points in the first theoretical quantiles, and divergence mainly in the third positive quantile. A one sample t-test was performed and indicated that the results for the factor level „focus" are significant (t = 5,21, df = 201, p < 0,001).

No focus: Regarding the 529 apices on non-focused constituents, the distribution of pitch accent deviation is more broad and ranged from -3000 to 3000 ms. The slopes (Figure 62) are less steep than in all other deviation graphs.

A flatter plateau can be observed between -1200 and 1200 ms deviation, containing 375 data points (70,9%). 178 pitch accents (35%) were found in the interval of -500 and 500 ms deviation. The peak of the graph is around 500 ms, meaning in pitch accents following the apex. Since five perfectly aligned apex-pitch accent pairs are situated on focused constituents, no perfect synchronisation was found on non-focused words. 273 accents (51,6%) followed the apex, 256 (48,4%) preceded it. The mean deviation of pitch accents was μ_{NF} = 22.54 ms and the standard deviation was σ_{NF} = 432,74 ms.

The histogram and qq-plot (Figure 62 and 63) illustrate do not suggest a normal distribution, with divergence in the qq-plot of data points in all theoretical quantiles. Similar to information status, a one sample t-test showed that the results for non-focused constituents are not significant (t = 1,2, df = 528, p = 0,2293).

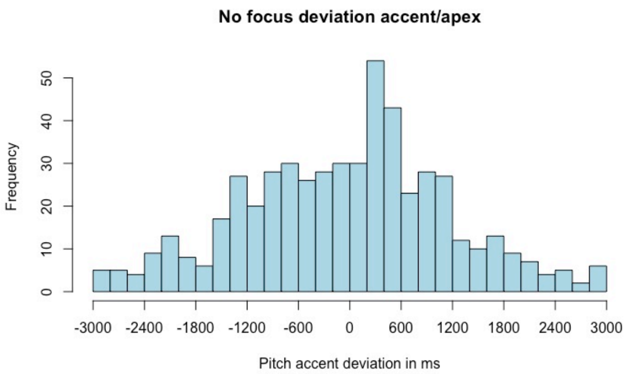


Figure 62: *Histogram on the deviation between apices and their nearest pitch accents on unfocused phrases in ms.*

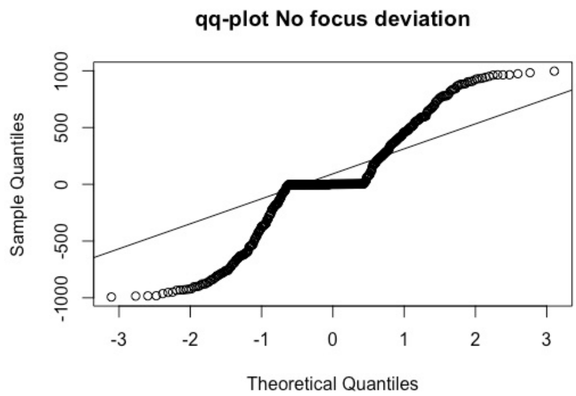


Figure 63: *qq-plot for the deviation of pitch accents from apices on unfocus phrases to visually test the normal distribution of the data set.*

Summary: To summarise this section, focused constituents show a smaller standard deviation of pitch accents from non-referential apices and more precise synchronisation at 0 ms. Non-focused referents have an average deviation closer to full synchronisation, which can be observed since the mean deviation of focused constituents is further towards the pitch accent following the apex. Frequency peaks of the histograms show an opposite tendency: Here, focused phrases have a peak closer to the 0 ms point, while on non-focused phrases, pitch accent deviations peak around 500 ms behind the apex. The mean deviation of focused phrases is more shifted to the right, the mean deviation of unfocused phrases to the left compared to the mean of the general distribution. Table 9 gives an overview over the exact values of this observation.

	Focus	No Focus	General Distribution
Mean	80,766 ms	22,54 ms	38,63 ms
Distance from general mean	42,14 ms	-16,09 ms	0 ms
Standard Deviation	220,2 ms	432,74 ms	385 ms
Peak	0 - 100 ms	400 ms	0 ms
p-value t-test	< 0,001	0,2293	< 0,01

Table 9: *Summary of means and deviations under focus presence and absence.*

The box plots in Figure 64 illustrate this difference: The median of the „focus“ condition is slightly higher than the one of „No focus“. In addition, the 95% quantile and therefore the variance of deviations of „focus“ is smaller than for „No focus“, and there are more data points outside the 95% quantile in the „No focus“ condition. A two sample t-test comparing focus vs. no-focus phrases indicates the significance of this variation (t = 2,3955, df = 671,63, p < 0,05).

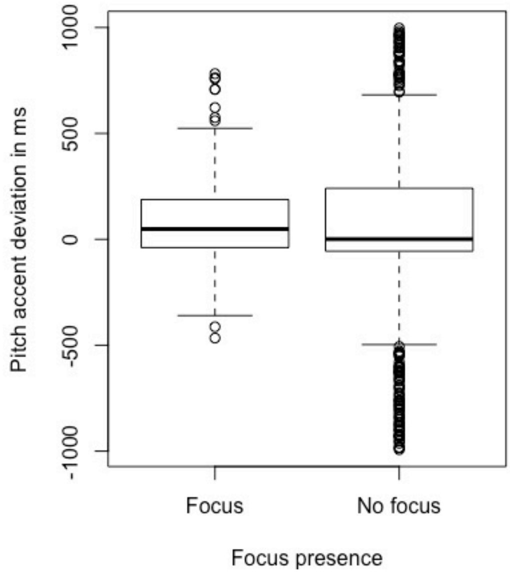


Figure 64: *Boxplots of the differing deviations of pitch accents to apices in presence or absence of focus.*

6. Discussion

In this chapter, the results of the corpus study on the influence of IS on the prosody-gesture interface are discussed and related to theoretical and empirical approaches on this topic. Initially, the results specific to each analysis are discussed in sections 6.1 and 6.2. They are then connected and generally discussed in section 6.3. Section 6.4 provides a reflection on the limitations of this study and its methods and section 6.5 gives a perspective for further research on the prosody-gesture interface, focusing on non-referential gestures and IS.

6.1 Distribution Analysis

The distribution analysis gave insight in the alignment of pitch accents and apices on the level of the lexical word. It has shown that IS influences the likelihood of an apex appearing in company of a pitch accent. Apices appeared the most often on more prominent referents („new“ referents and focus). Additionally, the apices appeared more accurately on more prominent referents. In less prominent referents („given“ and „accessible“ information status & non-focused constituents), the apices were more likely to be displaced and only the stroke overlapped with the referent. The occurrence of an overlapping stroke indicates that the apex does not align. Apart from this, the majority of apices did not occur on IS referents, only „new“ referents had a positive correlation with gesture apices. All other information status levels and focus tended to be not accompanied by a gesture. This is reasonable, because the purpose of non-referential gestures is to form and direct the discourse, and apices tend to synchronise with words that fulfil the same purpose, rather than IS referents that contribute to the content of an utterance and rarely guide the discourse. In addition, the results show that the absence of a gesture correlated with pitch accents and IS referents, reinforcing the observation that a non-referential gesture tends not to align with IS referents. This correlation is also supported by the big gap of occurrences between gestures on the one hand and pitch accents and IS referents on the other hand. Since there were more than four times as many pitch accents and IS material, gesture occurrences are more rare compared to the other factors. This is in accordance to observations by IM & BAUMANN 2020, who claim that either pitch accents and gestures occur together or only pitch accents are present on IS material. The difference between their study and the present study is that in this study more than half of the apices were not accompanied by a pitch accent while this case was rarely found in IM &

Baumann. This could be a result from the missing annotation of pitch accents on non-IS material, but it is also likely to be influenced by the non-referential nature of the investigated gestures. Im & Baumann did not distinguish the different types of gestures. The observation that so many apices did not accompany accents in the present corpus study might therefore also be an indication of them being situated on a) discourse forming speech material, which is less prosodically prominent and b) speech interruptions, as a tool to navigate the discourse even when no spoken language input can be provided (no accent can be produced during silence). These results give reason to assume a more complementary effect of non-referential gestures on speech than a redundant supporting or imitating function.

Apart from the general results on gesture occurrence and absence, the examination of pitch accents, gestures and IS revealed interactions. It has been addressed that „given“ and „new“ referents behave similar in terms of pitch accent distribution and apex occurrence across all data in this study, which is untypical for the two categories considering their differing retrievability (see Im & Baumann 2020 for the interaction of gestures and prosodic prominence and Pierrehumbert & Hirschberg 1990, Féry & Kügler 2008, Kügler & Calhoun 2020 for prosodic prominence). While this behaviour could be a result of the task for the interlocutors of the corpus (cf. section 6.4.1), it is striking that the prominence difference is more clear considering only referents on which an apex is produced: more apices are produced on „new“ referents, more apex displacement on „given“ referents (more stroke overlaps) and most importantly, the pitch accent distribution is more adopted to the prominence scales (more L+H* for „new“ and more L* for „given“). „New“ is thus clearly more prominent and „given“ is less prominent when an apex is produced. This signals that apex presence increases the prominence accuracy and not only IS influences the two modalities. This observation applies only partly to „accessible“ referents, which have far less apices than „given“ referents, but show a less prominent pitch accent distribution than „new“ referents. It seems that „accessible“ referents behave like outliers, distinct from the remaining IS referents and more behaving like non-IS material. This might be a result from the type of referents included in this category. „Accessible“ referents include pronouns and more referents that have been introduced into the conversation already, rather than referents being retrievable by semantic connections. Those referents were often produced to recall information, possibly by the interlocutor with the „follower“ function and were as a result less prominent. Additionally, an influence of apices on information status was not found to the same extend on focus. The distribution of pitch accents did not change dependent on the production of an apex. This indicates that focus in general is less sensitive to the presence of an apex in this corpus. Focus distribution did behave untypical in the corpus with many „given“ referents being in focus, which is not in accordance with e.g. FÉRY & KÜGLER 2008 or BAUMANN & RÖHR 2015. Another untypical observation concerning the focus categories was that they did not differ in the corpus study in terms of prominence and alignment with gestures. It is especially surprising that contrastive focus was not more prominent than new-information focus which is what PIERREHUMBERT & HIRSCHBERG 1990 and KATZ & SELKIRK 2011 suggest among others. To summarise, apart from the details explained in the last paragraph, „focus“, „given“ and „new“ seem to behave similar in this dataset as do „no focus“, „no information status“ and „accessible“.

One influence of information status on the pitch accent-gesture alignment shows a more indirect relation with prominence through the measurement of accuracy of apex alignment. Apex alignment has already been discussed to be influenced by information status, generally showing more apices aligning with more prominent referents. But taking the pitch accent types into account as well, alignment accuracy seems to be influenced by IS. It appears that less prominent referents have the most accurate apex alignment with pitch accents with lower prominence („given“ with L*) and more displaced apices with more prominent accents in this corpus. Similarly, more prominent referents (applying to „accessible“ and „new“ here) had better apex alignment with more prominent accents, being most accurate with L+H*. This alignment of apices increases linearly in the more prominent pitch accents, showing an interaction of prominence (prosodically and informativeness) and gestures. This interaction seems to take place on another layer that might be able to connect and increase the accuracy of pitch accent and information structural prominence.

The influence of focus on the pitch accent-gesture alignment seems to be less prominent in these data, even though the apex displacement was the dominating form of gesture occurrence on non-focused constituents, while it was apices for focus. The effect of focus on the synchronisation of accents and apices was that apex alignment was comparable across all pitch accent types, only minor increase or decrease of apex alignment was found. This is different for non-focused constituents, where apex alignment differed strongly between the different pitch accent types. Nevertheless, no trend along the prominence scale was detected. Apex alignment was weak for the moderately prominent H*+L accents and increased towards both less prominent and more prominent accents. This might show that the assignment of focus contributes to the alignment consistency of gestures and pitch accents. Unfocused constituents accordingly seem not to be „controlled" in their alignment between pitch accents and gestures. This indicates that the trigger for alignment-consistency of pitch accents and apices is focus.

With regard to the first research question, the distribution analysis found effects of IS on the synchronisation of prosody and gestures in a multidimensional manner. IS influenced gesture occurrence, concerning the general tendency of non-referential gestures to occur apart from IS referents, but also the tendency of apices to align with more prominent material. In addition, IS influenced the accuracy of alignment of pitch accent types and gesture apices. It is important to note, that it seems that the relation is bidirectional, meaning that gestures also have an impact on the correlation of IS and pitch accents. Finally, while prominence is a factor for the alignment of gestures and prosody, this data set and the occurrence of apices do not strictly follow the prominence hierarchy of pitch accents as well as of IS and as a result the pitch accent prominence scale.

6.2 Temporal Synchronisation Analysis

While the distribution analysis examined the occurrence of gestures and the impact of IS, it was limited to investigating alignment of pitch accents and apices on the lexical level by the presence of an actual apex on a referent vs. an overlapping stroke (apex displaced before or behind the referent) or the absence of a gesture. The temporal synchronisation analysis was able to provide more detail on the accuracy of synchronisation between pitch accents and apices of gestures. This is especially valuable since the distribution analysis showed the tendency of apices to avoid IS.

While the goal was to investigate the presence and position of apices on different IS categories and with different pitch accent types, this analysis started from the position of apices and searches for the nearest pitch accent. The reason for this is that apices are rarer than pitch accents in the corpus. Therefore, an analysis with pitch accents as a base searching for apices would lead to a lot higher deviations. This is not representative, since pitch accents and apices could pair across intonation phrase boundaries. This could not be accounted for by any semantic connection. For the same reason, data points with a deviation of more than three seconds were excluded, since the semantic, pragmatic or discourse connection is too far, even though it was not controlled for a content wise connection of pitch accents and apices.

It was found that on average, apices and their nearest pitch accent appeared very close to each other. The mean distance between the two was only slightly more than one frame (38 ms) with the pitch accent coming behind the apex. This means that the pitch accent and apex synchronised well within one word. The average length of a word in this corpus are 270 ms. LOEHR 2012 reports a perfectly synchronous average alignment at 0 ms deviation. In this study, the average deviation was less synchronous, but in accordance with claims by MCNEILL 1992 in his phonological synchrony rule. It states that „the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech.“ (MCNEILL 1992, p. 26). This can additionally be observed in the present results since apices preceded pitch accents more often than they followed them. However, apices also followed the pitch accent in about 45% of the cases, thus the phonological synchrony rule was not followed very strictly. This might result from a crucial factor of this analysis: the observation that apices were found absent from IS referents and pitch accents. This can also be explained by their function in the discourse. An example for this a sentence that was started and interrupted after an IS referent. During the interruption a non-referential gesture was produced as a signal of turn keeping, this easily leads to an apex occurring multiple seconds behind their nearest pitch accent. While a considerable amount of divergence from synchronisation is found for the apices, which could be a result from the non-referential nature of the gestures, the over all synchronisation seems to be balanced on a value in line with the synchronisation approaches by MCNEILL 1992 and LOEHR 2012.

A second factor in this analysis is the (range of) deviation between pitch accents and gestures, which was counted in a range between pitch accents preceding and following apices 3000 ms at most. It can be observed that smaller deviations are more frequent than higher deviations. Higher deviation mean worse synchronisation and the distribution of deviations showed, that apices prefer to have a pitch accent nearby over the two modalities having a greater distance to each other. This is a sign for the tendency of synchronisation of prosody and gestures.

The results of this thesis indicate an impact of IS on the temporal synchronisation between the modalities, which is a new finding in gestural research. In order to interpret the results, it is important to note that pitch accents were only annotated on constituents that were coded for IS. This might increase the deviation on non-IS apices. Nevertheless, also in those apices with a small deviation (thus putatively pitch accent and apex being produced on the same word), differences between the IS levels and non-IS were detected. Therefore, the consideration of all data for the IS influence, regardless of their deviation range, is justified. This analysis found that apex synchronisation improves on IS referents compared to non-IS material. This became obvious in the distribution of deviation in two ways: First, in the range of deviation, which was a lot higher for non-IS material. It showed that the accuracy of synchrony improved in presence of IS, as can be seen on the standard deviations of the different categories. In addition, the peak frequency of deviations was close to 0 ms on all IS categories and shifted more than 500 ms (in both directions) on non-IS constituents. „Accessible“ had a noticeably higher standard deviation than the remaining IS categories, another indication for these referents as outliers in the present study.

The change of mean deviation is more complex across the different levels and parameters of IS. Those results can be interpreted with consideration of the phonological synchrony rule. While in general, the apex was produced around one frame before its nearest pitch accent, the mean deviations shifted under certain levels of IS. This means that the apex was produced on a referent coded for that IS category. For information status, „given“ and „new“ referents showed a higher mean deviation than the general distribution. These referents behaved similar in shifting towards the right, meaning that the pitch accent follows a longer time after the apex. Constituents not coded for information status only showed a little shift of mean deviation towards the left, leading to a smaller deviation between accents and apices. On „accessible“ referents, the mean deviation was shifted strongly to the left such that on average, the pitch accent even preceded the apex. Similar observations can be made for focus: non-focus material shifted the mean deviation slightly towards 0 ms, while focused constituents showed a higher deviation with pitch accents following the apex. In addition, „given“ showed the highest change in mean deviation, followed by „accessible“, „new“ had a smaller shift of mean deviation, „No Info“ and „No Focus“ showed the smallest changes in mean deviation. This indicates that accent-apex synchronisation is influenced by IS, since non-IS material changed less. Within the information status categories, less prominent statuses had a higher change of deviation than the prominent category „new“. However, „given“ and „accessible“ shift the deviation into different directions. This could be interpreted as „given“ obeying to the phonological synchrony rule while „accessible“ does less. Since the mean deviation includes all deviations on that data set, it means that „given“ and „new“ referents have more apices preceding the pitch accent than the general distribution and more than „accessible“ referents. The preceding of the apex is what is demanded by the phonological synchrony rule. „Accessible“ referents have an increased amount of data points in which the apex follows the pitch accent, which is not in accordance with the rule. Assuming that „accessible“ is an outlier in this data set, not behaving in line with its designated prominence, a similar behaviour is found in general between IS and non-IS. IS referents seem to obey the phonological synchrony rule more than non-IS material. This is also reflected in the significance of the inferential tests that were performed in the temporal synchronisation analysis. The shifted deviation values were significant for „new“, „given“ and „focus“, meaning that in those categories, the mean values differ significantly from 0. In „accessible“, „No Info“ and „No Focus“ referents, the tests indicated that the values did not differ significantly from 0. This further clarifies the influence of the phonological synchrony rule. The rule played a bigger role for more prominent material, which differed significantly from 0, in the direction that apices tended to precede pitch accents more. This synchrony rule had a smaller influence on the non-IS categories and „accessible“ referents, which showed a higher amount of pitch accents preceding the apex, which is reflected in the non-significant divergence from 0.

Another factor indicating a difference in the accuracy of accent-apex synchrony between IS referents and non-IS material is the degree of approximation to a normal distribution. A distribution closer to a normal distribution shows the accumulation of pitch accents and apices at a certain deviation. The deviation distributions were similar to a normal distribution for the IS categories (weakest with „accessible“), and further away from a normal distribution in the non-IS categories „No Info“ and „No Focus“. This shows that the presence of IS facilitates the precision of the alignment between pitch accents and apices, while the deviation on non-IS material appears to be more coincidental. As a result, it can be stated that higher prominence (through IS presence) increases the accuracy of the temporal prosody-gesture alignment.

In connection to this, the divergence from a normal distribution was primarily present on pitch accent-apex pairs with a high deviation. These items are the ones with a weaker semantic connection. That they appear more frequent than expected in a normal distribution could be another indication towards the discourse function of non-referential gestures. Non-referential gestures have been mentioned to appear apart from IS referents, and possibly align better with function words. This explains the favoured distance between apices and pitch accents, since the latter ones were only annotated on IS referents, which do not represent function words.

It has also been observed in the data in several cases, that apices were assigned the same „nearest pitch accent". This can be responsible for a subset of the alignments with a high deviation since naturally, all the apices in such a cluster cannot have the same distance to the nearest pitch accent. The clustering of apices is a result of the specific characteristic of the non-referential gestures, that they adopt to the speech rhythm and therefore show repetitive occurrence in a short time. While it could be assumed to find multiple pitch accents on such a gesture cluster, this was not the case in this study since pitch accents were only annotated on IS material in the annotation for the present study.

With regard to the second research question, these results show a general synchronisation of pitch accents and apices in line with the proposal of McNeill 1992 on phonological synchrony and empirical studies that investigated similar prosody gesture synchronisation (e.g. Loehr 2012). Pitch accents and apices in the present corpus study aligned within one word on average and more specifically with an average distance of only one frame. An influence of IS on this synchronisation was also detected. The synchronisation was more precise on IS referents for both information status and focus than on non-IS material. Thus, IS facilitates the alignment of apices of non-referential gestures and pitch accents. As a more detailed result, a smaller range of deviation and a more accurate behaviour along the phonological synchrony rule are the indicators of IS effects, being situated on IS material. This also means that prominence is a navigator of accent-apex synchrony accuracy both between IS and non-IS referents as well as between the information status levels (considering that „accessible" behaves like an outlier in this data set).

6.3 General Discussion

The results of both the distribution analysis and the temporal synchronisation analysis suggest that IS has an impact on the synchronisation of prosodic prominence in form of pitch accent and the apices of non-referential gestures. This was observed on the level of the lexical word as well as on the level of intonation. Information structural prominence (determined by newness and informativeness of the referents) facilitates the synchronisation of the two investigated prominence modalities, pitch accents and the smallest gestural components, which was discussed in detail in sections 6.1 and 6.2. Both pitch accents and apices interact with both IS parameters, information status and focus, by themselves. The general occurrence of pitch accents and apices is not completely synchronous, since the distribution analysis shows a preference of either an accent or an apex present, not both together and the temporal synchronisation indicates a high deviation of pitch accents both preceding and following an investigated apex. However, the presence of an IS referent improves this alignment, with a more sorted alignment on the word level and a smaller range of deviation on the intonation level.

On the other hand side, gesture apices were found to not be likely to align with IS constituents: more than half of all apices appeared in parts of the conversation that were not coded for IS. The distance between the modalities was partly greater than the length of a word, such that only half of the strokes belonging to the „displaced" apices overlapped with the IS referents. This might be an indication that apices of non-referential gestures do not tend to accompany material conveying IS information, but are produced on other speech material, like putatively constituents that convey discourse (directing) information. In general, in the spontaneous speech corpus, the majority of IS referents was accented, even less prominent referents. This is in contrast to other studies, where „given" referents are frequently unaccented or have lower prosodic prominence (KÜGLER & CALHOUN 2020, KRIFKA 2008, BAUMANN et al. 2006, BAUMANN & RÖHR 2015 among many others). This could be a result from the task that was given to the interlocutors in the corpus, as direction description and memory play a major role in the conversations. IS referents being accompanied by an apex are more likely to be accented in this study (apart from „accessible" referents, which include pronouns), which implies that gestural apices are likely to be accompanied by a pitch accent in line with IM & BAUMANN 2020. This can only directly be seen in a subset of the dataset, in which apices appeared on IS referents, since these constituents are the ones annotated for apices and pitch accents. Non-IS material was not annotated for pitch accents and therefore does not make a statement in this direction.

The most important influences of IS were the following: The different levels of information status facilitate the alignment of pitch accents and apices dependent on prominence. It was observed that more prominent and new referents show a better accent-apex synchronisation with more prominent accents, while less prominent referents have a better alignment with less prominent accents. This alignment was measured considering the percentage of aligning apices and overlapping strokes, which indicate a displacement of the apex away from the referent. However, prominence cannot be rated as a linear increasing factor for the alignment of pitch accents and apices. „Accessible“ referents received the least prominence (prosodically and gesturally) in this specific task-oriented spontaneous speech setting. With regard to the influence of focus, the results showed that the different focus types did not have a relevantly differentiating impact on the results, apart from the observation that contrastive focus seemed to receive less prominence from both modalities. This is in contrast to theoretical models (PIERREHUMBERT & HIRSCHBERG 1990, BARTELS & KINGSTON 1994, KATZ & SELKIRK 2011), where it is stated that a contrastive feature of focus phrases increases their prominence. Therefore, the distinction of the presence or absence of focus is more important for this study. The main observation for „focus“ vs. „no-focus“ was that this parameter behaved very similar to information status in terms that the more prominent category, „focus“, showed the same effects as the more prominent information status referents and the other way around. The difference between the two focus factors became obvious in the temporal synchronisation analysis, where not only the deviation was smaller for focused constituents than for non-focused, but no deviation was only found on focused constituents. This is a difference to information status, where perfect alignment is found on all categories once. This difference is also mirrored in the significance of the compared deviations of the information status levels and focus presence. Information status differences to accent-apex alignment were not significant, while the deviation difference between focus and unfocused phrases was significant. This is an indication for a higher importance of the IS parameter „focus“ than „information status“ for temporal synchronisation of pitch accents and apices. The parameter „information status“ was found to have a bigger influence on the distribution analysis than „focus", which is expressed by the occurrence of gestures on the different factor levels as well as the accuracy of alignment between accents and apices on the word level. Both observations were less impacted by focus.

The most relevant aspects of these observations are discussed in further detail in the following sections and afterwards the limitations of the corpus study and questions for future research are presented.

6.4 Limitations of the study

6.4.1 Task influence

The SaGA corpus helpful to receive insight to the alignment of pitch accents and apices in spontaneous speech. Since the corpus is a comprehensive audiovisual spontaneous speech corpus, it provides gesture type annotations already and allows for all important further annotation and data extraction, it is very valuable for this analysis. Nevertheless, the recordings were not the optimal type of conversation for the purpose of this analysis. Although the conversations represented spontaneous speech, the interlocutors had to perform a certain task that led to unnatural components of their speech. This might lead to a bias in the results. The task was that one interlocutor was guided through a virtual reality town environment for more than eight minutes, since it contains five stops on the landmarks and rides of about 60 seconds in between. This makes it hard to memorise and recall all details, which is important since the second conversation partner afterwards had to find their way through the town by themselves. The difficult task containing direction description and stressing the memory of both the describer and the follower lead to hesitation or interruption of speech, to longer speech breaks, to a lot of repetition and emphasis of every detail. This is reinforced by the aim to fulfil or succeed in the task.

This influences the behaviour of all factors of this analysis. With regard to non-referential gestures, it can be observed that during a longer interruption of speech, gestures are often produced, putatively as a sign of turn keeping. This is one factor that led to a high deviation (range) in the temporal synchronisation analysis. It is also possible that the high amount of referential gestures in the corpus is a result from the task as visualisation of the objects by hands could be used as a tool to facilitate memorisation.

Intonational phrases were harder to delimit or identify as a result of the task influence, since several speaking parts only consisted of one word or were discontinued. This was primarily important for the labelling of IS but also for the distinction of pitch accents from phrase and boundary tones. With regard to the utilisation of pitch accents, one effect probably initiated by the conversation task was that the vast majority of referents was accented, which is untypical for the differentiation of prominence. The accentuation could have been used as a tool to emphasise the relevant information, which applies to most of the information since every detail had to be remembered. This could also be the reason for the observation that the distribution of most pitch accent types did not vary as expected across different information structural levels.

The task had effects on IS as well. It is likely that the comparable amount of „given“ and „new“ referents is due to the high rate of contrast and repetition that is used to a) recall the route and b) distinguish the landmarks from the remaining environment. The missing difference between „given“ and „new“ with regard to prosodic prominence could also be a result from the importance of all referents, also „given“ ones, despite their different level of newness and informativeness. While the results of focus do not seem to have been influenced by the task as much as information status, the material that was assigned focus was influenced. Instead of mainly referring nouns being focused, the direction giving part elicited that directions (like *links* „left“, *rechts* „right“, *geradeaus* „straightforward“ or *entlang* „along“) were often focused or even contrasted, since they were the most important/ informative part of the sentence, but did not receive the highest prominence (contrary to what is typical for focused constituents).

As a conclusion, it has to be mentioned that while these effects contort the data, the recordings were not collected specifically for this analysis or even for phonological research, but rather originally for physical features and components of gestures and their general characteristics. Investigating spontaneous speech naturally includes a higher amount of unexpected behaviour than controlled experimental studies, since disturbing factors cannot easily be ruled out. Overall, the interlocutors made use of pitch accents and gestures intuitively and spontaneously and did not manipulate the IS parameters. Since the corpus provides a large data set, untypical observations did not overweigh, and observations on the influence of IS on the accent-apex alignment could be made. Therefore, the SaGA corpus is suitable for a first exploration on the prosody-gesture interface in consideration of IS.

6.4.2 Further limitations

During the annotation and analysis of this corpus, further challenges were encountered. The corpus was recorded twelve years ago. Nowadays, technology has advanced enormously so that greater amount of detail and flexibility in the setup of the VR environment would be possible. The town environment was built up very uniformly (see Figure 65) and the landmarks a) did not look very distinct from one another and b) also sounded similar / displayed similar buildings (churches & chapel).

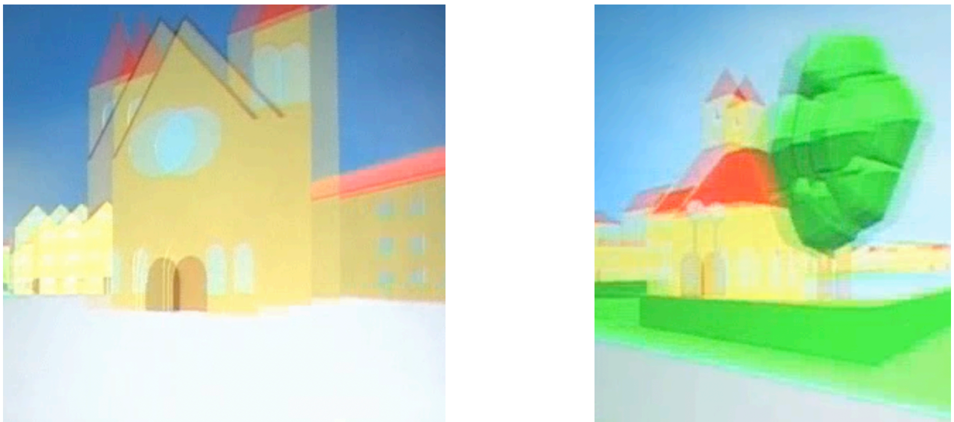


Figure 65: *Two similar looking buildings from the VR town. Left: church to the right of the church square, right: chapel landmark.*

This might be a reason for problems during the conversations, like landmarks that were interchanged, long speech breaks appeared (hesitation, remembering) and turn taking did not work out naturally. Since virtual reality modelling improved significantly during the last decade, it might be helpful to collect a new corpus with an adapted task for recordings and an improved VR environment: More distinct landmarks with more prominent characteristics and paths (with features) in between them as well as a shorter trip would presumably help to distinguish and navigate through the environment, especially keeping in mind that the task's purpose is more to direct the conversation and less to be fulfilled correctly.

The annotation of intonation provided the challenge, that the spontaneous speech data is less consistent than experimentally controlled speech. It sometimes contained creaky voice, volume inconsistencies and both interlocutors talking at the same time, which made it difficult to identify the correct pitch accents in some cases. To assure a better voice quality and accuracy of the data, a new recording should take a high quality audio recording into account. A second intonation issue is the choice of a labelling system for intonation. The choice of the GToBI system over other renomated intonation annotation systems for German like DIMA (Deutsche Intonation: Modellierung und Annotation; KÜGLER et al. 2015, 2019, 2022) or KIM (Kiel Intonation Model; NIEBUHR 2019) resulted from the applicability to other languages and the use of the autosegmental-metrical framework. The KIM uses another intonational framework, the DIMA system is suitable for German, but it is customised to this language. Since extensions and comparisons of this study to Catalan are planned, GToBI was used in this study to be able to use its existing relative system(s) - ToBI (SILVERMAN et al. 1992 a,b) - for comparable data in other studies. While GToBI in general provides a very good base for pitch accent annotation, and is also easily adaptable to other languages for a cross-linguistic comparison, it lacks a direct indication of prosodic prominence apart from ordering the pitch accents on a prominence scale. This would have been helpful to receive another measurement for prominence. The DIMA annotation system uses level tones to indicate different prominences (intensity, duration) on possibly comparable f0 contours. Using DIMA and especially the prominence tone labelling could make it possible to explore a more differentiated influence of IS on the prosody-gesture alignment, since equally „prominent“ pitch accents on the Prominence Scale could receive different level tones. This might allow a more in-depth differentiation between otherwise similar rated „given“ and „new“ referents based on level tones: „given“ referents could be more often associated with level 1 prominence, whereas „new“ referents often could receive prominence of level 2. Additionally, pitch accent annotation has to be extended to the full material of the corpus, not only on IS material to get a more precise information on the temporal alignment of pitch accents and non-referential gesture apices, which were found to avoid IS referents.

Another challenge was the annotation of IS. While it has already been mentioned that the RefLex annotation scheme would provide a more differentiated and precise look on IS to improve separating referential and lexical retrievability, two further topics have to be addressed. The first one concerns the dialogue nature of the conversations. Delimitation of the information status types follows conventions (at least for „given“ and „accessible“), which are not empirically funded and which were not investigated for the applicability from monologues to conversations in a one-to-one relation. For this study, the conventions for monologues were adapted. Possibly, prosodic and multimodal prominence could contribute to a better delimitation of the information status types. Finally, pronouns provided a big challenge in the full analysis of this corpus. While GÖTZE et al. 2007 treat them as „accessible“, which was followed in this thesis as well, other approaches (eg. RIESTER & BAUMANN 2017, BAUMANN, p.c.) classify them as „given“. This decision has a strong influence on the results since almost all pronouns were unaccented and had a low chance of apex alignment. This is one reason for the decision to exclude pronouns unaccompanied by an accent or apex, but should be elaborated in greater detail in further research.

6.5 Research Perspective

This study provided an insight into the mediation of IS on the alignment of prosody and non-referential gestures. It is a basis for further research on the adaptation and extension of the prosody-gesture interface. As has been mentioned in the previous section 6.4.2, initial improvements would be to reannotate the data including boundary tones as well as following DIMA and RefLex annotation systems for intonation and IS. This would allow for a more differentiated analysis of the data and might lead to new realisations and a better understanding of the role of prominence in this interface. In addition, conducting an inter-annotator reliability calculation would improve the accuracy of annotation in such a big dataset. A repetition of recording data using up-to-date technology would also make sense, to receive better control of the disturbing factors, for example by only allowing one interlocutor to speak or by shortening the task.

Exploring the corpus showed that gestures, especially iconic and deictic ones, could have an influence on the information status of a referent. Gestures could increase the retrievability of a referent by forming its features or pointing to it, even when the referent is not mentioned directly in speech. In the present study, IS was determined completely based on speech. A systematic investigation using prominence amongst other factors measuring retrievability would be a good follow up on this study. It would address a possible bidirectional relation between gestures and IS.

Another perspective on the research topic of this thesis raises the question whether non-referential gestures could also have different degrees of prominence. While intonation and IS have clear levels of prominence and express them by determined tools, no such distinctions have been established for gestures until now and mainly the presence or absence and position of gestures has been considered. It would be interesting to explore, whether temporal alignment, or possibly the length or amplitude of gestures or even their morphology are indicators for prominence. This study has found a correlation between the synchronisation of pitch accents & gestures and prominence determined by newness and informativeness. However, clear „levels" of gestural prominence could not be distinguished. However, a hypothesis of future research could be that the direction of the gesture is an indication for prominence, in a way that a downwards movement is more associated with lower prominence and upwards movement indicates higher prominence, similar to the system of intonation. This question underlines the need to explore more different aspects of gestures and to overwrite the picture of a one-dimensional modality. In the basic of gesture research it was assumed that gestures are more simple than speech, like established in McNeill 1992.

A third field for gestural research that is of relevance for future investigations is the distinction between referential and non-referential gestures. The two categories have been established early in gestural research (eg. McNeill 1992, Kendon 2004). Their most crucial difference is that referential gestures have a semantic connection to speech which non-referential gestures do not. Non-referential gestures display discourse functions and contribute to the development of the discourse, but it is not impossible that referential gestures do the same. Their non-semantic components can fulfil similar functions, at least to some extent. It would be interesting to explore discourse functions of referential gestures, by which factors they are expressed and how non-referential gestures set themselves apart. In this context, it would also be interesting to explore the alignment of non-referential gestures with function words. A hypothesis might be that these discourse directing words and gestures show a closer alignment since they express similar discourse functions. Since referential gestures were a lot more frequent than non-referential gestures in the SaGA corpus, it would be helpful to develop a method to elicit non-referential gestures in a natural way or to find out if and why non-referential gestures are less frequent than referential ones. This would provide insight in the exact functions of non-referential (and referential) gestures.

These are additional interesting and important questions resulting from the conclusions of this thesis. Nevertheless, there are many more open issues and questions yet to be solved on this new and promising multimodal interface.

7. Conclusion

This thesis has investigated the interaction of intonation, non-referential gestures and Information Structure in a German audiovisual spontaneous speech corpus. The questions that were addressed concern the occurrence of gestures on the level of their apex and of gestures in combination with pitch accents (of varying prosodic prominence), as well as the temporal alignment of gestures and pitch accents. Both occurrences and alignment were examined for their variation under different levels of IS, namely information status and focus, as representations of linguistic prominence.

An extensive statistical analysis has shown that IS influences both the frequency of occurrence of apices on IS referents as well as the accuracy of synchronisation between pitch accents and apices of gestures. A distribution analysis has found a general tendency of non-referential gestures to not appear on IS referents, which might be a result from their non-referential nature showing no semantic connection to speech. However, if they accompanied IS referents, prominence influenced a) the frequency of gesture occurrence on the referent, newer and more informative referents being more likely to be accompanied by the apex of a gesture and b) the likelihood of alignment of the apex with each pitch accent type - on more prominent referents, apices aligned better with more prominent pitch accents and vice versa. As a side effect, the presence of apices was found to increase the accuracy of alignment of pitch accents and information status along the pitch accent prominence scale. An analysis of the temporal synchronisation has shown that apices of gestures and pitch accents tend to align with each other, preferring a small distance between them. In general, this is in accordance with established phonology-gesture synchrony rules, since apices preceded rather than followed their nearest pitch accent. IS facilitated this synchronisation, resulting in less deviation between the two factors and and a more accurate consideration of the phonological synchrony rule. The effect of the specific Information Structural types however was not sorted straightforwardly along prosodic prominence. Between the two IS parameters, the word level distribution of apices was stronger influenced by the information status categories, while focus had a bigger impact on the temporal synchronisation of pitch accents and apices.

Throughout the whole analysis, it appeared that „accessible“ referents do not integrate into the prominence relation in this data, showing similarities to less prominent non-IS material. In addition, the investigated conversations being based on a direction giving task yielded further unexpected results with regard to prominence, since also less prominent material was accented and the distribution of pitch accents did not follow typical prominence patterns.

To conclude, this thesis suggests that gestures and pitch accents tend to align following different synchronisation principles. This synchronisation is influenced by linguistic prominence by IS, though not completely straightforward and one-dimensional. Gaining these basic insights on the interaction of intonation, gestures and IS opens the doors for more in depth exploration of the role of prominence in gestural research and the discourse functions of non-referential gestures.

References

Alibali, M. W. 2005. Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation* 5. 307–331.

Bartels, Christine & John Kingston. 1994. Salient pitch cues in the perception of contrastive focus. In P. Bosch & R. van der Sandt (eds.), *Focus and Natural Language Processing. Proceedings of the Journal of Semantics Conference on Focus, IBM Working Papers, TR-80.94.006*, ??

Baumann, Stefan & Martine Grice. 2006. The intonation of accessibility. *Journal of Pragmatics* 38(10). 1636–1657.

Baumann, Stefan, Martine Grice & Susanne Steindamm. 2006. Prosodic Marking of Focus Domains - Categorical or Gradient? 301–304.

Baumann, Stefan & Arndt Riester. 2013. Coreference, lexical givenness and prosody in German. *Lingua* 136, http://www.sciencedirect.com/science/article/pii/S002438411300168X.

Baumann, Stefan & Christine T. Röhr. 2015. The perceptual prominence of pitch accent types in German. 384.

Benson, D. F. 1985. Aphasia. In K. Heilman & E. Valenstein (eds.), *Clinical neuropsychology*. New York: Oxford Univ. Press.

Bergmann, Kirsten & Stefan Kopp. 2006. Verbal or visual: How information is distributed across speech and gesture in spatial dialog. 90–97.

Blumstein, Sheila E., William Milberg & Robin Shrier. 1982. Semantic processing in aphasia: evidence from an auditory lexical decision task. *Brain and Language* 17. 301–315.

Boersma, Paul & David Weenink. 2022. *Praat: doing phonetics by computer [Computer program]: Version 6.2.09 (http://www.praat.org/)*.

Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word* 14. 109–149.

Cangemi, Francesco & Martine Grice. 2016. The importance of a distributional approach to categoriality in Autosegmental-Metrical accounts of intonation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1).

Chafe, Wallace L. 1976. Givenness, Contrastiveness, Definiteness, Subjects and Topics. In Charles N. Li (ed.), *Subject and Topic*, 27–55. New York: Academic Press.

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English.* New York: Harper and Row.

Clark, Herbert H. 1992. *Arenas of language use.* Chicago: University of Chicago Press.

Clark, Herbert H. 1996. Communities, Communalities. *Rethinking linguistic relativity* 17. 324.

Cochet, Hélène & Jacques Vauclair. 2014. Deictic gestures and symbolic gestures produced by adults in an experimental context: Hand shapes and hand preferences. *Laterality: Asymmetries of Body, Brain and Cognition* 19(3). 278–301.

Cole, Jennifer & Jeremy Steffman. 2022. The primacy of the rising/non-rising dichotomy in Amercian English intonational tunes. In Oliver Niebuhr (ed.), *Proceedings of TAI 1*, 122–126. ISCA Archive.

Connell, Bruce. 2001. Downdrift, downstep, and declination. In *Proceedings of Typology of African Prosodic Systems Workshop*. Bielefeld: Bielefeld University, Germany.

ELAN. 2021. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive.

Féry, Caroline. 2010. Syntax, information structure, embedded prosodic phrasing, and the relational scaling of pitch accents. In Nomi Erteschik-Shir & Lisa Rochman (eds.), *The sound patterns of syntax*, 271–290. Oxford: Oxford University Press.

Féry, Caroline. 2017. *Intonation and Prosodic Structure* (Key Topics in Phonology). Cambridge: Cambridge University Press.

Féry, Caroline, Gisbert Fanselow & Manfred Krifka. 2007. *The Notions of Information Structure*. SFB 632 Information Structure.

Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4). 680–703.

Féry, Caroline & Hubert Truckenbrodt. 2005. Sisterhood and tonal scaling. *Studia Linguistica* 59(2-3). 223–243.

Genzel, Susanne & Frank Kügler. 2011. Phonetic Realization of Automatic (Downdrift) and Non-automatic Downstep in Akan. *ICPhS*. 735–738.

Goldsmith, John A. 1990. *Autosegmental and metrical phonology.* Oxford, UK, Cambridge, Mass., USA: B. Blackwell.

Goodluck, Helen. 1991. *Language acquisition: A linguistic introduction.* Basil Blackwell.

Götze, Michael, Thomas Weskott, Cornelia Endriss, Ines Fiedler, Stefan Hinterwimmer, Svetlana Petrova, Anne Schwarz, Stavros Skopeteas & Ruben Stoel. 2007. Information structure. In Stefanie Dipper, Michael Götze & Stavros Skopeteas (eds.), *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics, and Information Structure* (Interdisciplinary Studies on Information Structure (ISIS) 7), 147–187. Potsdam: Universitätsverlag Potsdam.

Grabe, Esther. 1998. Pitch accent realization in English and German. *Journal of Phonetics* 26(3). 129–143.

Grice, Martine & Stefan Baumann. 2002. Deutsche Intonation und GToBI. *Linguistische Berichte* 191. 267–298.

Grice, Martine, Stefan Baumann & Ralf Benzmüller. 2005. German Intonation in Autosegmental-Metrical Phonology. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 55–83. Oxford: Oxford University Press.

Grice, Martine, D. R. Ladd & Amalia Arvaniti. 2000. On the place of phrase accents in intonational phonology. *Phonology* 17(2). 143–185.

Grice, Martine, Matthias Reyelt, Ralf Benzmüller, Jörg Mayer & Anton Batliner. 1996. Consistency in Transcription and Labelling of German Intonation with GToBI 3. 1716–1719.

Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

Halliday, Michael A. K. 1967. *Intonation and grammar in British English.* The Hague: Mouton.

Hockett, Charles F. & Charles D. Hockett. 1960. The origin of speech. *Scientific American* 203(3). 88–97.

Hoff, Erika. 2013. *Language development.* Cengage Learning.

Im, Suyeon & Stefan Baumann. 2020. Probabilistic relation between co-speech gestures, pitch accents and information status. *Proceedings of the Linguistic Society of America* 5(1). 685–697.

Im, Suyeon, Jennifer Cole & Stefan Baumann. 2018. The probabilistic relationship between pitch accents and information status in public speech. *ISCA*. 508–511.

Jackendoff, Ray S. 1972. *Semantic Interpretation in Generative Grammar.* Cambridge, MA: MIT Press.

Jannedy, Stefanie & Norma Mendoza-Denton. 2005. Structuring Information through Gesture and Intonation. In Shinichiro Ishihara, Michaela Schmitz & Anne Schwarz (eds.), *Interdisciplinary Studies on Information Structure* (ISIS 3), 199–244. Postdam: Universitätsverlag Potsdam.

Jun, Sun-Ah. 2005. Prosodic Typology. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, 430–458. Oxford: Oxford University Press.

Karttunen, Lauri & Stanley Peters. 1975. Conventional Implicature in Montague Grammar. 266–278.

Katz, Jonah & Elisabeth O. Selkirk. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4). 771–816.

Kember, Heather, Jiyoun Choi, Jenny Yu & Anne Cutler. 2021. The Processing of Linguistic Prominence. *Language & Speech* 64(2). 413–436.

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary R. Key (ed.), *The Relationship of Verbal and Nonverbal Communication*, 207–227. Berlin: Mouton.

Kendon, Adam. 2004. *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kratzer, Angelika & Elisabeth Selkirk. 2007. Phase theory and prosodic spellout: The case of verbs. *The Linguistic Review* 24(2-3). 93–135.

Krifka, Manfred. 2008. Basic notions of information structure. *Acta Linguistica Hungarica* 55(3). 243–276.

Kügler, Frank. 2018. Optional accentuation of pronouns in German. *ISCA*. 671–675.

Kügler, Frank, Stefan Baumann, Bistra Andreeva, Bettina Braun, Martine Grice, Jana Neitsch, Oliver Niebuhr, Jörg Peters, Christine T. Röhr, Antje Schweitzer & Petra Wagner. 2019. Annotation of German Intonation: DIMA compared with other annotation systems. 1297-1301.

Kügler, Frank, Stefan Baumann & Christine T. Röhr. 2022. Deutsche Intonation, Modellierung und Annotation (DIMA): Richtlinien zur prosodischen Annotation des Deutschen. In Cordula Schwarze & Sven Grawunder (eds.), *Transkription und Annotation* *gesprochener Sprache und multimodaler Interaktion: Konzepte, Probleme, Lösungen*, 23–54. Tübingen: Narr Francke Attempto Verlag.

Kügler, Frank & Sasha Calhoun. 2020. Prosodic Encoding of Information Structure: A typological perspective. In Carlos Gussenhoven & Aoju Chen (eds.), *The Oxford Handbook of Language Prosody*, 453–467. Oxford: Oxford University Press.

Kügler, Frank, Bernadett Smolibocki, Denis Arnold, Stefan Baumann, Bettina Braun, Martine Grice, Stefanie Jannedy, Jan Michalsky, Oliver Niebuhr, Jörg Peters, Simon Ritter, Christine T. Röhr, Antje Schweitzer, Katrin Schweitzer & Petra Wagner. 2015. DIMA - Annotation guidelines for German intonation. 317.

Ladd, D. R. 1980. T*he Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. 1986. Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook* 3. 311–340.

Ladd, D. R. 1996. *Intonational Phonology.* Cambridge: Cambridge University Press.

Ladd, D. R. 2008. *Intonational Phonology,* 2nd edn. Cambridge: Cambridge University Press.

Lambrecht, Knud. 1994. *Information structure and sentence form*. Cambridge: Cambridge University Press.

Leonard, Thomas & Fred Cummins. 2009. Temporal alignment of gesture and speech.

Leonard, Thomas & Fred Cummins. 2011. The temporal relationship between beat gestures and speech. *Language and Cognitive Processes* 26(10). 1457–1471.

Liberman, Mark & Alan Prince. 1977. On stress and linguistic rythm. *Linguistic Inquiry* 8(2). 249–336.

Loehr, Daniel P. 2004. *Gesture and Intonation*. Washington: Georgetown University PhD Thesis.

Loehr, Daniel P. 2007. Aspects of rhythm in gesture and speech. *Gesture* 7(2). 179–214.

Loehr, Daniel P. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 3(1). 71–89.

Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp & Hannes Rieser. 2010. The Bielefeld Speech and Gesture Alignment Corpus (SaGA). *ELRA*. 92–98.

McClave, Evelyn Z. 1994. Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research* 23(1). 45–66.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

McNeill, David & Elena Levy. 1982. Conceptual representations in language activity and gesture. *Speech, place and action*. 271–295.

Meyer, D: Schvaneveldt, R. 1971. Facilitation in recognizing pairs of words: Evidence of a dependance between retrieval operations. *Journal of Experimental Psychology* 90. 227–234.

Nespor, Marina & Irene Vogel. 1989. On clashes and lapses. *Phonology* 6(1). 69–116.

Niebuhr, Oliver. 2019. Pitch Accents as Multiparametric Configurations of Prosodic Features - Evidence from Pitch-accent Specific Microrhythms in German. In Aarhus University (ed.), *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn*, 321–351.

Pagano, R. R. 2010. *Understanding statistics in the behavioral sciences,* 9th edn. Australia, Belmont: CA Thomson Wadsworth.

Pierrehumbert, Janet B. 1980. *The phonology and phonetics of English intonation*. Cambridge, Massachusetts: MIT PhD Thesis.

Pierrehumbert, Janet B. & Julia Hirschberg. 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. In Philip R. Cohen, Jerry Morgan & Martha E. Pollack (eds.), *Intentions in Communication*, 271–311. Cambridge: MIT Press.

Prather, P., E. Zurif & T. Love. 1992. Time course of lexical access in fluent and nonfluent aphasia. In *Academy of Aphasia*.

R Core Team. 2013. *R: A Language and Environment for Statistical Computing.* Vienna, Austria.

Rasch, D. & V. Guiard. 2004. The robustness of parametric statistical methods. *Psychology Science* 46. 175–208.

Reyelt, Matthias, Martine Grice, Ralf Benzmüller, Jörg Mayer & Anton Batliner. 1996. *Prosodische Etikettierung der Deutschen mit ToBI.*

Riester, Arndt & Stefan Baumann. 2017. *The RefLex Scheme – Annotation Guidelines*. SFB 732, Universität Stuttgart. http://elib.uni-stuttgart.de/bitstream/11682/9028/1/RefLex-SinSpec14.pdf.

Rohrer, Patrick L., Ingrid Vilà-Giménez, Júlia Florit-Pons, G. Gurrado, Núria Esteve-Gibert, Ada Ren, Stefanie Shattuck-Hufnagel & Pilar Prieto. 2020. *The MultiModal MultiDimensional (M3D) labeling system for the annotation of audiovisual corpora: Gesture Labeling Manual.* UPF Barcelona.

Rooth, Mats. 1985. *Association with Focus*. Amherst: University of Massachusetts PhD thesis.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116.

Roth, Wolff-Michael. 2002. From action to discourse: The bridging function of gestures. *Cognitive Systems Research* 3. 535–554.

Schwarzschild, Roger. 1999. Givenness, AvoidF and other constraints on the placement of accent. *Natural Language Semantics* 7(2). 141–177.

Selkirk, Elisabeth. 1981. On prosodic structure and its relation to syntactic structure. 111–140.

Selkirk, Elisabeth O. 1996. The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, 187–213. Mahwah, NJ: Erlbaum.

Sgall, Petr, Eva Hajičová & Jarmila Panevová. 1986. *The meaning of the sentence in its semantic and pragmatic aspects.* Dordrecht: Reidel.

Shattuck-Hufnagel, Stefanie, Mari Ostendorf & Ken Ross. 1994. Stress shift and early pitch accent placement in lexical items in Amercian English. *Journal of Phonetics* 22(4). 357–388.

Shattuck-Hufnagel, Stefanie, Ada Ren, Mili Mathew, Ivan Yuen & Katherine Demuth. 2016. Non-referential gestures in adult and child speech: Are they prosodic? *ISCA*. 836–839.

Shattuck-Hufnagel, Stefanie, Yelena Yasinnik, Nanette Veilleux & Margaret Renwick. 2007. A method for studying the time-alignment of gestures and prosody in American English: 'Hits' adn pitch accents in academic-lecture-style speech. In Anna Esposito, Maja Bratanic, Eric Keller & Maria Marinaro (eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue*, 34–44. Amsterdam: IOS Press.

Silverman, Kim, Mary E. Beckman, John Pitrelli, Mari Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert & Julia Hirschberg. 1992a. ToBI: A standard for labeling English prosody. *ISCA*. 867–870.

Silverman, Kim, Eleonora Blaauw, J. Spitz & John Pitrelli. 1992b. Prosodic comparison of spontaneous speech and read speech. 1299–1302.

Stalnaker, Robert. 1978. Assertion. In Peter Cole (ed.), *Pragmatics* (Syntax and semantics 9), 315–332. New York: Academic Press.

Truckenbrodt, Hubert. 2002. Upstep and embedded register levels. *Phonology* 19(1). 77–120.

Truckenbrodt, Hubert. 2004. Final lowering in non-final position. *Journal of Phonetics* 32(3). 313–348.

Truckenbrodt, Hubert. 2006. Phrasal Stress. In Edward K. Brown (ed.), *Encyclopedia of language and linguistics*, 2nd edn., 572–579. Amsterdam: Elsevier.

Truckenbrodt, Hubert. 2007. The syntax-phonology interface. In De Lacy, Paul V. (ed.), *The Cambridge Handbook of Phonology*, 435–456. Cambridge: Cambridge University Press.

Uhmann, Susanne. 1991. *Fokusphonologie: Eine Analyse deutscher Intonationskonturen im Rahmen der nicht-linearen Phonologie.* Tübingen: Niemeyer.

Vilà-Giménez, Ingrid & Pilar Prieto. 2021. The Value of Non-Referential Gestures: A Systematic Review of Their Cognitive and Linguistic Effects in Children's Language Development. *Children (Basel, Switzerland)* 8(2). 148.

Wilcox, Rand R. 2012. *Introduction to robust estimation and hypothesis testing,* 3rd edn. Amsterdam, Boston: Academic Press.

Zerbian, Sabine & Marlene Böttcher. 2019. Stressed pronouns in German mono- and bilingual speech. 2640–2644.

Appendix A - Road map to Information Structure annotation according to GÖTZE et al. 2007

Information Status - p. 160-161

Q1: Has the referent been mentioned in the previous discourse?

- yes: label expression as giv! - no: go to Q2!

Q2: Is the referent a physical part of the utterance situation?

- yes: label expression as acc! - no: go to Q3!

Q3: Is the referent accessible (1) via some kind of relation to other referents in the previous discourse, (2) from assumed world knowledge, or (3) by denoting a group consisting of accessible or given discourse referents?

- yes: go to Q4! / label expression as acc - no: label expression as new!

Q4: Does the referring expression denote a group consisting of accessible or given discourse referents?

- yes: label element as acc! - no: go to Q5!

Q5: Is the referent inferable from a referent in the previous discourse by some relation as specified in section 2.2.5 under 'Inferable (acc-inf)'?

- yes: label element as acc! - no: go to Q6!

Q6: Is the referent assumed to be inferable from assumed world knowledge?

- yes: label element as acc!

Focus - p. 185

Q1: Is the sentence a declarative or a non-declarative one?

- if non-declarative (imperative, question): go to Q3 - if declarative: go to Q2

Q2: Does the utterance complete an explicit wh-question?

- Yes: the constituent which is congruent to the wh-word is to be annotated “nf-sol” - No: go to Q3

Q3: Does a constituent of the utterance (or the utterance as a whole) evoke the notion of contrast to another constituent in previous context?

- Yes: annotate it for “cf” – for further annotation go to Q4 - No: go to Q5

Q4: Does the context enable you to further specify the contrastive relation according to the inventory given in 4.3.2?

- Yes: annotate according to the inventory given in 4.3.2. - No: restrict the annotation to “cf”

Q5: Which part of the utterance reveals the new and most important information in discourse? Try to identify the domain by asking implicit questions as done in the example in 4.2.2!

- annotate the identified costituent or domain as “nf-unsol”

Q6: Is it possible to add to the utterance a formula like “It is true / It is not true ...”, “Is it true / Is it not true ...?” to the respective proposition without changing its meaning/function within the discourse?

- Yes: annotate it as “cf-ver” according to 4.3.2.5 - No: no additional specification is necessary

Q7: Does the sentence contain a focus operator?

- Yes: annotate the constituent that is bound by it for “+op” - No: no additional specification is necessary

Appendix B - Complete R-Script for Descriptive and Inferential Statistics in this thesis

```
library(ggplot2)
library(caret)
library(corrplot)

# Chapter 4
# DataMAthesisAlt
dataAlt <- read.table (file.choose(), header=TRUE, sep = ";")

# Stacked bar plot Pitch accent & Information Status mit Pronomen
dataAlt$Pitch.accent <- factor(dataAlt$Pitch.accent , levels=c("unaccented", "L*", "!H*", "H+L*", "H*+L",
"H*", "L*+H", "L+H*"))
dataAlt$Information.Status <- factor(dataAlt$Information.Status , levels=c("giv", "acc", "new"))
PitchInfoAlt <- data.frame(table(dataAlt$Pitch.accent, dataAlt$Information.Status))
names(PitchInfoAlt) <- c("PitchAccent","InformationStatus","Percent")
ggplot(PitchInfoAlt, aes(x=InformationStatus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chapter 4.1
# DataMAthesisDis
dataDis <- read.table (file.choose(), header=TRUE, sep = ";")

# Distribution of gestures
dataDis$GestureGen <- factor(dataDis$GestureGen , levels=c("NoGesture", "Apex", "Stroke"))
plot(dataDis$GestureGen, col = "purple", names = c("None", "Apex", "Stroke" ), main = "Gestures")
dataDis$Discourse.Gesture <- factor(dataDis$Discourse.Gesture , levels=c("ApexNoIS", "ApexonIS",
"StrokeonIS"))
plot(dataDis$Discourse.Gesture, col = "purple", names = c("Apex, no IS", "Apex on IS", "Stroke on IS" ), main
= "Gesture parts")

# Distribution of pitch accents
dataDis$Pitch.accent <- factor(dataDis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*",
"L+H*"))
plot(dataDis$Pitch.accent, col = "purple", names = c("None", "L*", "!H*", "H*+L", "H*", "L+H*"), main =
"Pitch Accents")

# Distribution of information status
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("giv", "acc", "new"))
plot(dataDis$Information.Status, col = "orange", names = c("Given", "Accessible", "New"), main = "Information
Status")

# Distribution of focus
dataDis$FocusGen <- factor(dataDis$FocusGen , levels=c("NoFocus", "Focus"))
plot(dataDis$FocusGen, col = "orange", names = c("No focus", "focus"), main = "Focus general")
dataDis$Focus <- factor(dataDis$Focus , levels=c("nf", "cf"))
plot(dataDis$Focus, col = "orange", names = c("new-information", "contrastive"), main = "Focus categories")


# Chapter 4.1.1
# Pitch accents and information status
# Percent stacked bar plots pitch accents and information status
dataDis$Pitch.accent <- factor(dataDis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*",
"L+H*"))
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("giv", "acc", "new"))
PitchInfo <- data.frame(table(dataDis$Pitch.accent, dataDis$Information.Status))
names(PitchInfo) <- c("PitchAccent","InformationStatus","Percent")
ggplot(PitchInfo, aes(x=InformationStatus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chi-Square Test pitch accents and information status + unaccented
```

```
datapi.CHI <- dataDis[ ,c(2,4)]
dataPI.CHI=table(datapi.CHI)
CHIpi=chisq.test(dataPI.CHI)
CHIpi
CHIpi$res
corrplot(CHIpi$res, is.corr = FALSE)

# Chi-Square Test pitch accents and information status - unaccented
dataDis$Pitch.accent <- factor(dataDis$Pitch.accent , levels=c("L*", "!H*", "H*+L", "H*", "L+H*"))
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("giv", "acc", "new"))
data2pi.CHI <- dataDis[ ,c(2,4)]
data2PI.CHI=table(data2pi.CHI)
CHIpi2=chisq.test(data2PI.CHI)
CHIpi2
CHIpi2$res
corrplot(CHIpi2$res, is.corr = FALSE)

# Pitch accents and focus
# Percent stacked bar plots Pitch accents and focus absence/presence
dataDis$Pitch.accent <- factor(dataDis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
dataDis$FocusGen <- factor(dataDis$FocusGen , levels=c("NoF", "Foc"))
PitchFocus <- data.frame(table(dataDis$Pitch.accent, dataDis$FocusGen))
names(PitchFocus) <- c("PitchAccent","Focus","Percent")
ggplot(PitchFocus, aes(x=Focus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chi-Square Test pitch accents and focus general
datapfg.CHI <- dataDis[ ,c(2,6)]
dataPFG.CHI=table(datapfg.CHI)
CHIpfg=chisq.test(dataPFG.CHI)
CHIpfg
CHIpfg$res
corrplot(CHIpfg$res, is.corr = FALSE)

# Percent stacked bar plots Pitch accents and focus categories
dataDis$Focus <- factor(dataDis$Focus , levels=c("nf", "cf"))
PitchFocus2 <- data.frame(table(dataDis$Pitch.accent, dataDis$Focus))
names(PitchFocus2) <- c("PitchAccent","Focus","Percent")
ggplot(PitchFocus2, aes(x=Focus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chi-Square Test pitch accents and focus categories
data2pf.CHI <- dataDis[ ,c(2,5)]
data2PF.CHI=table(data2pf.CHI)
CHI2pf=chisq.test(data2PF.CHI)
CHI2pf
CHI2pf$res
corrplot(CHI2pf$res, is.corr = FALSE)


# Chapter 4.1.2
# Gestures and information status
# Percent stacked bar plots Gestures and information status
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("NoInfo", "giv", "acc", "new"))
dataDis$Discourse.Gesture <- factor(dataDis$Discourse.Gesture , levels=c("NoGesture", "ApexNoIS", "StrokeonIS", "ApexonIS"))
GestureInfo <- data.frame(table(dataDis$Discourse.Gesture, dataDis$Information.Status))
names(GestureInfo) <- c("Gesture","InformationStatus","Percent")
ggplot(GestureInfo, aes(x=InformationStatus, y=Percent, fill=Gesture)) + geom_col(position = "fill")

# Chi-Square Test gestures and information status
datagi.CHI <- dataDis[ ,c(3,4)]
dataGI.CHI=table(datagi.CHI)
CHIgi=chisq.test(dataGI.CHI)
```

```
CHIgi
CHIgi$res
corrplot(CHIgi$res, is.corr = FALSE)

# Gestures and focus
# Percent stacked bar plots gestures and focus
dataDis$Focus <- factor(dataDis$Focus , levels=c("NoFocus", "nf", "cf"))
dataDis$Discourse.Gesture <- factor(dataDis$Discourse.Gesture , levels=c("NoGesture", "ApexNoIS", "StrokeonIS", "ApexonIS"))
GestureFocus <- data.frame(table(dataDis$Discourse.Gesture, dataDis$Focus))
names(GestureFocus) <- c("Gesture","Focus","Percent")
ggplot(GestureFocus, aes(x=Focus, y=Percent, fill=Gesture)) + geom_col(position = "fill")

# Chi-Square Test gestures and focus
datagf.CHI <- dataDis[ ,c(3,5)]
dataGF.CHI=table(datagf.CHI)
CHIgf=chisq.test(dataGF.CHI)
CHIgf
CHIgf$res
corrplot(CHIgf$res, is.corr = FALSE)


# Chapter 4.1.3
# Pitch accents and gestures
# Stacked bar plots pitch accents and gestures
dataDis$Pitch.accent <- factor(dataDis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
dataDis$GestureGen <- factor(dataDis$GestureGen , levels=c("NoGesture", "Stroke", "Apex"))
PitchGesture <- data.frame(table(dataDis$Pitch.accent, dataDis$GestureGen))
names(PitchGesture) <- c("PitchAccent","Gesture","Percent")
ggplot(PitchGesture, aes(x=Gesture, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chi-Square Test pitch accents and gestures
datagp.CHI <- dataDis[ ,c(2,7)]
dataGP.CHI=table(datagp.CHI)
CHIgp=chisq.test(dataGP.CHI)
CHIgp
CHIgp$res
corrplot(CHIgp$res, is.corr = FALSE)


# Chapter 4.2
# Chapter 4.2.1 gestures given
dataDisApex <- subset(dataDis, dataDis$GestureGen == "Apex")
summary(dataDisApex)
dataDisAis <- subset(dataDis, dataDis$Discourse.Gesture == "ApexonIS")
summary(dataDisAis)

# Pitch accent Distribution with apex
dataDisApex$Pitch.accent <- factor(dataDisApex$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
plot(dataDisApex$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"), main = "Pitch Accents: Apex produced")
dataDisAis$Pitch.accent <- factor(dataDisAis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
plot(dataDisAis$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"), main = "Pitch Accents: Apex produced on IS")

# Information status Distribution with apex
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("giv", "acc", "new"))
plot(dataDis$Information.Status, col = "orange", names = c("Given", "Accessible", "New"), main = "Information Status")
dataDisAis$Information.Status <- factor(dataDisAis$Information.Status , levels=c("giv", "acc", "new"))
```

```
plot(dataDisAis$Information.Status, col = "orange", names = c("Given", "Accessible", "New"), main =
"Information Status: Apex produced non IS")

# Focus Distribution with apex
dataDis$FocusGen <- factor(dataDis$FocusGen , levels=c("NoF", "Foc"))
plot(dataDis$FocusGen, col = "orange", names = c("No Focus", "Focus"), main = "Focus")
dataDisApex$FocusGen <- factor(dataDisApex$FocusGen , levels=c("NoF", "Foc"))
plot(dataDisApex$FocusGen, col = "orange", names = c("No Focus", "Focus"), main = "Focus: Apex produced")
dataDisAis$FocusGen <- factor(dataDisAis$FocusGen , levels=c("NoF", "Foc"))
plot(dataDisAis$FocusGen, col = "orange", names = c("No Focus", "Focus"), main = "Focus: Apex produced on
IS")

# Stacked bar plot pitch accents and information status --> apex on IS produced
dataDisAis$Information.Status <- factor(dataDisAis$Information.Status , levels=c("giv", "acc", "new"))
dataDisAis$Pitch.accent <- factor(dataDisAis$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L",
"H*", "L+H*"))
AisPitchInfo <- data.frame(table(dataDisAis$Pitch.accent, dataDisAis$Information.Status))
names(AisPitchInfo) <- c("PitchAccent","InformationStatus","Percent")
ggplot(AisPitchInfo, aes(x=InformationStatus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# CHi-Square Test pitch accents and information status --> apex on IS produced
Aisdatapi.CHI <- dataDisAis[ ,c(2,4)]
AisdataPI.CHI=table(Aisdatapi.CHI)
AisCHIpi=chisq.test(AisdataPI.CHI)
AisCHIpi
AisCHIpi$res
corrplot(AisCHIpi$res, is.corr = FALSE)

# Stacked bar plot pitch accents and focus --> apex on IS produced
dataDisAis$FocusGen <- factor(dataDisAis$FocusGen , levels=c("NoF", "Foc"))
AisPitchFocusGen <- data.frame(table(dataDisAis$Pitch.accent, dataDisAis$FocusGen))
names(AisPitchFocusGen) <- c("PitchAccent","Focus","Percent")
ggplot(AisPitchFocusGen, aes(x=Focus, y=Percent, fill=PitchAccent)) + geom_col(position = "fill")

# Chi-Square Test pitch accents and focus --> apex on IS produced
Aisdatapfgen.CHI <- dataDisAis[ ,c(2,6)]
AisdataPFgen.CHI=table(Aisdatapfgen.CHI)
AisCHIpfgen=chisq.test(AisdataPFgen.CHI)
AisCHIpfgen
AisCHIpfgen$res
corrplot(AisCHIpfgen$res, is.corr = FALSE)
## hier fehlt noch mit kategorien der nicht signifikante


# Chapter 4.2.2 information status given
dataDisGiv <- subset(dataDis, dataDis$Information.Status == "giv")
summary(dataDisGiv)
dataDisAcc <- subset(dataDis, dataDis$Information.Status == "acc")
summary(dataDisAcc)
dataDisNew <- subset(dataDis, dataDis$Information.Status == "new")
summary(dataDisNew)

# Pitch accent distribution --> information status specified
dataDisGiv$Pitch.accent <- factor(dataDisGiv$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L",
"H*", "L+H*"))
plot(dataDisGiv$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"),
main = "Pitch Accents: 'Given'")
dataDisAcc$Pitch.accent <- factor(dataDisAcc$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L",
"H*", "L+H*"))
plot(dataDisAcc$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"),
main = "Pitch Accents: 'Accessible'")
dataDisNew$Pitch.accent <- factor(dataDisNew$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L",
"H*", "L+H*"))
```

```
plot(dataDisNew$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"),
main = "Pitch Accents: 'New'")

# Gesture distribution --> information status specified
dataDis$Discourse.Gesture <- factor(dataDis$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDis$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main = "Gestures")
dataDisGiv$Discourse.Gesture <- factor(dataDisGiv$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDisGiv$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main =
"Gestures: 'Given'")
dataDisAcc$Discourse.Gesture <- factor(dataDisAcc$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDisAcc$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main =
"Gestures: 'Accessible'")
dataDisNew$Discourse.Gesture <- factor(dataDisNew$Discourse.Gesture , levels=c("ApexonIS",
"StrokeonIS"))
plot(dataDisNew$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main =
"Gestures: 'New'")

# Pitch accents & gestures --> information status specified
# Stacked bar plot giv
GGesturePitch <- data.frame(table(dataDisGiv$Discourse.Gesture, dataDisGiv$Pitch.accent))
names(GGesturePitch) <- c("Gesture","PitchAccent","Percent")
ggplot(GGesturePitch, aes(x=PitchAccent, y=Percent, fill=Gesture)) + geom_col(position="fill")

# Chi-Square Test giv
dataGivgp.CHI <- dataDisGiv[ ,c(2,3)]
dataGivGP.CHI=table(dataGivgp.CHI)
CHIGivgp=chisq.test(dataGivGP.CHI)
CHIGivgp
CHIGivgp$res
corrplot(CHIGivgp$res, is.corr = FALSE)

# Stacked bar plot acc
AcGesturePitch <- data.frame(table(dataDisAcc$Discourse.Gesture, dataDisAcc$Pitch.accent))
names(AcGesturePitch) <- c("Gesture","PitchAccent","Percent")
ggplot(AcGesturePitch, aes(x=PitchAccent, y=Percent, fill=Gesture)) + geom_col(position="fill")

# Chi-Square Test acc
dataAccgp.CHI <- dataDisAcc[ ,c(2,3)]
dataAccGP.CHI=table(dataAccgp.CHI)
CHIAccgp=chisq.test(dataAccGP.CHI)
CHIAccgp
CHIAccgp$res
corrplot(CHIAccgp$res, is.corr = FALSE)

# Stacked bar plot new
NGesturePitch <- data.frame(table(dataDisNew$Discourse.Gesture, dataDisNew$Pitch.accent))
names(NGesturePitch) <- c("Gesture","PitchAccent","Percent")
ggplot(NGesturePitch, aes(x=PitchAccent, y=Percent, fill=Gesture)) + geom_col(position="fill")

# Chi-Square Test new
dataNewgp.CHI <- dataDisNew[ ,c(2,3)]
dataNewGP.CHI=table(dataNewgp.CHI)
CHINewgp=chisq.test(dataNewGP.CHI)
CHINewgp
CHINewgp$res
corrplot(CHINewgp$res, is.corr = FALSE)


# Chapter 4.2.3 focus given
dataDisFoc <- subset(dataDis, dataDis$FocusGen == "Foc")
summary(dataDisFoc)
dataDisNF <- subset(dataDis, dataDis$FocusGen == "NoF")
summary(dataDisNF)
```

```
# Pitch accent distribution --> focus specified
dataDisNF$Pitch.accent <- factor(dataDisNF$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
plot(dataDisNF$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"), main = "Pitch Accents: No Focus")
dataDisFoc$Pitch.accent <- factor(dataDisFoc$Pitch.accent , levels=c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"))
plot(dataDisFoc$Pitch.accent, col = "purple", names = c("Unaccented", "L*", "!H*", "H*+L", "H*", "L+H*"), main = "Pitch Accents: Focus")

# Gesture distribution --> focus specified
dataDis$Discourse.Gesture <- factor(dataDis$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDis$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main = "Gestures")
dataDisNF$Discourse.Gesture <- factor(dataDisNF$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDisNF$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main = "Gestures: No Focus")
dataDisFoc$Discourse.Gesture <- factor(dataDisFoc$Discourse.Gesture , levels=c("ApexonIS", "StrokeonIS"))
plot(dataDisFoc$Discourse.Gesture, col = "purple", names = c("Apex on IS", "Stroke on IS" ), main = "Gestures: Focus")

# Pitch accents and gestures --> focus specified
# Stacked bar plot No Focus
NFGesturePitch <- data.frame(table(dataDisNF$Discourse.Gesture, dataDisNF$Pitch.accent))
names(NFGesturePitch) <- c("Gesture","PitchAccent","Count")
ggplot(NFGesturePitch, aes(x=PitchAccent, y=Count, fill=Gesture)) + geom_col(position="fill")

# Chi-Square Test No Focus
dataNFgp.CHI <- dataDisNF[ ,c(2,3)]
dataNFGP.CHI=table(dataNFgp.CHI)
CHINFgp=chisq.test(dataNFGP.CHI)
CHINFgp
CHINFgp$res
corrplot(CHINFgp$res, is.corr = FALSE)

# Stacked bar plot Focus
FocGesturePitch <- data.frame(table(dataDisFoc$Discourse.Gesture, dataDisFoc$Pitch.accent))
names(FocGesturePitch) <- c("Gesture","PitchAccent","Count")
ggplot(FocGesturePitch, aes(x=PitchAccent, y=Count, fill=Gesture)) + geom_col(position="fill")

# Chi-Square Test Focus
dataFocgp.CHI <- dataDisFoc[ ,c(2,3)]
dataFocGP.CHI=table(dataFocgp.CHI)
CHIFocgp=chisq.test(dataFocGP.CHI)
CHIFocgp
CHIFocgp$res
corrplot(CHIFocgp$res, is.corr = FALSE)


# Chapter 4.3
# Stacked bar plot of occurrences of pitch accents, apices and together under information status
dataDis$Information.Status <- factor(dataDis$Information.Status , levels=c("NoInfo", "giv", "acc", "new"))
SumStatus <- data.frame(table(dataDis$Summary, dataDis$Information.Status))
names(SumStatus) <- c("TotalDistribution","InformationStatus","Percent")
ggplot(SumStatus, aes(x=InformationStatus, y=Percent, fill=TotalDistribution)) + geom_col(position="fill")

# Chi-Square Test occurrences of pitch accents, apices and together under information status
dataSum.CHI <- dataDis[ ,c(4,8)]
dataSUM.CHI=table(dataSum.CHI)
CHIsum=chisq.test(dataSUM.CHI)
CHIsum
CHIsum$res
corrplot(CHIsum$res, is.corr = FALSE)
```

```
# Stacked bar plot of occurrences of pitch accents, apices and together under focus
dataDis$FocusGen <- factor(dataDis$FocusGen , levels=c("NoF", "Foc"))
SumFocus <- data.frame(table(dataDis$Summary, dataDis$FocusGen))
names(SumFocus) <- c("TotalDistribution","Focus","Percent")
ggplot(SumFocus, aes(x=Focus, y=Percent, fill=TotalDistribution)) + geom_col(position="fill")

# Chi-Square Test occurrences of pitch accents, apices and together under focus
dataSum.CHI <- dataDis[ ,c(8,6)]
dataSUM.CHI=table(dataSum.CHI)
CHIsum=chisq.test(dataSUM.CHI)
CHIsum
CHIsum$res
corrplot(CHIsum$res, is.corr = FALSE)




# Chapter 5
dataTemp <- read.table (file.choose(), header=TRUE, sep = ";")

# Chapter 5.1 general
# histogram general deviation
hist(dataTemp$Dist3sec, breaks = 81, main = "Grouped distance accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-4000, 4000, 8))
mean(dataTemp$DistanceMS)
sd(dataTemp$DistanceMS)

# test normal distribution and significance general deviation
qqnorm(dataTemp$Dist3sec, main = "qq-plot general distribution")
qqline(dataTemp$Dist3sec)
t.test(dataTemp$DistanceMS, mu = 0)

# Chapter 5.2 information status
dataTempGiv <- subset(dataTemp, dataTemp$InformationStatus == "giv")
summary(dataTempGiv)
dataTempAcc <- subset(dataTemp, dataTemp$InformationStatus == "acc")
summary(dataTempAcc)
dataTempNew <- subset(dataTemp, dataTemp$InformationStatus == "new")
summary(dataTempNew)
dataTempNoi <- subset(dataTemp, dataTemp$InformationStatus == "NoInfo")
summary(dataTempNoi)

# Information status: "given" deviation
hist(dataTempGiv$Dist3sec, breaks =31, main = "'Given' referents deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-500, 750, 5))
mean(dataTempGiv$DistanceMS)
sd(dataTempGiv$DistanceMS)

# test normal distribution and significance deviation "given"
qqnorm(dataTempGiv$DistanceMS, main = "qq-plot 'given' referents deviation")
qqline(dataTempGiv$DistanceMS)
t.test(dataTempGiv$DistanceMS, mu = 0)

# Information status: "accessible" deviation
hist(dataTempAcc$Dist3sec, breaks =31, main = "'Accessible' referents deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-2000, 2000, 8))
mean(dataTempAcc$DistanceMS)
sd(dataTempAcc$DistanceMS)

# test normal distribution and significance deviation "accessible"
qqnorm(dataTempAcc$DistanceMS, main = "qq-plot 'accessible' referents deviation")
```

```
qqline(dataTempAcc$DistanceMS)
t.test(dataTempAcc$DistanceMS, mu = 0)

# Information status: "new" deviation
hist(dataTempNew$Dist3sec, breaks =31, main = "'New' referents deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-1000, 750, 7))
mean(dataTempNew$DistanceMS)
sd(dataTempNew$DistanceMS)

# test normal distribution and significance deviation "new"
qqnorm(dataTempNew$DistanceMS, main = "qq-plot 'new' referents deviation")
qqline(dataTempNew$DistanceMS)
t.test(dataTempNew$DistanceMS, mu = 0)

# Information status: "no information status" deviation
hist(dataTempNoi$Dist3sec, breaks =31, main = "No information status deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-3000, 3000, 10))
mean(dataTempNoi$DistanceMS)
sd(dataTempNoi$DistanceMS)

# test normal distribution and significance deviation "no information status"
qqnorm(dataTempNoi$DistanceMS, main = "qq-plot No information status deviation")
qqline(dataTempNoi$DistanceMS)
t.test(dataTempNoi$DistanceMS, mu = 0)

# Comparison between all levels
boxplot(dataTemp$DistanceMS ~ dataTemp$InformationStatus, xlab = "Information status levels", names = c("Given", "Accessible", "New", "No info"), ylab = "Pitch accent deviation in ms")
ANOVA <- aov(dataTemp$DistanceMS ~ dataTemp$InformationStatus, data = dataTemp)
summary(ANOVA)

# Chapter 5.3 focus
dataTempFoc <- subset(dataTemp, dataTemp$FocusGen == "Focus")
summary(dataTempFoc)
dataTempNF <- subset(dataTemp, dataTemp$FocusGen == "NoFocus")
summary(dataTempNF)

# Focus deviation
hist(dataTempFoc$Dist3sec, breaks =31, main = "Focus deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-1000, 2000, 6))
mean(dataTempFoc$DistanceMS)
sd(dataTempFoc$DistanceMS)

# test normal distribution and significance deviation "focus"
qqnorm(dataTempFoc$DistanceMS, main = "qq-plot focus deviation")
qqline(dataTempFoc$DistanceMS)
t.test(dataTempFoc$DistanceMS, mu = 0)

# No focus deviation
hist(dataTempNF$Dist3sec, breaks =31, main = "No focus deviation accent/apex", xlab = "Pitch accent deviation in ms", col = "lightblue", xaxp = c(-3000, 3000, 10))
mean(dataTempNF$DistanceMS)
sd(dataTempNF$DistanceMS)

# test normal distribution and significance deviation "focus"
qqnorm(dataTempNF$DistanceMS, main = "qq-plot No focus deviation")
qqline(dataTempNF$DistanceMS)
t.test(dataTempNF$DistanceMS, mu = 0)

# Comparison focus
boxplot(dataTemp$DistanceMS ~ dataTemp$FocusGen, xlab = "Focus presence", names = c("Focus", "No focus"), ylab = "Pitch accent deviation in ms")
t.test(dataTemp$DistanceMS ~ dataTemp$FocusGen)
```

Appendix C - Residuals of all Chi-Square tests in this thesis

- **Pitch accents and information status including unaccented words**

X-squared = 408.24, df = 10, p-value < 2.2e-16

	Given	Accessible	New
Unaccented	-6.2215442	16.2194364	-6.7467584
L*	0.6963754	-1.6556073	0.6254093
!H*	0.5650205	0.8963755	-1.3107963
H*+L	0.9761600	-1.0851559	-0.1264325
H*	2.2056327	-2.2737276	-0.4303262
L+H*	-2.3560710	-3.2705113	5.0865279

- **Pitch accents and information status without unaccented words**

X-squared = 47.161, df = 8, p-value = 1.429e-07

	Given	Accessible	New
L*	0.16472716	-0.29248943	0.05082340
!H*	-0.02299026	2.57114863	-1.91469021
H*+L	0.65932956	-0.30322339	-0.45056532
H*	1.42324523	-0.31025564	-1.23215945
L+H*	-2.79660896	-2.12199563	4.48049481

- **Pitch accents and focus presence**

X-squared = 853.84, df = 5, p-value < 2.2e-16

	No Focus	Focus
Unaccented	19.583435	-17.644188
L*	-3.329241	2.999563
!H*	-1.899909	1.711771
H*+L	-4.401236	3.965405
H*	-1.709505	1.540222
L+H*	-7.126051	6.420395

- **Pitch accents and focus categories**

X-squared = 26.767, df = 5, p-value = 6.332e-05

	new-information focus	contrastive focus
Unaccented	0.4821261	-1.1285223
L*	-0.1070427	0.2505571
!H*	-0.6462874	1.5127780
H*+L	-0.3945322	0.9234897
H*	-0.6271803	1.4680537
L+H*	1.7090322	-4.0003665

- **Gestures and information status**

X-squared = 2078.3, df = 9, p-value < 2.2e-16

	No Information status	Given	Accessible	New
No Gesture	-10.44248789	5.10944864	-0.03250714	3.79902565
Apex No IS	36.53698591	-11.99007684	-9.43209261	-11.61838470
Stroke on IS	-3.79323124	-3.18478564	12.50578091	-3.58003862
Apex on IS	-4.06984299	-0.80514519	0.65037330	3.82831382

- **Gestures and information status**

X-squared = 710.34, df = 6, p-value < 2.2e-16

	No Focus	new-information focus	contrastive focus
No Gesture	-6.4647395	5.3226766	2.3668598
Apex No IS	17.8576262	-14.7956153	-6.3209665
Stroke on IS	4.7079051	-4.1711322	-1.0333086
Apex on IS	-2.9125291	2.7756122	0.1824477

- **Gestures and pitch accents**

X-squared = 2577.5, df = 10, p-value < 2.2e-16

	NoGesture	Stroke	Apex
Unaccented	-21.1379662	14.5678147	39.9009305
L*	3.7993676	-2.9993810	-6.9585393

!H*	3.7232960	-2.4697558	-7.0821525
H*+L	1.5912178	-0.3441308	-3.4250167
H*	4.5922871	-2.9033486	-8.8150557
L+H*	3.5582573	-2.9953405	-6.4126253

- **Pitch accents and information status when an apex is produced**

X-squared = 62.708, df = 10, p-value = 1.11e-09

	Given	Accessible	New
Unaccented	-2.25270830	6.17923812	-2.52843178
L*	1.97121499	-1.77555495	-0.37602752
!H*	-0.07832098	-0.75570865	0.60388425
H*+L	0.87289748	0.16012815	-0.84108230
H*	-0.04396203	-1.30744090	0.96853754
L+H*	-0.69170359	0.02441931	0.55864190

- **Pitch accents and focus when an apex is produced**

X-squared = 38.738, df = 5, p-value = 2.681e-07

	No Focus	Focus
Unaccented	4.6282951	-3.2967927
L*	-1.1394516	0.8116456
!H*	-0.6961766	0.4958953
H*+L	-1.4383739	1.0245717
H*	0.2210289	-0.1574417
L+H*	-0.6139700	0.4373386

- **Pitch accents and gesture on information status given**

X-squared = 3.0482, df = 5, p-value = 0.6925

	Apex on IS	Stroke on IS
Unaccented	0.43946350	-0.59446065
L*	0.77750669	-1.05173041
!H*	0.04584356	-0.06201241
H*+L	-0.27261874	0.36877035

H*	-0.40992240	0.55450050
L+H*	-0.18733013	0.25340077

- **Pitch accents and gesture on information status accessible**

X-squared = 58.657, df = 5, p-value = 2.302e-11

	Apex on IS	Stroke on IS
Unaccented	-3.573186	2.477937
L*	1.469726	-1.019227
!H*	1.906445	-1.322083
H*+L	1.469726	-1.019227
H*	3.313147	-2.297605
L+H*	2.812307	-1.950282

- **Pitch accents and gesture on information status new**

X-squared = 2.7731, df = 5, p-value = 0.7349

	Apex on IS	Stroke on IS
Unaccented	0.39668175	-0.69843030
L*	-0.43757432	0.77042909
!H*	0.14959806	-0.26339456
H*+L	-0.42643288	0.75081257
H*	0.01381069	-0.02431622
L+H*	0.35088007	-0.61778811

- **Pitch accents and gesture on no focus**

X-squared = 66.124, df = 5, p-value = 6.551e-13

	Apex on IS	Stroke on IS
Unaccented	-4.37030266	3.52859312
L*	2.08813719	-1.68596711
!H*	1.93781425	-1.56459599
H*+L	-0.08840972	0.07138222
H*	2.84261573	-2.29513494
L+H*	2.17387442	-1.75519156

- **Pitch accents and gesture on focus**

X-squared = 0.24464, df = 5, p-value = 0.9986

	Apex on IS	Stroke on IS
Unaccented	-0.078796885	0.121772047
L*	-0.158100519	0.244327219
!H*	-0.016543056	0.025565500
H*+L	0.020261022	-0.031311215
H*	-0.009191055	0.014203780
L+H*	0.200573889	-0.309965210

- **Occurrences of pitch accents, apices and both together under information status**

X-squared = 2142.9, df = 9, p-value < 2.2e-16

	Accent & Apex	Accent & no Apex	no Accent & Apex	None
No Information status	-3.6178663	-10.6210987	35.2476639	-3.3162607
Given	-0.2201584	5.1979357	-12.1865589	-5.0243660
Accessible	-1.3449812	-0.1588245	-7.6820263	17.1624833
New	4.4529771	3.9649724	-11.7195211	-5.8751971

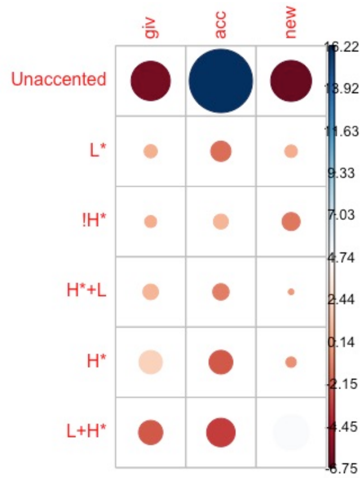
- **Occurrences of pitch accents, apices and both together under focus**

X-squared = 808.72, df = 3, p-value < 2.2e-16

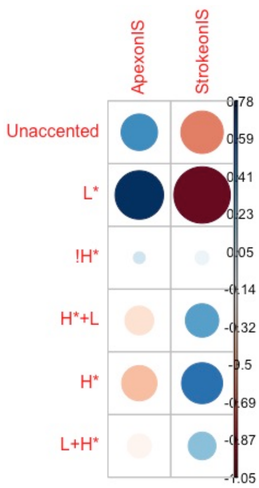
	Accent & Apex	Accent & no Apex	no Accent & Apex	None
No Focus	-3.922402	-6.643041	18.136820	7.610208
Focus	3.533987	5.985215	-16.340824	-6.856608

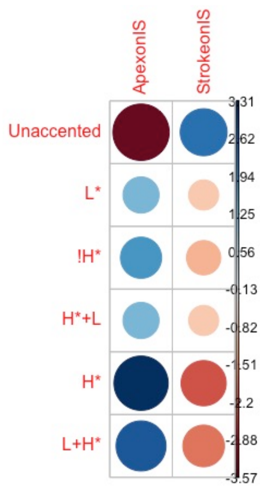
Appendix D - Additional plots that were not included in the results

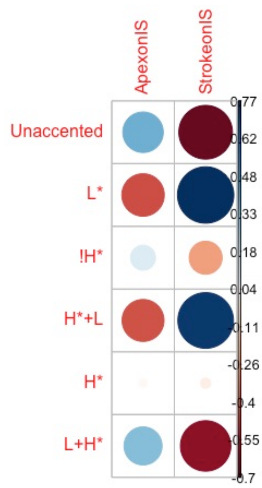
- Correlation plot on the correlation of pitch accents and information status levels including unaccented words, which had a correlation much stronger than all remaining correlations. In connection to Figure 22/ p. 45.



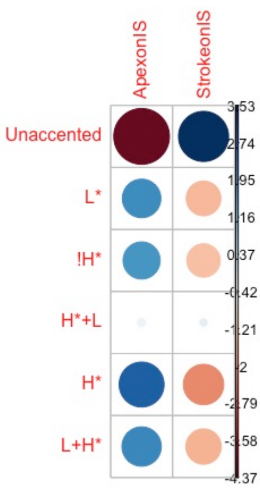
- Correlation plots to the alignment of pitch accents and gesture apices under different levels of information status. Left: „Given“, center: „Accessible“, right: „New“. Figure 42/ p. 55.

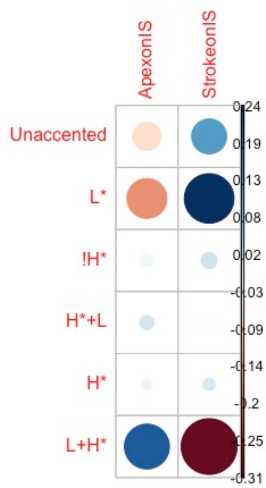






- Correlation plots to the alignment of pitch accents and gesture apices under focus absence (left) and presence (right). Figure 45/ p. 57.





Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die im Literaturverzeichnis angegebenen Quellen benutzt habe. Ich habe die Regeln der guten wissenschaftlichen Praxis eingehalten.

Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder noch nicht veröffentlichten Quellen entnommen sind, sind als solche kenntlich gemacht.

Die Zeichnungen oder Abbildungen in dieser Arbeit sind von mir selbst erstellt worden oder mit einem entsprechenden Quellennachweis versehen.

Diese Arbeit ist in gleicher oder ähnlicher Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Frankfurt, den 29.08.2022

Unterschrift