Prosodic Decoding and Encoding of Referential Givenness in Adults with Autism Spectrum Disorders

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophischen Fakultät
der Universität zu Köln
im Fach Phonetik

vorgelegt von

Martina Krüger

geb. am 2. Juli 1984

in Wesel (NRW), Deutschland

Köln, April 2018

Erste Referentin: Prof. Dr. Martine Grice

Zweiter Referent: Prof. Dr. Dr. Kai Vogeley

Dritter Referent: PD Dr. Stefan Baumann

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Prosody plays an essential role in human social communication, where it serves to enhance or modify the meaning of what is said. Speakers use prosody to express multiple functions simultaneously - grammatical, pragmatic, and affective - making it a highly complex feature of speech.

Successful communication requires that all participants involved share the ability to correctly perceive and interpret the prosody of another person's speech as well as to use prosody appropriately to convey meaning to others.
In Autism Spectrum Disorders (ASD), in which social communication and interaction are impaired (American Psychiatric Association, 2013), unusual prosody has been frequently reported since the earliest descriptions by Kanner (1943) and Asperger (1944). However, in spite of extensive documentation of prosody as a feature of impaired communication in ASD, the research of prosody in ASD is limited with regard to the following critical points:
First, the literature on prosodic deficits in ASD provides little differentiation between studies that investigated children with ASD and those that investigated adults. In fact, the majority of research focussing on language impairments in ASD evaluated children rather than adults, and these studies have frequently been referred to when characterising language impairments in ASD in general. Results of studies with children with ASD might, however, not be indicative for ASD in adulthood, since many aspects of language continue to develop until adulthood, among them prosody (Chen, 2011a; Cruttenden, 1985; Diehl, Bennetto, Watson, Gunlogson, and McDonough, 2008; Berman, 2008; Nippold, 2007).

Second, many of the studies used subjective ratings rather than objective mea-

sures to describe the prosody of individuals with ASD, or they used poorly defined prosodic categories (see also Green and Tobin, 2009). There has not therefore been a great deal of research that has investigated prosody in ASD within phonological models of prosody.
Third, tasks often involved unnatural settings, read speech, or imitation. All of these aspects can lead to problems in prosody comprehension and production.
Fourth, some functions of prosody still remain unexplored, such as the encoding and decoding of referential givenness, a key function of prosody to indicate who or what a speaker is talking about by highlighting words that are new to the discourse (and to the listener) and attenuating those that are not. Marking referential givenness appropriately and understanding the cues to referential givenness both depend on the cognitive status of the referent as represented in the mental representations of the discourse participants. This suggests that mentalising capacities are required in this process, which makes this aspect of prosody especially interesting to investigate in individuals with ASD, a group that reportedly exhibits impaired mentalising skills.

This thesis aims to contribute to the current knowledge base of prosodic ability in individuals with ASD by counteracting these points of criticism in the following ways:
First, an extensive literature review will shed light on studies investigating prosodic deficits in ASD. Studies that investigated children with ASD will be reviewed separately from studies that investigated adults with ASD in order to identify which aspects of prosody are affected in which population.

Second, two experiments aim to contribute to our understanding of how adults with ASD encode and decode referential givenness.
The perception study investigates the ability of adults with ASD to perceive and interpret aspects of prosodic marking of referential givenness. This involves estimating to what extent referents are given in the current state of the discourse, based on the prosody with which the words and sentences are produced.
The production study complements the findings of the perception experiment by exploring the ability of adults with ASD to encode referential givenness in a story-telling-task. This task provides a structured elicitation of referential expressions and their prosodic marking in spontaneous speech. Within this task, participants refer to characters in a series of pictures in which new referents (characters) are introduced, maintained (a character stays in the foreground of the picture) and reactivated after a period of inactivity (a character reappearing in the foreground).

Both studies use the Autosegmental-Metrical framework to analyse prosody with a view to providing a phonologically informed account of the prosodic features that are prevalent in the speech of individuals with ASD as compared to neurotypical controls.

Outline

In order to provide the necessary background for the experiments, part I of the thesis introduces relevant aspects of prosody, referential givenness and autism.
In chapter 2, phonetic and phonological aspects of prosody are described (sections 2.1 and 2.2 respectively). Section 2.2 includes an introduction to Autosegmental-Metrical theory and the GToBI annotation system. The subsequent section provides an overview of the different functions that prosody can serve and how they are categorised into grammatical, pragmatic, and affective functions.
Chapter 3 is devoted to the notion of referential givenness. The first section (3.1) aims to characterise the notion of referential givenness within the more general framework of information structure. For that purpose, the three most commonly assumed dimensions of information structure are identified (information status, aboutness, and informativeness). The following section explores the concepts of identifiability and activation and how these are manifested in the choice of referring expression. Section 3.3 deals with referential givenness and how referring expressions are marked prosodically.
Chapter 4 deals with the characteristics of autism as a spectrum disorder. Section 4.1 provides a brief introduction into the history of autism and its current established classification. Section 4.2 constitutes a detailed literature review of research investigating prosodic impairments in autism. Studies in this section are divided into those concerning the perception of prosody and those concerning the production of prosody. Furthermore, studies that investigated children are distinguished from those that investigated adults. Each sub-section concludes with a synoptic table that summarises all studies mentioned.

Part II constitutes the empirical part of the thesis and presents two experiments that aim to investigate the decoding and encoding of referential givenness in autism.
Chapter 5 presents a perception experiment that tested the ability of adults with ASD to perceive aspects of prosodic marking of referential givenness.
Chapter 6 presents a production experiment that explored the ability of adults

with ASD to encode referential givenness in a story-telling-task.

Finally, part III comprises a summary of the results, the final conclusions and an outlook on future research.

Part I

BACKGROUND

Chapter 2

Prosody

The speech signal is made up of consecutive speech sounds (vowels and consonants) that form syllables, words, and sentences. These sounds are usually referred to as segments. However, there is more to speech than segments: The way in which words and sentences are produced involves features that usually span over stretches longer than one segment. These features, considered to operate above the segmental level, are called *suprasegmental* or *prosodic* features (e.g. Lehiste, 1970, 1976; Cruttenden, 1997). Prosody can be described as "a structure that organises the phonetic form of an utterance into larger units [...] and assigns prominence to units within this structure" (Diehl, Friedberg, Paul, and Snedeker, 2015: 868; see also Liberman and Prince, 1977).
The prosodic form of an utterance is shaped by the lexical content and syntactic structure of that utterance, its context in discourse as well as the speaker's intentions, emotional state and attitude, which is why it contains cues to syntactic, semantic, pragmatic, and affective interpretation of speech at once (see Diehl et al., 2015). Furthermore, prosodic features can be part of a word's lexical entry in many languages (e.g. lexical tone in languages such as Standard Chinese, Vietnamese, Cantonese, etc.; and lexical stress, as in English and German). Here the focus will be primarily on aspects of prosody which are assigned at the level of the utterance (also referred to as "postlexical").

It has been argued that the terms "prosody" and "intonation" have to be discriminated in a way that the term intonation refers to a specific part of prosody only, namely pitch variation (see, e.g. 't Hart, Collier, and Cohen, 1990), while the term prosody is used to refer to all suprasegmental phenomena. However, in-

tonation can also be understood in a broad sense, as an equivalent to prosody as defined above (see, e.g. Baumann, 2006; Grice, 2006). Grice (2006: 778) argues:

> "However, this distinction between prosody and intonation is rather artificial, since the terms are often used interchangeably - not only in more traditional phonetic models such as the British School (Crystal, 1969; Cruttenden, 1997), but also within phonological models of intonation which embrace the autosegmental-metrical framework."

This thesis adopts this view, although the preferential term used here will be "prosody".

The following section 2.1 describes the phonetic properties of prosody in general and of prosodic prominence in particular. The basic principles of the phonological theory adopted in this thesis are introduced in section 2.2. Finally, section 2.3 provides an overview of the different functions that prosody can serve.

2.1 Phonetic Properties of Prosody

Prosody is a complex vocal signal that includes variations of a large number of prosodic features. The three most commonly and consistently mentioned features when describing the acoustic and physical dimensions of prosody are fundamental frequency (F_0), intensity, and duration[1] (e.g. Ladd, 2008).
F_0 reflects the periodicity in the acoustic signal. It is primarily regulated through changes of the rate of vibration of the vocal folds. The opening and closing cycles of the vocal folds that result in their vibration are (quasi-) periodic, each period corresponding to one opening and closing cycle. F_0 is measured and expressed in Hertz (Hz), which correlates with the frequency of opening and closing cycles of the vocal folds per second.
The F_0 range of individual speakers depends on the length and mass of their vocal folds. Male speakers usually exhibit an F_0 range between 50 and 250 Hz, while female speakers, who usually have thinner and shorter vocal folds, exhibit a range between 120 and 480 Hz (Laver, 1994). Within these ranges, speakers can, to a large extent, actively control F_0 by means of vocal fold tension and subglottal

[1] Other spectral characteristics such as voice quality have also been regarded as aspects of prosody (e.g. Campbell and Mokhtari, 2003).

air pressure[2]. For example, higher vocal fold tension and higher subglottal air pressure result in a higher rate of vibration of the vocal folds and therefore in a higher F_0.
"Pitch" is the psychophysical or perceptual counterpart of F_0, corresponding to how high or low a speaker's voice sounds and whether the voice is going up or down (e.g. Cruttenden, 1997). Perceived pitch is often deemed to be the primary channel of prosody (Grice, 2006). The overall pattern of pitch movements is commonly also referred to as the "speech melody". In practice, the notion "pitch" is often used to refer to F_0 itself, which is why these two terms are used synonymously in this thesis.
Since F_0 is dependent on the periodicity of the vocal fold vibration, it can only be realised and measured in voiced segments. However, a discontinuous F_0 track in the acoustic signal can be perceptually complemented by listeners, which creates the impression of a continuous pitch movement (see Gilles, 2005; Nooteboom, 1997; Cruttenden, 1997). Furthermore, microprosodic perturbations caused by consonants as involuntary side-effects of articulation do not contribute to the perceived speech melody (Silverman, 1987; Nooteboom, 1997).

Duration is a multifaceted aspect of prosody that can be applied to numerous domains of speech. In general, it applies to speech sounds and silences, directly reflecting the timing and temporal patterning of articulatory processes of a speaker. Duration is measured in time units such as, e.g., milliseconds (ms). Perceptually, duration can be described as the length of time during which a listener hears a speech unit or a pause (Cruttenden, 1997). When investigating prosodic aspects, duration is most frequently measured for syllables, where it influences their perceived prominence as well as the perceived rythmical structure of speech (along with F_0 and intensity). This will be discussed in more detail below. Furthermore, the modulation of duration plays an important part in marking prosodic structure, e.g. syllables that occur at the ends of prosodic phrases are lengthened compared to the same syllables in medial position.

Finally, the intensity (or amplitude) of speech represents the amount of acoustic energy of sounds or sequences of sounds. A higher amount of acoustic energy results in a higher intensity, which in turn is perceived as a louder unit of speech, while lower intensity is perceived as a softer unit of speech (Cruttenden, 1997). Hence, the perceptual correlate of intensity is loudness. Intensity can be regulated

[2]Subglottal air pressure also affects the amplitude of the vibration of the vocal folds.

by means of subglottal air pressure and articulatory effort.

Table 2.1: Perceptual, acoustic, and articulatory correlates of prosody (adapted from Baumann, 2006: 12; building on Uhmann, 1991: 109).

Perception *scale*	**Acoustics** *measure*	**Articulation**
Pitch *high - low*	F_0 *Hertz (Hz)*	quasi-periodic vibration of vocal folds
Length *long - short*	Duration *Milliseconds (ms)*	timing of articulatory gestures
Loudness *loud - soft*	Intensity/Amplitude *Decibel (dB)*	articulatory effort, subglottal air pressure

The perceptual, acoustic, and articulatory correlates of prosody are summarised in table 2.1. While these three aspects of prosody are present to differing degrees in many of the world's languages, all three play a major role in West Germanic languages such as English and German, in which they contribute to the division of speech into chunks (phrasing) and influence the speech signal with regards to the relative prominence of syllables in an utterance (highlighting). The phenomena of phrasing and highlighting and their phonetic implementation in West Germanic languages will be introduced below.

Phrasing

Speakers use prosody to organise their speech into chunks. This phenomenon is called "phrasing" and applies to multiple levels that represent the prosodic structure of utterances. Prosodic structure includes domains such as, e.g., syllables, words, or varying sizes of sequences of words (Cruttenden, 1997).
One domain that is prosodically defined is the "intonation phrase" (also referred to as, inter alia, "intonational phrase" (Pierrehumbert and Beckman, 1988), "intonation-group" (Cruttenden, 1997), "tone group" (Halliday, 1967), etc.). Each utterance comprises at least one intonation phrase which is hallmarked by a perceptually coherent pitch contour and contains at least one prominent element and a final local pitch movement on unaccented syllables (according to the Autosegmental-Metrical theory, see section 2.2). Furthermore, pauses, fi-

nal syllable lengthening, a change in voice quality[3], decreasing intensity, and/or phrase-initial anacrusis can indicate the edges of intonation phrases (Cruttenden, 1997; Chafe, 1994).
The more salient these prosodic cues are, the stronger the perceived boundary. These differences in perceived boundary strength, among other things, have led to a number of proposals of different types of intonation phrases, namely large ones and small ones, or major ones and minor ones, respectively. For example, Beckman and Pierrehumbert (1986) and Pierrehumbert and Beckman (1988) suggest that a large intonation phrase can consist of one or more smaller "intermediate" phrases, making "two degrees of depth available for an intonational boundary" (Gussenhoven, 2004: 125). Other proposals of two different kinds of intonation phrases include major and minor tone groups (Trim, 1959) transcribed as single and double bar boundaries (O'Connor and Arnold, 1973).
While phrasing is often influenced by the syntactic structure of utterances, syntax cannot always predict the prosodic choices that speakers make to structure their speech. Semantic and pragmatic factors as well as utterance length and speaking rate also influence phrasing (Shattuck-Hufnagel and Turk, 1996).
More details about the prosodic marking of intonation phrases and their definition within the Autosegmental-Metrical framework will be given in section 2.2.

Highlighting

Speakers use prosody to highlight certain parts of their speech. Highlighting is a relative concept, meaning that an element is perceived as highlighted or prominent if it is more salient than other elements around it. At the word level, the domain of highlighting in West Germanic languages is the syllable. Its perceived prominence depends on its acoustic features in comparison to the acoustic features of surrounding syllables.

The term prominence is connected to the notions of both "stress" and "accent", which have been variously defined and used by different authors. In this thesis, a differentiation between abstract prominence at word level ("lexical stress" or "word stress") and concrete prominence at utterance level ("postlexical stress" and "pitch accent") will be made, following Grice and Baumann (2007), who, in turn, follow the British school approach (e.g. Crystal, 1969).
Accordingly, at word level, a syllable can have the abstract phonological property

[3] e.g. phrase-final creaky voice (laryngealisation or "fry").

of lexical stress (Ladd, 2008) as determined in the lexicon. For example, the word "increase" can either be a noun ("**in**crease[4]) or a verb (in**crease**), depending on whether its lexical stress is determined on the first or on the second syllable, respectively. Lexical stress can be seen as providing the potential for concrete prominence at the utterance level or as attracting postlexical prominence, respectively (see, e.g. Bolinger, 1958; Weinreich, 1954; Lehiste, 1970). In other words, at utterance level, a lexically stressed syllable can become postlexically stressed, or even pitch accented if the word containing that syllable is emphasised by the speaker.

In contrast to abstract prominence at word level, prominence at utterance level is not a binary property as it is generally felt that a mere distinction between prominent and non-prominent elements in speech does not adequately capture the observed phenomena (Terken and Hermes, 2000). Rather, different *degrees* of (postlexical) prominence are assumed. Following Grice and Baumann (2007), there are at least four different degrees of prominence at utterance level (see also table 2.2): The weakest degree of prominence can be ascribed to unstressed syllables, i.e. syllables that have no postlexical stress (or accent) at the utterance level. Postlexically stressed syllables are articulated louder, longer and more strongly, with less vowel reduction in comparison to unstressed syllables. If a syllable has additional tonal movement on or near it, it can be referred to as a pitch accented syllable. Finally, a *nuclear* pitch accent is defined as the last pitch accent in an intonation phrase and is usually perceived as the most prominent one in that phrase (Crystal, 1969; Ayers, 1996; Jagdfeld and Baumann, 2011; Baumann, 2014). Therefore, the strongest degree of prominence can be ascribed to syllables carrying the last pitch accent (= nuclear pitch accent) in a phrase.

In addition to the four prominence categories mentioned in table 2.2, the local shape of a pitch accent as well as its alignment with the accented syllable also contribute to a syllable's perceived prominence (for German: Kohler, 1991b; Niebuhr, 2009; Baumann, 2014; Baumann and Röhr, 2015; for English: Knight, 2008). Baumann and Röhr (2015) investigated the perceptual prominence of seven (nuclear) pitch accent types in German (from the GToBI inventory, see section 2.2.2). They found that three tonal dimensions contribute to the perceived prominence of pitch accents, namely the direction of a pitch movement, the relative pitch height, and the degree of pitch excursion. More precisely, they found that rising pitch accents are perceived as more prominent than falling

[4]Throughout this thesis, stressed and accented syllables will be indicated in **bold**.

Table 2.2: Prominence at word and utterance level, with increasing degree of prominence from top to bottom (adapted from Grice and Baumann, 2007).

	Word Level *abstract*	**Utterance Level** *concrete*
Syllable σ	No lexical stress	**No (postlexical) stress/accent**
	Lexical stress potential for concrete prominence	**Postlexical stress** A syllable with postlexical stress is louder, longer and more strongly articulated than an unstressed syllable
		(Prenuclear) pitch accent An accented syllable (i.e. a syllable bearing a pitch accent) has additional tonal movement on or near it
		Nuclear pitch accent The nuclear syllable is the last pitch accent in an intonation phrase, usually perceived as the most prominent one in the phrase.

pitch accents (see also Baumann, 2014), that high pitch accents are perceived as more prominent than downstepped or low accents (see also Gussenhoven and Rietveld, 1988; Ladd and Morton, 1997), and that an increase of excursion size of an accent-lending pitch movement increases the perceived prominence of the accented syllable (see also 't Hart et al., 1990). These findings and corresponding GToBI pitch accent types are illustrated in figure 2.1.
In addition, it has been argued for English that later pitch peaks can perceptually substitute higher pitch peaks, as they can create the same prominence-lending effect (Gussenhoven, 2004).

To conclude, a differentiation between abstract prominence at word level (lexical stress) and concrete prominence at utterance level (postlexical stress and pitch accent) has been made. Furthermore, at utterance level, (postlexical) stress can be distinguished from (pitch) accent, in the sense that (postlexical) stress involves properties that can be related to greater force of articulation, e.g. increased duration and intensity, while pitch accentuation requires an additional tonal movement on postlexically stressed syllables. This implies that pitch is an important prominence-lending cue that accounts for the difference between (postlexically)

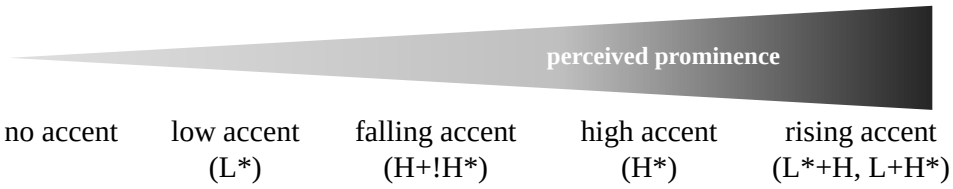


Figure 2.1: Prominence scale of accent types, based on results from Baumann and Röhr (2015).

stressed syllables and (pitch) accented syllables. Furthermore, the shape and alignment of tonal movements with accented syllables also contribute to their perceived prominence.

2.2 Phonological Aspects of Prosody

2.2.1 Basic Principles of Autosegmental-Metrical Theory

Autosegmental-Metrical theory (henceforth: AM theory) will provide the framework of analysis for this thesis. The theory itself has its origins in Pierrehumbert's (1980) analysis of American English intonation which, in turn, was influenced by earlier work on both autosegmental (e.g. Goldsmith, 1976) and metrical (Liberman, 1975; Liberman and Prince, 1977) phonology as well as by the work of Bruce (1977) on Swedish. An introduction to the AM theory is provided by Ladd (1996), who also coined the term "Autosegmental-Metrical" (see also Shattuck-Hufnagel and Turk, 1996; Beckman, 1996).
The AM theory is a phonological theory of intonational structure that connects prominence and phrasing with the phonologically relevant tonal events of a tune. The autosegmental aspect of AM theory involves the idea that the tune is independent of the text and that suprasegmental and segmental features should be represented on different tiers. The metrical aspect of the theory is concerned with the hierarchical organisation of units and their prominence relations within and between prosodic domains of different sizes (prosodic constituents within a prosodic hierarchy, see, e.g. Grice, 2006). Accordingly, the central idea of AM theory is that while the tune can vary independently of the segmental material, it is also sensitive to certain structural specifications.

Intonation models within the AM theory analyse continuously varying F_0 contours phonologically as sequences of discrete levels of high (H) and low (L) tonal events that can be combined in multiple ways, with transitions or interpolation of F_0 between them. According to the original analysis by Pierrehumbert (1980), these tonal events can either be pitch accents, phrase accents or boundary tones[5] that are associated with certain points in the segmental string. However, the tonal events are organised autonomously from the segmental tier, which reflects Pierrehumbert's observation that "not only can the same text have many different melodies, the same melody can occur on many different texts" (Pierrehumbert, 1980: 8).

Pitch accents serve a highlighting function and are associated with (post)-lexically stressed syllables. They consist of single H or L tones, or, in cases where the pitch accent is characterised by rapid local F_0 movement instead of just a local maximum or minimum, of a combination of two tones resulting in a bitonal pitch accent. Bitonal pitch accents are indicated by a "+" sign between the tones. A starred notation is used in pitch accents to indicate the association between a tone and the lexically stressed syllable of the accented word (e.g. H* for a monotonal accent or L*+H for a bitonal accent). While there is no theoretical distinction between prenuclear and nuclear pitch accents within most AM models, nuclear pitch accents are nonetheless regarded as pragmatically the most important pitch accent in a phrase (see Baumann, 2006).
Boundary tones serve a delimiting function and are primarily associated with right edges of phrases. In the original framework of Pierrehumbert (1980), the Intonation Phrase (IP) was the only intonationally defined constituent that was assigned a boundary tone, marked with "%" (e.g. H% for a high boundary tone). Phrase accents are the third type of tonal event in the tonal inventory of the original analysis of Pierrehumbert (1980). In this analysis, phrase accents constitute an additional tone between the last pitch accent of an IP and its boundary tone. They are always monotonal and are marked with a "-" after the tone (e.g. H-). In a revised version of Pierrehumbert's original account, Beckman and Pierrehumbert (1986) redefined the phrase accent[6] as constituting the boundary tone of an additional intonationally defined constituent, the Intermediate Phrase (ip), which is ranked immediately below the Intonational Phrase (IP) (Gussenhoven, 2002).

[5] In languages like English or German. In languages with lexically specified pitch features, the tonal events might have different functions (Ladd, 2008).

[6] Today, a modified concept of phrase accents claims that they can be secondarily associated with postnuclear stressed syllables (Grice, Ladd, and Arvaniti, 2000).

Since each major IP is made up of one or more minor ips, both the boundary tone of a minor ip (the phrase accent) and the boundary tone of a major IP always coincide at the end of a major IP as in example (1).

(1) [[Hannah will win]$_{ip}$ [won't she]$_{ip}$]$_{IP}$

As a result, according to the revised version of Pierrehumbert's account, boundary tones of IP's are annotated as a combination of a phrase accent, which is at the same time an intermediate phrase boundary ("-") and an intonation phrase boundary ("%"), resulting in an annotation such as, e.g. L-H% or L-L%.

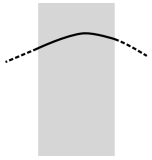
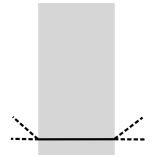
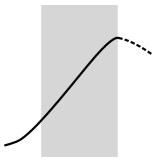
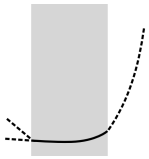
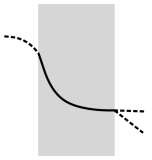
2.2.2 GToBI

The "Tones and Break Indices" (ToBI) framework is a widely used phonological intonation description system that is based on AM theory (Grice and Baumann, 2002). It was first developed as a transcription system for American English (see Silverman et al., 1992), but ToBI standards have since then been developed for several other languages as well. The tonal inventory of a ToBI system is usually adjusted to the tonal structure of the language it is developed for, so that it comprises tonal events that have been found to be phonologically distinctive in that language.
The model used to describe prosodic phenomena in the experiments below is the "German Tones and Break Indices" (GToBI) system for standard German. It was developed between 1995 and 1996 (Grice, Reyelt, Benzmüller, Mayer, and Batliner, 1996; Reyelt, Grice, Benzmüller, Mayer, and Batliner, 1996) and has since been slightly modified (see Grice and Baumann, 2002; Grice, Baumann, and Benzmüller, 2005; Grice and Baumann, 2016). A cross labeller consistency test showed that labellers were able to use GToBI consistently and to learn GToBI within a short period of time from printed training materials[7] and accompanying sound files with little or no individual coaching (Grice et al., 2005).

A GToBI annotation minimally consists of a text tier for the orthographic transcription of words and a tone tier for tonal events such as pitch accents and boundary tones (Grice and Baumann, 2016). A third tier for break indices that indicate the perceived strength of phrase boundaries is optional, and will not be

[7]The GToBI training materials are currently available via the GToBI homepage at http://www.gtobi.uni-koeln.de/

Table 2.3: GToBI pitch accents: Labels, stylised intonation contours (from Grice, Baumann, Ritter, and Röhr, n.d.) and explanations (according to Grice, Baumann, and Benzmüller, 2005; Grice and Baumann, 2016). Accented syllables are shaded in gray.

Pitch Accents		
H*		Peak accent, may be preceded by a shallow rise; accented syllable is perceived as relatively high; unmarked 'default' accent.
L*		Low accent, may be preceded by a shallow fall; accented syllable is perceived as low.
L+H*		Sharp rise from low up to peak accent, peak is often late in the accented syllable; accented syllable is perceived as high.
L*+H		Valley accent plus rise, low target within the accented syllable followed by a rise starting late in the accented syllable and reaching its peak in the following syllable (or later); accented syllable is perceived as low.
H+!H*		Early peak accent, high tonal target on the syllable immediately preceding the accented syllable followed by a local pitch fall onto the accented syllable.

further discussed in this thesis.

As in other AM models, the GToBI inventory postulates two tone levels, H for high and L for low. The levels are determined in relation to a speaker's pitch range, with H tones established in the upper three quarters of the range. L tones on the other hand lie within the lower quarter of the range (Grice and Baumann, 2016). H and L tones can be combined to form five different pitch accent types (see table 2.3).

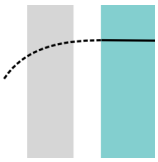
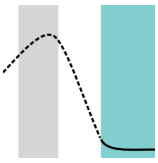
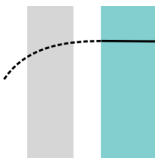
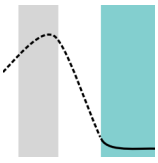
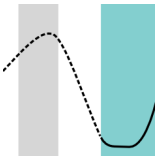
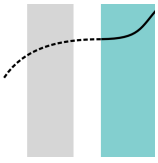
Note that in table 2.3, only one early peak accent is specified (H+!H*), while

the original GToBI pitch accent inventory included a second early peak accent (H+L*) which wass characterised by a downstep to the bottom of the speaker's range (Grice et al., 2005). However, production and perception experiments have up to this point failed to find support for a categorical distinction between these two early peak accents (Rathcke and Harrington, 2010; Grice, Baumann, and Jagdfeld, 2009). Therefore, the two early peak accents are conflated into a single category H+!H*, as suggested by Grice et al. (2009).

The inventory of boundary tones in GToBI is shown in table 2.4, divided into boundary tones of intermediate phrases and boundary tones of intonation phrases. The boundary tones of intonation phrases are annotated as combinations of intermediate and intonation phrase boundaries. In cases where the tonal target of the intermediate phrase boundary and the intonation phrase boundary are the same, the second tonal target is omitted (e.g. what would be H-H% is annotated as H-%).
Apart from final boundary tones, it is also possible to transcribe initial boundary tones in cases of an exceptionally high beginning of an IP (mid or low onsets are not explicitely marked, they are regarded as the default case) by using %H.

It is possible to modify H tones (of both pitch accents and boundary tones) using operators in the form of "downstep" and "upstep". If a high tone is considerably lower than a preceding high tone (but not as low as an L tone), it is considered to be downstepped and marked with "!" before the downstepped tone (e.g. !H*). An upstepped tone, indicated by "^" (e.g. ^H*) is considered to be higher than a preceding H tone.
The concept of downstep is not to be confused with the effect of declination, a gradual slight decrease of F_0 over the phrase. According to Pierrehumbert (1980), declination is a phonetic effect due to a decreasing amount of air in the lungs of a speaker towards the end of an utterance, while downstep is a phonological effect that is under the control of the speaker.

Table 2.4: GToBI phrase / boundary tones of intermediate and intonation phrases: Labels, stylised intonation contours (from Grice, Baumann, Ritter, and Röhr, n.d.) and explanations (according to Grice, Baumann, and Benzmüller, 2005; Grice and Baumann, 2016). Accented syllables are shaded in gray (in all cases: H*), boundary tones are coloured in turquoise.

Phrase/Boundary Tones *of intermediate phrases (ip)*		
H-		A high plateau (or a slight rise) to the middle of the speaker's pitch range; boundary tone is perceived as high.
L-		Terminal boundary tone low in the speaker's range; perceived as low.
Boundary Tones *of Intonation Phrases (IP)*		
H-%		High plateau until the end of the phrase; similar to tonal contour of H-, but boundary is perceived as stronger.
L-%		Terminal boundary tone low in the speaker's range, may be followed by a drop to extra low; L-% is often lower than L-.
L-H%		Terminal falling-rising contour if preceded by an H tone, otherwise low tonal target followed by a rise generally to the mid of the speaker's range on the last syllable.
H-ˆH%		High plateau (like H-%) with a sharp rise on the last syllable of the phrase, often to a point very high in the speaker's pitch range.

2.3 Functions of Prosody

As mentioned in the previous sections, prosody involves categorical decisions about highlighting and phrasing, as well as about the type of pitch accent and boundary tone that is used. Moreover, speakers can vary pitch height and range as well as the exact timing and shape of the pitch contour. Speakers can also adjust their speaking rate, voice quality, or other prosodic features that, in turn, alter the way in which something is said. All of these prosodic aspects are used by speakers, sometimes consciously, sometimes unknowingly, to transmit meaning or information beyond the literal content of what is said. The various communicative functions that can be conveyed by prosody range from the marking of lexical tone to the expression of emotions and can roughly be divided into grammatical, pragmatic, and affective functions (e.g. Shriberg et al., 2001). Another way of categorising the functions that prosody can serve is their division into linguistic (including grammatical and pragmatic) and paralinguistic (including affective) functions[8] (Grice and Baumann, 2007). Figure 2.2 provides an overview of a

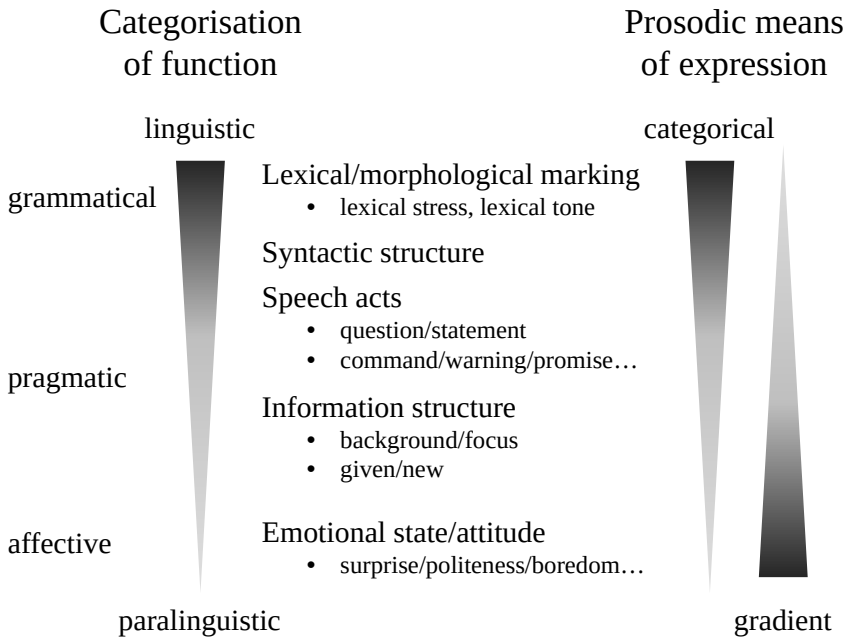


Figure 2.2: Categorisation of prosodic functions (adapted from Grice and Baumann, 2007: 14), with modifications and additions from the author of this thesis).

[8]It should be noted that prosody also conveys extralinguistic information such as information about the age, gender, or physical state of a speaker, but this aspect will not be further discussed in this thesis.

suggested gradual categorisation of a representative sample of prosodic functions with regards to their prosodic means of expression. The figure shows that prosodic functions can be arranged on a continuum from linguistic to paralinguistic, or from grammatical through pragmatic to affective. It is important to note that it is impossible to clearly define where one category ends and the next one starts, and most prosodic functions are situated between the different functions or are difficult to categorise. Furthermore, language structure and language use are interdependent, which is why the domains (grammar and pragmatics in particular) can be expected to influence and affect each other (see also Lambrecht, 1994: 4).

The fact that the same acoustic cues can convey grammatical, pragmatic and affective information simultaneously makes prosody a very complex aspect of language. However, this does not mean that a given prosodic impairment will equally affect all prosodic aspects. As Shriberg et al. (2001) points out, "research on individuals with unilateral brain damage (Emmorey, 1987; Heilman, Bowers, Speedie, and Coslett, 1984; Luks, Nusbaum, and Levy, 1998) suggests that these three functions of prosody can be differentially affected. Thus, their neurological organisation may be, at least to some degree, functionally independent." (Shriberg et al., 2001: 1098). This observation indicates that in general, prosodic impairments can be limited to certain aspects of communication, while other aspects may remain intact. Therefore, the following sections will give a more detailed overview over grammatical, pragmatic, and affective prosodic functions.

2.3.1 Grammatical Functions

Grammatical functions of prosody include signalling lexical, morphological and syntactic contrasts. These aspects are represented by a set of discrete and formal rules and are in some cases even specified in the lexicon of a language (e.g. lexical tone, lexical stress). Grammatical aspects of prosody are mostly obligatory in the sense that they have to be employed in order to resolve ambiguity and to make a message clear and comprehensible. Furthermore, the prosodic means of expressing grammatical or "highly" linguistic functions are often (but not necessarily) found to be of a categorical nature, as, for example, the presence or absence of lexical stress, or the presence or absence of a phrase boundary to signal a certain syntactic structure.

In tone and pitch accent languages[9], categorical tonal contrasts at word level can determine the meaning of words or grammatical categories. For instance in Standard Chinese, the words *huā* ('flower') and *huà* ('speech, language') are composed of the same segments, but the former has high level pitch ("Tone 1") and the latter has sharply falling pitch ("Tone 4") which makes them differ in meaning (see Ladd, 2008: 6).
An example of a language that uses a categorical tonal contrast to determine a grammatical category, namely tense, is the West African (Niger Congo) language Bini, where a low tone marks present tense and high or high-low tones mark past tense (see Crystal, 1987: 172; as cited in Grice and Baumann, 2007).

In some languages, prosody can also signal a change in grammatical class from noun to verb. In English, the word "permit" can have lexical stress on the first or the second syllable. When the stress falls on the first syllable (**per**mit), the word is recognised as a noun, and when the stress falls on the second syllable (per**mit**), it is recognised as a verb (Ladd, 2008). A comparable mechanism that is mediated by lexical stress is the disambiguation of compound nouns from adjective-noun sequences (e.g. **green**house vs. green **house**).

Another grammatical function of prosody is the phenomenon of phrasing, which has already been introduced in section 2.1. It primarily involves the presence or absence of boundary tones and pauses to subdivide the stream of words into phrases. Phrasing is used to disambiguate in certain cases between two different syntactic structures and/or their semantic interpretation (for an overview, see Hirschberg, 2015), as in the following example (2) from Roach (1991: 174):

(2) Those who sold quickly made a profit.
    a. [Those who sold] [quickly made a profit].
    b. [Those who sold quickly] [made a profit].

In example (2a.), the phrase break after "sold" leads to the interpretation that profit was made quickly by those who sold, while in example (2b.), the phrase break after "quickly" leads to the interpretation that profit was made by those who sold quickly.

[9]The difference between tone languages (e.g. Standard Chinese, Cantonese) and pitch accent languages (e.g. Swedish, Japanese) is that the former have contrastive tone on almost all syllables, whilst the latter restrict their tonal contrasts to specific syllables, which bear a pitch accent (Grice and Baumann, 2007: 7).

Phrasing is also required to identify compounds in lists, such as, e.g., "[thirty-eight] [three]" versus "[thirty] [eight] [three]". These types of phrasing are considered a grammatical function of prosody that guide the analysis of the syntactic structure of utterances. However, syntactic structure and prosodic phrasing, while being strongly related, do not necessarily have to always correspond exactly (Grice and Baumann, 2007). Furthermore, it should be noted that the context can also contribute to resolving syntactic ambiguity, which is why speakers will not make such distinctions all of the time (Grice and Baumann, 2007: 8). Moreover, in certain contexts, phrasing can also be an optional stylistic device.

Another function of prosody that is often considered a grammatical one is the differentiation between questions and statements. This proposition requires some clarification as to what is meant by "questions" and "statements". It is a widespread practice to use these terms both to refer to a communicative function (speech act) at utterance level and to categories of syntactic form at sentence level (sentence modality). However, a distinction of the two concepts is crucial with regards to the role of prosody. Following Lyons (1977) and Huddleston (1994), the terms "interrogative" and "declarative" will be used when referring to sentences with a particular grammatical structure and syntactic form (sentence modality), while the terms "question" and "statement" will be used when referring to utterances with a particular communicative function (speech acts). Hence, sentence modality is primarily determined by grammatical/syntactic form, while speech acts are often determined by prosody only. For example, rising pitch contours often signal questions, while falling pitch contours tend to characterise statements (Gussenhoven, 2004). The following example (3) illustrates the difference between sentence modality and speech act with regards to syntactic and prosodic marking.

(3)

	SENTENCE	MODALITY	PITCH CONTOUR	SPEECH ACT
a.	Is the sun shining?	*interrogative*	*rising*	*question*
b.	The sun is shining.	*declarative*	*falling*	*statement*
c.	The sun is shining?	*declarative*	*rising*	*question*

The syntactic structure in (3a.) determines the sentence as interrogative. When uttered by a speaker, it is usually produced with a rising pitch contour to signal that a question is being asked, but pitch is not the sole medium to indicate this, because word order also determines the interrogatory nature of the sentence.

On the other hand, the syntactic structure in (3b.) and (3c.) is identical, indicating a declarative sentence. However, the communicative function of (3b.) and (3c.) differs, depending on the pronounced pitch contour. A falling pitch contour indicates a statement (3b.), while a rising pitch contour indicates a confirmation-seeking question (3c.). The confirmation-seeking question (sometimes also referred to as "declarative question", see, e.g., Cruttenden, 1997: 155) is a subtype of the polar question[10] that exhibits a declarative syntax.
In English, confirmation-seeking questions are typically solely determined by prosody, while other types of polar questions (such as the one in (3a.)) are determined by word order (and content questions are determined by interrogative words such as "who", "where", etc.). However, some languages rely exclusively on prosody in order to indicate whether an utterance is a polar question or a statement (e.g. Italian, see Dryer, 2013; Cangemi and D'Imperio, 2013).
Thus, it appears that in many languages, prosody provides crucial cues to whether an utterance is a (polar) question or a statement, that is, it determines the type of speech act and the communicative function intended by a speaker. Speech acts are commonly considered to fall under the scope of pragmatics. However, it has been suggested that rises or high pitch in questions are a language universal[11] (Cruttenden, 1997; Gussenhoven, 2004; Ohala, 1983, 1984). Although it is not the case that it is a universal, it is nonetheless common, although the way this high pitch or rise is implemented is highly language specific, suggesting that grammaticalisation is underway. Therefore, the prosodic function of signalling question/statement contrasts can be seen as falling between being grammatical and pragmatic (while the prosodic marking of other types of speech acts (e.g. warnings, promises, commands) more clearly fall under the scope of pragmatics).

2.3.2 Pragmatic Functions

Pragmatic aspects of prosody concern all aspects that contribute to how language is used in context and when communicating with others. This includes, for example, the encoding of speech acts, the marking of information structure, or the regulation of turn-taking. When speakers use prosody for pragmatic purposes, they are also following certain rules, but these rules often involve intuitiveness and are less formal, less clearly defined and more context-dependent. Therefore,

[10] Polar questions are ones to which the expected answer is the equivalent of 'yes' or 'no' (and which are thus sometimes called "yes-no questions", see Dryer, 2013).
[11] But see exceptions reported in Williamson, 1979.

pragmatic prosody can be categorised as serving a linguistic function that operates somewhere between grammatical rules, context-dependency, and a speaker's intuition. Compared to grammatical functions, pragmatic functions more often involve gradient and nuanced aspects of prosody, such as variation in pitch height or contour to, for example, signal what is important and less important in an utterance.
While signaling certain pragmatic aspects is important for successful communication, it is not obligatory for a speaker to pay attention to these aspects. In other words, pragmatically neutral communication would still be possible, but less effective, less social, and restricted to a basic exchange of information.

As mentioned above, speakers make use of prosody to signal certain kinds of speech acts which reflect their communicative intentions. For example, the same one-word utterance "Coffee" can be uttered in various ways (see example (4)), expressing either a command (4a.), criticism (4b.), an offer (4c.), a wish (4d.), or doubt (4e.) and so forth.

(4) a. Coffee! *(Now!)*
b. Coffee! *(So unhealthy.)*
c. Coffee? *(Would you like some?)*
d. Coffee! *(That's what I need.)*
e. Coffee? *(Since when do you like coffee?)*

Thus, the same expression can have different meanings in different communicative situations and in different contexts, and prosody is one of the main channels used to determine the communicative intention of a speaker.

Prosody also plays a role in the conversational interaction between speakers. It regulates turn-taking in a way that speakers indicate prosodically that they are about to finish speaking and that it is someone elses turn to speak. At the same time, speakers can also indicate prosodically that they will continue speaking and thus avoid being interrupted.

An important pragmatic function of prosody is the marking of information structure (see, e.g., Halliday, 1967; Chafe, 1976; Prince, 1981; Lambrecht, 1994, among others). Speakers use prosody to emphasise units of information that require the attention of the listener and attenuate those that do not. This includes the prosodic marking of givenness of entities in discourse as well as the division of

utterances into focus and background elements (Grice and Baumann, 2007: 8). The concept of givenness, which will be introduced in detail in chapter 3, reflects the cognitive representation of individual entities in the interlocuters's minds. In West Germanic languages, new entities are commonly assumed to be marked by a pitch accent, while given entities are not. Furthermore, the type of pitch accent can signal different degrees of givenness[12] (Baumann and Grice, 2006). The following example (5) shows how accentuation encodes givenness of the referent "book" (the syllables that are carrying (nuclear) pitch accents are in bold).

(5) I bought an expensive **book** yesterday, because I **need**ed the book and it was on **sale**.

In example (5), the first mention of "book" will most likely receive a (nuclear) pitch accent, because it constitutes a new referent that is introduced into the discourse. The second mention of "book", on the other hand, will usually not receive a pitch accent, because it has been mentioned in the previous sentence. The third reference to the book ("it") will usually also not receive a pitch accent, and the use of a pronoun provides an additional cue to the given status of the referent. This example shows that the cognitive activation of a referent (here: a book) is reflected in its prosodic marking[13].

The other aspect of information structure involving prosodic marking is the concept of focus-background structure, reflecting which parts of an utterance are important and which ones are less important in an utterance, as shown in the following question context[14] in example (6):

(6) Q: What did you do yesterday?

A: I [bought a **book**]$_{\mathrm{F}}$ yesterday.

Example (6) shows the division of the answer into an informative part of the message (focus - within squared brackets) and an uninformative part (background).

[12] Additional cues to the degree of givenness of an entity can be the choice of referential expressions (e.g. using indefinite noun phrases (NPs) for new referents and pronouns for given referents, see also chapter 3).

[13] But note that the status of the first mention of "book" in example (5) is, in fact, what Lambrecht (1994) would refer to as "unidentifiable". See chapter 3 for a detailed discussion of the difference between identifiability and activation.

[14] The question context is merely used here to illustrate the focus-background relation. In naturally occuring question-answer contexts, elliptic answers are more common (see also Braun, 2005).

In this example, the informative part of the answer comprises several elements and is therefore commonly referred to as "broad" focus. In broad focus conditions, the pitch accent is often realised on one element within the focus-domain only[15]. On the other hand, narrow focus structures exhibit a more direct relationship between accentuation and focus, as shown in the following examples (7) and (8):

(7) Q: What did you buy yesterday?
A: I bought [a **book**]$_{\mathrm{F}}$ yesterday.

(8) Q: When did you buy a book?
A: I bought a book [**yesterday**]$_{\mathrm{F}}$.

While there is a relation between focus and newness on the one hand and background and givenness on the other, newness and givenness apply to the degree of cognitive activation or familiarity of discourse referents, while focus and background apply to the informativeness of elements of a proposition with respect to the discourse context. This means that even a referent that has been mentioned before can be new information (see example (9), adapted from Krifka, 2007: 32).

(9) A: What would you like to drink, tea or coffee?
B: I want [**tea**]$_{\mathrm{F}}$.

In example (9), the referent "tea" is mentioned in the question, which is why it can be classified as given in the answer. However, "tea" is in focus and therefore accented, because it contains newly provided information.
Hence, focus prosody can override activation prosody (Baumann, 2006: 82), in the sense that accentuation does not only depend on the degree of activation of a referent, but also on the choice of a speaker to present an entity as particularly newsworthy and to highlight it irrespective of its activation status (Baumann, 2006: 88). Therefore, the role of prosody in marking information structure is bilateral and can determine both the degree of activation of a referent and the informativeness or newsworthiness of elements within an utterance.

[15]This phenomenon is often referred to as "focus projection" (e.g. Uhmann, 1991; Selkirk, 1984).

2.3.3 Affective Functions

Affective aspects of prosody are deliberately employed by speakers and entirely depend on their intentions, emotions and attitudes. Affective prosody therefore resides at the intuitive, paralinguistic end of the spectrum (see figure 2.2) and can be seen as highly context-dependent. However, this does not mean that the additional meaning provided by affective prosody is not important. In fact, according to Pike (1945), "the hearer is frequently more interested in the speaker's attitude than in his words - that is, whether a sentence is 'spoken with a smile' or with a sneer [...]" (Pike, 1945: 22).
While grammatical and pragmatic prosody determine or contribute towards the meaning of the message, affective prosody conveys information about the affective state of a speaker. Affective prosody is usually expressed by gradient means, which accords with the gradient nature of emotions (e.g. more excited versus less excited speech) and attitudes. For example, speakers may speak louder when they are angry.
Affective prosody is involved in conveying a speaker's general emotional state (e.g. happiness, sadness, anger etc.), as well as in conveying a speaker's attitude (e.g. politeness, confidence, sincerity).

Furthermore, speakers can use prosodic means to vary their personal speaking style when talking to peers, to young children or infants, or to people of higher social status. For example, speakers of a range of English varieties have been observed to make use of a phenomenon referred to as "Uptalk", involving a final rise in pitch at or near the end of the phrase, to express personal and / or social identity (Warren, 2016).

Chapter 3

Referential Givenness

3.1 Reference and Information Structure

One of the main goals of conversation is the exchange of information. Interlocuters achieve this goal by formulating propositions[1] about individuals, events, things, and other entities. In doing so, they constantly *refer* to these entities in discourse in order to make statements about them. The way in which speakers do this depends on many factors, which will be discussed in this chapter.

Discourse is not usually composed of a random set of utterances but of a structured series of information made up of referents and propositions that develops into a coherent whole (Foley, 2006). This process of structuring information in discourse is guided by a number of principles which can be subsumed under the notion of "information structure" (or "information packaging", see, e.g., Chafe, 1976). It incorporates several concepts about how speakers organise the information they want to communicate. The full range of approaches to this fairly broad discipline cannot be covered within the scope of this thesis, but this section will provide an overview of the basic concepts of information structure that are necessary for a proper understanding of the concept of "referential givenness" and how it is employed in this thesis.

The three most common subconcepts under the notion of information structure refer to a) the cognitive status of information in the minds of the interlocuters

[1]Following Lambrecht (1994), this thesis will use the term "proposition" as a neutral expression with regard to the question of truth value.

(given - new), b) what the utterance is about and what the speaker says about it (topic - comment) and c) the informativeness and importance of elements with respect to the discourse context (background - focus).
While these concepts involve distinct notions, they are also correlated and partly overlapping. This might be one reason for the terminological incongruence of these levels in the literature (see also Baumann, 2006: 55). What can be said is that all levels of information structure relate in some way to the distinction between new elements and given elements, but they do so in different ways (Gundel, 2003). The three levels will be outlined below with the intention to tease them apart but also to point out where they overlap.

3.1.1 Information Status: Given - New

The given - new dimension involves the cognitive representation (or accessibility) of information in the interlocuters' minds with regard to the current discourse environment (Baumann, 2006: 36). This concept assigns a certain information status to elements with respect to the speaker's belief about the hearer's knowledge. Information status can be a categorical feature (given or not given (i.e. new), see, e.g., Schwarzschild, 1999), or a scale that expresses different degrees of givenness (e.g. Prince, 1981; Gundel, Hedberg, and Zacharski, 1993; Chafe, 1976; Lambrecht, 1994). In general, given elements are assumed by the speaker to already reside in the mind of the listener, while new elements are not.
The distinction between given and new elements can be independently applied to an entire idea of an event or state or just to the entities that participate in events and states (see Chafe, 1994: 71). In other words, propositions can be new or given, and so can individual referents. For example, in the sentence

(10) I bought an expensive book yesterday,

the event of the speaker having bought an expensive book can be analysed as conveying new information within a discourse that, in some way, adds to the knowledge of the listener. On the other hand, the referents themselves (*I, book*) can also be analysed in terms of their information status: The referent *book* can be analysed as new in some sense, while the referent expressed by *I* (the speaker) can be analysed as already given, because of the obvious presence of the speaker in the discourse environment.

The former analysis of the information status of entire ideas of events or states is considered a global view of the given - new distinction, whereas the latter one that focusses on the information status of individual referents is considered a local view (see Chafe, 1994: 71). Gundel (2003) refers to these as "relational givenness" and "referential givenness", respectively.
This thesis will adopt the local view of referential givenness and therefore limit the domain of information status to (mental representations of) referents (for a discussion see Lambrecht, 1994: 110 ff). This very concrete view has the advantage of being able to observe the direct impact of aspects of givenness on the linguistic realisation of referents. Nevertheless, the influence of *relational* givenness will also be considered where necessary.
In order to better understand how referential and relational givenness are anchored in the framework of information structure, the remaining two subconcepts of information structure mentioned above will be presented briefly, before the influence of referential givenness on the prosodic and morphosyntactic marking of referring expressions will be explained in detail.

3.1.2 Aboutness: Topic - Comment

The topic - comment (or "theme - rheme") concept is related to where incoming information is stored in the memory of interlocuters. It is based on the view that utterances are about something or someone (the topic, see, e.g., Kuno, 1972; Reinhart, 1981; Lambrecht, 1994) with the intention to increase the listeners' knowledge about the respective referent (comment). Krifka (2008: 265) adopts the following definition of topic:

> "The topic constituent identifies the entity or set of entities under which information expressed in the comment constituent should be stored in the content."

The topic, therefore, constitutes the element under which new information provided by the comment will be stored. In English and German, the first part of sentences often expresses the information that has topic (or theme) function, and the last part often expresses the information that has the comment (or rheme) function[2].

[2]However, it has been argued that topics do not necessarily have to occur in the first part of sentences in German (e.g. Lambrecht, 1994; Vallduví and Engdahl, 1996).

One prevalent claim is that topics correspond to new information and comments to given information. However, this claim must be viewed critically with regard to the distinction between relational and referential givenness as discussed above. In the following example (11), both comment parts represent new information in the sense that knowledge is added about the topic (Hannah). However, the second comment also contains given information in the sense that the pronoun "him" refers to a referent that has been previously mentioned (her dog).

(11) Q: What about Hannah?
A: $[\text{She}]_{\text{topic}}$ $[\text{will bring her dog.}]_{\text{comment}}$
$[\text{She}]_{\text{topic}}$ $[\text{likes to take him everywhere.}]_{\text{comment}}$

Thus, example (11) shows that comments correspond to new information only insofar as they constitute the new part of a proposition that will add to the knowledge of the listener (relational givenness). However, they can also contain items that have been mentioned before (referential givenness).

On the other hand, topics frequently represent referents that are, in some way, accessible for listeners, which ensures that they know what a proposition is about and where to "store" it. However, topics can also represent referents that are introduced to the discourse for the first time, so-called "newly-introduced" topics (Givón, 1983: 9). Hence, topics do not necessarily have to be given or identifiable[3]. In fact, according to Givón (1983), topics can be newly-introduced, returned to (referred to as "topic shift"), or maintained (continuous topic). This will be discussed in more detail below.

Topics are generally referential (although not every referent is a topic). Thus, topicality has to be taken into account when investigating the linguistic representation of referents. The existence of a close relationship between the marking of topicality and the cognitive accessibility of a corresponding referent has already been established in numerous studies (see, e.g., Chafe, 1976; Prince, 1981; Lambrecht, 1994). It is mediated by factors such as topic continuity, interference from other topics, as well as the distance to its previous occurrence in discourse (Givón, 1983). Most of these factors apply to all referents in a discourse, not

[3]But note that this is an issue of debate. Lambrecht (1994), for example, argues that comments about unidentifiable referents are unacceptable. Halliday (1967) and Reinhart (1981), on the other hand, claim that predications about topics can be processed by listeners even if they have never heard of the referent before (see also Braun, 2005).

only to those that are topical. These factors can be subsumed under the notion of *Identifiability* and *Activation* of referents (Lambrecht, 1994), two concepts that will be introduced in section 3.2. Accordingly, the influence of identifiability and activation of referents on their prosodic and morpho-syntactic marking will be discussed in detail in that section as well.

However, a unique aspect of topicality that will become relevant for the production experiment presented in chapter 6 is the process of topic shifting. A topic shift occurs if the topic of the current utterance is different from the topic of the previous utterance. In that case, the focus of attention shifts from the previous referent to another referent. In order to guide this shift of attention, speakers have been observed to employ similar strategies to those found for introducing new referents into the discourse (see section 3.2 below), which is why topic shift is often also referred to as topic *re*activation or topic *re*introduction (Hendriks, Koster, and Hoeks, 2014; Fossard et al., 2018). Thus, the three most important status types for topical referents can be described as *new* (newly-introduced), *reactivated* (shift), and *given* (maintained).

3.1.3 Informativeness: Background - Focus

As briefly mentioned in section 2.3.2 of chapter 2, the background - focus structure reflects a speaker's choice as to which part of a sentence (s)he evaluates to be informative or newsworthy. The division of sentences into focus and background therefore indicates the level of communicative importance of entities. Those entities that are in focus represent the important or newsworthy parts of an utterance, while the less important or uninformative parts of an utterance constitute the background.
The evaluation of whether a certain constituent is informative or not often involves its relationship to previously mentioned information. However, this does not mean that given information is always in the background or that new information is automatically focussed. Rather, the focussed part of an utterance is the one that contains information which, at the time of the utterance, cannot be presupposed or expected by the listener. In other words, something that is less predictable has more information content (Roach, 1991). This implies that, similar to what has been discussed in the topic - comment section, also previously mentioned information can be focussed.

For Germanic languages (e.g. English or German), focus structure is mainly marked by means of prosody, that is, the focussed part of a sentence is emphasised prosodically while the backgrounded part is attenuated. Focus is argued to attract the nuclear pitch accent of a phrase, while material in the background tends to be deaccented (see, e.g., Ladd, 2008; Gussenhoven, 1983). However, the use of certain syntactic constructions (e.g. cleft sentences) as well as word order may also indicate that a particular constituent is in focus.

The notion of focus can be divided into different sub-structures, depending on a) the size or expansion of the focus domain (broad versus narrow focus, see, e.g., Ladd, 1980) and b) the type of focus as determined by its semantic-pragmatic purpose (e.g. contrastive focus). In addition to broad, narrow, and contrastive focus, there are other focus structures (e.g. verum focus, dual focus, second-occurence focus, etc.), which will not be covered in this thesis.
Broad focus can extend over several constituents or even over a whole sentence. Broad focus structures usually have very little or no information in the background, and are typically "all-new-sentences" or "out-of-the-blue-utterances" (see, e.g., Uhmann, 1991: 2). They often occur at the beginning of a new topic of conversation, or as an answer to an unspecific question as in example (12), where the whole answer is in focus as indicated by the square brackets. The syllable carrying the nuclear pitch accent is indicated in bold face.

(12) Q: What happened?
A: [Hannah gave her friend some **can**dy.]$_{\mathrm{F}}$

In narrow focus structures, only individual constituents are in focus. Example (13) represents a narrow focus structure by providing a context in which only the noun phrase ("some candy") is in focus, while the rest of the sentence is in the background[4].

(13) Q: What did Hannah give to her friend?
A: Hannah gave her friend [some **can**dy.]$_{\mathrm{F}}$

Contrastive focus describes cases in which one element is in direct contrast to a previously mentioned element, or a choice from a set of (limited) alternatives. In

[4]It should be noted that the background can also be omitted, for example the answer could simply be "some candy".

example (14), the word "candy" is in direct contrast to "fruit" from the context question, thus it is in contrastive focus. Contrastive elements usually involve a more emphatic prosodic marking, for example through the use of a more prominent pitch accent type, than non-contrastive elements do.

(14) Q: Did Hannah give her friend some fruit?
A: Hannah gave her friend some [**can**dy.]$_{\mathrm{F}}$

Note that in examples (12), (13), and (14), the nuclear pitch accent always falls on the same syllable[5]. Constituents in narrow and contrastive focus naturally attract the nuclear pitch accent due to the intention of the speaker to highlight their communicative importance. For example, if the context question of example (13) is changed to "Who did Hannah give some candy to?", the nuclear pitch accent would fall on the now narrowly focussed element "friend". However, this does not explain the nuclear pitch accent placement in broad focus structures, in which whole sentences can be in focus. Here, the phenomenon of "focus projection", which was introduced in section 2.3.2 of chapter 2, applies, so that one element receives the nuclear accent in substitution for all focussed elements. In example (12), the syllable "can" of the word "candy" receives the nuclear pitch accent, which makes this element the *focus exponent*. However, this does not exclude the possibility that other elements of the same broad focus domain can also be marked with (prenuclear) pitch accents. In fact, this is often the case (see, e.g., Uhmann, 1991: 221).
Although focus will not be investigated in this thesis, its influence on the prosodic marking of referential givenness will be considered in the experiments.

3.1.4 Summary

Information structural phenomena have been shown to operate on different levels that all contribute to the way in which speakers organise the information they want to communicate. Information is generated through the relation between something new and something given. The cognitive aspect of givenness plays a crucial role in all levels of information structure, as the distinction between given

[5]However, the different focus structures might differ gradually in terms of their prosodic realisation. The nuclear accent in (12) is often a less prominent one than the one in (13) which, in turn, is often a less prominent one than the one in (14), see, e.g., Grice, Ritter, Niemann, and Roettger (2017).

and new information can be applied to both the content of a message (relational givenness) as well as to the individual referents of each sentence (referential givenness).
Hence, referential givenness (whether a referent referred to has been recently mentioned or is newly introduced to the discourse) should be distinguished from its "informativeness" (how newsworthy and how unpredictable it is) as well as from its "aboutness" (whether a given sentence is about that referent or about something / someone else). This refers to the discrepancy between temporary cognitive states of discourse referents and the pragmatic roles that referents can play within propositions (topic and focus, see Lambrecht, 1994 : 76).
The terms "referential givenness", "information status" and "referent status" will be used synonymously in this thesis to refer to the cognitive representation of referents. This includes instances of reactivation of a referent. Thus, the "reactivated" referent status means that a referent is given in discourse but promoted to the foreground in order to regain topicality.

3.2 Referantial Givenness: Identifiability and Activation

In order to make a reference successful, a speaker has to 1) make sure that the listener can identify the entity that the speaker is talking about and 2) assess the degree of activation of an already identifiable referent in the mind of the listener in order to avoid misunderstanding and confusion. Cognitive approaches commonly account for these two levels of referential givenness by distinguishing between 1) "Identifiability" and 2) "Activation" (Chafe, 1994; Lambrecht, 1994)[6].
Identifiability and activation reflect the current salience of entities in discourse. At this point it should be noted that discourse *salience* is often also referred to as discourse *prominence*. In this thesis, however, the term "prominence" is exclusively used to refer to the degree to which a set of prominence-lending cues are used to highlight or emphasise an item (mostly prosodic prominence). The salience of entities will therefore be determined by refering to their identifiability and activation throughout this thesis.

[6]See also Gundel's (1985) "topic-identifiability" and "topic-familiarity".

The entities that a speaker refers to are the *discourse referents*[7], and the words or linguistic expressions a speaker uses to refer to them are *referring expressions*. Referring expressions can be any expression used in an utterance to unambiguously refer to something or someone (e.g. full NPs, proper names, pronouns), making them a uniquely identifiable entity among other potential referents.

While the identifiability of a referent is based on the general shared knowledge between a speaker and a listener, the activation of a referent is based on the mental representation of referents during one individual ongoing discourse. Identifiability can also be matched to long-term memory (does a listener know (or have a mental representation of) the referent the speaker is going to talk about at all?) while activation can be matched to short-term memory ((when) has the speaker mentioned a referent in that particular conversation before?). These two levels have also been discussed by Prince (1981), who refers to them as the "givenness as shared knowledge" level which is comparable to identifiability, and the "givenness as saliency" level which is comparable to activation[8].

According to Lambrecht (1994), a referent is unidentifiable if it is not yet represented in the listener's mind and if it cannot be referred to deictically. Thus, "an identifiable referent is one for which a shared representation already exists in the speaker's and the hearer's mind [...], while an unidentifable referent is one for which a representation exists only in the speaker's mind" (Lambrecht, 1994 : 77). If a referent is unidentifiable, a speaker has to create a new discourse representation of that referent for the listener. Lambrecht compares this to "the establishment of a new referential "file" in the discourse register, to which further elements of information may be added in the course of the conversation [...]" (Lambrecht, 1994 : 77). This is reminiscent of the previously discussed notion of topic as the entity under which information should be stored, which reveals the close relationship between referents and their topicality.
Following the terminology of Prince (1981), an unidentifiable referent can also be referred to as "brand-new". In order to create such a brand-new discourse representation, the speaker makes use of a "linguistic description, which can then be anaphorically referred to in subsequent discourse" (Lambrecht, 1994: 77).
An *anaphor* is an item that refers back to an earlier item, its antecedent. For

[7]This thesis will mostly be concerned with discourse referents as mental representations of entities in contrast to referents expressing a presupposition of existence (see Lambrecht, 1994).

[8]Prince (1981) actually discusses three levels of givenness, the third one being the "givenness as predictability/recoverability" level which is comparable to the focus - background concept that has been discussed before.

example, in the sentence "Hannah smiled because she was happy", the pronoun "she" is an anaphor that refers back to the antecedent "Hannah".
Pronouns are the most common anaphor, but NPs or whole sentences can also be anaphoric. Another type of anaphor is the so-called "zero anaphor". In this case a null element (or the absence of a linguistic item) is the anaphor, as in the sentence "Hannah was happy and Ø smiled".

In English and German (and many other languages), unidentifiable referents are commonly introduced with indefinite NPs (Ariel, 1988). Identifiable referents, on the other hand, are usually signalled with some form of anaphoric reference[9]. This is shown in example (15). The referring expressions in question are underlined:

(15) I bought an expensive book yesterday, because I needed the book and it was on sale.

In (15), a brand-new referent is introduced with the expanded indefinite full NP "an expensive book". Now that the referent is activated in the listener's mind, subsequent referring expressions for that referent are anaphoric and can be, inter alia, a definite NP ("the book"), or, even more likely, simply a pronoun ("it") to refer to the same book.

In fact, once a referent is identifiable, it *must* from that point on be referred to with an anaphoric reference[10]. This is shown in example (16), where only (a) and (b), but not (c) would be appropriate (as indicated by the hash mark #):

(16) I bought an expensive book yesterday.
  (a) How expensive was it?
  (b) How expensive was the book?
  (c) # How expensive was a book?

On the other hand, anaphoric expressions necessarily depend on a referent being identifiable in order to be interpretable.
Whether or not a referent is identifiable depends on the shared knowledge between interlocuters about what has been said before, what happend before, what is

[9]But see Lambrecht (1994: 79-87) for a discussion of the correlation between definiteness and identifiablilty.
[10]This requirement does not hold for non-specific and generic indefinite NPs (see Lambrecht, 1994 : 89).

happening at that moment (see Givón, 1983), as well as on whether the referent is saliently present in the speech setting. In the following example (17), the referent will be identifiable for the listener because of the shared knowledge between the listener and the speaker as well as because of the saliency of the referent in the speech setting. The context for the example could be as follows: Two colleagues are at work and one colleague is summoned by their superior. After she returns from the office of the superior, she says to her colleague:

(17) He is so mean.

Her colleague, the listener, will correctly identify the pronoun *he* as referring to their superior. The context is sufficient in order to identify the referent, because the referent is salient and part of the shared knowledge of the interlocuters. However, in a different context where, later that day, the colleague who was summoned by her superior talks to a friend, the sentence in example (17) would most likely not be interpretable by the friend. Instead, a sentence as in example (18) would more likely be used:

(18) My boss is so mean.

These examples show that the cognitive state of a discourse referent in the mind of the speaker and the listener is reflected in the corresponding referring expressions that are used. However, this is not simply a dichotomous distinction between unidentifiable and identifiable referents. Once a referent is identifiable, it can take on different activation states. One reason for this is discourse dynamics: The "main characters" change within a discourse as there are constant shifts of attention (i.e. topic shifts) from one referent to another. Furthermore, as Chafe (1987 : 22ff) notes, "only a very small amount of [. . .] information can be focussed on, or be 'active' at any one time".

Both Chafe (1994 : 73) and Lambrecht (1994 : 94) establish three activation states that an identifiable referent can take on, namely active, semiactive (or accessible), and inactive, respectively.
The relationship between identifiability and activation is illustrated in figure 3.1, which is a simplified version of a diagram from Lambrecht (1994).

A referent that is active is one that is currently in a person's focus of consciousness. A referent that is semiactive is in a person's peripheral consciousness

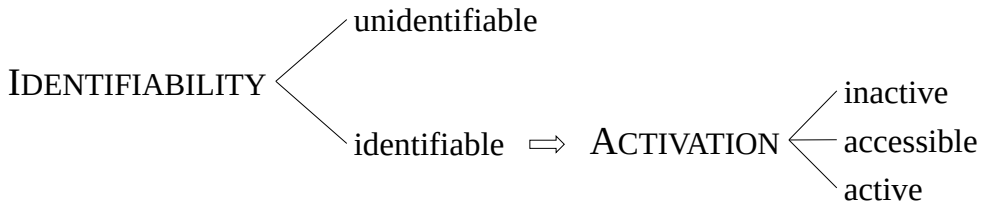


Figure 3.1: The relationship between identifiability and activation of referents (adapted from Lambrecht, 1994: 109).

(but not directly focussed on). A referent that is inactive is in a person's long-term memory, but neither focussed on nor peripherally active (Chafe, 1987: 22 ff). Semiactive or accessible referents can either be textually accessible, when they were active at an earlier point in the discourse and their activation has depreciated. Or they can be inferentially accessible, when they are derivable from the preceding contextual frame (e.g. the anaphor "the teacher" can be assumed to be inferable from an antecedent "school"). According to Lambrecht (1994), a third variant is what he refers to as "situationally" accessible, when a referent is physically present in the discourse environment and can be referred to deictically.

Just as there is a correlation between the (un)identifiability of a referent and the (in)definitness of its referring expression, there is also a correlation between the activation state of a referent and the explicitness of the referring expression used. Furthermore, other aspects of the linguistic realisation of a referent, such as its prosodic marking (see section 3.3), are also directly connected to both the identifiability and the activation state of a referent.
Chafe (1994: 74) explains this correlation in terms of an activation *cost* that a speaker has to invest in order to transfer a referent[11] from its previous state into an active state.
As illustrated in figure 3.2, the activation cost is highest (as indicated by the longest arrow) if a referent is in an inactive state. The degree of givenness of such an inactive item is *new*. A referent that becomes activated from a semiactive state is *accessible*. Lastly, the activation cost is lowest if a referent is already active, which is when its degree of givenness is *given*.

[11] It should be noted that Chafe (1987) applies these activation states to "concepts", or later, "ideas" (Chafe, 1994: 80), which include referents, but also events and states (see also discussion above about different domains of givenness). While acknowledging Chafe's view, this thesis will focus on activation states of *referents*.

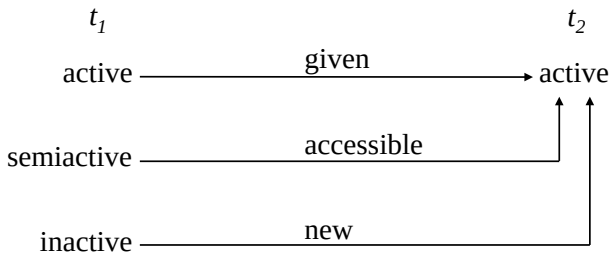


Figure 3.2: Activation states, activation costs, and time (adapted from Chafe, 1994: 73).

Thus, an increased activation is directly correlated with a decreased effort when mentioning the referent in that the more active a referent is, the less articulatory effort and phonological material is used. There is an inverse relation between the explicitness of an anaphoric expression (in terms of descriptive, lexical, and phonological material) and the degree of activation of the associated referent. This makes the relationship between the activation state of a referent and its linguistic realisation an iconic one (see Lambrecht, 1994: 96). Chafe (1987: 26) describes this correlation as follows:

> "Those concepts which are already active for the speaker, and which the speaker judges to be active for the hearer as well, are verbalised in a special way, having properties which have often been discussed in terms of "old" or "given" information. The general thing to say is that given concepts are spoken with an attentuated pronunciation. The attenuation involves, at the very least, weak stress. Typically, though not always, it involves either pronominalisation or omission from verbalisation altogether."

This generalisation is captured in a number of scales or hierarchies of givenness (Ariel, 1990; Gundel et al., 1993; Prince, 1981; Givón, 1983, among others). Some of these hierarchies show fine-grained differences between linguistic expressions and their corresponding accessibility or cognitive state. They differ in terms of their description levels, as well as in terms of what cognitive states they define or assume and whether these are continuous or discrete. However, they usually agree on the fact that more elaborate types of referring expressions, such as full noun phrases (including determiners and modifiers), are associated with referents that are less active in a discourse, while shorter forms, pronouns and zero anaphors are associated with more active referents. A simplified hierarchy of referring expressions with regard to their associated cognitive state and givenness

is illustrated in figure 3.3.

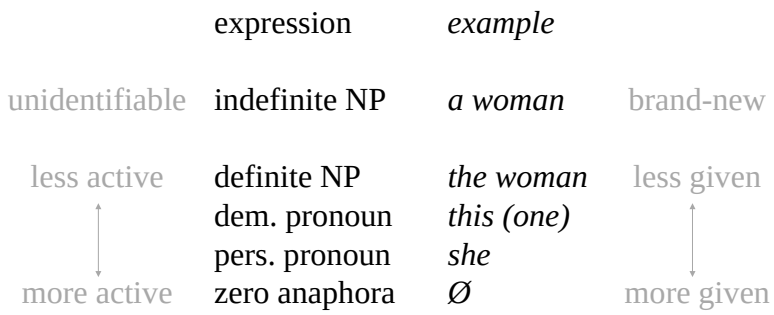


Figure 3.3: Hierarchy of referring expressions with examples, and corresponding cognitive state and givenness.

The scale in figure 3.3 shows how the effort that has to be put into activating a referent in the mind of a listener is manifested in the explicitness and complexity of the corresponding referring expression.

It is important to note that the cognitive activation of a referent, once introduced, will not infinitely stay active. If a referent has not been mentioned for a while, the mental representation for it declines in activation and therefore becomes less accessible. Another reason why a conceptual referent may become less accessible is interference from other referents which are mentioned (Gernsbacher, 1990; Arnold, 1998). Activation ceases especially when another referent becomes the topic in a discourse. In order to then reactivate a recent non-topic referent, a referring expression with more descriptive content and a more prominent prosody has to be used (Givón, 1983) in order to resolve potential ambiguity.

As mentioned before, not only the choice of referring expression, but also its prosodic marking, contribute to signalling which referent is being talked about, whether it is identifiable and how accessible it is. The way in which prosody is used to mark referential givenness is explained in the following section 3.3.

3.3 Prosodic Marking of Referential Givenness

As discussed above, speakers select referring expressions on the basis of their representation of the interlocutor's knowledge (e.g. indefinite NPs for unidentifiable referents, pronouns for active referents, etc.). Crucially, they also select a prosodic marking of that referring expression on the basis of this representation.

For West Germanic languages, new referents have been shown to be marked with (nuclear) pitch accents, whereas given referents are not (e.g. Allerton, 1978; Terken and Hirschberg, 1994; Féry and Kügler, 2008; Cruttenden, 2006; Röhr and Baumann, 2010). Furthermore, the type of pitch accent provides important cues for different degrees of referential givenness (Pierrehumbert and Hirschberg, 1990; Kohler, 1991a; Chen, Den Os, and De Ruiter, 2007; Schumacher and Baumann, 2010; Baumann and Grice, 2006; Röhr and Baumann, 2010, 2011) in a way that newer, or less activated, referents are often marked with more prominent accent types than more given or activated referents.
These findings are in line with Chafe's (1994) activation cost model and provide evidence for the assumption that the higher the assumed activation of a referent in the listener's mind, the lower the prosodic prominence required for the listener to decode it (Baumann, Röhr, and Grice, 2015). Thus, the direct correlation of effort and activation concerning the choice of referring expression also holds for their prosodic marking.

Furthermore, an unidentifiable or brand-new referent is necessarily prosodically prominent, since "lack of [prosodic] prominence is reserved for constituents with active referents" (Lambrecht, 1994: 105). This is in line with the assumption that brand-new referents require the highest effort (or cause the highest activation cost) because they still need to be added to the memory representation of the listener. On the other hand, attenuated prosody can signal that the linguistic expression should be mapped onto an existing discourse referent (Birch and Clifton, 1995).

In fact, prosody can also be used as the sole cue to resolve referential ambiguities, as shown in example (19):

(19) Hannah helped her sister with the dishes.
And then **she** helped **her** with the homework.

When the speaker emphasises the two pronouns in (19) prosodically through pitch accents, the intent is to express that the sister helped Hannah with the homework. If the pronouns were not emphasised prosodically, the sentence would be interpreted as meaning that Hannah also helped the sister with the homework. This shows that prosodic cues can be used by speakers to (re)orient the listener's attention between two (potentially ambiguous) referents as in example (19), in which more prominent prosody has to be used to (re-)activate referents and switch

between *Hannah* and *her sister*.
Therefore, prosodic highlighting is not only necessary to introduce brand-new referents into discourse, but also to reactivate them and promote them back into the focus of attention. Hence, not only the identifiability and activation of referents influence their prosodic marking, but also their reactivation due to interference from other referents.
However, it should be kept in mind that, as demonstrated in section 2.3.2 of chapter 2, focus prosody can override activation prosody in the sense that speakers have a choice to present an entity as particularly newsworthy and to highlight it irrespective of its activation state.

The connection between perceived prominence of pitch accents and aspects of prosodic meaning that has been presented in this section is compatible with Gussenhoven's (2004) "Effort Code". The Effort Code covers a majority of the attributes that are involved in manifestations of prosodic prominence. This code states that the energy expenditure during speech production can be varied by speakers, who are assumed to produce certain parts of speech with more effort to highlight their importance or newness for the listener. The higher the pitch on a lexically stressed syllable (which implies increased effort in speech production), the higher its perceived prominence and the newer (or less accessible) the referent.

Appropriate prosodic marking of information substantially facilitates comprehension, whereas prosodic inappropriateness will slow or distort comprehension (Nooteboom and Terken, 1982; Birch and Clifton, 1995; Chen, 2010). Furthermore, "the interaction between [...] [information] structure and accenting can play a major role in discourse comprehension by influencing the mapping process between linguistic expressions and discourse entities" (Birch and Clifton, 1995: 386).

3.4 Summary

This chapter gave an overview of the three levels of information structure that contribute to the way in which speakers organise the information they want to communicate, namely givenness, aboutness and informativeness.
In particular, it was shown that both context and shared knowledge lead to different cognitive representations of information, which, in turn, have formal

correlates in linguistic structure and prosody.

The focus of this chapter was on the notion of referential givenness, which involves the cognitive representation of referents in the interlocuters' minds with regard to the current discourse environment. The two central aspects of referential givenness are identifiability and activation. Linguistically, referential givenness has been shown to be expressed in the form of referring expression as well as in its prosodic marking.

A referent in an active status is formally expressed through pronominalisation and, prosodically, through weak prominence or the absence of accentuation. A referent with an inactive status, on the other hand, requires a higher activation cost, which usually manifests in higher prosodic prominence and the use of more phonological material (e.g. full NPs instead of pronouns).
Furthermore, a high activation cost is also required to reactivate referents and switch between them in order to avoid ambiguity.

Successful referential communication has been linked to perspective-taking skills in the typically developing population (Volden, Mulcahy, and Holdgrafer, 1997; Roberts and Patterson, 1983). However, it is not always clear whether the referential choices that speakers make are listener-driven or simply context-driven, as both the listener and the speaker usually have access to the same discourse context (Hendriks et al., 2014: 392). There is evidence that referential choices and their prosodic marking are, at least to some degree, listener-driven (see, e.g., Hendriks et al., 2014; Galati and Brennan, 2010; for a review see Brennan and Hanna, 2009). Some have argued for models in which the early, rapid processing is susceptible to egocentric speaker-only knowledge, and more effortful listener-driven adjustments emerge relatively later (e.g. Bard et al., 2000).
Listeners have been shown to use the cues to referential givenness provided by speakers. For example, deaccented nouns have been found to be interpreted as anaphoric (Dahan, Tanenhaus, and Chambersc, 2002), accented nouns have been found to be interpreted as newer than unaccented nouns (Baumann, Röhr, and Grice, 2015), and the appropriate prosodic marking of information has been found to facilitate comprehension and enhance communication (Birch and Clifton, 1995).
It can thus be concluded that perspective-taking capacities are, at least to a certain degree, required to encode and decode referential givenness appropriately (Clark and Murphy, 1982; Galati and Brennan, 2010; Kaland, Swerts, and Krah-

mer, 2013). The following chapter will be concerned with a population that is known to have impaired perspective-taking abilities, namely individuals with autism.

Chapter 4

Autism

4.1 Classification and Characteristics

Autism is a neurodevelopmental disorder characterised by persistent deficits in both social interaction and communication, associated with repetitive, restricted behaviours, interests, and / or activities (DSM-V[1] (American Psychiatric Association, 2013), ICD-10[2] (World Health Organization, 1992)).

According to the ICD-10, which is the mandatory classification system for clinical diagnoses in Germany, the following subgroups can be defined: Individuals that show all of the core symptoms mentioned above and a low IQ (< 70) are diagnosed with *early infantile autism* (F84.0). Individuals with normal or above average intelligence who experienced a delay in language acquisition are also diagnosed with early infantile autism (F84.0), but their condition is often referred to as *high-functioning autism* (HFA), although this is not an official term used in the ICD-10. Individuals with an IQ> 70 and without delays in language acquisition and cognitive development are diagnosed with *Asperger Syndrome*[3] (AS (F84.5)). Finally, *atypical autism* (F84.1) is diagnosed if not all of the core symptoms occur and if impairments appear after the age of three (see Krämer, Gawronski, and Vogeley, 2016). This classification system is controversial. In particular, the claim that AS and (high-functioning) autism are two distinct diagnostic categories has

[1] Diagnostic and Statistical Manual of Mental Disorders, 5th ed.

[2] International Statistical Classification of Diseases and Related Health Problems, 10th revision.

[3] or *Asperger's Syndrome*. Wing (1981) first referred to the syndrome as "Asperger's Syndrome", but Frith (1991) later states that "both terms are widely used." (Frith, 1991: 1).

been a matter of debate and is probably mainly an artefact of the two separate, independent primary descriptions of what is currently assumed to be the same underlying disorder (according to the DSM-V (American Psychiatric Association, 2013)).

In the 1940s, Leo Kanner, a child psychiatrist from Baltimore, and Hans Asperger, a pediatrician from Vienna, independently published case studies of children with atypical patterns of behavior. The symptoms and characteristics that both authors described were, in many ways, similar. In a comparison of the two accounts, Wing (1991) identified several "striking similarities between the children described by the two authors" (Wing, 1991: 96), such as, e.g. "social isolation, egocentricity and lack of interest in the feelings or ideas of others". Also, both authors identified diverse communication deficits such as "the lack of use of language for interchange with others" (Wing, 1991: 96), pedantic speech, poor eye contact and "peculiar vocal intonation" (Wing, 1991: 96). Both authors noticed repetitive patterns of activities and stereotyped behaviours as well as rejection of changes in routines and environment. Also, both said that the condition was more prevalent in males than females.
Furthermore, both Kanner (1943) and Asperger (1944) used the term "autistic" to refer to the conditions that they described, referring to Bleuler's (1911) concept of "Autism" that he had originally established to describe the withdrawal from reality in people with schizophrenia.

Ever since parts of the work of Asperger were translated into English (Wing, 1981, Frith, 1991) his "Autistic Psychopathy" was referred to as "Asperger Syndrome". The work of Kanner, on the other hand, was referred to as "early infantile autism" or "classic autism". Despite many similarities between the descriptions, there have been attempts to discriminate one condition from the other (e.g. van Krevelen, 1971).
Although Asperger (1968) himself acknowledged that the two descriptions show striking similarities in both central aspects as well as in small subtleties, he also pointed out some differences. He described Kanner's early infantile autism as being comparable to a psychotic condition while his own cases were not. Furthermore, he claimed that the children he described had less severe early symptoms, an earlier language acquisition and better language skills than those described by Kanner (1943).

On the other hand, Asperger (1968) emphasised the fact that the impairments

in the children he described were simply not as severe as the impairments in children with Kanner's early infantile autism (Asperger, 1968: 141). From this, Wing (1981: 122) concludes:

> "[...] the impression gained [...] is that, although there are some differences, the syndromes are more alike than unalike. The variations could be explained on the basis of the severity of the impairments [...]."

Initially, Wing's (1991) suggestion of an "autistic continuum" that includes AS was neglected in the field (Wing, 2000). Instead, the differences that Asperger (1968) had emphasised (no history of language delay, later and less severe manifestation of impairments in AS) had a direct influence on the diagnostic criteria for autism and later became the central distinguishing features between AS and (high-functioning) autism when AS was included in the DSM-IV (1994) and the ICD-10 (1992).

However, research failed to provide cogent evidence to differentiate between the established subgroups and instead suggests that autism in general and AS in particular are one underlying nosological disorder, manifesting in a spectrum with common primary deficits and individual differences (e.g. Frazier et al., 2012; Kamp-Becker et al., 2010; Lord et al., 2012; Mayes, Calhoun, and Crites, 2001; Miller and Ozonoff, 2000; Sanders, 2009).

This assumption has been incorporated into the DSM-V, which redefines autism as a spectrum disorder (*Autism Spectrum Disorder* (ASD)) that includes AS, with the occurence and severity of deficits varying greatly from one individual to another. The spectrum ranges from individuals with "low-functioning" autism and associated learning difficulties who need support throughout life, to "high-functioning" autonomous individuals with normal or above average intelligence (IQ$>$ 70). The upcoming revision of the ICD (ICD-11, expected in 2018) will align with the DSM-V with regard to including AS in ASD. Accordingly, the term ASD will be used in this thesis to refer to all autistic disorders including AS. Additional information about anterior diagnostic subgroups will be given where necessary.

The current estimated prevalence of ASD ranges from about 0.62% (1 in 161; Elsabbagh et al., 2012) up to 1.89% (1 in 53; Kim et al., 2011). The official report of the Centers for Disease Control and Prevention (CDC) estimate the

current prevalence of ASD to be 1.5% (1 in 68; Christensen et al., 2016).

4.2 Autism and Language

One of the core symptoms of ASD is related to deficits in communication. Language impairments have been a central element in descriptions of ASD since the earliest observational reports. Both Kanner (1943) and Asperger (1944) noted peculiarities in the way that the verbal children they described used language and produced speech.

The manifestations of language impairments in ASD are diverse and, much like other deficits of ASD, range from mild to severe, occuring in a variety of combinations. If language is present, it has been found to be one-sided rather than reciprocal and preferably used for instrumental rather than social purposes. Individuals with ASD have also been found to be either over-informative or under-informative when providing information (Paul and Cohen, 1984; Dahlgren and Sandberg, 2008; Loveland, Tunalia, Mcevoy, and Kelley, 1989). A literal use and interpretation of language has also been frequently reported (see Kalandadze, Norbury, Nærland, and Næss (2016) for a recent review).
Another frequent observation is the fact that in high-functioning verbal adults with ASD, the rule-based, formal, grammatical aspects of language are usually intact while some of the more intuitive aspects of language such as pragmatic and affective aspects are impaired (Landa, 2000). Prosody is of obvious relevance to this subdivision of impaired aspects in ASD as it is a medium that serves grammatical, pragmatic as well as affective functions in language. The issue of what is meant by "formal" and "intuitive" has been discussed in more detail in section 2.3 of chapter 2 and will be returned to in sections 4.2.1 and 4.2.2 below.

In the literature, findings of language impairments in children, adolescents and adults with ASD are often reported without a clear distinction between these groups, leading to seemingly conflicting and diverging results within the group of individuals with ASD. In fact, the majority of research focussing on language impairments in ASD evaluated children rather than adults, and these studies have frequently been referred to in characterising language impairments in ASD in general. While those studies provide important insights into asynchronies in language development and impaired aspects of language in childhood ASD, the

results might not always be indicative for ASD in adulthood: Many aspects of language continue to develop until adulthood (see, e.g., Berman, 2008; Nippold, 2007), among them prosody (Chen, 2011a; Cruttenden, 1985; Diehl et al., 2008; Shport and Redford, 2014; Wells, Peppé, and Goulandris, 2004).
For example, typically developing children aged 11-12 years still show difficulty using prosody to disambiguate compound nouns from NPs (Vogel and Raimy, 2002). Thus, variability and asynchronies of prosodic development are common and have to be taken into account:

> "[...] the age of acquisition of a specific prosodic ability may vary; levels of ability in a specific skill vary across children; and competence in different modes (comprehension and expression of prosody) may become evident at different ages (Wells et al., 2004: 776)."

For ASD in particular, it has been reported that language skills in children with ASD show accelerated improvements after an initial delay (Gernsbacher, Morson, and Grace, 2015, see also DeMyer et al., 1973) and that children with ASD with lower inital overall language skills compared to typically developing peers show significant improvement of these skills over a nine-year period from 10 to 19 years of age (Cariello et al., 2011 as cited in Gernsbacher et al., 2015). Even children with ASD at the early stages of acquiring language who show very unusual language use such as reverse pronouns and echolalia make considerable advances later in language development (Tager-Flusberg, 1996). Also, compared to children with developmental receptive language disorders, children with ASD evinced significantly more improvement in verbal IQ and receptive language over a period of sixteen years (Mawhood, Howlin, and Rutter, 2000). In fact, early differences in language ability of children with ASD (with vs. without delay in language acquisition) can diminish over time (Gilchrist et al., 2001; Howlin, 2003). Overall, it has been shown that developmental change in communicative deficits in individuals with ASD until adulthood is prevalent (Tager-Flusberg, 1996).

Children with ASD have shown improvement with age of prosodic abilities, too (Lyons, Schoen Simmons, and Paul, 2014). Prosodic abilites correlate both with chronological age and verbal mental age in children with ASD (Peppé, McCann, Gibbon, O'Hare, and Rutherford, 2007), suggesting ongoing prosodic development through childhood in ASD. Studies investigating receptive prosodic skills in children with ASD from different age groups (6-13 years (Peppé et al., 2007) vs. 7-16 years (Järvinen-Pasley, Peppé, King-Smith, and Heaton, 2008b) vs. 14-21

years (Paul, Augustyn, Klin, and Volkmar, 2005)) observed differing patterns of prosodic impairment and ability. Järvinen-Pasley, Peppé, et al. (2008) ascribed the divergent results to the fact that the youngest group of children "had not yet acquired the necessary skills" (Järvinen-Pasley, Peppé, et al., 2008b: 1334) and Peppé et al. suggest that "maturation alone may result in improved receptive prosody" (Peppé et al., 2007: 1022-23). Nevertheless, some prosodic impairments seem to persist into adulthood and show little improvement over time (Shriberg et al., 2001; McCann and Peppé, 2003).

These findings indicate that the acquisition of language in general and of prosody in particular is often delayed rather than deviant in children with ASD and early language impairments are therefore not necessarily informative about later functioning in ASD. Accordingly, in this thesis, studies evaluating impairments in childhood will be reviewed seperately from studies evaluating adults. Furthermore, this thesis will focus on studies investigating language in verbal individuals with ASD with normal or above average IQ. Accordingly, the conducted experiments that this thesis reports on in chapters 5 and 6 also examined adults with normal or above average IQ.

This approach will ensure to factor out impairments that could be ascribed to an incomplete language development, underdeveloped cognitive capacities or intellectual disability rather than being an inherent element of language in ASD that persists into adulthood. Furthermore, pragmatic deficits can be isolated more easily when the faculty of speech and grammatical skills are intact (Baltaxe, 1977).

The following sections will provide an overview of research investigating perception (section 4.2.1) and production (section 4.2.2) of prosodic aspects in children and adults with ASD. At the end of each section, a table will summarise the findings for each group of participants for perception (table 4.1 for children and table 4.2 for adults) and for production (table 4.3 for children and table 4.4 for adults).

4.2.1 Perception of Prosody

Understanding others is crucial for successful and efficient communication. Prosody is an important aspect for the perception of speech, as it provides the listener with cues for the segmentation and interpretation of the speech signal and at the

same time provides information about the speaker, such as, for example, his/her affective state (Baltaxe and Simmons, 1985: 96).
In general, the ability to accurately process and interpret auditory information has been found to be difficult for individuals with ASD, especially in speech. In her extensive review of research investigating auditory processing in ASD, O'Connor (2012) specifies two main approaches that address atypical processing of auditory information in ASD, namely a) "social theories" focussing on impairments in social cognition and b) "general" theories focussing on general differences in information processing.

The *social* theories are based on the hypothesis that ASD involves deficits in *mentalising* (or *Theory of Mind (ToM)*) which denotes the ability to spontaneously ascribe mental states such as intentions, knowledge and beliefs to others and to understand, interpret or predict their behaviour and emotions (Happé, 2015). This approach can account for the fact that while individuals with ASD are able to use explicit verbal information, they often fail to process non-verbal (prosodic) cues in social encounters (see next section). With respect to prosody perception this would mean that prosodic aspects of speech signalling the emotions, opinions or thoughts of a conversational partner would be more difficult for an individual with ASD to interpret than those serving rule-based functions (but see Tager-Flusberg (2000) for a discussion about lexical and grammatical aspects also being affected by mentalising deficits).

The *general* theories include the executive dysfunction (ED) hypothesis as well as the theory about weak central coherence (WCC). The ED hypothesis (Griffith, Pennington, Wehner, and Rogers, 1999; Robinson, Goddard, Dritschel, Wisley, and Howlin, 2009; Friedman et al., 2006) suggests that individuals with ASD might have problems in cognitive domains including planning, attention shifting, problem solving, and working memory (Segal, Kaplan, Patael, and Kishon-Rabin, 2017: 68). The WCC account (Frith, 1989; Happé and Frith, 2006) refers to an information-processing style of people with ASD that involves an "enhanced processing of local information or detail" (O'Connor, 2012: 837) while paying less attention to global information and information in context. In the linguistic domain, WCC may explain impairments of pragmatic aspects of language, which are context-dependent. The WCC approach also includes the hypothesis that in ASD, the perception of low-level auditory information is enhanced, while the perception of more complex auditory information is impaired (Bertone, Mottron, Jelenic, and Faubert, 2005; Mottron, Dawson, Soulières, Hubert, and Burack,

2006).

Speech sounds include multiple sources of complexity, such as fundamental frequency, formants, and rapidly changing consonant sounds, making them very complex. In fact, speech prosody itself can already be regarded as a complex auditory signal that includes variations of fundamental frequency, intensity, duration of syllables, and pauses (see chapter 2). Therefore it could be assumed that the perception of isolated basic components of prosody such as pitch, loudness, or segregated rhythmic patterns would be enhanced in ASD, while the perception of prosodic cues that are spectrally and temporally more complex as in natural speech could show impairments. This proposition is supported by the fact that musical abilities in individuals with ASD have been reported to be spared or even enhanced (Molnar-Szakacs and Heaton, 2012) while aspects involving speech prosody have often been observed to be impaired in ASD (DePriest, Glushko, Steinhauer, and Koelsch, 2017). For example, DePriest et al. (2017) report impaired processing of prosodic phrase boundaries along with preserved processing of musical phrase boundaries in individuals with ASD. The high prevalence of absolute pitch and musical savants among individuals with ASD (Happé, 1999; Hermelin, 2001) and enhanced performance in pitch naming and pitch memory tasks (Heaton, 2003) further support this theory (see also Chevallier, Noveck, Happé, and Wilson, 2009).
An enhanced pitch discrimination ability has been suggested to not always be advantageous for speech perception (Mayer, Hannent, and Heaton, 2016). For example, Heaton, Davis, and Happé (2008) reported that in childhood, a man with high-functioning ASD and absolute pitch used to have difficulty understanding that the same word, when spoken by a male or a female speaker, was not a different word.

Since perceptual performance not only depends on stimuli properties (simple versus complex) but also on task demands, it would be expected that low-level categorisation, detection or discrimination tasks are less affected than evaluation or comprehension tasks that require interpretation of prosodic cues in context and in relation to other cues. This, in turn, also supports the ED hypothesis.

Both the social theories and the general theories have been substantiated by results from perception experiments. It is therefore assumed that impairments in auditory processing in ASD originate from a complex interplay between mentalising deficits, alternative processing mechanisms and weak central coherence. The

next sections will provide evidence for impaired perception of prosodic aspects of speech in children and adults with ASD.

Perception of Prosody in Children with ASD

Studies investigating children with ASD have found impairments in various areas of prosody perception and comprehension. Most work on the perception of prosody in children with ASD has been concerned with the recognition of emotion (Brooks and Ploog, 2013; Chevallier, Noveck, Happé, and Wilson, 2011; Diehl and Paul, 2013; Kjeelgaard and Tager-Flusberg, 2013; Lindner and Rosén, 2006; McCann, Peppé, Gibbon, O'Hare, and Rutherford, 2007; Paul et al., 2005; Peppé et al., 2007; Van Lancker, Cornelius, and Kreiman, 1989). Most of these studies (except Chevallier et al., 2011) found difficulties identifying or processing emotions through prosody in children with ASD, suggesting mentalising deficits. However, older adolescents with ASD (12-16 years) were able to recognise very basic emotions (happiness, anger, sadness) from prosody (Heikkinen et al., 2010) as well as their typically developed peers did.
Some of the studies that investigated emotion recognition in ASD report an overreliance on verbal / lexical content of speech, especially when information from the prosodic and the lexical domains was conflicting or competing (Lindner and Rosén, 2006; Kjeelgaard and Tager-Flusberg, 2013; Van Lancker et al., 1989).
In the study by Paul et al. (2005), adolescents with ASD seemed able to discriminate between excited and calm speech, but they did not seem to base their judgements on all available prosodic cues but exclusively on speech-rate, which shows their ability to adopt alternative strategies when approaching such a task. Moreover, the strategy of focussing on one aspect of the speech signal conforms with the hypothesis that individuals with ASD pay less attention to global and complex auditory information and exhibit enhanced processing of detail (WCC theory).

Impaired prosody perception in children with ASD goes beyond recognising emotions conveyed. The recognition of pragmatic functions conveyed by prosody has also been found to be difficult. This includes impaired perception of accentuation when it serves to mark contrastive focus (Paul et al., 2005; Peppé et al., 2007; Diehl and Paul, 2013). In the study conducted by Paul and colleagues (2005), participants were asked to listen to sentences with a contrastive pitch accent and to choose the sentence that preceded the sentence they heard. The following ex-

ample shows one of the sentences they heard (with a contrastive pitch accent on *bank*) and the two sentences out of which one had to be chosen as the preceding context sentence (Paul et al., 2005: 216):

(20) Go in front of the **bank**, I said.
  a. I waited for you out back.
  b. I waited for you at the grocery store.

In this example, the correct preceding sentence is (20b), because the pitch accent on *bank* is evoked by a contrast to the *grocery store*. If the contrastive pitch accent had been realised on the word *front*, then (20a) would have been the correct answer. Children with ASD made significantly less correct choices in this task, pointing towards difficulties in understanding contrastive accentuation.

Paul et al. (2005) also detected a tendency towards impairments of lexical stress perception (discriminating, e.g., re**call** from **re**call) but this finding failed to reach significance. A few years later Chevallier et al. (2009) also found no differences between adolescents with ASD and controls regarding perception of lexical stress discriminating the same kinds of minimal pairs in sentences. Along these lines, Grossman, Bemis, Plesa Skwerer, and Tager-Flusberg (2010) found that children and adolescents with ASD successfully used lexical stress to disambiguate compound nouns (e.g. "**green**house" vs. "green **house**").
These findings indicate that children with ASD are able to decode stress when used in a context-independent manner (lexical stress) but show impairments when stress, or, more precicesly, accentuation, is used in a context-dependent manner, e.g. for signalling contrast, an aspect of information structure.

Most studies found no differences in perceiving prosodic cues to syntactic parsing in children and adolescents with ASD when compared to control groups (Paul et al., 2005; Peppé et al., 2007; Chevallier et al., 2009; Diehl and Paul, 2013). When prosodic boundaries provided syntactic information, participants in both groups were equally able to resolve ambiguities such as, e.g. "Ellen, the dentist is here" vs. "Ellen, the dentist, is here" (Paul et al., 2005) or "dragonfly and carrot" vs. "dragon, fly and carrot" (Chevallier et al., 2009) when explicitly asked to disambiguate. However, in the study with the youngest participants (Peppé et al., 2007, children aged 6-13), both groups performed roughly at chance level on the chunking task, which might have masked group differences (Diehl et al., 2008) and indicates that the ability for syntactic parsing through prosody was not fully

developed in either group yet.
In contrast, another study with participants with a slightly older age range (7-16 years) found that children with ASD performed worse than matched controls in such a chunking task (Järvinen-Pasley, Peppé, et al., 2008b), suggesting that age and general development plays an important role here (see Diehl et al., 2015). Furthermore, it indicates that studies with children which did not find group differences might have suffered from the fact that either some control participants had not yet fully acquired the ability to use prosodic cues to parse sentences or that their task design was more explicit and therefore easier to fulfill.

Interestingly, a study that used a less explicit task (Diehl et al., 2008) detected difficulties in children and adolescents with ASD using prosody to disambiguate sentence meanings such as in the following examples (21a) and (21b):

(21) a. [Put the dog] [in the box on the star].
b. [Put the dog in the box] [on the star].

It should be noted that in this study, the age range of participants was between 11 and 19 years, which indicates that the impairment of this aspect of prosodic decoding might persist into adulthood.
Compared to the other chunking tasks mentioned above, this task was more difficult, as it required the participants to comprehend, interpret and execute the instructions, rather than to just discriminate two stimuli. Furthermore, it tested a more syntactically challenging ambiguity than the paradigm used in, e.g., Chevallier et al. (2009), who investigated compounds ("dragonfly") versus split compounds ("dragon, fly").

In the previous section it was hypothesised that basic components of prosody such as isolated pitch or loudness would be enhanced while complex prosodic cues that would have to be interpreted in context would be impaired in individuals with ASD. In fact, research showed enhanced discrimination of voice pitch[4] in children with ASD (Heaton, Hudry, Ludlow, and Hill, 2008; Järvinen-Pasley, Wallace, Ramus, Happé, and Heaton, 2008; Järvinen-Pasley, Pasley, and Heaton, 2008). Järvinen-Pasley and colleagues (2008a) found that high-functioning children with ASD (aged 7-16) outperformed control participants in matching sentence pitch contours to visual representations of those contours, while they showed difficulties in the associated comprehension task. In a subsequent study, Järvinen-Pasley and

[4] And of *musical* pitch as mentioned in the previous section.

colleagues (2008c) found that children with ASD preferred to match sentences according to pitch contours rather than to pictures summarising the content of the sentence, suggesting a reduced tendency to process language meaningfully in context.

It is interesting to note that the same children with ASD who outperformed control participants in matching sentence pitch contours to visual analogues exhibit difficulties discriminating questions from statements by means of pro-sody (Järvinen-Pasley, Peppé, et al., 2008; Peppé et al., 2007). They showed selective attention to lexical content rather than information provided by prosody and rated sentences without "Wh-" question words as statements regardless of their prosodic realisation. Instead of taking the prosodic information into account, they relied on lexical cues they were familiar with (question words).
Children with ASD were also found to take longer to respond and take the turn during dialogues (Heeman, Lunsford, Selfridge, Black, and van Santen, 2010), which contributes to the assumption that they are less sensitive to question and turn-taking cues.

However, studies investigating older groups of adolescents with ASD (Chevallier et al. (2009), 11-17 years; Paul et al. (2005), 14-21 years) found that their participants with ASD were able to discriminate questions from statements by means of prosody. Furthermore, less complex one-word utterances were correctly identified as questions or statements by children with ASD aged 8 and 9 years (Filipe, Frota, Castro, and Vicente, 2014).

The studies mentioned in this section are summarised in table 4.1 below.

Table 4.1: Overview: Prosody perception in children and adolescents with ASD.

FORMAL ↔ *INTUITIVE*

Grammatical — *Pragmatic* — *Affective*

Function	***Details***	***Impairments found***	***No impairments found***
Lexical Stress			Chevallier et al. (2009) Grossman et al. (2010) Paul et al. (2005)
Syntactic Structure	Implicit > explicit tasks	Diehl et al. (2008) Järvinen-Pasley, Peppé, et al. (2008)	Chevallier et al. (2009) Diehl and Paul (2013) Paul et al. (2005) Peppé et al. (2007)
Speech Acts	Younger > older participants Complex > one-word utterances	Järvinen-Pasley, Peppé, et al. (2008) Peppé et al. (2007)	Chevallier et al. (2009) Filipe et al. (2014) Paul et al. (2005)
Turn-taking		Heeman et al. (2010)	
Information Structure	Contrast only	Diehl and Paul (2013) Paul et al. (2005) Peppé et al. (2007)	
Emotional State	Complex > basic Younger > older participants	Brooks and Ploog (2013) Kjeelgaard and Tager-Flusberg (2013) Lindner and Rosén (2006) McCann et al. (2007) Paul et al. (2005) Van Lancker et al. (1989) Järvinen-Pasley, Peppé, et al. (2008) Peppé et al. (2007)	Chevallier et al. (2011) Heikkinen et al. (2010)

Perception of Prosody in Adults with ASD

Compared to children with ASD, adults with ASD are generally found to exhibit less severe impairments when processing prosody. This can mostly be attributed to the fact that in adulthood, language and cognitive development are complete and do not interfere with language abilities anymore. Another frequent observation, especially in the domain of perception, is the fact that adults with ASD often develop compensation strategies[5] to cope with social encounters. In that case, learned rules often replace an intuitive understanding of social cues (Livingston and Happé, 2017).
While these strategies might be sufficient in certain contexts and especially in controlled experimental settings, they cannot always support performance in unpredictable, novel, or fast-paced settings and tend to break down in situations where more complex or implicit processing skills are required, such as everyday real life social interactions.
Experimental tasks that imitate such settings have shown to successfully bypass compensation strategies. For example, while adults with ASD are able to pass explicit simple ToM tasks (e.g. false belief tasks), they exhibit deficits in implicit (Schneider, Slaughter, Bayliss, and Dux, 2013; Senju, 2013) and more naturalistic and dynamic tasks with limited time constraints (Brewer, Young, and Barnett, 2017). This has also been validated for tasks involving the processing of prosodic cues to interpret emotions: A "Reading the Mind in the Voice" task (Rutherford, Baron-Cohen, and Wheelwright, 2002) that suffered from ceiling effects and limited sensitivity yielded a better reliability and validity after the task was made more complex (Golan, Baron-Cohen, Hill, and Rutherford, 2007).
These findings emphasise both the importance of task design and the significance of studies that, despite possible compensation, find consistent impairments in adults with ASD when processing prosody. These findings can provide reliable evidence that some deficits remain even when language development is complete and compensation mechanisms are involved.

Adults with ASD are capable of recognising basic emotions such as happiness, sadness, or anger (O'Connor, 2007; Stewart, McAdam, Ota, Peppé, and Cleland, 2013), but exhibit difficulties recognising complex and situation-based emotions (e.g. confusion, resignation, gratitude) from prosodic cues (Golan et al., 2007;

[5]But it should be noted that even young children with ASD may develop compensation strategies.

Hesling et al., 2010; Kleinman, Marciano, and Ault, 2001; Rutherford et al., 2002; Rosenblau, Kliemann, Dziobek, and Heekeren, 2017). The intensity of the cues also plays a role, in a way that high-intensity (exaggerated) basic emotions have been found to be easily recognisable for adults with ASD, while the same basic emotions with low intensity and subtle cues were not (Globerson, Amir, Kishon-Rabin, and Golan, 2015). Individuals with ASD were also found to rate the intensity of emotions as less emotionally intense than a control group (Gebauer, Skewes, Hørlyck, and Vuust, 2014).
Furthermore, when the emotion conveyed through prosody is incongruent[6] with respect to the semantic content of a sentence, individuals with ASD relied more on the semantic cues and less on the prosodic cues than controls (Stewart et al., 2013).

Impaired prosody perception in adults with ASD goes beyond recognising complex and / or subtle emotions. In general, according to a recent study, adults with ASD show a reduced sensitivity when making judgements about stress patterns relative to controls (Kargas, López, Morris, and Reddy, 2016). Participants with ASD had difficulties detecting stress pattern differences in word pairs such as, e.g. "**au**ditory" and "au**di**tory". However, there was considerable variation of performance on the task within the ASD group. Speech production abnormalities (as measured by the Autism Diagnostic Observation Schedule (ADOS) score (Lord et al., 2000)) in the ASD group correlated with sensitivity to syllable stress in a way that participants with speech abnormalities were less sensitive to syllable stress differences. This implies that adults with ASD who do not exhibit obvious speech production abnormalities (such as, e.g., atypically slow or rapid speech) are sensitive to syllable stress.

Globerson et al. (2015) investigated the perception of contrastive prosody in adults with ASD. Participants listened to sentences such as "There are <u>birds</u> in the park" with a contrastive pitch accent on "birds" and had to choose one of the following interpretations:

1. There **are** birds in the park, as opposed to "there are **no** birds in the park."
2. There are **birds** in the park, as opposed to **monkeys**.
3. There are birds in the **park**, as opposed to the **zoo**.

[6] In communicative situations, emotion conveyed through prosody and the semantic content of an utterance are usually congruent, unless a non-literal interpretation is desired, for example when expressing irony or sarcasm. Concurrently, understanding irony has been shown to be difficult for people with ASD (e.g. Martin and McDonald, 2004).

Globerson et al. (2015) found that adults with ASD were able to detect contrastive accentuation in this explicit task in which multiple choice answers gave a limited set of possible interpretations. Other studies investigating the perception of prosody signalling aspects of information structure in adults with ASD are not known[7].

However, Bishop (2017) and Hurley and Bishop (2016) have shown that "autistic traits" in the general population can have an effect on the perception of prosodic cues signalling different focus sizes. Bishop (2017) found that the absence of prenuclear accents in sentences had a less reliable priming effect for narrow focus structures in individuals with higher scores on the Autism Quotient (AQ) communication subscale, which indicates reduced communication abilities. Concurrently, Hurley and Bishop (2016) investigated the interpretation of "only" in sentences such as "he only dried the bowls", where three distinct interpretations of the sentence are possible, depending on its prosodic realisation: 1. he dried nothing else, 2. he did nothing else or 3. he did nothing else to the bowls. Hurley and Bishop (2016) found that the presence of prenuclear accents on the verb as well as lower relative prominence on the object in general triggered a broad focus interpretation (2. he did nothing else). Interestingly, participants with more autistic traits did not show this pattern, which indicates that they did not rely on prosody when interpreting the meaning of "only". Instead, they tended to engage in a semantic processing of "only", favoring [1. he dried nothing else] in all conditions. This study thus provides further evidence for an over-reliance on propositional content and a reduced attention paid to prosodic cues in individuals with ASD.

Adults with ASD have been reported to exhibit impairments perceiving prosodic cues to phrasing, too. For Greek, Martzoukou, Papadopoulou, and Kosmidis (2017) showed that individuals with ASD had problems using prosody to resolve structural ambiguities of sentences as in example (22):

(22) While she was sewing $_{(,)}$ the buttons ...
  a. ...she slipped on the floor.
  b. ...slipped on the floor.

In this example, the absence of a prosodic phrase boundary after the word "sewing" signals a continuation of the sentence as in (22a.), with "the buttons"

[7]But see Paul et al. (2005), who found impaired perception of contrastive accentuation in adolescents with ASD. This study has been discussed in the previous section.

constituting the object of the subordinate phrase. On the other hand, the presence of a prosodic boundary after the word "sewing" signals a continuation of the sentence as in (22b.), with "the buttons" constituting the subject of the main phrase. The ASD group performed similar to the control group on the object condition, i.e. in cases without prosodic boundaries, indicating they correctly chose (22a.) as the appropriate continuation of the sentence. However, the ASD group performed significantly worse in the subject condition and had problems resolving the ambiguity of the first part of the sentence when a phrase boundary signalled that "the buttons" constituted the subject of the following main phrase. Moreover, the ASD group manifested slower reaction times in all conditions. These findings show that adults with ASD face slight difficulties with the decoding of prosody when it is used to indicate syntactic structure and phrasing.

In the tone language Cantonese, where prosody determines the meaning of words through lexical tones, native speakers with ASD were found to perform similarly to controls in a discrimination task (Cheng, Lam, and To, 2017). Interestingly, the experience of speaking and understanding a tone language may have modulated the pitch processing mechanism in all Cantonese speakers that participated in the task, as they exhibited significantly higher discrimination ability in non-speech stimuli than in pseudo-syllables. This indicates that enhanced pitch perception is prevented when listening to (pseudo-)speech in order to be able to categorise pitch variation according to its phonological representations. In this case, an enhanced pitch perception of, e.g., a one semitone difference could impede a successful classification of pitch movements into lexical tone categories.

In a battery of tests, Hesling et al. (2010) investigated the ability of eight male adults with ASD to use prosodic cues serving grammatical, pragmatic, as well as affective functions. The tests involved prosodic cues indicating 1. sentence form, 2. phrasing, 3. contrast, and 4. emotional state in French. In the sentence form task, participants had to distinguish an offer from a statement by listening to single words (food items) and deciding whether the speaker "was asking them if they want some" or not. In the phrasing task, participants had to distinguish structures such as "twenty-four, twelve" from "twenty, four, twelve". In the contrast task, participants had to identify which item "had been forgotten" in sentences such as "I wanted **bread** and apples". Finally, in the emotional state task, participants listened to single words (food items) and had to decide whether the speaker liked or did not like a food item.
Individuals from the ASD group had significantly lower scores in all tasks. How-

ever, this preliminary study has two limitations: First of all, the individuals with ASD that participated in these experiments might not have had the underlying skills required to complete the tasks. This was detected in two preceding form tasks that did not involve meaning processing to test the general auditory discrimination abilities in the groups. The stimuli for this task were laryngograph signals taken from the recordings for the subsequent communicative function tasks. The participants with ASD scored significantly lower than the control group on these control tasks, suggesting that they could not reliably discriminate the underlying auditory signals that later constituted the four communicative function tasks. The second limitation was the fact that the two groups differed considerably in measures of verbal IQ (mean verbal IQ in ASD group: 89 vs. in control group: 128.33). This could presumably account for differences both in the control tasks as well as in the communicative function tasks.

Finally, three of the papers mentioned in the previous section concerning impairments in childhood investigated participants from age ranges that sometimes included young adults. For example, Paul et al. (2005) investigated individuals aged between 14 and 21, and found impairments regarding the perception of contrastive accentuation. Furthermore, Chevallier et al. (2009) investigated individuals aged between 11 and 17 and found no impairments in the perception of lexical stress, chunking, and sentence form. And lastly, Diehl et al. (2008) investigated individuals aged between 11 and 19 and found impairments in an implicit chunking task. Because the age ranges of these studies span into adulthood, the findings can also be informative of prosodic deficits in adults with ASD.
Therefore, these studies are included again in the following table 4.2, which summarises all studies mentioned in this section. The three studies that investigated children and adolescents but whose age limits reach into adulthood are marked with asterisks.

Table 4.2: Overview: Prosody perception in adults with ASD. Asterisks indicate studies that mainly investigated adolescents, but with an age range reaching into adulthood.

FORMAL ↔ INTUITIVE

Grammatical / Pragmatic / Affective

Function	***Details***	***Impairments found***	***No impairments found***
Lexical Tone			Cheng et al. (2017)
Lexical Stress	Impaired only in association with speech production abnormalities	Kargas et al. (2016)	(Paul et al. (2005)*)
Syntactic Structure	Implicit > explicit tasks	(Diehl et al. (2008))* Hesling et al. (2010) Martzoukou et al. (2017)	(Paul et al. (2005))*
Speech Acts		Hesling et al. (2010) Peppé et al. (2007)	(Chevallier et al. (2009))* (Paul et al. (2005))*
Information Structure	Focus	Bishop (2017) Hurley and Bishop (2016)	
	Contrast	Hesling et al. (2010) (Paul et al. (2005))*	Globerson et al. (2015)
Basic Emotional States	Higher > lower cue intensity	Globerson et al. (2015)	O'Connor (2007) Stewart et al. (2013)
Complex Emotional States		Golan et al. (2007) Hesling et al. (2010) Kleinman et al. (2001) Rosenblau et al. (2017) Rutherford et al. (2002)	

4.2.2 Production of Prosody

Atypical prosody can be one of the first noted characteristics of language in individuals with ASD (Kargas et al., 2016). Prosodic characteristics and impairments of the speech of individuals with ASD have been found to be diverse and highly variable. When Asperger first described the "tone of voice" of the children that he examined, he already made a note of individual differences:

> "The abnormalities differ, of course, from case to case. Sometimes the voice is soft and far away, sometimes it sounds refined and nasal but sometimes it is too shrill and ear-splitting. In yet other cases, the voice drones on in a sing-song and does not even go down at the end of a sentence." (Asperger, 1944 as translated in Frith, 1991: 70)

Some studies report that about half of their participants with ASD exhibited prosodic deficits, while the other half did not (see Paul et al., 2005; Shriberg et al., 2001[8]). Moreover, McCann et al. (2007) and Peppé et al. (2007) found evidence of deficits for each participant with ASD in at least one area of prosody functioning, but not all participants had deficits in the same areas. This suggests individual differences regarding the areas of prosodic impairment as well.
While individual differences observed within studies provide substantial evidence for the heterogeneity of prosodic deficits in individuals with ASD, conflicting or contradictory findings across studies can mostly be ascribed to methodological differences. For example, while some studies that investigated autistic speech have relied on subjective appropriateness ratings, others have applied more objective measures to describe the prosody of individuals with ASD. Also, speech elicitation techniques range from interview corpora with spontaneous speech samples, over speech imitation tasks, to read speech, which makes it difficult to compare the findings. Furthermore, as mentioned before, studies investigating children with ASD and studies investigating adults with ASD have often been compared or generalised over, which can also lead to the misconception that studies provided "conflicting" results.
Nevertheless, individual differences that were found *within* studies indicate that individual differences can be expected when investigating prosodic aspects of speakers with ASD.

[8]It should be noted that Shriberg et al. (2001) investigated individuals in a large age range that included children as well as adults (10-50 years), which might, to some extent, account for the differences.

Possible explanations as to why individuals with ASD exhibit prosodic impairments in speech production can be derived from the same psychological theories that have been introduced above (section 4.2.1) as possible explanations for impairments in (speech) *perception* in ASD (Groen, Zwiers, van der Gaag, and Buitelaar, 2008).
For example, mentalising deficits, which account for problems understanding certain aspects expressed through prosody, can also provide a possible explanation for the fact that speakers with ASD exhibit deficits marking information structure appropriately for their listeners. In fact, all listener-driven aspects of prosody can be ascribed to mentalising deficits. Accordingly, researchers have started to utilise the assumption that individuals with ASD have mentalising deficits by choosing them as subjects in order to identify listener-driven aspects of speech (see, e.g., Lake, Humphreys, and Cardy, 2011).

The WCC theory (Frith, 1989; Happé and Frith, 2006) does not only account for difficulties in perceiving context-dependent prosody, but can also explain difficulties producing appropriate prosody in context. For high-functioning individuals with ASD, it is often not a problem to produce formally and grammatically correct speech, but to use language appropriately in context. This includes a mismatch between the words spoken, the context, and the prosody used by a speaker (Diehl et al., 2008). Such prosodic impairments in the pragmatic domain can be one of the main barriers to social acceptance (Peppé and McCann, 2003; Shriberg et al., 2001), as they apply to every social situation.

The following sections will provide an overview over prosodic impairments in the speech of children and adults with ASD.

Production of Prosody in Children with ASD

The speech prosody of autistic children has been described as "improperly modulated, dull, and wooden and as having a singsong quality" (Baltaxe and Simmons, 1985: 104) or as monotonic, machine-like, exaggerated, or sing-song (Rapin, 1991).
More recently, exaggerated pitch range has also been measured acoustically in children with ASD in comparison to matched controls in a number of studies (Sharda et al., 2010; Bonneh, Levanon, Dean-Pardo, Lossos, and Adini, 2011; Nadig and Shaw, 2012; Diehl, Watson, Bennetto, McDonough, and Gunlogson,

2009; Hubbard and Trauner, 2007). Diehl et al. (2009) found differences between the ASD group and the control group regarding F_0 variation in a story-telling task. Participants with ASD had more within-subject F_0 variation in their narratives than controls.
Hubbard and Trauner (2007) measured increased pitch range in speakers with ASD when they were instructed to imitate emotional prosody. Nadig and Shaw (2012) additionally examined how listeners perceived the acoustic differences between the groups regarding pitch range and asked listeners to give their overall impression of each speaker's prosody on a scale from one to four, where four indicated "normal" and one indicated "atypical". Some (but not all) children with ASD were perceived as producing atypical prosody.

Green and Tobin (2009) identified three different subgroups of speakers in their ASD group: Those with narrow, wide or typical pitch ranges, reflecting individual differences within speakers with ASD. Baltaxe (1984) also found that children in their ASD group had either very narrow or very wide pitch ranges. Bonneh et al. (2011) reported that their ASD group was much more variable with respect to individual mean pitch and pitch range than the control group, documenting increased variability and individual differences concerning pitch in speakers with ASD.

Usually, greater pitch range would be expected in more involved or more emotional speech, which is not what autistic speech is perceived to be. A possible explanation as to why increased pitch range has nevertheless been measured acoustically in these studies is that pitch range is increased in an atypical manner or at unusual positions: More extreme pitch range or pitch variation might be produced but placed arbitrarily in a phrase, thus being non-meaningful, unnatural, or odd (see Nadig and Shaw, 2012). Green and Tobin (2009) also mention that the use of prosody in ASD is sometimes described as "disorganised".
Hubbard and Trauner (2007) measured pitch range in an imitation task in which children with ASD had to repeat sentences spoken by a model speaker with emotional prosody (happy, sad or angry). They found that repetitions of the children with ASD were measured to have a larger pitch range and an "anomalous location of maximum pitch" (Hubbard and Trauner, 2007: 171). Nevertheless, subjective ratings of these repetitions showed that some of the children of the ASD group were as successful at encoding the target emotion as the control group was, indicating their general ability to imitate prosodic cues signalling emotions. The task did not require processing of emotions in order to express them, but could

be completed by simply mimicking what has been said.
In a second task, the participants were instructed to complete stories that were intended to evoke a clear emotional response (happy, sad or angry). In this free-response task, the performance of the speakers with ASD that had previously been successfully mimicking emotional prosody was much less reliable. The subjects from the ASD group occasionally used explicit statements about their intended emotions instead (e.g. "This makes me really angry" or "I'm so happy"), suggesting that they might have been aware of their difficulties conveying emotions non-verbally.
Peppé et al. (2007) investigated the ability of children with ASD to produce affective prosody indicating whether they liked or disliked a food item. The liking responses of children with ASD tended to be judged as disliking, and vice versa, and their responses were judged as significantly more ambiguous than those of the control group. This provides further evidence for the assumption that aspects of affective prosody production are impaired in children with ASD.
This assumption is further supported by reports from Ricks (1975) and Lord (1996) as cited in Shriberg et al. (2001). They observed that it is more difficult for parents of children with ASD to identify the emotional content of prespeech vocalisations than it is for parents of children with mental retardation or normal development (Shriberg et al., 2001: 1099).
Thus, research clearly shows that children with ASD exhibit difficulties producing affective prosody and that their prosody in general has often been found to be disorganised.

The picture is not so clear with regard to the ability of children with ASD to place pitch accents appropriately in different information-structural conditions. Performance of children with ASD seems to depend on a) their age and level of language development and b) on the aspect of information structure under investigation.
Baltaxe and Guthrie (1987) investigated whether children with ASD (3-12 years) are able to appropriately place nuclear accents in broad focus conditions (nuclear accent on the last lexical item of a sentence, see, e.g., *the nuclear stress rule* (Chomsky and Halle, 1968)). Nuclear accent placement in sentences with broad focus was found to be defective in children with ASD, but the control group also exhibited high levels of errors, suggesting that the assignment of nuclear accents in broad focus conditions had not yet been fully acquired by all children. Therefore, group differences in this study can be attributed to developmental differences

(see section 4.2 for a discussion of this bias).

Fine, Bartolucci, Ginsberg, and Szatmari (1991) investigated nuclear accent placement in participants that were older (7-32 years) than those tested by Baltaxe and Guthrie (1987). They analysed accent placement in samples of spontaneous speech from interviews. This time, the groups were similar regarding accent placement in broad focus conditions. This suggests that unmarked accentuation patterns like those used for broad focus structures were not yet fully developed in younger children that participated in the study by Baltaxe and Guthrie (1987), but were found to be intact in ASD at a later age. Interestingly, Fine et al. (1991) found differences between groups when accentuation was used in narrower contexts or conveying contrastive meaning. Furthermore, participants with ASD tended to place accents on function words instead of content words significantly more often than the control group, thus employing accentuation that is not communicatively useful.

Shriberg et al. (2001) also report that speakers with ASD showed inappropriate pitch accent placement within sentences, although it is unclear which focus structures were affected or what kind of words erroneously received pitch accents in which contexts. Consequently, in this study, it is not discernible whether inappropriate accent placement was mapped onto the discourse or referential context (see also Nadig and Shaw, 2015).

Baltaxe (1984) investigated contrastive pitch accent placement in children with ASD in structured contexts that elicited contrastive focus and found that the ASD group had a significantly lower number of correct accent placements on contrastive elements than the control group. Moreover, subjects with ASD in this study misassigned accents to function words or assigned accents to more than one element, which is inaccurate when marking contrast. Findings of Paul et al. (2005) further confirm impairments in the prosodic marking of contrast in children and adolescents with ASD.

Peppé et al. (2007) also found that prosodic marking of contrast was impaired in children with ASD (aged 6-13). The task was to correct a commentator during a simulated football game with animals. When the commentator said a sentence like "The red cow's got the ball", the ASD group showed a greater tendency to place the contrastive accent on the colour adjective ("No, the **red** sheep's got it"), even though the contrast should have been between the cow and the sheep. The correct response would have a contrastive pitch accent on the noun but the participants with ASD tended to emphasise the redundant adjective. Furthermore, responses of the ASD group were significantly more ambiguous than those

of the control group.
However, Diehl and Paul (2013) used the same task and found no significant differences between groups for older participants (8-16 years). They also conducted acoustic analyses of the responses, but these were computed over entire utterances (e.g., "No, the red sheep's got it") instead of over parts that would provide information about relative acoustic emphasis on constituents.
Two studies that provide detailed acoustic analyses of contrastive prosody in comparable tasks report conflicting results: Van Santen, Prud'hommeaux, Black, and Mitchell (2010) used a modified version of the task used by Peppé et al. (2007) and Diehl and Paul (2013) on young children (4-8 years) and measured peaks in F_0 patterns across prosodic minimal pairs (e.g., "**red** sheep" versus "red **sheep**"), as well as duration and amplitude of target syllables. They found that the groups marked the minimal pairs differently with regard to their variation of duration (but no differences were found for variation of amplitude or pitch). When all acoustic measures were used in a linear discriminant analysis, the groups could be reliably differentiated. Furthermore, correct contrastive accentuation was higher in the control group than in the ASD group. Nadig and Shaw (2015) on the other hand found appropriate contrastive prosody on the correct constituents in older children with ASD aged 8-14 years and no significant group differences.

Thus, the available evidence shows that unmarked accentuation patterns that are less context-dependent such as those used for broad focus structures are only problematic for very young children but not for older children with ASD. However, accentuation that is dependent on contextual information, including contexts evoking contrastive meaning, has been found to be impaired in a number of studies. Acoustic measures did not always confirm an impairment, but the way in which contrastive prosody was measured was not documented. In general, younger children with ASD were found to exhibit more severe impairments in prosodic marking of contrast than older children.

Apart from prosodic marking of contrast, prosodic marking of other aspects of information structure (e.g. givenness) have not been so frequently investigated in ASD. For children with ASD, only one study by McCaleb and Prizant (1985) investigated the prosodic marking of referring expressions encoding new and old information in children with ASD. The study found that children with ASD were able to use accentuation to signal new information. However, an interesting observation was that "[...] the subjects encoded old information almost as frequently

as they encoded new information" (McCaleb and Prizant, 1985: 237). It should be noted that this finding included referential echolalic utterances that were counted as old information. Echolalia is a frequent characteristic of verbal children with ASD, but it is unclear in how far these automated repetitions are informative of prosodic encoding of old information. Nonetheless, not all of the utterances categorised as old information were echolalic. Therefore, although this study had some limitations (no control group, echolalia), it is still regarded as noteworthy because it constitutes pioneering work in the area of information structure research in autism.

Baltaxe (1977) also reported that adolescents with ASD exhibit difficulties in "foregrounding and backgrounding of information" and in differentiating between given and new information, but her analysis focussed exclusively on syntactic marking of information (e.g. anaphoric pronouns for given information etc.). Baltaxe (1977) gives the following example (23) (underline added by the author of this thesis) from an interview as evidence for the impairment (the speaker that answers is the one with ASD):

(23) Q: Have you ever seen a lovely lady that you thought you might like to have as your wife?
A: No, I haven't seen a lovely lady like that but I am going to keep meeting lots of nice looking nice lovely ladies close to my age hopefully.

In the answer, the NP "lovely lady" is repeated twice, where anaphoric pronouns would have been more appropriate, since the NP has already been mentioned in the question and therefore constitutes given information. This behaviour adds redundant and irrelevant information to the discourse, which creates the impression of oddness. Even though Baltaxe (1977) did not investigate the prosodic marking of referential expressions, her results indicate that adolescents have difficulties expressing givenness appropriately.

The fact that pitch accent placement was found to be impaired in specific contexts but intact in structures that were context-independent raises the question whether the production of lexical stress is also unimpaired in speakers with ASD. Paul, Bianchi, Augustyn, Klin, and Volkmar (2008) investigated the ability of speakers with ASD (aged 7-28) to imitate stress patterns on nonsense syllables and found subtle but significant differences between the autistic and control groups. For example, the difference in syllable duration between stressed and unstressed syllables was significantly larger in the control group than in the ASD

group.
Paul et al. (2005) report that adolescents with ASD (aged 14-21) showed deficits in producing lexical stress to disambiguate nouns from verbs (**re**call vs. re**call**) in a reading task. On the other hand, according to Grossman et al. (2010), children with ASD (aged 7-18) were able to produce lexical stress to disambiguate compound nouns from adjective-noun sequences (e.g. "**green**house" vs. "green **house**"). However, participants with ASD often produced exaggerated pauses between the relevant syllables in this study. Shriberg et al. (2001), in turn, found no group differences regarding lexical stress in participants between 10 and 50 years of age.
The different methodologies (reading task versus spontaneous speech, acoustic measures versus subjective judgements, groups with large age ranges) make it difficult to carve out whether the ability to produce lexical stress correctly is impaired in children with ASD.

Phrasing and the occurrence of disfluencies have been investigated in speech of children with ASD by a number of studies. Fine et al. (1991) found that children with ASD were able to use boundary tones adequately to indicate ends of utterances in spontaneous speech[9].
In read speech, children with ASD tended to produce boundary tones more frequently than the comparison groups, placing a boundary tone after almost every pitch accent (Fosnot and Jun, 1999). Thurber and Tager-Flusberg (1993) found a reduced frequency of nongrammatical pauses in children with ASD in a story telling task. Nongrammatical pauses were defined as those silences that occur within groups of grammatically related words (e.g. within NPs) as opposed to pauses that occur between phrases. Nongrammatical pauses are assumed to reflect aspects of cognitive demand such as lexical decisions, and a lack of nongrammatical pauses in the ASD group was interpreted as a lack of communicative investment when telling stories. A more recent study demonstrated that disfluencies can mark the introduction of new information to the discourse and can, therefore, be regarded as an important signal to listeners (Arnold, Fagnano, and Tanenhaus, 2003).
Heeman et al. (2010) found longer silent pauses for children with ASD than for their typically developing peers, especially when they were expected to take the turn (e.g. after a question). Feldstein, Konstantareas, Oxman, and Webster

[9]However, the fact that in this study, the comparison group comprised individuals with psychiatric diagnoses and social problems might not have been ideal.

(1982) also found pauses to be asynchronous in conversations between children with ASD and their parents or the experimenter. The children with ASD did not achieve temporal synchrony in their interactions with others, suggesting that turn-taking was disordered.

Lastly, the ability to use prosody to differentiate between questions and statements has also been investigated in children with ASD. Although it was found that in general, children with ASD were able to use prosody for that purpose, their produced utterances were perceived as sounding odd by listeners (Filipe et al., 2014). Furthermore, Fosnot and Jun (1999) report that four children with ASD aged 7-14 tended to produce (confirmation-seeking) questions with low boundary tones in a reading task, thus not prosodically distinguishing between questions and statements. Peppé et al. (2007) found that young children with ASD sounded as if they were asking questions when statements were required and their responses were often judged as ambiguous. On the other hand, Paul et al. (2005) and Diehl and Paul (2013) found no differences between groups in older participants.

In sum, several aspects of prosody have been shown to be impaired in children with ASD. A lot of the studies mentioned above conducted global measures of F_0 and other acoustic properties or relied on subjective ratings of appropriateness. Both global measures and subjective ratings might run the risk of not entirely capturing prosodic deficits and possible systematic impairments. In the last decade, research has begun to combine phonetic measures at targeted relevant parts of utterances with a phonological description of prosody, for instance by applying the AM and ToBI frameworks. For example, Green and Tobin (2009) analysed pitch contours in read and spontaneous speech of Hebrew-speaking children with ASD and found a limited repertoire of edge tone patterns as well as a more frequent use of a "monotonous accent" that they described as a high accent (H* according to IH-ToBI[10]). Conversely, the control group was described as showing a greater number and larger degree of variation in pitch, "creating a more diverse and flexible sounding prosody" (Green and Tobin, 2009: 314). Repetitive use of a limited range of pitch accents in children and adolescents with ASD has also been reported by Edelson, Grossman, and Tager-Flusberg (2007). They report that the ASD group favoured a "downward pitch slope" while the control group favoured "rising-falling" and more "complex" slopes than the ASD group, con-

[10]Israeli Hebrew Tones and Break Indices (Green and Tobin, 2009.

firming constrained prosodic patterns in speakers with ASD.
Accordingly, Demouy et al. (2011) found that children with ASD had difficulties imitating rising, but not falling or "floating" pitch contours. These approaches bear the potential to provide systematic and comparable descriptions of prosodic characteristics in ASD and can lead to a better understanding of the prosodic behaviour of speakers with ASD (see Green and Tobin, 2009). For this reason, this thesis will follow the same path and apply the AM framework to analyse prosody of speakers with ASD.

All studies that were mentioned in this section are summarised in table 4.3 below.

Table 4.3: Overview: Prosody production in children and adolescents with ASD.

	Function	***Details***	***Impairments found***	***No impairments found***
FORMAL / *Grammatical*	*Lexical Stress*		Grossman et al. (2010) Paul et al. (2005; 2008)	Shriberg et al. (2001)
FORMAL / *Grammatical*	*Syntactic Structure*		Fosnot and Jun (1999) Thurber and Tager-Flusberg (1993)	Diehl and Paul (2013) Fine et al. (1991) Peppé et al. (2007)
FORMAL / *Pragmatic*	*Speech Acts*	Younger > older participants	Filipe et al. (2014) Fosnot and Jun (1999) Peppé et al. (2007)	Diehl and Paul (2013) Paul et al. (2005)
FORMAL / *Pragmatic*	*Information Structure*	Broad Focus	Baltaxe and Guthrie (1987) Shriberg et al. (2001)	Fine et al. (1991)
FORMAL / *Pragmatic*	*Information Structure*	Contrast	Baltaxe (1984) Fine et al. (1991) Paul et al. (2005) Peppé et al. (2007) Van Santen et al. (2010)	Diehl and Paul (2013) Nadig and Shaw (2015)
FORMAL / *Pragmatic*	*Information Structure*	Givenness	McCaleb and Prizant (1985)	
INTUITIVE / *Affective*	*Emotional States*	Original > Mimicked	Lord (1996); Ricks (1975) Peppé et al. (2007)	Hubbard and Trauner (2007)
INTUITIVE / *Affective*	*General Prosodic Properties*	Pitch Contours	Demouy et al. (2011) Edelson et al. (2007) Green and Tobin (2009)	
INTUITIVE / *Affective*	*General Prosodic Properties*	Pitch Range	Baltaxe (1984) Bonneh et al. (2011) Diehl et al. (2009) Green and Tobin (2009) Hubbard and Trauner (2007) Nadig and Shaw (2012) Sharda et al. (2010)	

Production of Prosody in Adults with ASD

A majority of the early studies investigating production of prosodic aspects in ASD focussed on children, since autism was assumed to mainly affect language development and to manifest in a delay of aspects of language acquisition. Furthermore, a delay in language development as well as distinctive vocal characteristics have been considered one of the earliest appearing markers of a possible ASD diagnosis (Fusaroli, Lambrechts, Bang, Bowler, and Gaigg, 2017), which makes their investigation in early childhood particularly interesting. Moreover, just as for perception, production deficits in adults with ASD might sometimes be concealed through compensational strategies in order to manage and cope with the challenges of social interactions.

One of the first studies that investigated older subjects with ASD found that some of their participants between 14 and 21 years still had prosodic deficits including abnormalities in pitch, rhythmic elements and stress despite high IQ and verbal IQ measures (Simmons and Baltaxe, 1975). This was taken as evidence that prosodic impairments in ASD can persist into adulthood, which initiated research investigating prosodic abilities in adults with ASD.
There are still much fewer studies investigating production of prosodic aspects in adults with ASD than there are investigating children with ASD. However, some of the papers that have already been mentioned in section 4.2.2 had broad age ranges and their findings are therefore also relevant for the group of adults with ASD, which is why these studies will be briefly mentioned in this section again.

First of all, pitch range in general has been found to be atypical in speakers with ASD compared to the pitch range of control speakers (Diehl et al., 2009; Hubbard and Trauner, 2007; DePape, Chen, Hall, and Trainor, 2012; Kaland et al., 2013). For example, Kaland et al. (2013) showed that naive listeners perceived the speech of speakers with ASD as less dynamic (= more monotonous) than the speech of a control group. There were also differences between speakers within the groups concerning how dynamic their speech was perceived, which confirms individual differences in pitch range characteristics of speakers with ASD.
Some individuals have been found to exhibit either a larger pitch range or a smaller pitch range than control speakers (DePape et al., 2012), which is consistent with findings of studies that investigated children with ASD (Baltaxe, 1984; Bonneh et al., 2011; Green and Tobin, 2009). Furthermore, larger pitch range has also been found in some adults with ASD who speak Cantonese, a tone language

(Chan and To, 2016), confirming both the universality of atypical prosody across languages as well as individual differences among speakers with ASD.

Edelson et al. (2007) investigated pitch contours of speech from individuals with ASD (aged 8-19 years) who retold short stories. They report significant group differences for pitch contours, "with the ASD group favoring a downward slope, and the TD [(= typically developed)] group favouring a rising-falling slope" (Edelson et al., 2007: 44). The TD group also used significantly more "complex" slopes involving both rises and falls while the ASD group produced more "flat" utterances.

Affective aspects of prosody have also been found to be atypical in adults with ASD. Hubbard, Faso, Assmann, and Sasson (2017) found that when asked to produce emotional prosody (e.g. angry, happy, sad), adults with ASD exaggerated their productions, using increased pitch range, longer durations and greater intensity. When listeners were asked to identify the intended emotions, the emotions from sentences produced by individuals with ASD were more accurately identified than those of control speakers, probably due to their exaggeration. However, at the same time, they were also rated as sounding less natural.

Fine et al. (1991) and Shriberg et al. (2001) investigated individuals aged between 7 and 50 years regarding general accentuation patterns in conversational speech. Both studies report inappropriate accent placement. Fine et al. (1991) found that while accent placement was unimpaired in broad focus conditions, it was impaired in narrower contexts or conveying contrastive meaning. Shriberg et al. (2001) did not give information about the contexts in which accent placement was inappropriate.
The broad age ranges make the findings hard to interpret with respect to the developmental trajectory (see Nadig and Shaw, 2015). Furthermore, the results are based on subjective ratings of whether a word sounded highlighted or not. Nevertheless, more context-dependent prosody seems to be more problematic than prosody in broad focus structures (as has been reported for children with ASD as well, see section 4.2.2).

DePape et al. (2012) investigated the ability of adults with ASD to encode new information in narrow focus prosodically in question-answer pairs using the "Under the Shape" game (Chen, 2011b). In this game, two referents are presented simultaneously but one of them is hidden under a rectangle. While the picture with the rectangle over the referent is presented, the experimenter says: "Look! A

bed. It looks like someone is painting the bed. **Who** is painting the bed?" Then the shape disappears to reveal the second referent, for example a rabbit with a paintbrush in his hand. The participant is prompted to answer the question with a full sentence like "The rabbit is painting the bed." The experimenter can ask "who" or "what" questions, depending on whether the narrow focus is supposed to occur in the sentence-initial position as a subject or in the sentence-final position as an object.

DePape et al. (2012) found individual differences within the ASD group, in a way that those with higher language abilities as measured by a vocabulary test showed a different behaviour from those with moderate language abilities.
The subgroup with ASD and higher language abilities failed to vary pitch significantly to differentiate between focussed and unfocussed words at any position in the sentence, despite their general use of large pitch variation (see DePape et al., 2012: 10).
The other subgroup of participants with ASD and moderate language abilities failed to differentiate prosodically between focussed and unfocussed words in sentence-final positions, where the control group showed significantly larger pitch falls for words in focus than for unfocussed words but the participants with ASD did not. Apart from that, this ASD subgroup appropriately marked information structure prosodically, but to a lesser extent than the control group, exhibiting smaller pitch excursions. This could be perceived as monotonous speech and be interpreted by listeners as a lack of interest or as being uninvolved in conversation. The subgroup with higher language abilities did use more prosodic variation than controls but this variation was not useful to listeners.
Both behaviours are likely to impede the ability of people with ASD to communicate effectively and successfully.

Kaland et al. (2013) compared the production of contrastive prosody in adults with ASD and typically-developing adults and the extent to which they account for their listeners in a referential communication task in Dutch. The participants had to give instructions to listeners (e.g. "Put the blue triangle on the banana") and the movable objects could be distinguished by either their color or their shape in order to elicit contrast on either the adjective or the noun, respectively. Two successive instructions were either uttered to the same listener or to different listeners, in order to be able to compare situations in which the speaker and listener have either the same or different perspectives on the information. The noun phrases were analysed in terms of F_0 differences between the focussed

and the unfocussed words. It was found that the control group produced larger differences between focussed and unfocused words than the ASD group, suggesting that contrastive accentuation was more clearly realised by the control group than by the ASD group. However, there was no effect of listener (same/different) on the prosodic marking of the referents in either group, suggesting that neither group adjusted their prosodic marking when addressing a different listener, at least with regard to F_0.
However, in the second part of the study, the collected NPs (e.g. "the blue triangle") were presented to three intonation experts to obtain prominence (strength of the accent) ratings. In general, contrastive prosody of both the control group and the ASD group was perceived as less prominent when a different listener was addressed, showing that both groups produced contrastive prosody more clearly when addressing the same listener, thus taking the perspective of the listener into account. Even though this effect was not measured acoustically in the data, listeners were able to perceive these prominence differences in both groups. This could be explained by the fact that the only acoustic measure taken was F_0, while duration and intensity also contribute to the perceived prominence of syllables.
Interestingly, in the ASD group, the contrastive word was not always perceived as the most prominent word: In cases where speakers with ASD addressed a different listener and the noun was supposed to receive the contrastive accent, the adjective was often perceived as more prominent than the noun, which could reflect perspective-taking difficulties in speakers with ASD.

In a battery of tests, Hesling et al. (2010) investigated the ability of eight male adults with ASD to use prosody serving grammatical, pragmatic, as well as affective functions. The same study also investigated prosody perception and has therefore already been mentioned in section 4.2.1.
The tests involved producing prosody to indicate 1. sentence form, 2. phrasing, 3. contrast, and 4. emotional state in French. In the sentence form task, participants had to produce one-word sentences (food items) as questions (offering the food item to someone) and as statements (naming the food item). In the phrasing task, participants were asked to produce structures such as "twenty-four, twelve" versus "twenty, four, twelve". In the contrast task, participants had to indicate which item "had been forgotten" in sentences such as "I wanted **bread** and apples". Finally, in the emotional state task, participants were asked to name food items and produce the emotion symbolised by a smiley (happy smiley = like the food; sad smiley = dislike the food).

Individuals from the ASD group had significantly lower scores in all tasks. However, as already mentioned in section 4.2.1, this preliminary study had limitations regarding differences in measures of verbal IQ between the groups (mean verbal IQ in ASD group: 89 versus in control group: 128.33). This presumably accounts for (some of) the group differences in the tasks.

All studies that were mentioned in this section are summarised in table 4.4 below.

Table 4.4: Overview: Prosody production in adults with ASD. Asterisks indicate studies that mainly investigated adolescents, but with an age range reaching into adulthood.

	Function	***Details***	***Impairments found***	***No impairments found***
FORMAL / *Grammatical*	*Lexical Stress*		(Grossman et al. (2010))* (Paul et al. (2005; 2008))*	(Shriberg et al. (2001))*
	Syntactic Structure		Hesling et al. (2010)	
	Speech Acts		Hesling et al. (2010)	
Pragmatic	*Information Structure*	Focus	DePape et al. (2012) (Shriberg et al. (2001))*	(Fine et al. (1991))*
		Contrast	DePape et al. (2012) (Fine et al. (1991))* Hesling et al. (2010) Kaland et al. (2013) (Paul et al. (2005))*	
INTUITIVE / *Affective*	*Emotional States*		Hesling et al. (2010) Hubbard et al. (2017)	
	General Prosodic Properties	Pitch Contours	Edelson et al. (2007)	
		Pitch Range	Chan and To (2016) DePape et al. (2012) (Diehl et al. (2009))* (Hubbard and Trauner (2007))*	

4.2.3 Summary

Individuals with ASD, whether children or adults, have been shown to present prosodic impairments in several aspects of perception and production.
In the domain of perception, impaired prosody has been found for affective, pragmatic, as well as grammatical aspects of prosody. However, more studies confirmed impairments in affective and pragmatic aspects, while grammatical aspects have been more often found to be intact (see table 4.1 and table 4.2).
When comparing the findings of the different studies it becomes apparent that task design, heterogeneity of participants, and type of prosodic analysis significantly contribute to the outcome of results. Furthermore, when children were investigated, their age was an important factor that influenced their performance, which was not the case when adults were investigated, where language development is complete. However, for adults, verbal IQ has been shown to play an important role.

Impaired processing of emotions conveyed by prosody has been identified for ASD in several studies, with more severe impairments in children than in adults. For example, adults exhibited no difficulties recognising basic emotions, but showed impairments when complex paradigms and complex stimuli were used.
Studies investigating the perception of pragmatic prosody have mainly focussed on the perception of contrastive prosody, which has been invariably shown to be impaired in children with ASD. The picture is less clear for adults with ASD, who were shown to be able to detect and interpret contrastive accentuation in an explicit multiple choice task (Globerson et al., 2015) but not in two other studies (Hesling et al., 2010; Paul et al., 2005). Two studies that investigated aspects of focus perception in individuals with "autistic traits" found atypical priming effects of prosody in this population (Bishop, 2012; Hurley and Bishop, 2016), confirming that autistic traits can impede the perception of prosody serving pragmatic functions such as the marking of information structure.

When prosody was used to signal phrasing or sentence form, impairments have been found in both children and adults with ASD for implicit and more complex tasks, while the more explicit and simple tasks revealed less difficulties. Furthermore, studies that investigated younger participants more often found impairments in the perception of prosody signalling sentence form than studies that investigated older participants.

The perception of lexical stress has been found to be unimpaired in children with ASD. To date, only one paper is known that investigated the perception of lexical stress in adults with ASD (Kargas et al., 2016). They found impairments in those participants with ASD that also exhibited severe speech production abnormalities, but not in the remaining participants with ASD, which suggests that the perception of lexical stress is also unimpaired in high-functioning adults with ASD and unimpaired verbal abilities.

Lastly, the perception of lexical tones has been shown to be unimpaired in adults with ASD (Cheng et al., 2017). These findings provide further evidence for the assumption that in ASD, the perception of grammatical aspects of prosody are less impaired than pragmatic and affective aspects.

To date, no studies are known that investigated the perception of prosody when used to signal givenness.

In the domain of production, impaired prosody has also been found for affective, pragmatic, as well as grammatical aspects of prosody in both children and adults with ASD. Furthermore, general prosodic properties were often described as being atypical, regardless of prosodic function. For example, pitch range has been found to be atypical in both children and adults with ASD. When measured acoustically, pitch range has more often been found to be larger than smaller in speakers with ASD compared to controls, although their speech was subjectively often described as "monotonous". An explanation for this discrepancy could be that acoustic measurements have mostly been executed globally instead of at targeted relevant parts of utterances. Targeted acoustic measurements on the other hand have shown that pitch range is increased at unusual positions, thus sounding disorganised and odd (Nadig and Shaw, 2012; Green and Tobin, 2009). So far, there is no conclusive evidence for an acoustic vocal marker that reliably predicts ASD (Fusaroli et al., 2017).

The production of emotional prosody has been found to be impaired in speakers with ASD. For example, they tended to exaggerate the emotions they were intending to convey or they were unable to reliably distinguish different emotions by means of prosody. Furthermore, they tended to use explicit statements about their intended emotions instead (Hubbard and Trauner, 2007; Hesling et al., 2010).

As in perception, studies investigating the production of pragmatic prosody have

mainly focussed on the prosodic marking of contrast. Impairments have been found in both children and adults with ASD. Studies that used targeted acoustic measures instead of subjective judgements report that while contrastive accentuation was often intact in speakers with ASD, they tended to produce less clear and less reliable contrastive accentuation (Kaland et al., 2013), which could be interpreted by listeners as being less involved in or committed to a conversation. While the appropriate use of contrastive prosody relies on contextual information, broad focus structures mainly involve unmarked accentuation patterns that are (mostly) predictable by grammatical rules. Once these rules were learned, accentuation in broad focus structures has been found to be unimpaired in ASD at a later age, at least when accent *placement* (i.e. the presence or absence of prosodic prominence) was investigated. Only a few studies have applied AM theory and the ToBI framework to achieve a systematic phonological analysis of pitch contour differences, or accent *types*, in the speech of individuals with ASD. Fosnot and Jun (1999) used ToBI for a qualitative analysis of read speech produced by children with ASD. However, their analysis was limited to rising versus falling contours in questions versus statements.
Green and Tobin (2009) have demonstrated that such an analysis can lead to a better understanding of prosodic behaviour of individuals with ASD. For children with ASD, pitch contours in general were found to be limited to a small repertoire of patterns, with less variation and a preference for "monotonous" accents such as H* (Green and Tobin, 2009; Edelson et al., 2007; Demouy et al., 2011).
However, to date, no known studies have looked at pitch accent types in relation to the function they were intended to convey.

Grammatical aspects of prosody have been shown to be impaired in (young) children with ASD. Phrasing, sentence form and lexical stress have been reported to be impaired in adults with ASD, too, but the papers that reported these impairments all have major limitations by either investigating and pooling both children and adults with ASD (Shriberg et al., 2001; Grossman et al., 2010; Paul et al., 2005; Paul et al., 2008) or by exhibiting significant group differences in verbal IQ, a factor that has been shown to influence prosodic ability in individuals with ASD (Hesling et al., 2010). Therefore, more research is needed in order to draw conclusions about possible impairments in grammatical aspects of prosody in adults with ASD.

In general, not all individuals with ASD were found to have similar prosodic impairments. In a majority of the studies discussed above, the performance

of some participants with ASD overlapped with the performance of controls, with individual differences in both groups. Nevertheless, evidence for a cross-language universality of atypical prosody in ASD has been provided by studies that investigated tone languages, but only for prosody production (Chan and To, 2016), not perception (Cheng et al., 2017).

A lot of the studies that were discussed in the sections above reported conflicting results. As mentioned above, this can mostly be attributed to methodological differences. Furthermore, conflicting results are not surprising, given the many functions of prosody and the many factors influencing prosodic characteristics. Moreover, ASD is characterised by a heterogeneity of clinical features and their severity, thus individual differences in prosodic ability is to be expected.
The following two experiments aim to add to the previous findings by investigating the perception and production of prosodic aspects of referential givenness, an understudied aspect of information structure in ASD.

Part II

Empirical Studies: Prosodic Decoding and Encoding of Referential Givenness in Autism

Introduction to Part II

This part of the thesis comprises two experiments that are concerned with the ability of adults with ASD to both decode and encode referential givenness prosodically.
The literature review of studies investigating the perception (section 4.2.1) and production (section 4.2.2) of prosody in individuals with ASD has shown that difficulties are prevalent in a number of areas. These include grammar, pragmatics, and affect. One of the areas that has been observed to be particularly affected involves receptive and expressive difficulties when prosody is used to mark information structure (focus, contrast). However, givenness, a related aspect of information structure, has to date not been investigated in adults with ASD.
Moreover, investigations of prosody of autistic speech within the AM framework are still rare (Green and Tobin, 2009). So far, studies have focussed either on more general subjective descriptions, simple dichotomous analyses (e.g. accent vs. no accent), or appropriateness ratings of naive listeners. In the current study, both the perception and production experiments use the AM framework to analyse prosody with a view to describing, comparing and contrasting prosodic aspects of individuals with ASD and neurotypical individuals in a more structured way and to making the results available to the AM community.

In chapter 5, a perception experiment is presented that focusses on the ability of adults with ASD to decode prosodic cues when making judgements about the referential givenness of referents in discourse. Adults with ASD have been reported to be less sensitive to prosodic marking of a related aspect of information structure, namely contrast (DePape et al., 2012; Hesling et al., 2010; Kaland et al., 2013). Paired with general problems in pragmatic processing and limited attention paid to non-verbal cues, this observation would predict that adults with ASD are less able than healthy controls to use prosodic cues when interpreting the givenness of a referent.

Chapter 6 aims to complement the findings of the perception experiment by investigating the ability of individuals with ASD to encode givenness of referents in a cooperative story-telling task. The task is designed to elicit spontaneous but structured speech. This method retains most of the benefits of natural conversation while still providing controlled, quantifiable and comparable samples of referring expressions and their prosodic marking.

The appropriate prosodic marking of referential givenness requires an assessment of the information status of a referent in the mind of the listener, which is why individuals with ASD, a group that reportedly exhibit impaired mentalising skills, are expected to show patterns of highlighting and attenuation that are different from those of typically developed adults.

Chapter 5

Perception Experiment

5.1 Method

All aspects of this study have been approved by the local ethics committee of the Faculty of Medicine at the University of Cologne and were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.

5.1.1 Participants

There were two groups of participants for this experiment, one experimental group consisting of individuals with ASD, and the other one consisting of typically develped control persons.
The ASD group comprised 37 native German speakers with ASD (23 male, aged 23-54). Participants had all been diagnosed in adulthood with ICD-10: F84.5 (AS) and were recruited from the Autism Outpatient Clinic at the Department of Psychiatry at the University of Cologne (Germany). As part of a systematic assessment implemented in the clinic, diagnoses were made independently by two different specialised clinicians corresponding to ICD-10 criteria and supplemented by an extensive neuropsychological assessment.
The control group comprised 37 typically developed native German speakers (24 male, aged 21-65) who were recruited from the general population. They were selected from a larger pool of participants from a previous perception study (Röhr and Baumann, 2011; Röhr, 2013; Baumann, Röhr, and Grice, 2015) to match the

Table 5.1: Age (in years) and gender information for both groups of participants. Results of a two-sample t-Test for age and a Chi-square test for gender (p-values).

	ASD		**Control**		p-value
Age	mean = 42.5	(±9.5)	mean = 38.4	(±14.4)	0.1552
Gender	m = 23	f = 14	m = 24	f = 13	0.8092

ASD group regarding gender and age. Statistical tests revealed no significant differences between groups regarding gender (Pearson Chi-square test: $p = 0.8092$) and age (two-sample t-Test: $p = 0.1552$). The demographics of both groups are summarised in table 5.1 and information about individual subjects can be found in Appendix A.1.
All subjects reported normal hearing abilities with no known hearing impairments. Participants were excluded if they did not complete the whole online survey that the perception experiment was based on, if they had rated all stimuli with an identical rating, or if their mother tongue was not German.

On the basis of current occupation and professional qualifications, it was estimated that 76% of participants in the ASD group and 65% in the control group had the German "Abitur" (the general qualification for university entrance). All remaining participants in the ASD group and all but four in the control group had successfully finished secondary school. With regard to IQ levels, this indicates that all participants in the ASD group and at least 33 (out of 37) subjects in the control group had an average or above average intellectual capacity.

5.1.2 Materials

This perception study was originally designed for experiments reported on in Röhr and Baumann (2011), Röhr (2013), Baumann, Röhr, and Grice (2015) and Röhr (2016). The stimuli for the perception study were generated using selected parts of read speech of typically developed adults from a separate production study (also reported on in Röhr and Baumann, 2011; Röhr, 2013; Baumann, Röhr, and Grice, 2015; Röhr, 2016). Both the stimuli and the perception study design were kindy provided by the authors.

The stimuli consisted of several small vignettes containing ten different tar-

get words that represent discourse referents. The target words were bi- and tri-syllabic feminine common nouns (*Ballade* 'ballad', *Banane* 'banana', *Dame* 'lady', *Lawine* 'avalanche', *Rosine* 'raisin') and proper nouns (*(Dr.) Bahber/Bieber*, *Janina*, *Nina*, *Romana*), each with lexical stress on the penultimate syllable and a comparable segmental structure that consisted of mainly voiced segments. Of the target words, four nouns were inanimate, the remaining six were animate (persons). In total, 11 of the 28 target sentences contained inanimate nouns. The word frequency differed considerably across the target words as estimated using the Subtlex Corpus (Brysbaert et al., 2011, the log word frequency values for each target word are presented in table A.3 in Appendix A.2). Neither word frequency nor animacy were specifically controlled for, as the materials were taken from a production study concentrating on the phonetic/phonological properties of the words. However, the frequency scores and the animacy of each noun were taken into account in the statistical analysis.

Each target word was embedded in different contexts in order to elicit four different types of information status, namely "new", "inferentially accessible", "given-displaced" (given but with two or three sentences of intervening textual material), and "given" (mentioned immediately prior to the target sentence). The focus structure of the target sentences was held constant where possible in order to keep its influence on the prosodic marking of the target words to a minimum. Figure 5.1 provides an example of the reading material that was used[1] to elicit prosodic realisations of different degrees of givenness.

In target sentence (a) of Context 1, the target word "banana" is mentioned for the first time and is not derivable from the previous context sentence. It is therefore still inactive in the minds of speaker and listener (Chafe, 1994) and can be regarded as (discourse-)new. After two to three sentences of intervening textual material with a change in topic, "banana" is repeated in target sentence (b). Here, the target word "banana" is no longer fully activated and is regarded as given-displaced.

Context 2 sets up a scenario (zoo, monkey, feeding), from which the target word "banana" in target sentence (c) is inferentially accessible. That is, the target word has not been explicitly mentioned before but is derivable from the preceding contextual frame.

In sentence (d) of Context 3, the target word "banana" is a repetition of an antecedent in the immediately preceding context sentence. In contrast to sentence

[1] See Röhr (2016) for the reading material of all target words.

(b), this target word is already fully activated and thus regarded as given.

Context 1	**(a) new**	**(b) given-displaced**
"Was hätten Sie gerne?". **(a) "Ich nehme die Banane mit"**, antwortet Thomas dem Obsthändler. Normalerweise ernährt er sich sehr ungesund und isst zwischendurch ständig Süßigkeiten. Außerdem treibt er fast nie Sport und wenn doch, dann am liebsten Minigolf. **(b) Er steckt sich die Banane ein**. Lecker sieht die Banane aus. Vielleicht wird er demnächst öfter welche kaufen. *"What would you like?" **(a) "I'll take the banana with me"**, says Thomas to the fruit merchant. He usually eats very unhealthily and he is always eating sweets between meals. He hardly ever plays sport, and if he does he prefers minigolf. **(b) He pockets the banana**. The banana looks delicious. Maybe he'll buy them more often in future.*		
Context 2	**(c) inferentially accessible**	
Thomas darf heute im Zoo seinen Lieblingsaffen füttern. Voller Vorfreude wird er sich gleich auf den Weg zu ihm machen. **(c) Er steckt sich die Banane ein**. Vorhin war er dafür extra noch auf dem Markt beim Obsthändler. *Today Thomas is allowed to feed his favourite monkey in the zoo. With great anticipation he's about to set off (for the zoo). **(c) He pockets the banana**. He's just been to the green grocer's at the market especially to get one.*		
Context 3	**(d) given**	
Thomas hat gerade auf dem Markt eine Banane gekauft. **(d) Er steckt sich die Banane ein**. In Zukunft möchte er sich viel gesünder ernähren. *Thomas has just bought a banana at the market. **(d) He pockets the banana**. In the future he wants to eat much more healthily.*		

Figure 5.1: Sample reading material for the target word *Banane* with English translation. The target sentences are printed in bold face and the target words are underlined. Arrows indicate relationships between words that alter the information status of the target word. Adapted from Röhr (2016).

Perception Stimuli

For the perception study, seven isolated target sentences (without their corresponding contexts) were selected for each of the four types of information status (new, accessible, given-displaced, given). Specifically, sentences were chosen that reflected the full range of five nuclear pitch accent types (H*, !H*, H+!H*, H+L*,

L*)[2] on the target words, and the remaining two instances were sentences in which the target words received either a low prenuclear accent, or no accent at all (Ø). All target sentences had a prenuclear rising accent on the finite part of the verb (H* or L+H*) and a sentence-final low boundary tone (L-%).
The stimuli comprised sentences spoken by ten speakers in total (seven female and three male). An overview of the distribution of stimuli among the different speakers and the corresponding target words is given in Appendix A.2, table A.4. Stimuli examples of target sentences with speech pressure waveform, F_0 contour, and GToBI annotation can be found in Röhr and Baumann (2011), Baumann, Röhr, and Grice (2015) and Röhr (2016). No adjustments of the original utterances were made, except for an equalisation of the amplitude of the test material.

5.1.3 Task

The perception experiment was conducted as an online questionnaire using the software package "SoSci Survey" from the online questionnaire "onlineFragebogen (oFb)" (Leiner, 2014). The link to the questionnaire was provided by e-mail, so that all participants were able to participate in the experiment at home.
The task was to evaluate how far an item/person referred to by the target word in a test sentence sounded as if it was known to the listener. That is, participants evaluated the degree of givenness of a referent after hearing a word spoken in an utterance. They listened to the utterances and subsequently rated the stimuli by placing a roll bar on a continuous line between two end-points, labelled as 'known' and 'new' on a visual analogue scale (VAS) without apparent discrete scaling. A screenshot of the setup is shown in figure 5.2.

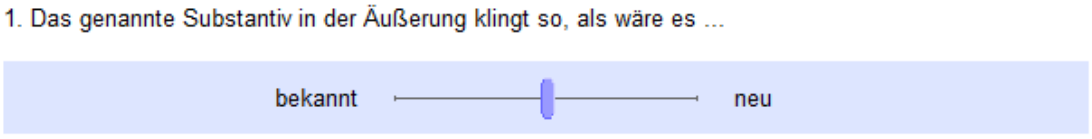


Figure 5.2: Screenshot of the web-based task that was displayed during the presentation of each stimulus. (English translation: The noun spoken in the utterance sounds as if it were... 'known'...'new'), implemented with the "SoSci Survey" software (Leiner, 2014).

[2] L+H* and L*+H accents were not tested since the production data did not provide instances of these accent types for each type of information status. Furthermore, H+!H* and H+L* accents were pooled for the analysis (see Rathcke and Harrington, 2010; Grice et al., 2009).

Test sentences were played twice, separated by a pause of one second and triggered by the participant. Stimuli were not presented visually and could not be played again. The responses were encoded as interval data with the lowest value (1) at the left pole 'known' and the highest value (100) at the right pole 'new'. Participants rated each test sentence three times in randomised order, amounting to a total of 84 stimuli for evaluation in the main part of the experiment (seven target sentences * four types of information status * three repetitions).

A complete experimental sequence consisted of (1) an introduction and description of the task, (2) an anonymous questionnaire to obtain personal data, (3) a practice section with seven prototypical stimuli, (4) the main section[3].

5.2 Results

In order to increase the statistical power of the dataset, the different accent types were initially pooled into two categories: The category of "nuclear accents" included target words with any type of nuclear accent, while the category of "no nuclear accents" included target words that were unaccented or that bore a low prenuclear accent, but not a nuclear one. This analysis was complemented by a descriptive analysis of the perception of different accent types in the two groups.

A linear mixed effects analysis with response value (ranging from 1 to 100) as the dependent measure and group (ASD vs. control), accentuation (nuclear accent vs. no nuclear accent), repetition (1st, 2nd or 3rd repetition), animacy of the words (animate vs. inanimate) and word frequency (log values from the Subtlex corpus (Brysbaert et al. 2011)) as fixed effects was performed. Interactions between group and accentuation, group and word frequency, group and repetition as well as group and animacy were also included in the model. Subject, speaker and target word were included as random intercepts. P-values are reported based on Likelihood Ratio Tests comparing a reduced model (without the effect in question) against the full model (including all effects). The random effects structure was constant across these comparisons.
Visual inspection of residuals revealed some violation of normality and homoscedasticity. These violations can be disregarded for this analysis due to the size of this data set (for a discussion on this issue see Faraway, 2004).

[3]The main section involved seven additional stimuli at the beginning and seven additional stimuli at the end of the main section that did not enter the analysis.

Accentuation

There was a main effect of accentuation indicating that, in general, both groups provided "newer" responses for target words bearing a nuclear accent than for target words that did not. The mixed effect model comparison showed a significant interaction of group and accentuation ($\chi^2(1) = 67.35,\ p < 0.001$), revealing that the influence of accentuation on ratings differed between the two groups. The ratings in the ASD group increased (= were interpreted as newer) from unaccented to accented target words only slightly, by 2.02 (SE = 1.46). In the control group, by contrast, the ratings increased substantially, by 17.07 (SE = 1.83)). Thus, the difference between ratings of accented versus unaccented words was significantly larger in the control group than in the ASD group (see figure 5.3).

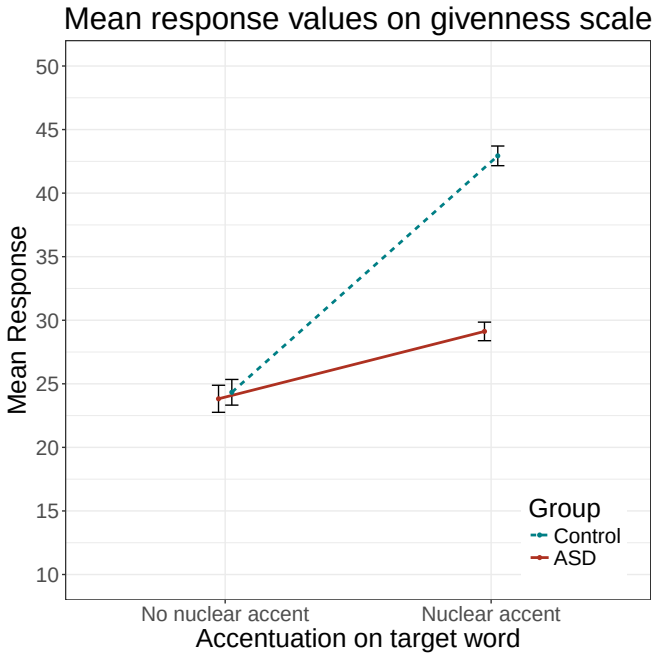


Figure 5.3: Distribution of ratings for target words with no nuclear accentuation and nuclear accentuation according to their mean response values, control and ASD groups.

With regard to the perception of different accent *types*, the two groups exhibited different patterns in their average ratings (see figure 5.4). The control group (figure 5.4, top) showed a trend to rate target words without nuclear accents (that is, target words that were either unaccented or bore a low prenuclear accent) as most given, followed by target words with low/falling nuclear accents, and target words with high/rising nuclear accents were rated as newest. This was not the case in the ASD group (figure 5.4, bottom), where the distribution of accent types on the givenness scale seems arbitrary: While participants from the ASD group

also rated target words without nuclear accents as most given, they rated target words with low L* accents as newest and target words with falling H+!H* accents as rather known, with rising accent types inbetween.
Thus, the ratings of the control group (roughly) reflect the inherent acoustic prominence of different accent types, in a way that the more prominent high/rising accents received the highest (= newest) ratings, while the less prominent falling and low accents received lower ratings and words without nuclear accents received the lowest (= most given) ratings.

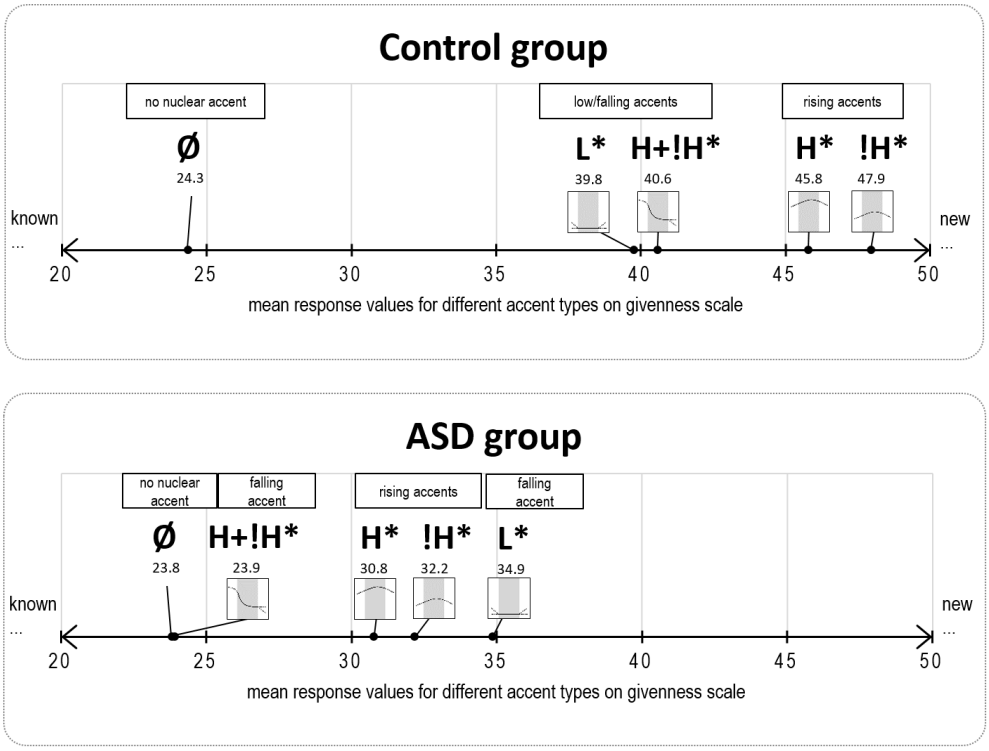


Figure 5.4: Distribution of nuclear accent types and no nuclear accents (Ø) on the givenness scale according to their mean response values for the control group (top) and ASD group (bottom).

Note that the scaling in figures 5.3 and 5.4 shows values up to 50% on the givenness scale which, in fact, provided a scale from 0 to 100. This means that both groups (on average) showed a limited use of the full scale, especially at the high end. A possible explanation for this is the fact that the target words were presented with a definite article (e.g. "the banana"), which signals that the information is not brand new. Singular nouns require an article, and in order to control for the effect of prosody alone, it was necessary to use the same one for all tokens. Since the target words consisted of both proper nouns and common nouns, the definite article was the only article available, since the indefinite article cannot

be used with proper nouns in German.

Interaction Group x Animacy

There was a significant interaction between group and animacy ($\chi^2(1) = 7.86$, $p = 0.005$). The ratings of the ASD group increased substantially, by 12.44 from inanimate to animate (SE = 1.3) while the ratings of the control group increased less strongly, by only 7.7 (SE = 1.69). This indicates that the category of animacy of the target words played a role in the givenness ratings of both groups, but more so in the ASD group than in the control group (see figure 5.5).

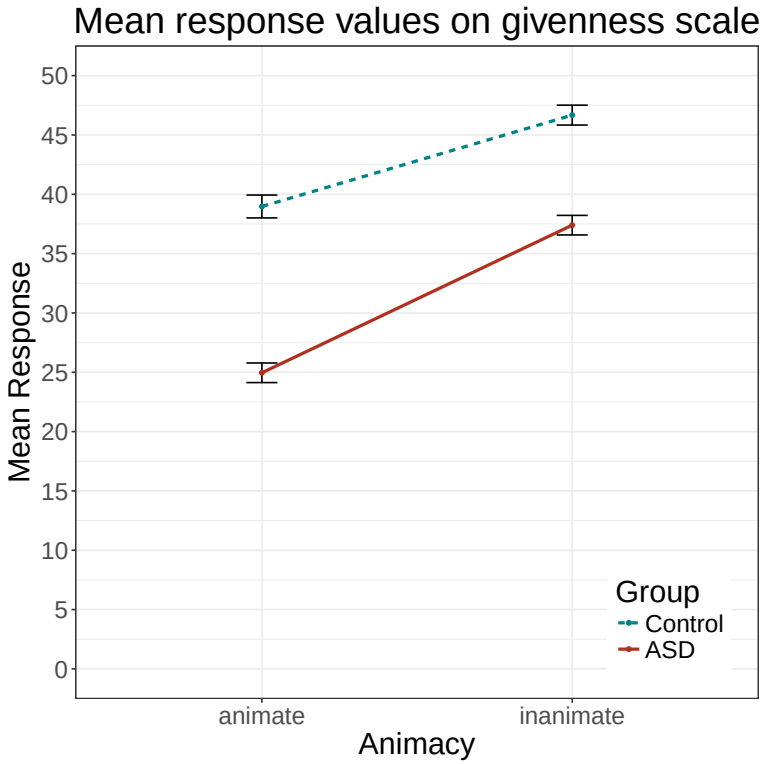


Figure 5.5: Distribution of ratings for animate and inanimate target words according to their mean response values, control and ASD groups.

Interaction Group x Word Frequency

There was a significant interaction between group and log word frequency ($\chi^2(1) = 8.95$, $p = 0.003$). The ratings of the ASD group decreased slightly, by 2.34 per log frequency (SE = 1.71), while the ratings of the control group did not decrease substantially (decrease of 0.09 per log frequency (SE = 0.76)).

Thus, word frequency affects the rating differently across the two groups. Unlike the control group, the ASD group rated words that are frequently used in speech (e.g. "banana") as more given than words that are less frequent (e.g. "avalanche"), irrespective of their prosodic characteristics. The mean ratings of

both groups for each target word are shown in figure 5.6.

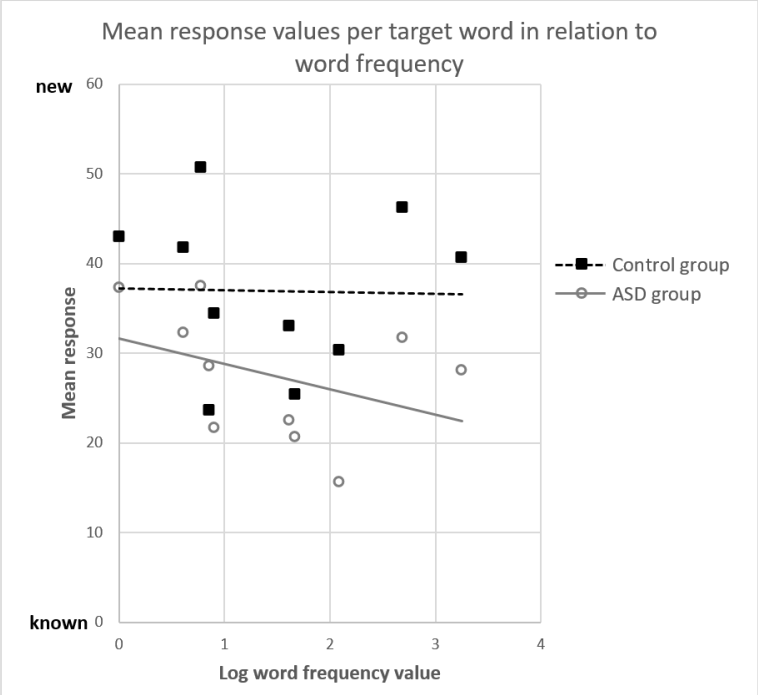


Figure 5.6: Mean ratings of each target word and the corresponding log word frequency according to the Subtlex corpus.

Interaction Group x Repetition

No significant interaction of group and repetition was found ($\chi^2(1) = 2.78$, $p = 0.095$). Repetition as a factor in itself was not significant ($\chi^2(1) = 1.61$, $p = 0.2052$).

The results are summarised in table 5.2, informally reflecting the trends in the statistical analysis.

Table 5.2: Summary of sensitivity to different properties of the target word in the two groups (– lack of sensitivity, + weak sensitivity, ++ strong sensitivity).

	Accentuation *main effect*	**Animacy**	**Word frequency**
Control	++	+	-
ASD	+	++	+
Rating effect	Accented is newer (Control >ASD)	Animate is newer (ASD > Control)	Infrequent is newer (ASD only)

5.3 Discussion

The results of this perception experiment generally confirmed the prediction based on other aspects of information structure that adults with ASD do not take into account prosodic cues to the same extent as control adults when interpreting the givenness of a referent.
Both groups rated words bearing a nuclear accent as sounding newer than words without one. However, the difference in ratings for the ASD group for words with a nuclear accent in comparison to words without one was significantly smaller than in the control group.
Furthermore, adults with ASD did not appear to take into account differences between accent types when rating the information status of a referent: The control group tended to rate target words with more prominent nuclear accents (e.g. high accents) as newest, target words with less prominent nuclear accents (e.g. low accents) as less new (in between new and given) and target words without nuclear accents as most given. By contrast, the ASD group did not show this prominence-driven pattern.
Thus, the prosodic marking of referents had a smaller effect on their perceived degree of givenness in individuals with ASD than in control subjects. This indicates that individuals with ASD paid only limited attention to prosody when making judgements about the givenness of items in discourse as compared to the control group.

Results suggest that instead of attending to prosody, the ASD group potentially relied on two compensation strategies relating to the lexical properties of the stimuli, namely on word frequency and animacy.

First, the ASD group rated more frequent words as known, and less frequent words as less known, regardless of the way the word or sentence was pronounced. Hence, words they will have heard frequently over their lifespan (e.g. “banana”) were rated as more known, while less frequent words and names that they might not have heard as often (e.g. “avalanche”) were rated as less known (newer). In doing this, they prioritised their own generic knowledge of a word and their personal relatedness to it, resulting in higher (i.e. newer) ratings for less familiar words. In the control group, word frequency did not have an influence on the ratings.

Second, results indicate that both groups were to some extent sensitive to ani-

macy, in a way that words referring to inanimate words (e.g. "raisin") tended to be rated as more given than words with an animate referent (e.g. "lady"). Crucially, this effect was stronger in the ASD group than in the control group, corroborating the fact that the lexical properties of words themselves could play a greater role than their prosodic realisation for individuals with ASD.
Animacy is a generic property of a word and its meaning: Words that are inanimate, such as "banana" or "avalanche", refer to objects that we know, while words that are animate, especially proper nouns referring to humans, might imply a relationship with a person and evoke a question like "Do I (personally) know Janina?". It thus seems reasonable to conclude that rating inanimate words as more "known" than animate words reflects one's own knowledge of the referent (the object or the person), rather than providing an estimate of the speaker's intention to convey the information status of this referent.
Hence, instead of taking what could be perceived as the speaker's intention into account, the task is solved by relying on personal knowledge, a behaviour that can be attributed to mentalising deficits.

Furthermore, since prosody, and in particular accentuation, varies in a context dependent way, whereas animacy and frequency of words is context independent, the results of this perception experiment can be taken to support findings in other domains in terms of underlying cognitive mechanisms, corroborating the concept of WCC (Frith, 1989; Happé and Frith, 2006, see also chapter 4). It appears that instead of making sense of information from prosody "intersubjectively", individuals with ASD attempt to recruit intellectual or "objective" compensation mechanisms, using lexical properties of words that are invariable to calculate the degree of givenness of an object or person referred to with this word.

This study confirms an over-reliance on propositional content and personal knowledge in ASD participants which has been shown in other studies dealing with the perception of prosody in both linguistic and affective functions as well, for example when judging emotions (Lindner and Rosén, 2006). Moreover, studies that did not investigate prosody in particular found over-reliances on verbal content in individuals with ASD. For example, Kuzmanovic, Schilbach, Lehnhardt, Bente, and Vogeley (2011) detected a reduced sensitivity to nonverbal cues when ASD subjects had to consider conflicting information from nonverbal and verbal domains at the same time during impression formation.
In general, extreme literalism and a very formal use of language are properties that people with ASD are known for. They are often reported to interpret words

too literally, which poses one of the recurrent problems they face when communicating with others (Dewey and Everard, 1974). Overly literal interpretation of language was also found in autistic children (Mitchell, Russell, and Saltmarsh, 1997). Reported difficulties interpreting non-literal or figurative speech complete this picture, showing a preference of individuals with ASD for understanding and interpreting speech literally. For example, understanding metaphor or irony was found to be difficult for subjects with ASD (Happé, 1993).

The detected insensitivity of the ASD group to prosody cannot be ascribed to hearing deficits or difficulties in being able to detect changes in pitch or the differences between accented and unaccented syllables acoustically. All participants in this study reported normal hearing abilities. Furthermore, as mentioned in section 4.2.1 of chapter 4, studies have shown that people with ASD do not have difficulty in perceiving parameters signalling accentuation such as pitch movement, loudness and intensity. Consequently, participants with ASD are able to hear prosodic cues if the task explicitly requires paying attention to this feature in speech. However, their performance is impaired if stimuli become more complex and/or if the task requires interpretation (O'Connor, 2012). Singh and Harrow (2014) confirm that with ASD subjects, sensitivity to what they refer to as "surface detail", an example being prosody, appears to be highly task-dependent.
In the current study, subjects were not directly asked to attend to prosody, but were asked to interpret it. The problems encountered by the ASD subjects in this study may therefore lie either in the interpretation of the speaker's intentions or in the complexity of the speech material (see also the ED hypothesis discussed in section 4.2.1 of chapter 4) both of which have been shown to lead to impaired performance in perception experiments with this population (O'Connor, 2012; Singh and Harrow, 2014; Tager-Flusberg, 2000).
The deficits in prosody perception observed in the ASD group are thus most likely not caused by an inability to perceive the acoustic cues that convey a certain function, but the result of difficulties in paying attention to the cues and to interpreting them.

An alternative way of processing speech may be an important factor underlying the social and communicative difficulties in autism, and the results of this study support this suggestion. Being aware that individuals with ASD pay less attention to prosodic information is crucial for the comprehensive understanding of possible communicative difficulties in persons with ASD. A lower sensitivity to prosodic cues can lead to misunderstandings on various levels during social com-

munication and interaction. In order to be able to understand the intentions of the speaker and to react appropriately in social contexts, the ability to perceive the level of givenness of a mentioned referent is indispensable.
In addition, situations with ambiguous or conflicting information might not be solvable by listeners who do not take the information status of referents into account. For example, a conversation as in example (24) is not fully comprehendible without taking the prosody of the answer into account (example adapted from Baumann and Roth, 2014):

(24) A: Have you met your cousin?
  a. I met Tamara **yesterday**.
  b. I met **Tamara** yesterday.

Depending on the accentuation in the response, "Tamara" could either be the cousin and the speaker met her yesterday (when "Tamara" is unaccented as in (24a.), or Tamara could be someone else, not the cousin (when "Tamara" bears the nuclear accent, as in (24b.)). If prosody is not taken into account in situations like that, misunderstandings can easily arise.

All of the above indicate problems understanding that communication is about intended rather than surface meaning (Tager-Flusberg, 2000) and that prosody and context play a crucial role in transmitting meaning or information beyond the literal content of what is said.

5.4 Summary and Conclusion

In this perception experiment participants listened to utterances with target words that differed in their degree of givenness. The task was to judge how far the item or person referred to by a target word sounded as if it was known or new to the listener. Instead of attending to the prosody of spoken words, individuals with ASD chose a different strategy in order to complete the task: They based their judgement more on their personal knowledge of the words and less on their pronunciation, as opposed to the control persons, who prioritised prosodic cues.

Results provide further substantiation of both the social theories and the general theories of auditory processing in ASD. The impairments of prosodic processing

that were found in the ASD group most likely originate from a complex interplay between ToM deficits, alternative processing mechanisms (ED) and WCC, along with the development of compensation strategies (see also section 4.2.1 of chapter 4).

The findings point towards an impaired grasp of how different prosodic manifestations of utterances convey different meanings. This, in turn, could also account for atypical expressive prosodic abilites and communication skills, given the link between perception and production and given the effect that receptive prosodic deficits can have on prosodic expression (Peppé et al., 2007).

Lastly, this experiment has methodological implications for future studies on the prosody of information structure with this population of subjects: It was shown that in order to be able to investigate the influence of accentuation on the perceived level of givenness of a word, controlling for the lexical content is crucial when working with autistic subjects, as their tendency to rely on the lexical properties of words can override the prosodic prominence with which they are produced.

Chapter 6

Production Experiment

6.1 Motivation

The perception experiment (see chapter 5) revealed that individuals with ASD made significantly less use of prosody when assessing referential givenness than the control group. While this already provides important insights into difficulties in speech-based communication in ASD, the following production experiment aims to complement this finding by investigating the ability of individuals with ASD to encode referential givenness in a story-telling task.

Reports on adults with ASD exhibiting expressive difficulties when prosody is used to mark information structure have so far been limited to focus (DePape et al., 2012) and contrast (Kaland et al., 2013; DePape et al., 2012; Hesling et al., 2010). Only one study by McCaleb and Prizant (1985) investigated the prosodic marking of new and given referents, but the study investigated children with ASD aged 4 - 14. An interesting observation of that study was that "the subjects encoded old information almost as frequently as they encoded new information" (McCaleb and Prizant, 1985: 237).
This finding, together with reported difficulties regarding pragmatic aspects of language use in general, lead to the prediction that adults with ASD will exhibit impairments in the (prosodic) marking of referential givenness.
Furthermore, reportedly impaired mentalising skills in speakers with ASD potentially also limit the appropriate marking of referents, since this mechanism requires an assessment of the information status of a referent in the mind of the listener.

Lastly, the choice of referring expression (pronouns versus NPs) is also expected to be different in the ASD group as compared to the control group. This prediction is based on findings from three studies that investigated referential choices made by individuals with ASD during narratives. Baltaxe (1977) observed that adults with ASD often used referential expressions that were more specific than needed. Arnold, Bennetto, and Diehl (2009) reported that young children with ASD were significantly less likely to produce pronouns than their typically developing peers in some discourse contexts. Colle, Baron-Cohen, Wheelwright, and Van Der Lely (2008) also reported that speakers with ASD used less pronouns but more NPs than the control group when maintaining reference.

These findings indicate that both the prosodic marking of referents as well as the choice of referring expression might be impaired in speakers with ASD. The following study aims to complement these findings by exploring both prosodic variation and referential choices of adults with ASD when they encode referential givenness in a structured story-telling task.

This cooperative task elicits spontaneous but structured speech. Given the aim to investigate pragmatic aspects of prosody, such an approach is to be favoured over read speech, as read speech does not always reflect the pragmatic goals of a speaker (see, e.g., Speer, Warren, and Schafer, 2011).

6.2 Method

All aspects of this study have been approved by the local ethics committee of the Faculty of Medicine at the University of Cologne and were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments.
All participants gave their written informed consent prior to participating in the experiment.

6.2.1 Participants

The ASD group comprised 16 native German speakers with ASD (11 male, aged 25 - 55), who had all been diagnosed with ICD-10: F84.5 (Asperger Syndrome) and were recruited in the Autism Outpatient Clinic at the Department of Psychiatry

at the University of Cologne (Germany). As part of a systematic assessment implemented in the clinic, diagnoses were made independently by two different specialised clinicians corresponding to ICD-10 criteria and supplemented by an extensive neuropsychological assessment.
The control group comprised 16 typically developed native German speakers (4 male, aged 19 - 27) who were recruited from the student pool of the University of Cologne (Germany) and who were matched on verbal IQ to the ASD group.[1]

All participants completed the German version of the *Autism-Spectrum Quotient* (AQ) questionnaire, an instrument that has been developed by Baron-Cohen, Wheelwright, Skinner, Martin, and Clubley (2001) to measure autistic traits in adults with normal intelligence. The test consists of 50 items (statements) assessing aspects of social skills, imagination, communication, attention to detail and attention switching (e.g. "I would rather go to a library than to a party"). Each statement has to be evaluated by the participant by choosing one of four possible responses ("strongly agree", "slightly agree", "slightly disagree", "strongly disagree").
For evaluation, the answers are collapsed into "agree" vs. "disagree", resulting in scores of either 0 or 1 per item (1 if the response is characteristic of ASD). AQ scores therefore range from 0 to 50, with higher scores indicating more autistic traits.

All participants completed the *Wortschatztest* (WST, Schmidt and Metzler, 1992), a standardised German vocabulary test that not only exhibits high correlation to verbal intelligence, but also to general intelligence (Satzger, Fessmann, and Engel, 2002). Thus, it serves as a measure of both verbal and general IQ. The test consists of 42 items each containing one meaningful target word and five distractor non-words. Participants have to identify the target word (e.g. *Tilmad – Dailed – Laidel – Defain – Detail – Ailrod*).

Statistical tests (see table 6.1) revealed differences between the two groups for AQ scores (two-sample t-Test, $p > 0.001$), age (two-sample t-Test: $p > 0.0001$) and gender[2] (Pearson Chi-square test: $p = 0.03$). Crucially, there was no significant difference between the groups for (verbal) IQ as measured by WST (two-sample

[1] Initially, 31 participants were recorded for the control group. However, 15 of them had to be excluded from the analysis due to the following reasons: They were either acquainted with the confederate, or they stated that they suffered from Dyslexia, or they did not complete the AQ/WST tests.

[2] For information about individual speakers (e.g. AQ, WST, age), see Appendix B.1.

t-Test, p = 0.137).

Table 6.1: Means and standard deviations (sd) for age (in years), AQ scores and WST scores for both groups of participants and results of a two-sample t-Test (p-value).

	ASD		**Control**		
	mean	sd	mean	sd	p-value
Age	42.5	(±7.8)	22.1	(±2.7)	> 0.0001
AQ	41.4	(±3.3)	16.2	(±6.5)	> 0.0001
WST	116.9	(±12.6)	111.2	(±7.9)	0.137

6.2.2 Materials

The collaborative story-telling task was adapted from Fossard et al. (2018), who originally designed the task to explore effects of story coherence and complexity in referential choices of speakers with and without neurodegenerative and psychotic disorders. The task was also designed to investigate different discourse stages (introduction, maintenance, and shift of topic).

The task consisted of 18 stories, organised in three levels of increasing referential complexity: The level 1 stories featured one referent only, while the level 2 and level 3 stories featured two referents of different (level 2) or same (level 3) gender. Thus, both referential complexity (one referent versus two referents) and referential ambiguity (referents with same or different gender) varied between story levels. Furthermore, half of the stories followed a logical chronological development (logical stories), while the other half did not (non-logical stories). While the level 1 stories focussed on the introduction and subsequent maintenance of one referent only, the level 2 and level 3 stories involved a more complex structure. Fossard et al. (2018) defined three discourse stages for level 2 and level 3 stories, based on the six images composing each sequence, namely *introduction* of a new (i.e. unidentifiable) referent, *maintaining* this referent, and *shift* of the referent in focus in the image (see figure 6.1). Speech materials collected from these different discourse stages were analysed in terms of referent status (new, reactivated, and given) which takes into account both the information status and the topichood of a referent in order to determine instances of reactivation.

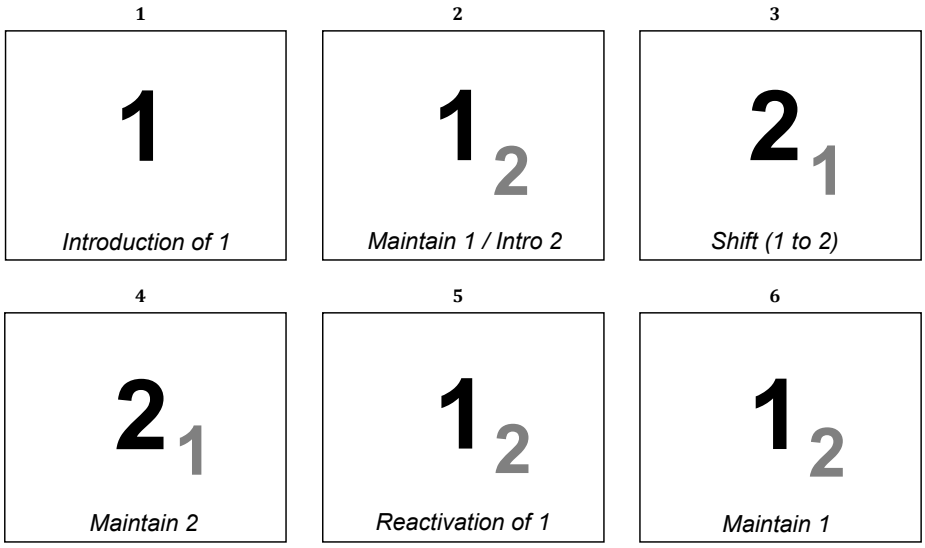


Figure 6.1: Structure of Level 2 - 3 stories with Introduction (picture 1), Maintenance (pictures 2, 4 and 6) and Shift (pictures 3 and 5) of the two referents.

Referents were considered as *new* when they were introduced for the first time, as *given* when they were maintained from one picture to the next, and as *reactivated* when mentioned after another character had taken the focus of the narration. This systematic pattern, which was kept constant throughout all level 2 and level 3 stories, is illustrated in figure 6.1.

6.2.3 Task and Recordings

Two participants of the ASD group (F01 and M01) were recorded in a quiet room at the Autism Outpatient Clinic at the Department of Psychiatry at the University of Cologne. The remaining 30 participants were recorded in a sound booth at the Institut für Linguistik - Abteilung Phonetik of the University of Cologne (Germany).

Participants were seated at a table opposite the confederate (author of this thesis). A screen was placed between the participant and the confederate to prevent eye contact and any other non-verbal communication. The participants were asked to narrate stories based on sequences of six ordered images. They were led to believe that they would be helping the confederate, who had the same six images in random order, to arrange them in the correct order. Participants were also

led to believe that the confederate was seeing the narrative sequences for the first time. For that purpose, a lab assistant handed over the materials to the participant in sealed envelopes.

In the first run, all logical stories were presented in randomised order. In the second run, all non-logical stories were presented in randomised order. In both runs, the very first story (a level 1 story) was kept the same for all participants to familiarise themselves with the task and with the structure of the stories.
Participants were instructed not to describe the pictures one by one or give the numbers of the pictures, but to try and narrate a cohesive story. Also, participants were instructed not to use proper names, nicknames or direct speech between the characters in order to ensure an appropriate elicitation of referring expressions.

6.2.4 Analysis

The recorded speech was segmented and annotated in Praat (Boersma, 2001) on eight different tiers (a summary of all annotation tiers is provided in table 6.4 further down). In the following, the specifications of each tier and the associated labels are briefly introduced. It should be noted that the information provided by some of the labels and tiers goes beyond the scope of what is analysed in this thesis, but will be consulted in future research (see also chapter 8 for an outlook). This is pointed out where applicable.

On tier 1, the speech was segmented into sections for each picture/scene. Tier 2 comprised orthographic annotations of each spoken word and of hesitation signals[3]. On tier 3, the positions of realised pitch accents and boundary tones were marked and their tonal configuration was categorised according to GToBI (Grice and Baumann, 2002; Grice et al., 2005).
While labels on tiers $1-3$ were applicable to all words uttered in a story, labels on tiers $4-8$ were applied to referring expressions referring to one or both of the characters of the story only[4], in order to achieve an elaborate analysis of the referents. Tier 4 incorporated the RefTop scheme developed by Cangemi, Kalbertodt, Krüger, and Grice (in prep.). This scheme merges information about the givenness (new or given (old)) and topichood (new or old and same or other) of referents (see

[3]However, hesitation signals were not evaluated in this thesis.

[4]In cases where referents appeared as zero anaphora (i.e. coordination of two verbs where the subject was not mentioned the second time), the corresponding verb was marked as the referring expression with all associated referent labels.

table 6.2).

Table 6.2: The RefTop annotation scheme (adopted from Cangemi, Kalbertodt, Krüger, and Grice (in prep.)).

		TOPIC		
		New	Old	
		—	Same	Other
GIV	New	**NN–**	—	NOO
	Old	**ON–**	**OOS**	OOO

Example (25) provides an English translation of a story that was told in which each label of the RefTop scheme occured.

(25) [A man]$_{\text{NN-}}$ is standing on stage and [starts]$_{\text{OOS}}$ juggling. [He]$_{\text{OOS}}$ is being watched by [a woman]$_{\text{NOO}}$. [She]$_{\text{ON-}}$ then also starts juggling, but with burning sticks. Then [she]$_{\text{OOS}}$ throws one of the sticks to [the man]$_{\text{OOO}}$. [He]$_{\text{ON-}}$ starts to juggle with the burning sticks and then [blows]$_{\text{OOS}}$ them out.

Tier 5 comprised information about the complexity and explicitness of the referring expressions used. The labels on this tier are composed of a capital letter to identify the referent (A for the first referent of the story, B for the second referent of the story, C if both referents are referred to simultaneously), plus a referential score indicating the referential complexity of the expressions used. The scores roughly reflect common hierarchies of accessibility/topicality (e.g. Givón, 1983; Ariel, 1990; Arnold, 1998) in a way that more complex referring expressions (e.g. expanded indefinite NPs) receive higher scores. The hierarchy and corresponding scores are displayed in table 6.3.

Tiers 6 and 7 featured a simplified version of the RefLex annotation scheme (Riester and Baumann, 2017; Baumann and Riester, 2012), with tier 6 representing the referential and tier 7 representing the lexical level of givenness. However, the analysis in this thesis will be limited to referential givenness as coded in the RefTop scheme mentioned above (see table 6.2) in order to be able to determine instances of reactivation. The reactivation of a given referent as topic would be expected to evoke a high level of prominence due to a high (re)activation cost and cannot be captured by means of the RefLex scheme alone[5].

[5]However, the information coded in the RefLex tiers will be used in a future analysis (see also chapter 8).

Table 6.3: Hierarchy of referring expressions with the corresponding values as used in tier 5 of the annotation, and examples in German with English translations.

value	referring expression	example
4,5	expanded indefinite NP	eine blonde Frau *a blonde woman*
4	indefinite NP	eine Frau *a woman*
3,5	expanded definite NP	die blonde Frau *the blonde woman*
3	definite NP	die Frau *the woman*
2	demonstrative pronoun	diese *this (one)*
1,5	relative pronoun	(eine Frau,) die *(a woman,) who*
1	personal pronoun	sie *she*
0	zero anaphora	(sie lachte und) bezahlte *(she smiled and) paid*

Finally, tier 8 was used to mark referents that occured in contrastive focus, an aspect that is beyond the scope of this thesis.
All annotation tiers with label examples are summarised in table 6.4.

The dataset was annotated by three trained annotators who were blind to the participants' diagnoses. GToBI annotations were made by a student assistant with expertise in GToBI analysis. The remaining tiers were annotated by a second student assistant and later checked and complemented by a Phonetician. Parts of the material were annotated independently by the author of this thesis and compared to the initial annotations. Ambiguous or unclear cases were discussed among the annotators and the resulting agreements were documented in order to achieve a valid reliable annotation on all tiers. An example of a fully annotated sentence is provided in figure 6.2.

The following analysis focussed on the prosodic marking of referring expressions. Cases of zero anaphora (coordination of two verbs where the subject is not mentioned the second time) were excluded from the prosodic analysis, so that only explicit mentions of referents were taken into account. Furthermore, only words that would typically be predistined to carry pitch accents were included (i.e. pronouns, nouns, adjectives, but not articles).

The GToBI transcriptions were simplified in a way that nuances of upstepped or

Table 6.4: Overview of annotation tiers used in Praat.

Nr.	tier name	description	example
1	Scene	Annotation of each scene/picture. "0" was used for introductions prior to picture 1.	0–6
2	Word	Orthographic annotation of spoken words and hesitation signals (HES).	ein mann HES
3	GToBI	Annotation of pitch accents and boundary tones according to GToBI guidelines. Prenuclear pitch accents were annotated in parentheses ().	L+H* (L*) H-%
4	RefTop	Five-way distinction merging information about referential givenness and topic. See table 6.2.	ON-OOS
5	RefScores	Combination of information about referent (A, B or C) and a hierarchic scale with higher scores indicating a higher referential complexity. See table 6.3.	A4,5 B3
6	Ref	Specification of the referential status of a referent according to the RefLex annotation scheme.	r-new r-given
7	Lex	Specification of the lexical status of a referent according to the RefLex annotation scheme.	l-new l-given
8	ALT	Annotation of referents that occured in contrastive focus.	ALT

downstepped H-tones within rising bitonal pitch accents (e.g. L+!H*, L+ˆH*) were analysed as their underlying canonical form (e.g. L+H*).

For the analysis, the information about referential givenness as provided by the RefTop scheme was simplified to obtain three different types of referent status referred to as "New", "Reactivated", and "Given" (see table 6.5): Referents with the status "New" were brand-new referents that were introduced as the new topic (NN-). Referents with the status "Reactivated" were already given in the discourse ("old") but re-introduced as the topic (ON-). Referents with the status "given" were given in the discourse and maintained as a topic (OOS). The remaining labels from the RefTop scheme (NOO and OOO) were not considered for the analysis.

In the following, a statistical analysis of accent categories (section 6.3.1) and accent types (section 6.3.2) on referring expressions is presented. Since the sto-

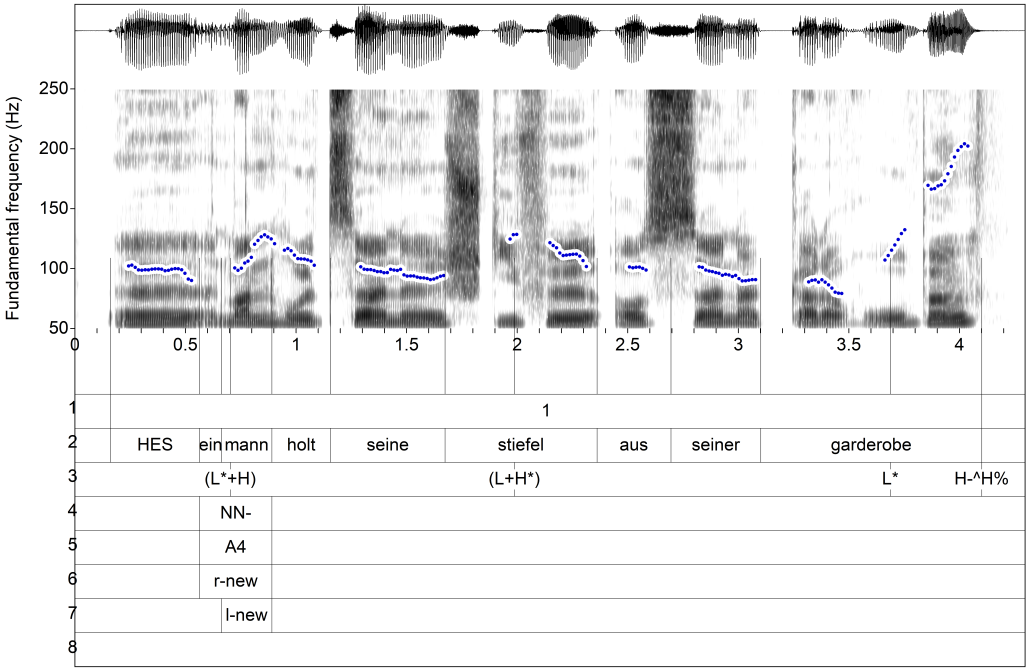


Figure 6.2: Praat annotation example with waveform (top panel), spectrogram (middle panel), F_0 contour (in blue) and eight annotation tiers (bottom panel). Example from control speaker M01.

Table 6.5: Simplification of RefTop labels leading to different types of referent status.

RefTop		*Referent Status*
NN-	=	*New*
ON-	=	*Reactivated*
OOS	=	*Given*

rytelling task is a corpus-based approach, the extent of inferential statistics that can be applied to this data is limited. Therefore, the findings of the inferential analysis are complemented by a descriptive analysis of certain aspects of the data in section 6.3.3.

6.3 Results

P-values were generated using likelihood ratio tests to compare full models (Poisson Regression) with *group* (ASD vs. control) as a fixed effect to the corresponding null models without the fixed effect. The total amount of produced referents was taken as an offset to adjust for differences in total numbers of mentions between the groups.
Logical and non-logical stories were taken together for the analyses. This decision was based upon the facts that a) all participants narrated non-logical stories in a coherent way, as if the stories were logical, and b) visual inspections of all generated graphs separated into logical versus non-logical stories exhibited no differences in accent distributions or referential choices.

6.3.1 Accent Categories

In a first analysis, the distribution of nuclear and prenuclear accents as well as cases where no accent was realised on a referent was compared in both groups. The accent placement patterns for each type of information status in each group are shown in figure 6.3.

According to the statistical analysis (p-values are provided in table 6.6), the relative distribution of accent categories (nuclear accent, prenuclear accent, no accent) across new and reactivated referents was similar in both groups (see also figure 6.3): New referents predominantly received an accent in both groups, the accent being either prenuclear (49% and 42%) or nuclear (32% and 38%). Reactivated referents were slightly less frequently marked with prenuclear and nuclear accents in both groups.
Given referents were predominantly unaccented in both groups, but significantly more often so in the control group (87%) than in the ASD group (78%), see p-values in table 6.6. This implies that, in sum, the ASD group used more accentuation on given referents than the control group did. However, the cases of nuclear and prenuclear accents on given referents were not significantly different and can therefore be regarded as similar in both groups.

Thus, an increase of (nuclear) accentuation was observed from given through reactivated to new in both groups, reflecting the increased activation cost for referents that either had to be reactivated or introduced. However, the ASD group

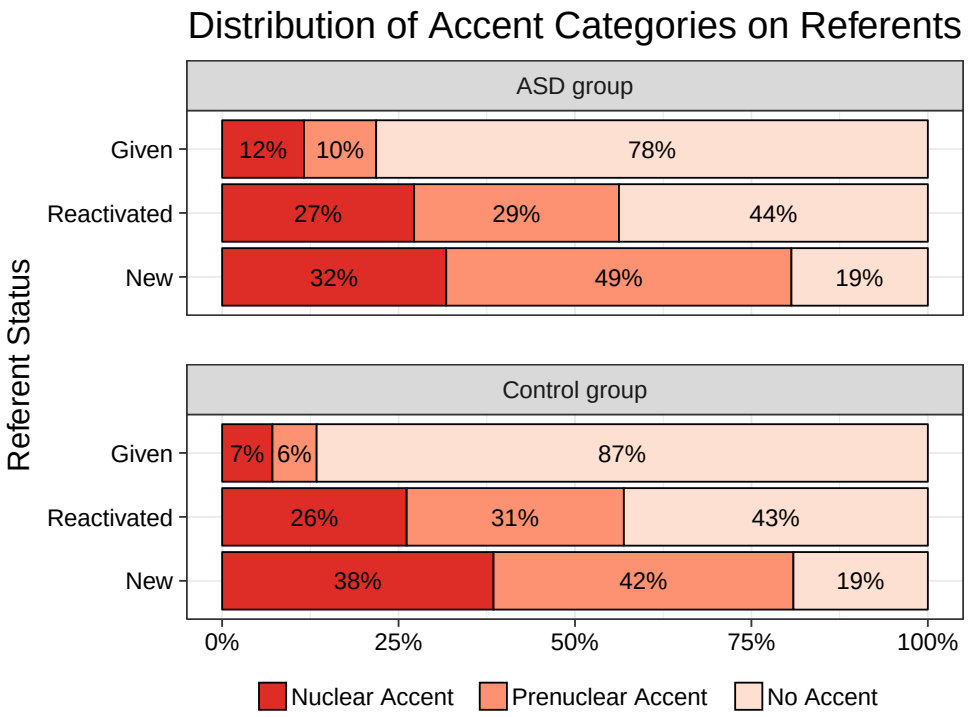


Figure 6.3: Relative distribution of accent categories on referents in all stories as produced by the ASD group (top) and the control group (bottom). Percentages are rounded.

Table 6.6: The effect of group on the distribution of accent categories as indicated by p-values for the different types of referent status. Italics and asterisks indicate significant group differences.

	Nuc. Accent	Prenuc. Accent	No Accent
		p-values	
Given	0.304	0.079	*0.021**
Reactivated	0.683	0.760	0.853
New	0.224	0.512	0.850

did not use deaccentuation as reliably as the control group did to prosodically attenuate given referents.

6.3.2 Nuclear Accent Types

In a second analysis, the different accent types that were used by speakers when they produced nuclear accents on referents were compared. The distribution of nuclear accent types for each level of information status in each group is shown in figure 6.4.

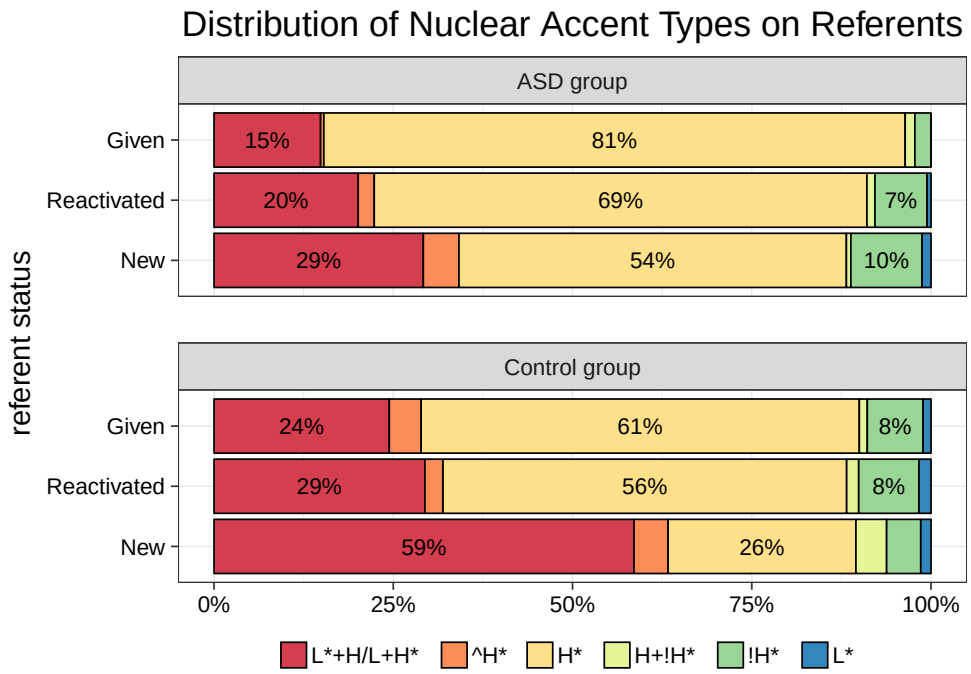


Figure 6.4: Relative distribution of nuclear accent types on referents in all stories as produced by ASD group (top) and control group (bottom). Percentages are rounded. Accent types are ordered according to their perceived prominence from left (most prominent) to right (least prominent), see Baumann and Röhr (2015).

Irrespective of condition, there was a significantly higher number of nuclear H* accents in the ASD group in comparison to the control group ($\chi^2(1) = 9.09$, $p = 0.003$).
Furthermore, the two groups differed with regard to their choice of nuclear accent type on referents in several conditions (see p-values in table 6.7)[6]. There were significant group differences for rising accents (L*+H/L+H*) and high accents (H*) on new referents, as well as for !H* accents on given referents. There were no group differences regarding the choice of nuclear accent type in any other condition.

Thus, while the control group predominantly used prominent rising accent types (L*+H/L+H*) on new referents (59%), the ASD group did so significantly less often (29%). Instead, the ASD group used H* accents significantly more often (54%) than the control group (26%) to mark new referents.
Reactivated referents were predominantly marked with H* accents in both groups,

[6]Accent types that are not included in table 6.7 did not occur often enough to be included in the statistical analysis.

Table 6.7: The effect of group on choice of accent type as indicated by p-values for the different types of referent status. Italics and asterisks indicate significant group differences.

	L+H*/L*+H	H*	!H*
	p-values		
Given	0.359	0.107	*0.028**
Reactivated	0.645	0.356	0.779
New	*0.011**	*0.007**	0.155

and the amount of rising accents on reactivated referents fell inbetween those of new and of given referents.
Given referents with nuclear accents predominantly received H* accents in both groups. However, there was a significant group difference regarding the use of (less prominent) !H* accents on given referents, with more occurences of !H* in the control group (8%) than in the ASD group (2%).

In sum, when referents were marked with nuclear accents, an increase in prominence of accent type was observed from given through reactivated to new in both groups. However, new referents were significantly more often marked with prominent rising accents in the control group than in the ASD group, who resorted to less prominent accents (H*) instead. Furthermore, while given referents were often marked with H* accents in both groups, the control group also used less prominent !H* accents more often than the ASD group.

6.3.3 Further Observations

Apart from the quantifiable results for accent placement and choice of accent type on referring expressions, other interesting differences between speakers with and without ASD regarding reference marking have been observed, some of which are presented in this section. These observations focus on choices of referring expressions and their (prosodic and lexical) marking.

The located differences between groups regarding prosodic emphasising of new referents (by using more prominent accent types) and prosodic attenuation of given referents (by using deaccentuation and less prominent accent types) can be substantiated by comparisons of individual speakers' consecutive mentions of the same referents.
The following figure 6.5 shows waveforms, spectrograms, F_0 contours and anno-

tations of four consecutive mentions of the same referent by speaker F01 from the ASD group. The first mention was the introduction of the referent, the second mention was a case of reactivation of the same referent, and the third and fourth mentions were cases of maintenance of the same given referent.
The ASD speaker used the same full expanded NP ("eine / die braunhaarige Person" ("*a / the brown-haired person*")) with no signs of attenuation: Each referring expression was uttered with comparable F_0 contours and a variation of nuclear or prenuclear accents that were in each case of the type H*. Furthermore, word durations did not decrease from new to given. The only apparent difference between new and given referents was the modification of the referring expression from an indefinite NP to a definite one.

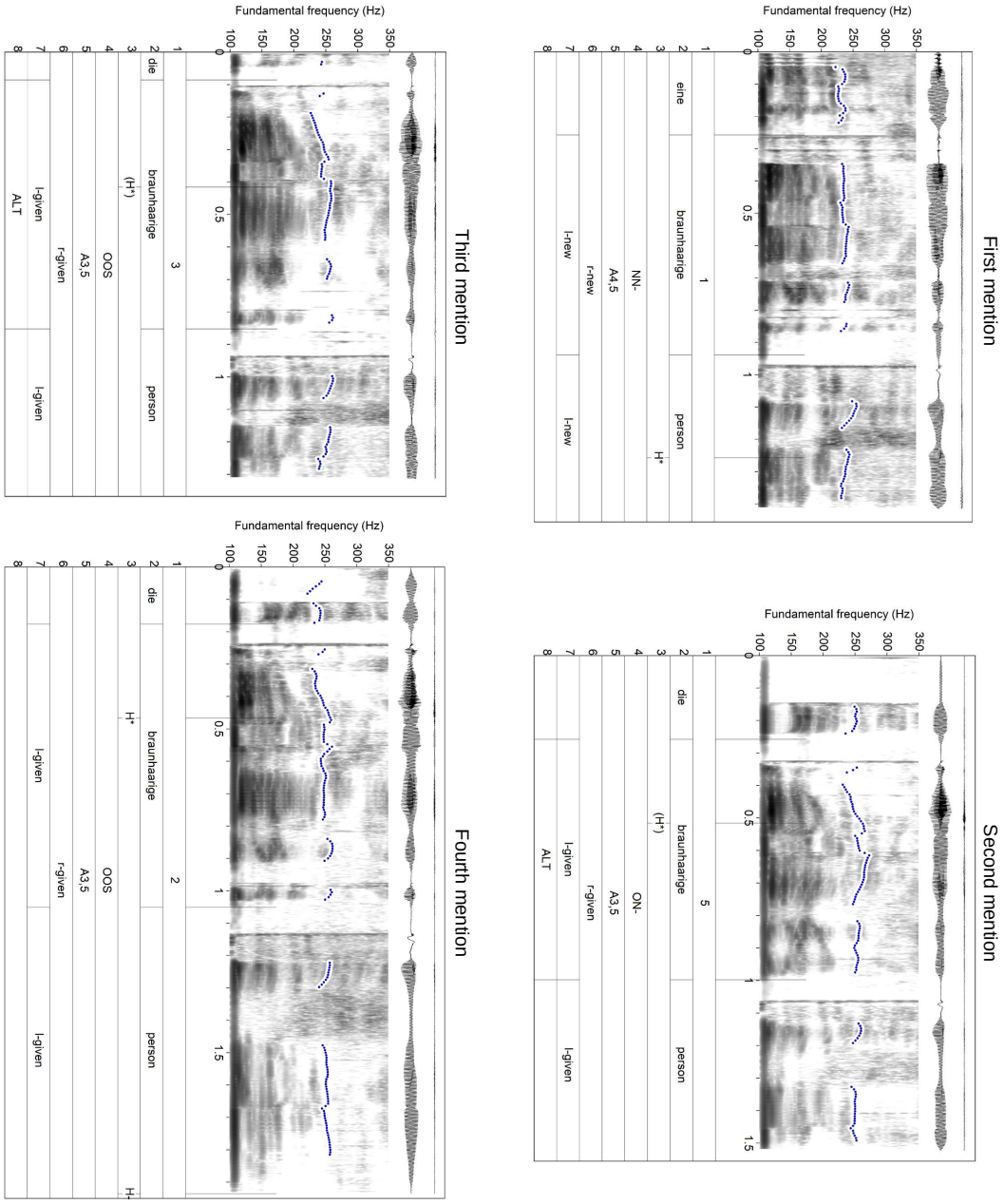


Figure 6.5: Example of four consecutive mentions of a referent by ASD speaker F01, story L2C, showing waveforms, spectrograms, F_0 contours and annotations.

For comparison, figure 6.6 shows three consecutive mentions of the same referent by speaker F11 from the control group. Compared to the ASD speaker, the control speaker attenuated the referring expressions of the second and third mention both prosodically and lexically (by choosing a less explicit NP). The prosodic attenuation is achieved through the use of a nuclear accent in the first mention and prenuclear accents in the consecutive mentions. Moreover, the reactivated referent received a rising (prenuclear) accent, while the given referent received a less prominent !H* (prenuclear) accent. Shorter word durations of the noun further contribute to the attenuation of the given referent.

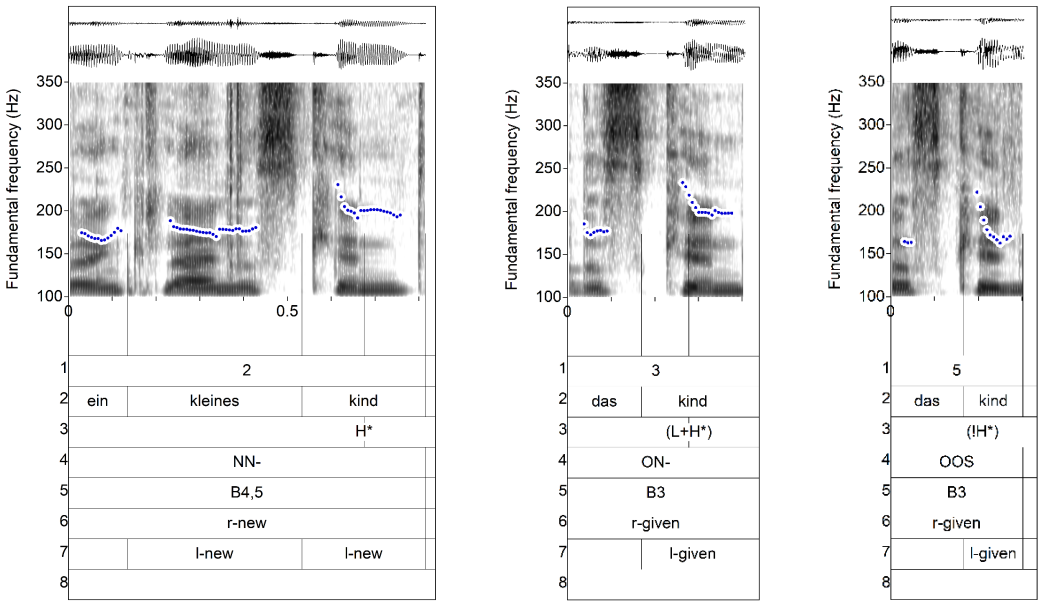


Figure 6.6: Example of first (left), second (middle) and third (right) mention of a referent by control speaker F11, story L2C, showing waveforms, spectrograms, F_0 contours and annotations.

It should be noted that it was challenging to provide comparable examples from speakers of the control group at all, as they mostly used pronouns when referring to given referents, unless these referents were in contrastive focus, in which case the expressions were not suitable to illustrate attenuation.

Speaker F01 from the ASD group represented the most extreme case of over-explicit reference marking. However, the tendency to use full NPs (instead of pronouns) for given referents has been noticed among several speakers from the ASD group. This is illustrated in figure 6.7, in which absolute numbers of per-

sonal pronouns, demonstrative pronouns, and full NPs used by all speakers for the different types of referent status are shown. The horizontal lines divide the bars up into counts for individual speakers.
Both groups showed an equal amount of NPs when introducing new referents. However, while the number of NPs observably decreased from new through reactivated to given in the control group, this was not the case in the ASD group, where the number of NPs remained constant throughout the different types of referent status.

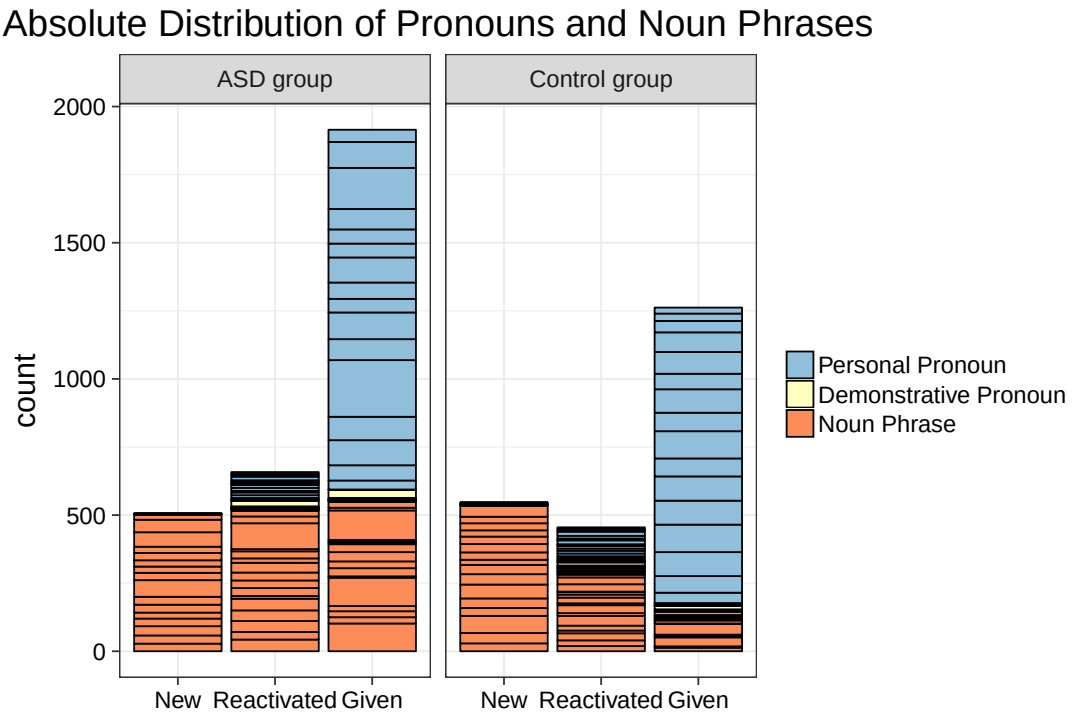


Figure 6.7: Absolute distribution of referential expressions (pronouns versus NPs) in all stories as produced by the ASD group (left) and the control group (right). Horizontal lines indicate counts for individual speakers.

The trend among ASD speakers to make less use of pronouns than control speakers when referring to given referents was particularly apparent in level 1 stories that focussed on the introduction and subsequent maintenance of one single referent only. This is illustrated in figure 6.8, in which diagrams of level 1 stories narrated by control speakers (left) and ASD speakers (right) are compared. The diagrams show the referring expressions used by the speakers throughout the stories. Each point in the diagram represents one reference made be a speaker, with the values corresponding to the hierarchy of referring expressions that was introduced in table 6.3 in section 6.2.4. Thus, higher values indicate more detailed and

more explicit forms of reference (e.g. expanded NPs), while lower values indicate more attenuated lexical forms (e.g. pronouns). The x-axis represents the time course of the individual pictures of the story.

The two speakers from the ASD group analysed in figure 6.8 were the ones who most frequently referred to given referents with (expanded) NPs (e.g. "the yellow-haired person"). The one on the top right is the same speaker (F01) that was analysed in figure 6.5 above. It should be noted that the two speakers represented the most extreme cases with regard to choices of referring expressions, and that their behaviour does not necessarily reflect the performance of all ASD speakers in this study.

For the control speakers, the first reference made in a story had high values (4.5 or 4). This means that they introduced the new referent of a story with indefinite NPs. The subsequent mentions of the referent had very low values (0 or 1) that indicate zero anaphora or pronouns, respectively.

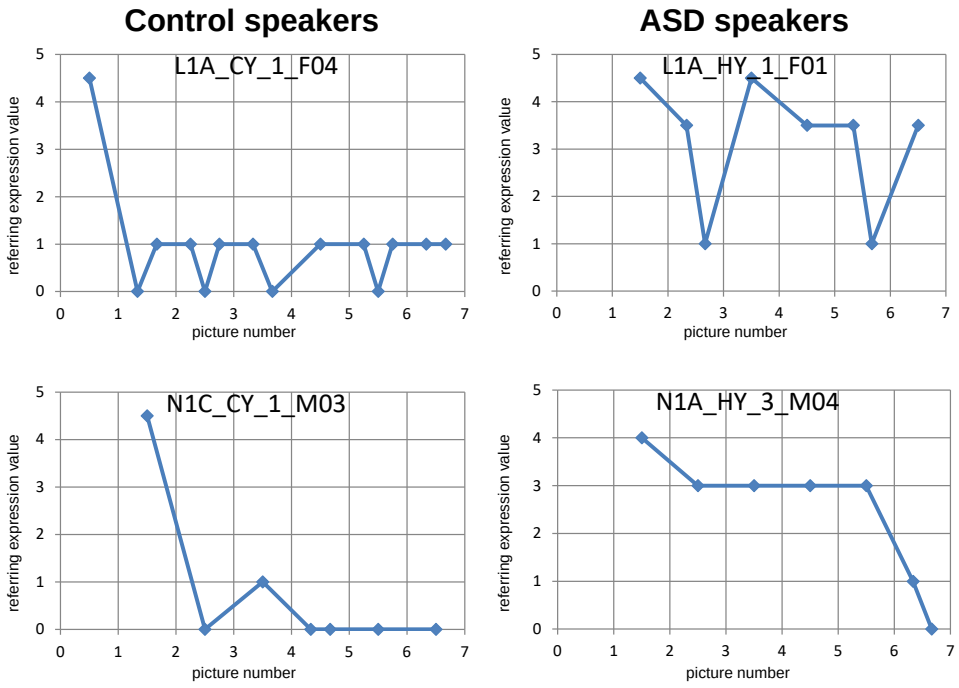


Figure 6.8: Diagrams of referring expressions used in a story by speakers of the control group (left) and the ASD group (right). Higher values indicate more complex referring expressions.

The diagrams for the speakers with ASD (on the right) look differently: While the new referent was also introduced with indefinite NPs (values 4.5 or 4), there was no immediate drop to pronouns or zero anaphora for subsequent mentions

of the referent. Instead, the speakers either alternated between pronouns and complex referring expressions (top right diagram of figure 6.8) or persisted in using definite NPs (value 3) throughout several pictures before finally using a pronoun at the end of a story (bottom right diagram of figure 6.8).
Hence, given referents were lexically attenuated by speakers of the control group, but not by some of the speakers of the ASD group, who were over-explicit when referring to given referents.

In sum, the observations presented in this section have provided further evidence for an impaired marking of referents in the ASD group, both in terms of emphasising new referents and attenuating given referents.
As mentioned above, it should be noted that these observations are anecdotal and do not reflect the performance of all ASD speakers. In fact, some of the speakers with ASD exhibited patterns that could be regarded as the opposite behaviour. For example, one speaker used pronouns to the extent that his references were ambiguous and unresolvable to the listener. Other ASD speakers exhibited similar patterns to those found in control speakers.
On the other hand, the performance of speakers within the control group was very consistent. The examples presented in this section represent typical behaviour of control speakers. Thus, the ASD group can be considered as more heterogeneous in comparison to the control group.

Nevertheless, the observations complement the results from the statistical analyses of sections 6.3.1 and 6.3.2 reported above, namely that there are differences between groups regarding the encoding of referential givenness. This seems to be the case not only in terms of prosodic marking of the referents, but also in terms of choice of referring expression to refer to the characters of the stories.

6.4 Discussion

The results of this production experiment generally confirmed the prediction that adults with ASD would show a different behaviour than typically developed adults when encoding referential givenness. Differences between groups were found for highlighting and attenuation of referents as well as for choices of referring expressions.
The differences were not as pronounced for the introduction of new referents as

they were for the maintenance of given referents. Both groups mostly introduced new referents with indefinite NPs and a (nuclear) pitch accent, the only difference being the choice of accent *type*, which tended to be a more prominent rising one (L+H*, L*+H) in the control group and a less prominent default one (H*) in the ASD group.
Given referents, however, were maintained with overspecific referring expressions such as expanded NPs by a majority of the speakers from the ASD group, together with a reduced use of pronouns as compared to the control group. Furthermore, given referents were not as frequently deaccented in the ASD group as they were in the control group. In cases where given referents were accented, the control group used less prominent !H* accents significantly more often than the ASD group.
Thus, the findings for both the prosodic marking and the choice of referring expressions suggest that speakers with ASD did not attenuate given information to the same extent as control speakers did.

The attenuation of previously mentioned information is assumed to depend on the speaker's assumption of the listener's knowledge state (e.g. Gundel et al., 1993; Clark and Marshall, 1981). In order to adhere to Grice's maxim of quantity, redundant information is to be avoided, and shorter and more effortless expressions (i.e. attenuated forms) are preferable in order to make communication efficient and economical, but only if the information provided allows the listener to identify the intended referent (see also section 3.2 of chapter 3).

However, it is not always clear whether the referential choices that speakers make are listener-driven or simply context-driven, as both the listener and the speaker usually have access to the same discourse context (Hendriks et al., 2014: 392). There is evidence that referential choices and their prosodic marking are, at least to some degree, listener-driven (see, e.g., Hendriks et al., 2014; Galati and Brennan, 2010, for a review see Brennan and Hanna (2009)). The present study supports this assumption in two ways. First, the task was designed in such a way that speakers were led to believe that their listeners did not know the stories and the correct sequence of the pictures, which is why they had to monitor the activation of each referent to avoid ambiguity and choose an appropriate referential marking to complete the task. Second, individuals with ASD, a group that reportedly exhibit impaired mentalising skills, perfomed differently from typically developed adults, which could be an indication that mentalising skills were of avail for the task.

However, the attenuation of previously mentioned information also seems to be susceptible to egocentric speaker-only knowledge (Bard et al., 2000). This might explain the fact that the ASD group managed to attenuate given referents to some extent (although less than controls), since they may have been doing so based on their own knowledge state.

On the other hand, the context-dependency of referential choices might have also had an influence on the performance of the speakers. For example, Arnold et al. (2009) assume that pronouns are more difficult to produce than full NPs because pronouns are context-dependent, whereas full NPs are not and can be used in all contexts. Speakers might thus avoid using pronouns in cognitively demanding situations (see also Hendriks et al., 2014) or when their ability to understand context is compromised, as is assumed to be the case in ASD (cf. WCC (Frith, 1989; Happé and Frith, 2006), see also section 4.2.1). The fact that speakers with ASD were found to use less pronouns is in line with this assumption. Other studies have also reported that references of speakers with ASD are more specific than needed (Colle et al., 2008; Arnold et al., 2009).

One of the most striking aspects of the prosody of speakers from the ASD group was their extensive use of H*. Irrespective of the referent status, the choice of H* was prevalent in the ASD group, unlike in the control group. While H* may be appropriate in certain contexts, persistent production of this type of accent suggests that a speaker is using prosody in an atypical manner, or is inattentive to pragmatic context (Diehl et al., 2009).
An extensive use of H* may in fact be a compensation strategy: Using a prosodic pattern that makes the referent neither too prominent nor too attenuated could be seen as a hedging, non-committal way of communicating. This is in line with the finding that speakers with ASD often used full NPs instead of context-dependent pronouns to refer to given referents. This behaviour, too, can be seen as a hedging strategy of residing uncommitted in communication, since NPs are acceptable in most contexts and using them reduces the risk of ambiguity.
An extensive use of the H* accent is in line with previous studies reporting that pitch contours in general were found to be limited to a small repertoire of patterns, with less variation and a preference for "monotonous" accents such as H* in children with ASD (Green and Tobin, 2009; Edelson et al., 2007; Demouy et al., 2011).

In social communication, the observed behaviour of the speakers with ASD in this

study might come across as a lack of narrative engagement. A more frequent use of neutral accents might lead to the impression that a speaker is not involved and sounds perfunctory. Less narrative engagement has also been found in a study investigating narratives of children with ASD (Capps, Losh, and Thurber, 2000).

The fact that the groups were not matched on chronological age or gender is certainly a limitation to this study. When investigating speech in children, chronological age is one of the key measures to control for ongoing language development. However, performance of adult participants is mainly predicted by verbal IQ and general language ability (Nadig, Vivanti, and Ozonoff, 2009). Furthermore, a study by Paul et al. (2008) that investigated speakers with a high age range between 7 and 28 years found that verbal IQ was the only measure that correlated significantly with accent production (while age did not). DePape et al. (2012) also report that in their study, those with higher language abilities (as measured by a vocabulary test) showed different prosodic patterns from those with moderate language abilites.
All participants of the present experiment had comparable verbal IQ measures and can therefore be regarded as equally able to perform the task. The effect of gender differences on the marking of referential givenness has not been investigated. Group effects due to gender disparity can therefore not be excluded.

However, despite these limitations, the results of this production experiment show interesting differences regarding the prosodic marking and choice of referring expression between speakers with and without ASD. The present findings contribute to the characterisation of the qualitative nature of speech and language impairments in autism.

6.5 Summary and Conclusion

In a picture-based cooperative story-telling task, the encoding of a referent's givenness was investigated in speakers with and without ASD. When new referents were introduced (or reactivated), adults with ASD were similar to typically developed adults in their pitch accent *placement*, but differed in their choice of accent *type*. On new referents, the ASD group produced accents which are less prominent and which have a non-committal nature, while the control group made greater use of more prominent accents.

The most pronounced group differences were found for the maintenance of given referents. Given referents were not as frequently deaccented in the ASD group as they were in the control group. Furthermore, in cases where given referents were accented, the control group used less prominent accents significantly more often than the ASD group.

Apart from differences in the prosodic marking of referents, differences in the choice of referring expressions, especially in the selection of either full NPs or pronouns to maintain given referents, were observed between groups. Given referents were maintained with overspecific NPs instead of pronouns by a majority of the speakers from the ASD group. The control group, on the other hand, used pronouns much more frequently to refer to given referents.

The findings for both the prosodic marking and the choice of referential expression suggest that speakers with ASD did not attenuate given information to the same extent as control speakers did. Moreover, two possible compensation strategies within the ASD group were identified: 1. An extensive use of the non-committal H* accent, irrespective of the referent status, and 2. an extensive use of NPs for given referents. Both behaviours can be interpreted as non-committal hedging strategies that are safe to use as their neutrality makes them suitable in a large number of contexts.

The selection of a referring expression along with an appropriate prosody is a highly context-dependent and listener-oriented aspect of language. The fact that speakers from the ASD group did not perform as reliably as speakers from the control group thus confirms both difficulties with aspects of central coherence (understanding information in context) as well as aspects of mentalising for listener-driven communication.

Taken together, these findings point towards a reduced ability in individuals with ASD to produce informative, committed, listener-oriented prosody and may offer an additional explanation as to why individuals with ASD are often faced with problems in social communication.

Part III

Final Conclusion and General Outlook

Chapter 7

Final Summary and Conclusion

The aim of this thesis was to contribute to the current knowledge of prosody competence in high-functioning individuals with ASD by investigating the perception and production of referential givenness in this population.

Previous studies have shown that the encoding and decoding of information structure by means of prosody poses an area that is particularly affected in adults with ASD. Paired with general problems in pragmatic aspects of speech, it was predicted that adults with ASD would be less able than typically developed adults to use prosody both when interpreting the givenness of a referent and when conveying referential givenness themselves. Both experiments confirmed this prediction.

In the perception experiment, participants were instructed to make judgements as to how far an item or person referred to by a target word sounded as if it was known or new. Results revealed that participants from the ASD group made significantly less use of prosody than participants from the control group: Although both groups rated words bearing a nuclear pitch accent as sounding newer than words without one, ASD subjects exhibited significantly smaller differences in ratings between words with and without nuclear accents.
Furthermore, unlike the control group, the ASD group did not appear to take into account differences between accent types: While the control group tended to rate words with more prominent nuclear accents as newer than words with less prominent nuclear accents, the ASD group did not show this prominence-driven pattern.
Instead of attending to the prosody of spoken words, individuals with ASD chose a different strategy in order to complete the task: They paid more attention to

lexical properties of the words instead (word frequency and animacy). Thus, they based their judgement more on their personal knowledge of the words and less on their pronunciation, as opposed to individuals from the control group, who prioritised prosodic cues.

In the production experiment, the ability of adults with ASD to encode referential givenness was investigated in a cooperative story-telling task. The findings for both the prosodic marking and the choice of referring expressions indicated that most of the speakers with ASD did not attenuate given information to the same extent as control speakers did. Given referents were not as frequently deaccented in the ASD group as they were in the control group. In cases where given referents were accented, the control group used less prominent accents significantly more often than the ASD group did. Additionally, given referents were maintained with overspecific NPs instead of pronouns by a majority of ASD speakers.
Furthermore, new referents were more frequently marked with prominent rising accents in the control group than in the ASD group, who resorted to less prominent accents (H*).
Interestingly, the ASD group made use of H* accents extensively in all conditions. Using a prosodic pattern that makes referents neither too prominent nor too attenuated could be seen as a hedging strategy of non-committal communication.

Taken together, the two experiments presented in this thesis provide further evidence for the assumption that pragmatic prosody represents an area of particular difficulty for individuals with ASD. The area of pragmatics requires both an understanding of context-dependent aspects of speech as well as the ability to take the mental states of others into account. Thus, the difficulties in inferring intentions of speakers through prosody and in using listener-oriented and context-dependent prosody can be taken to support theories of underlying cognitive impairments such as weak central coherence and mentalising difficulties.
While neither of the two experiments was specifically designed to tease apart the cognitive theories (ED, WCC, ToM, see also section 4.2.1 of chapter 4), their findings take an initial step towards understanding the influence of cognitive-linguistic ability of individuals with ASD on prosodic aspects of speech.

Observations from both experiments have also confirmed that individuals with ASD tend to employ compensation mechanisms to complete tasks. In the perception experiment, individuals with ASD attempted to recruit "objective" and context-independent lexical properties of the words instead of making sense of

information from prosody "intersubjectively" in order to calculate the degree of givenness of a referent.
In the production experiment, individuals with ASD chose noncommittal forms (definite NPs and H* accents) that fit several contexts rather than committing to more context-dependent forms and risk inappropriateness and ambiguity.
Hence, the tendency of individuals with ASD to develop compensation strategies both in structured tasks and in every-day social encounters has to be kept in mind when analysing their behaviour and drawing conclusions about their prosodic abilities. While at first view, their use of prosody might not always appear strikingly deviant, a profound and detailed analysis might reveal subtle differences that, in sum, can lead to the impression of a speaker being less involved in conversation.

The findings highlight the presence of prosodic deficits even in high-functioning adults with ASD and might help to better understand the difficulties encountered by people with ASD in speech-based communication and social encounters. The reported findings emphasise the need for diagnostic attention to prosody which can be a stigmatising barrier to social acceptance (Shriberg et al., 2001).
From a clinical viewpoint, it is important that aspects of prosody comprehension and production are included into the assessment and treatment of individuals with ASD. Prosodic deficits are rarely the focus of treatment. Yet, the demand for prosody intervention has been growing due to the increasing evidence for prosodic impairment in ASD.
Recently, there have been some attempts to develop intervention methods for prosody (Dunn and Harris, 2016; Rothstein, 2013). They predominantly concentrate on global features, such as loudness, pitch, and rhythm as well as on breathing, phonation and articulation to improve the overall intelligibility of speech. However, they also provide short excursions into descriptions of "pragmatic stress", mostly covering focus and contrast. These lessons could be supplemented by explanations about referential givenness in order to obtain holistic instructions about how to achieve appropriate "pragmatic stress". Furthermore, incorporating training of reference identification and comprehension for children with ASD can advance their understanding of peers and adults in everyday conversations at school and at home.

Lastly, the thesis has shown that AM theory can be a powerful tool to describe the prosodic characteristics of speakers with ASD and to compare them to those of typically developed speakers.

Chapter 8

Open Questions and Future Research

During the course of this thesis, a number of questions and ideas for future research have emerged.

First of all, the corpus that was recorded for the production experiment provides a valuable resource of spontaneous speech from high-functioning individuals with ASD. Many aspects of this data are still unexplored, such as the evalution of pauses and hesitation signals, the analysis of pitch range, or the investigation of turn-taking, to name but a few.
However, a new control group that is carefully matched on age and gender would have to be recorded in order to exclude possible group differences due to these factors. In the course of this, a re-analysis of referential givenness as in this thesis should be performed to ensure that the results reported were not influenced by differences of age or gender between the groups.

Furthermore, the following analyses that are related to the notion of referential givenness may be worth investigating: First, a detailed evaluation of the RefLex annotations, especially of instances of inferentially accessible referents (e.g. waiter in restaurant), would provide further insights into the ability of individuals with ASD to provide a nuanced prosodic marking of givenness of story characters.
Second, while the current analysis was limited to the referential givenness of the main characters of the story, a further analysis could focus on the prosodic marking of givenness of objects that appeared in the stories (shoes, a kite, a cake, pieces of wood, shovels, etc.). Given the detected sensitivity of individuals with ASD

to the animacy of referents, such an analysis would provide interesting insights into possible differences of prosodic marking between animate and inanimate referents.
Third, while contrastive prosody has been investigated in a considerable number of studies, results remain conflicting with regard to the ability of adults with ASD to appropriately signal contrastive focus prosodically in spontaneous speech. In order to gain further insights into this, the existing contrast annotations of the referents of the corpus could be investigated. Level 2 and level 3 stories provide suitable contexts for this analysis due to consistent contrasts between the two characters of the stories.

Previous research has suggested that in general, individuals with ASD exhibit less variation in their pitch contour repertoire. The corpus provides the opportunity to capture and analyse the pitch contour inventories of speakers with ASD in order to be able to provide an account of the characteristics of pitch contours of German ASD speakers.

Receptive and expressive prosodic skills have been reported to be correlated in subjects with ASD (e.g. Peppé et al., 2007). Unfortunately, this could not be tested in the current experiments. Since participation in the perception experiment was anonymous, it could not be retraced whether some of the participants from the perception experiment were the same as those that participated in the production experiment. Thus, no connection between individual performances in the perception and production tasks can be made. Therefore, how the production and perception of prosodic signals relate to each other in ASD is an empirically open question. It may be interesting to explore the link between prosody perception and production by running the two experiments with the same participants and by correlating their performance in the receptive and expressive tasks.

Another correlation that has been reported is the correlation between the ability to pass ToM tasks and to perform well on prosody perception tasks among individuals with ASD (see Kleinman et al., 2001; Hurley and Bishop, 2016; Bishop, 2017). This reflects the characterisation of autism as a spectrum disorder. Therefore, ToM-scores should be measured in future perception studies and should be correlated with the individual performance of subjects with and without ASD.
ToM-scores are available for the participants of the production experiment. It would be interesting to look into connections between individual ToM-scores and certain aspects of performance in the story-telling task, such as, e.g., the tendency

to use full NPs instead of pronouns for given referents or the ability to deaccent given information.

Lastly, it may be interesting to explore how the collected speech samples of individuals with ASD are perceived by naive (typically developed) listeners. A number of possibilites for interesting perception experiments arise from this. For example, the recorded stories could be played to listeners who would have to complete the same task as the confederate, namely to arrange the six pictures in the correct order. It would be interesting to analyse whether and how efficiently the listeners can complete the task and whether any misunderstandings would arise.
Another perception experiment could explore how prosodic aspects of the speech of individuals with and without ASD influence ratings of listeners regarding, e.g., the engagement, sympathy, mood, courtesy, etc. of a speaker.

More generally, it is hoped that future work on prosody in ASD will continue to benefit from the advantages of phonologial models of prosody.

Bibliography

Allerton, D. J. (1978). The Notion of 'Givenness' and its Relation to Presupposition and Theme. *Lingua*, *44*, 133–168.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Ariel, M. (1988). Referring and Accessibility. *Journal of Linguistics*, *24*, 65–87.

Ariel, M. (1990). *Accessing Noun Phrase Antecedents*. New York: Routledge.

Arnold, J. (1998). *Reference Form and Discourse Patterns* (PhD thesis, Stanford University).

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition*, *110*(2), 131–146.

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research*, *32*(1), 25–36.

Asperger, H. (1944). Die "Autistischen Psychopathen" im Kindesalter. *European Archives of Psychiatry and Clinical Neuroscience*, *117*(1), 76–136.

Asperger, H. (1968). Zur Differentialdiagnose des kindlichen Autismus. *Acta paedopsychiatrica*, *35*(4), 136–145.

Ayers, G. (1996). *Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments* (PhD dissertation, The Ohio State Unversity:).

Baltaxe, C. A. M. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology*, *2*(4), 176–180.

Baltaxe, C. A. M. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research*, *27*(1), 97–105.

Baltaxe, C. A. M. & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic, and autistic children. *Journal of Autism and Developmental Disorders*, *17*(2), 255–271.

Baltaxe, C. A. M. & Simmons, J. Q. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. B. Mesibov (Eds.), *Communication problems in autism* (pp. 95–125). Boston, MA: Springer US.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the Intelligibility of Referring Expressions in Dialogue. *Journal of Memory and Language*, *42*(1), 1–22.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5–17.

Baumann, S. (2006). *The Intonation of Givenness: Evidence from German*. Linguistische Arbeiten 508. Tübingen: Niemeyer.

Baumann, S. (2014). The Importance of Tonal Cues for Untrained Listeners in Judging Prominence. In *Proceedings of the 10th international seminar on speech production (issp)* (pp. 21–24). Cologne.

Baumann, S. & Grice, M. (2006). The Intonation of Accessibility. *Journal of Pragmatics*, *38*(10), 1636–1657.

Baumann, S. & Riester, A. (2012). Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 119–162). Interface Explorations 25. Berlin & New York: Mouton De Gruyter.

Baumann, S. & Röhr, C. T. (2015). The Perceptual Prominence of Pitch Accent Types in German. In T. S. C. f. I. 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)* (Vol. 298, pp. 1–5). Glasgow, UK: The University of Glasgow.

Baumann, S., Röhr, C. T., & Grice, M. (2015). Prosodische (De-)Kodierung des Informationsstatus im Deutschen. *Zeitschrift für Sprachwissenschaft*, *34*(1), 1–42.

Baumann, S. & Roth, A. (2014). Prominence and Coreference - On the Perceptual Relevance of F0 Movement, Duration and Intensity. In *Proceedings of the 7th International Conference on Speech Prosody* (pp. 227–231). Dublin, Ireland.

Beckman, M. E. (1996). The Parsing of Prosody. *Language and Cognitive Processes*, *11*, 17–67.

Beckman, M. E. & Pierrehumbert, J. B. (1986). Intonational Structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Berman, R. A. (2008). Developing Linguistic Knowledge and Language Use Across Adolescence. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (Chap. 17, pp. 347–367). Malden, MA: Blackwell Publishing Ltd.

Bertone, A., Mottron, L., Jelenic, P., & Faubert, J. (2005). Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. *Brain*, *128*(10), 2430–2441.

Birch, S. & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365–391.

Bishop, J. (2012). Information Structural Expectations in the Perception of Prosodic Prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning* (pp. 239–270). Berlin & New York: Mouton De Gruyter.

Bishop, J. (2017). Focus projection and prenuclear accents: evidence from lexical processing. *Language, Cognition and Neuroscience*, *32*(2), 236–253.

Bleuler, E. (1911). *Dementia Praecox oder Gruppe der Schizophrenien* (G. Aschaffenburg, Ed.). Handbuch der Psychiatrie Spez. Teil Abt, 1. Hälfte. Leipzig & Wien: Franz Deuticke.

Boersma, P. (2001). Praat a System for Doing Phonetics by Computer. *Glot International*, *5*(9/10), 341–345.

Bolinger, D. (1958). A Theory of Pitch Accent in English. *Word*, *63*, 109–149.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in human neuroscience*, *4*, 237.

Braun, B. (2005). *Production and perception of thematic contrast in German*. Contemporary studies in descriptive linguistics;9. Oxford: Lang.

Brennan, S. E. & Hanna, J. E. (2009). Partner-Specific Adaptation in Dialog. *Topics in Cognitive Science*, *1*(2), 274–291.

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring Theory of Mind in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *47*(7), 1927–1941.

Brooks, P. J. & Ploog, B. O. (2013). Attention to emotional tone of voice in speech perception in children with autism. *Research in Autism Spectrum Disorders*, *7*(7), 845–857.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, *58*(5), 412–424.

Campbell, N. & Mokhtari, P. (2003). Voice Quality: The 4th Prosodic Dimension. In *Proceedings of the 15th International Conference of Phonetic Sciences (ICPhS)*. Barcelona.

Cangemi, F. & D'Imperio, M. (2013). Tempo and the perception of sentence modality. *Laboratory Phonology*, *4*(1), 191–219.

Cangemi, F., Kalbertodt, J., Krüger, M., & Grice, M. (in prep.). The RefTop coding scheme (working title).

Capps, L., Losh, M., & Thurber, C. (2000). "The frog ate the bug and made his mouth sad": narrative competence in children with autism. *Journal of abnormal child psychology*, *28*(2), 193–204.

Cariello, A., Bigler, E. D., Tolley, S., Prigge, M. D., Neeley, E. S., Lange, N., ... Lainhart, J. E. (2011). A longitudinal look at expressive, receptive, and total language development in individuals with autism spectrum disorders. Paper presented at the International Meeting for Autism Research, San Sebastian, Spain. Retrieved from https://imfar.confex.com/imfar/2011/IMFAR_2011_Abstract.pdf

Chafe, W. (1976). Givenness, Contrastiveness, Definiteness, Subject, Topics, and Point of View. In C. Li (Ed.), *Subject and Topic* (pp. 25–56). New York: Academic Press.

Chafe, W. (1987). Cognitive Contstraints on Information Flow. In R. Tomlin (Ed.), *Coherence and Grounding in Discourse* (pp. 21–52). Amsterdam: John Benjamins Publishing Company.

Chafe, W. (1994). *Discourse, Consciousness, and Time*. Chicago: University of Chicago Press.

Chan, K. K. & To, C. K. (2016). Do Individuals with High-Functioning Autism Who Speak a Tone Language Show Intonation Deficits? *Journal of Autism and Developmental Disorders*, *46*(5), 1784–1792.

Chen, A. (2010). Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*, *120*(8), 1926–1939.

Chen, A. (2011a). The Developmental Path to Phonological Focus-Marking in Dutch. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 93–109). Dordrecht: Springer.

Chen, A. (2011b). Tuning information packaging: intonational realization of topic and focus in child Dutch. *Journal of Child Language*, *38*(05), 1055–1083.

Chen, A., Den Os, E., & De Ruiter, J. P. (2007). Pitch Accent Type Matters for Online Processing of Information Status: Evidence from Natural and Synthetic Speech. *The Linguistic Review*, *24*(2-3), 317–344.

Cheng, S. T. T., Lam, G. Y. H., & To, C. K. S. (2017). Pitch Perception in Tone Language-Speaking Adults With and Without Autism Spectrum Disorders. *i-Perception*, *8*(3).

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2009). From acoustics to grammar: Perceiving and interpreting grammatical prosody in adolescents with Asperger Syndrome. *Research in Autism Spectrum Disorders*, *3*(2), 502–516.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia*, *49*(3), 507–517.

Chomsky, N. & Halle, M. (1968). *The Sound Pattern of English*. New York: Harper and Row.

Christensen, D. L., Baio, J., Braun, K. V. N., Bilder, D., Charles, J., Constantino, J. N., ... Yeargin-Allsopp, M. (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. *MMWR. Surveillance Summaries*, *65*(3), 1–23.

Clark, H. H. & Marshall, C. R. (1981). Definite Reference and Mutual Knowledge. In A. K. Joshi, B. Webber, & I. Sag (Eds.), *Elements of Discourse Understanding* (pp. 10–63). Cambridge: Cambridge University Press.

Clark, H. H. & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology*, *9*(100), 287–299.

Colle, L., Baron-Cohen, S., Wheelwright, S., & Van Der Lely, H. K. J. (2008). Narrative discourse in adults with high-functioning autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, *38*(1), 28–40.

Cruttenden, A. (1985). Intonation comprehension in ten-year-olds. *Journal of Child Language*, *12*(3), 643–661.

Cruttenden, A. (1997). *Intonation* (2nd edition). Cambridge: Cambridge University Press.

Cruttenden, A. (2006). The De-accenting of Given Information: A Cognitive Universal? In G. Bernini & M. L. Schwartz (Eds.), *Pragmatic Organization of*

Discourse in the Language of Europe (pp. 311–355). The Hague: Mouton De Gruyter.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Dahan, D., Tanenhaus, M. K., & Chambersc, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314.

Dahlgren, S. & Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism*, *12*(4), 335–348.

Demouy, J., Plaza, M., Xavier, J., Ringeval, F., Chetouani, M., Périsse, D., ... Robel, L. (2011). Differential language markers of pathology in Autism, Pervasive Developmental Disorder Not Otherwise Specified and Specific Language Impairment. *Research in Autism Spectrum Disorders*, *5*(4), 1402–1412.

DeMyer, M. K., Barton, S., DeMyer, W. E., Norton, J. A., Allen, J., & Steele, R. (1973). Prognosis in autism: A follow-up study. *Journal of autism and childhood schizophrenia*, *3*(3), 199–246.

DePape, A.-M. R., Chen, A., Hall, G. B. C., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology*, *3*, 72.

DePriest, J., Glushko, A., Steinhauer, K., & Koelsch, S. (2017). Language and music phrase boundary processing in Autism Spectrum Disorder: An ERP study. *Scientific Reports*, *7*(1), 14465.

Dewey, M. A. & Everard, M. P. (1974). The near-normal autistic adolescent. *Journal of Autism and Childhood Schizophrenia*, *4*(4), 348–356.

Diehl, J. J., Bennetto, L., Watson, D., Gunlogson, C., & McDonough, J. (2008). Resolving ambiguity: A psycholinguistic approach to understanding prosody processing in high-functioning autism. *Brain and Language*, *106*(2), 144–152.

Diehl, J. J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics*, *30*(03), 385–404.

Diehl, J. J., Friedberg, C., Paul, R., & Snedeker, J. (2015). The use of prosody during syntactic processing in children and adolescents with autism spectrum disorders. *Development and Psychopathology*, *27*(3), 867–884.

Diehl, J. J. & Paul, R. (2013). Acoustic and perceptual measurements of prosody production on the profiling elements of prosodic systems in children by children with autism spectrum disorders. *Applied Psycholinguistics*, *34*(1), 135–161.

Dryer, M. S. (2013). Polar Questions. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from http://wals.info/chapter/116

Dunn, M. & Harris, L. (2016). *Prosody Intervention for High-Functioning Adolescents and Adults with Autism Spectrum Disorder*. London: Jessica Kingsley Publishers.

Edelson, L., Grossman, R., & Tager-Flusberg, H. (2007). Emotional prosody in children and adolescents with autism. Poster session presented at the annual international meeting for Autism Research, Seatle, WA. Retrieved from http://c.ymcdn.com/sites/www.autism-insar.org/resource/resmgr/docs/Meeting_Archives/IMFAR2007_Program.pdf

Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y. S., Kauchali, S., Marcin, C., ... Fombonne, E. (2012). Global prevalence of autism and other pervasive developmental disorders. *Autism Research*, *5*(3), 160–179.

Emmorey, K. D. (1987). The neurological substrates for prosodic aspects of speech. *Brain and Language*, *30*(2), 305–320.

Faraway, J. (2004). *Linear Models with R* (Second Ed.). CRC Press.

Feldstein, S., Konstantareas, M., Oxman, J., & Webster, C. D. (1982). The chronography of interactions with autistic speakers: An initial report. *Journal of Communication Disorders*, *15*(6), 451–460.

Féry, C. & Kügler, F. (2008). Pitch Accent Scaling on Given, New and Focused Constituents in German. *Journal of Phonetics*, *36*(4), 680–703.

Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, *44*(8), 1972–1981.

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The Use of Intonation to Communicate in Pervasive Developmental Disorders. *Journal of Child Psychology and Psychiatry*, *32*(5), 771–782.

Foley, W. A. (2006). Topic Chaining Constructions. In K. Brown (Ed.), *Encyclopedia of language & linguistics (second edition)* (pp. 773–777). Elsevier.

Fosnot, S. L. & Jun, S. (1999). Prosodic characteristics in children with stuttering or autism during reading and imitation. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 1925–1928).

Fossard, M., Achim, A. M., Rousier-Vercruyssen, L., Gonzalez, S., Bureau, A., & Champagne-Lavau, M. (2018). Referential Choices in a Collaborative Storytelling Task: Discourse Stages and Referential Complexity Matter. *Frontiers in Psychology*, *9*, 176. Retrieved from https://www.frontiersin.org/article/10.3389/fpsyg.2018.00176

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., ... Eng, C. (2012). Validation of Proposed DSM-5 Criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(1), 28–40.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.

Frith, U. (1989). *Autism: Explaining the enigma*. Oxford: Blackwell.

Frith, U. (1991). Asperger and his syndrome. In U. Frith (Ed.), *Autism and Asperger Syndrome* (pp. 1–36). Cambridge: Cambridge University Press.

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407.

Galati, A. & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language*, *62*(1), 35–51.

Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage: Clinical*, *6*, 370–378.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A., Morson, E. M., & Grace, E. J. (2015). Language development in autism. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 879–886). New York: Elsevier.

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M., & Couteur, A. L. (2001). Development and Current Functioning in Adolescents with Asperger Syndrome: A Comparative Study. *Journal of Child Psychology and Psychiatry*, *42*(2), 227–240.

Gilles, P. (2005). *Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterverweisung*. Berlin: Walter de Gruyter.

Givón, T. (1983). Introduction. In T. Givón (Ed.), *Topic continuity in discourse* (pp. 5–41). Amsterdam: John Benjamins.

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody recognition in adults with high-functioning autism spectrum disorders: From psychoacoustics to cognition. *Autism Research*, *8*(2), 153–163.

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The 'Reading the Mind in the Voice' test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106.

Goldsmith, J. A. (1976). An Overview of Autosegmental Phonology. *Linguistic Inquiry*, *2*, 23–68.

Green, H. & Tobin, Y. (2009). Prosodic analysis is difficult...but worth it: A study in high functioning autism. *International Journal of Speech-Language Pathology*, *11*(4), 308–315.

Grice, M. (2006). Intonation. In *Encyclopedia of Language & Linguistics* (2nd edition, Vol. 5, pp. 778–788). Oxford: Elsevier.

Grice, M. & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, *191*, 267–298.

Grice, M. & Baumann, S. (2007). An Introduction to Intonation - Functions and Models. In J. Trouvain & U. Gut (Eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice* (pp. 25–51). Trends in Linguistics. Studies and Monographs [TiLSM] 186. Berlin & New York: De Gruyter.

Grice, M. & Baumann, S. (2016). Intonation in der Lautsprache: Tonale Analyse. In U. Domahs & B. Primus (Eds.), *Handbuch Laut, Gebärde, Buchstabe* (pp. 84–105). De Gruyter.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 55–83). Oxford: Oxford University Press.

Grice, M., Baumann, S., & Jagdfeld, N. (2009). Tonal association and derived nuclear accents-The case of downstepping contours in German. *Lingua*, *119*(6), 881–905.

Grice, M., Baumann, S., Ritter, S., & Röhr, C. T. (n.d.). GToBI. Übungsmaterialien zur deutschen Intonation. Retrieved 2016, from www.gtobi.uni-koeln.de

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the Place of Phrase Accents in Intonational Phonology. *Phonology*, *17*(2), 143–185.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Consistency in Transcription and Labelling of German Intonation with GToBI. In *Proceedings of the 4th International Conference on Spoken Language Processing* (pp. 1716–1719). Philadelphia.

Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64*, 90–107.

Griffith, E., Pennington, B., Wehner, E., & Rogers, S. (1999). Executive functions in young children with autism. *Child Development*, *70*(4), 817–832.

Groen, W. B., Zwiers, M. P., van der Gaag, R.-J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: An integrative review. *Neuroscience & Biobehavioral Reviews*, *32*(8), 1416–1425. Special section: The European Workshop in Imagery and Cognition: Neurocognition and Visual Imagery.

Grossman, R. B., Bemis, R. H., Plesa Skwerer, D., & Tager-Flusberg, H. (2010). Lexical and Affective Prosody in Children With High-Functioning Autism. *Journal of Speech Language and Hearing Research*, *53*(3), 778–793.

Gundel, J. (2003). Information Structure and Referential Givenness/Newness: How Much Belongs in the Grammar? In S. Müller (Ed.), *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University* (pp. 122–142). Stanford, CA: CSLI Publications.

Gundel, J. K. (1985). 'Shared Knowledge' and Topicality. *Journal of Pragmatics*, *9*, 83–107.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive Status and the Form of Referring Expressions in Discourse. *Language*, *69*(2), 274–307.

Gussenhoven, C. (1983). Focus, Mode and The Nucleus. *Journal of Linguistics*, *19*(2), 377–417.

Gussenhoven, C. (2002). Phonology of Intonation. *Glot International*, *6*(9/10), 271–284.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. & Rietveld, A. C. M. (1988). Fundamental Frequency Declination in Dutch: Testing Three Hypotheses. *Journal of Phonetics*, *16*, 335–369.

Halliday, M. A. K. (1967). Notes on Transitivity and Theme in English: Part 2. *Journal of Linguistics*, *3*, 199–244.

Happé, F. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition*, *48*(2), 101–119.

Happé, F. (1999). Autism: Cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, *3*(6), 216–222.

Happé, F. (2015). Autism as a neurodevelopmental disorder of mind-reading. *Journal of the British Academy*, *3*, 197–209.

Happé, F. & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *36*(1), 5–25.

Heaton, P. (2003). Pitch memory, labelling and disembedding in autism. *Journal of Child Psychology and Psychiatry*, *44*(4), 543–551.

Heaton, P., Davis, R. E., & Happé, F. G. (2008). Research note: Exceptional absolute pitch perception for spoken words in an able adult with autism. *Neuropsychologia*, *46*(7), 2095–2098.

Heaton, P., Hudry, K., Ludlow, A., & Hill, E. (2008). Superior discrimination of speech pitch and its relationship to verbal ability in autism spectrum disorders. *Cognitive Neuropsychology*, *25*(6), 771–782.

Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., & van Santen, J. (2010). Autism and interactional aspects of dialogue. In *Proceedings of the 11th annual meeting of the special interest group on discourse and dialogue* (pp. 249–252). SIGDIAL '10. Tokyo, Japan: Association for Computational Linguistics.

Heikkinen, J., Jansson-Verkasalo, E., Toivanen, J., Suominen, K., Väyrynen, E., Moilanen, I., & Seppänen, T. (2010). Perception of basic emotions from speech prosody in adolescents with Asperger's syndrome. *Logopedics Phoniatrics Vocology*, *35*(3), 113–120.

Heilman, K. M., Bowers, D., Speedie, L., & Coslett, H. B. (1984). Comprehension of affective and nonaffective prosody. *Neurology*, *34*(7), 917–917.

Hendriks, P., Koster, C., & Hoeks, J. C. (2014). Referential choice across the lifespan: Why children and elderly adults produce ambiguous pronouns. *Language, Cognition and Neuroscience*, *29*(4), 391–407.

Hermelin, B. (2001). *Bright splinters of the mind : A personal story of research with autistic savants*. London: Jessica Kingsley.

Hesling, I., Dilharreguy, B., Peppé, S., Amirault, M., Bouvard, M., & Allard, M. (2010). The integration of prosodic speech in high functioning Autism: A preliminary fMRI study. *PLoS ONE*, *5*(7), e11571.

Hirschberg, J. (2015). Pragmatics and Prosody. In Y. Huang (Ed.), *Oxford Handbook of Pragmatics*. Published online. Oxford: Oxford University Press.

Howlin, P. (2003). Outcome in High-Functioning Adults with Autism with and Without Early Language Delays: Implications for the Differentiation Between Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *33*(1), 3–13.

Hubbard, D. J., Faso, D. J., Assmann, P. F., & Sasson, N. J. (2017). Production and perception of emotional prosody by adults with autism spectrum disorder. *Autism Research*. doi:10.1002/aur.1847

Hubbard, K. & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173.

Huddleston, R. (1994). The Contrast between Interrogatives and Questions. *Journal of Linguistics*, *30*(2), 411–439.

Hurley, R. & Bishop, J. (2016). Prosodic and individual influences on the interpretation of "only". In *Proceedings of the 8th international conference on speech prosody* (pp. 193–197). Boston. doi:10.21437/SpeechProsody.2016-40

Jagdfeld, N. & Baumann, S. (2011). Order Effects on the Perception of Relative Prominence. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* (pp. 958–961). Hong Kong.

Järvinen-Pasley, A., Pasley, J., & Heaton, P. (2008). Is the linguistic content of speech less salient than its perceptual features in autism? *Journal of Autism and Developmental Disorders*, *38*(2), 239–248.

Järvinen-Pasley, A., Peppé, S., King-Smith, G., & Heaton, P. (2008). The Relationship between Form and Function Level Receptive Prosodic Abilities in Autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1328–1340.

Järvinen-Pasley, A., Wallace, G. L., Ramus, F., Happé, F., & Heaton, P. (2008). Enhanced perceptual processing of speech in autism. *Developmental Science*, *11*(1), 109–121.

Kaland, C., Swerts, M., & Krahmer, E. (2013). Accounting for the listener: Comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *The Journal of the Acoustical Society of America*, *134*(3), 2182–2196.

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K.-A. B. (2016). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*. doi:10.1177/1362361316668652

Kamp-Becker, I., Smidt, J., Ghahreman, M., Heinzel-Gutenbrunner, M., Becker, K., & Remschmidt, H. (2010). Categorical and Dimensional Structure of Autism Spectrum Disorders: The Nosologic Validity of Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *40*(8), 921–929.

Kanner, L. (1943). Autistic disturbances of affective contact.

Kargas, N., López, B., Morris, P., & Reddy, V. (2016). Relations Among Detection of Syllable Stress, Speech Abnormalities, and Communicative Ability in Adults With Autism Spectrum Disorders. *Journal of Speech, Language, and Hearing Research*, *59*(2), 206–215.

Kim, Y. S., Leventhal, B. L., Koh, Y.-J., Fombonne, E., Laska, E., Lim, E.-C., ... Grinker, R. R. (2011). Prevalence of Autism Spectrum Disorders in a Total Population Sample. *American Journal of Psychiatry*, *168*(9), 904–912.

Kjeelgaard, M. M. & Tager-Flusberg, H. (2013). The perception of the relationship between affective prosody and the emotional content in utterances in children with autism spectrum disorders. *Perspectives on Language Learning and Education*, *20*, 20–32.

Kleinman, J., Marciano, P. L., & Ault, R. L. (2001). Advanced Theory of Mind in High-Functioning Adults with Autism. *Journal of Autism and Developmental Disorders*, *31*(1), 29–36.

Knight, R.-A. (2008). The Shape of Nuclear Falls and Their Effect on the Perception of Pitch and Prominence: Peaks vs. Plateaux. *Language and Speech*, *51*(3), 223–244.

Kohler, K. (1991a). Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics. In *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)* (Vol. 25, pp. 115–185).

Kohler, K. J. (1991b). A Model of German Intonation. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, *25*, 295–360.

Krämer, K., Gawronski, A., & Vogeley, K. (2016). Zur Diagnostik und Behandlung von Autismus-Spektrum-Störungen im Erwachsenenalter. *Fortschritte der Neurologie · Psychiatrie*, *84*(09), 578–588.

Krifka, M. (2007). Basic Notions of Information Structure. In C. Féry, G. Fanselow, & M. Krifka (Eds.), *Working Papers of the SFB632, Interdisciplinary Stud-*

ies on Information Structure (ISIS) (Vol. 6, pp. 13–55). Potsdam: Universitätsverlag Potsdam.

Krifka, M. (2008). Basic Notions of Information Structure. *Acta Linguistica Hungarica*, *55*, 243–276.

Kuno, S. (1972). Functional Sentence Perspective: A Case Study from Japanese and English. *Linguistic Inquiry*, *3*, 269–320.

Kuzmanovic, B., Schilbach, L., Lehnhardt, F. G., Bente, G., & Vogeley, K. (2011). A matter of words: Impact of verbal and nonverbal information on impression formation in high-functioning autism. *Research in Autism Spectrum Disorders*, *5*(1), 604–613.

Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

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

Ladd, D. R. (2008). *Intonational Phonology* (2nd edition). Cambridge: Cambridge University Press.

Ladd, D. R. & Morton, R. (1997). The Perception of Intonational Emphasis: Continuous or Categorical? *Journal of Phonetics*, *25*, 313–342.

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review*, *18*(1), 135–140.

Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Landa, R. (2000). Social language use in Asperger syndrome and high-functioning Autism. In A. Klin, F. R. Volkmar, & S. Sparrow (Eds.), *Asperger syndrome* (pp. 125–158). New York: Guilford Press.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1976). Suprasegmental Features of Speech. In N. J. Lass (Ed.), *Contemporary Issues in Experimental Phonetics* (pp. 256–239). New York: Academic Press.

Leiner, D. J. (2014). SoSci Survey (Version 2.4.00-i) [Computer Software]. Retrieved from https://www.soscisurvey.de

Liberman, M. Y. (1975). *The Intonational System of English*. New York: Garland Publishing.

Liberman, M. Y. & Prince, A. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry*, *8*, 249–336.

Lindner, J. L. & Rosén, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Aspergerś syndrome. *Journal of Autism and Developmental Disorders*, *36*(6), 769–777.

Livingston, L. A. & Happé, F. (2017). Conceptualising compensation in neurodevelopmental disorders: Reflections from autism spectrum disorder. *Neuroscience and Biobehavioral Reviews*, *80*, 729–742.

Lord, C. (1996). Autism diagnostic observation schedule. Unpublished manuscript, University of Chicago.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D., ... Bernier, R. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, *69*(3), 306–313.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The Autism Diagnostic Schedule - Generic: A standard measures of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*(3), 205–223.

Loveland, K. A., Tunalia, B., Mcevoy, R. E., & Kelley, M. L. (1989). Referential communication and response adequacy in autism and Down's syndrome. *Applied Psycholinguistics*, *10*(03), 301–313.

Luks, T. L., Nusbaum, H. C., & Levy, J. (1998). Hemispheric involvement in the perception of syntactic prosody is dynamically dependent on task demands. *Brain and Language*, *65*(2), 313–332.

Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.

Lyons, M., Schoen Simmons, E., & Paul, R. (2014). Prosodic development in middle childhood and adolescence in high-functioning autism. *Autism Research*, *7*(2), 181–196.

Martin, I. & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of Autism and Developmental Disorders*, *34*(3), 311–328.

Martzoukou, M., Papadopoulou, D., & Kosmidis, M.-H. (2017). The Comprehension of Syntactic and Affective Prosody by Adults with Autism Spectrum Disorder Without Accompanying Cognitive Deficits. *Journal of Psycholinguistic Research*, *46*(6), 1573–1595.

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and Developmental Receptive Language Disorder - a Comparative Follow-up in Early Adult Life. I: Cognitive and Language Outcomes. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 547–559.

Mayer, J. L., Hannent, I., & Heaton, P. F. (2016). Mapping the Developmental Trajectory and Correlates of Enhanced Pitch Perception on Speech Processing in Adults with ASD. *Journal of Autism and Developmental Disorders*, *46*(5), 1562–1573.

Mayes, S. D., Calhoun, S. L., & Crites, D. L. (2001). Does DSM-IV Asperger's Disorder Exist? *Journal of Abnormal Child Psychology*, *29*(3), 263–271.

McCaleb, P. & Prizant, B. M. (1985). Encoding of new versus old information by autistic children. *Journal of Speech and Hearing Disorders*, *50*(3), 230–240.

McCann, J. & Peppé, S. (2003). Prosody in autism spectrum disorders: a critical review. *International Journal of Language & Communication Disorders*, *38*(4), 325–350.

McCann, J., Peppé, S., Gibbon, F. E., O'Hare, A., & Rutherford, M. (2007). Prosody and its relationship to language in school-aged children with high-functioning autism. *International Journal of Language and Communication Disorders*, *42*(6), 682–702.

Miller, J. N. & Ozonoff, S. (2000). The External Validity of Asperger Disorder: Lack of Evidence from the Domain of Neuropsychology. *Journal of Abnormal Psychology*, *109*(2), 227–238.

Mitchell, P., Russell, H., & Saltmarsh, R. (1997). Overly literal interpretations of speech in autism: Understanding that messages arise from minds. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*(6), 685–690.

Molnar-Szakacs, I. & Heaton, P. (2012). Music: A unique window into the world of autism. *Annals of the New York Academy of Sciences*, *1252*(1), 318–324.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., & Burack, J. (2006). Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders*, *36*(1), 27–43.

Nadig, A. & Shaw, H. (2012). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders*, *42*(4), 499–511.

Nadig, A. & Shaw, H. (2015). Acoustic marking of prominence: How do preadolescent speakers with and without high-functioning autism mark contrast in an interactive task? *Language and Cognitive Processes*, *30*.

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research*, *2*(6), 334–347.

Niebuhr, O. (2009). F0-Based Rhythm Effects on the Perception of Local Syllable Prominence. *Phonetica*, *66*, 95–112.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults* (3rd edition). Austin, TX: Pro-Ed.

Nooteboom, S. (1997). The prosody of speech: melody and rhythm. *The Handbook of Phonetic Sciences*, 640–673.

Nooteboom, S. & Terken, J. M. B. (1982). What makes speakers omit pitch accents? An experiment. *Phonetica*, *39*, 317–336.

O'Connor, J. D. & Arnold, G. F. (1973). *Intonation of Colloquial English*. London: Longman.

O'Connor, K. (2007). Brief Report: Impaired Identification of Discrepancies Between Expressive Faces and Voices in Adults with Asperger's Syndrome. *Journal of Autism and Developmental Disorders*, *37*(10), 2008–2013.

O'Connor, K. (2012). Auditory processing in autism spectrum disorder: A review. *Neuroscience and Biobehavioral Reviews*, *36*(2), 836–854.

Ohala, J. J. (1983). Cross-Language Use of Pitch: An Ethological View. *Phonetica*, *40*, 1–18.

Ohala, J. J. (1984). An Ethological Perspective on Common Cross-Language Utilization of F0 of Voice. *Phonetica*, *41*, 1–16.

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and Production of Prosody by Speakers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–20.

Paul, R., Bianchi, N., Augustyn, A., Klin, A., & Volkmar, F. R. (2008). Production of syllable stress in speakers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *2*(1), 110–124.

Paul, R. & Cohen, D. J. (1984). Responses to Contingent Queries in Adults with Mental Retardation and Pervasive Developmental Disorders. *Applied Psycholinguistics*, *5*(4), 349–357.

Peppé, S. & McCann, J. (2003). Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clinical Linguistics and Phonetics*, *17*(4-5), 345–354.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and Expressive Prosodic Ability in Children With High-Functioning Autism. *Journal of Speech Language and Hearing Research*, *50*(4), 1015–1028.

Pierrehumbert, J. B. (1980). *The Phonetics and Phonology of English Intonation* (PhD thesis, MIT, Bloomington).

Pierrehumbert, J. B. & Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J. B. & Hirschberg, J. (1990). The Meaning of Intonational Contours in the Interpretation of Discourse. In P. R. Cohen, J. Morgan, & M. E. Pollak (Eds.), *Intentions in Communication* (pp. 271–311). Cambridge: MIT Press.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Prince, E. F. (1981). Toward a Taxonomy of Given-New Information. In P. Cole (Ed.), *Radical Pragmatics* (1981, pp. 223–256). New York: Academic Press.

Rapin, I. (1991). Autistic children: diagnosis and clinical features. *Pediatrics*, *87*(5 Pt 2), 751–760.

Rathcke, T. & Harrington, J. (2010). The variability of early accent peaks in Standard German. In C. Fougeron (Ed.), *Laboratory phonology 10* (pp. 533–556). De Gruyter Mouton.

Reinhart, T. (1981). Pragmatics and Linguistics: An Analysis of Sentence Topics. *Philosophica*, *27*(1), 53–94.

Reyelt, M., Grice, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Prosodische Etikettierung des Deutschen mit ToBI. In D. Gibbon (Ed.), *Natural Language and Speech Technology, Results of the Third KONVENS Conference* (pp. 144–155). Berlin & New York: Mouton De Gruyter.

Ricks, D. (1975). Vocal communication in pre-verbal normal and autistic children. In N. O'Connor (Ed.), *Language, cognitive deficits, and retardation* (pp. 245–268). London: Butterworths.

Riester, A. & Baumann, S. (2017). *The RefLex scheme - annotation guidelines*. University of Stuttgart. Retrieved from https://elib.uni-stuttgart.de/handle/11682/9028

Roach, P. (1991). *English phonetics and phonology* (2nd edition). Cambridge: Cambridge University Press.

Roberts, R. & Patterson, C. J. (1983). Perspective taking and referential communication: The question of correspondence reconsidered. *Child Development*, *54*, 1005–1014.

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with Autism Spectrum Disorders. *Brain and Cognition*, *71*(3), 362–368.

Röhr, C. T. (2013). Information Status and Prosody: Production and Perception in German. In *Working Papers of the SFB632, Interdisciplinary Studies*

on Information Structure (ISIS) (Vol. 17, pp. 119–141). Potsdam: Universitätsverlag Potsdam.

Röhr, C. T. (2016). *The Information Status of Nominal and Verbal Expressions: Intonational Evidence from Production and Perception in German*. Cologne: University of Cologne. Retrieved from https://kups.ub.uni-koeln.de/9146/

Röhr, C. T. & Baumann, S. (2010). Prosodic Marking of Information Status in German. In *Proceedings of the 5th International Conference on Speech Prosody* (Vol. 100019, pp. 1–4). Chicago.

Röhr, C. T. & Baumann, S. (2011). Decoding Information Status by Type and Position of Accent in German. In *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS)* (pp. 1706–1709). Hong Kong.

Rosenblau, G., Kliemann, D., Dziobek, I., & Heekeren, H. R. (2017). Emotional prosody processing in Autism Spectrum Disorder. *Social Cognitive and Affective Neuroscience*, 224–239.

Rothstein, J. (2013). *Prosody treatment program*. Austin, TX: Pro-Ed.

Rutherford, M., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the Mind in the Voice: A Study with Normal Adults and Adults with Asperger Syndrome and High Functioning Autism. *Journal of Autism and Developmental Disorders*, *32*(3), 189–194.

Sanders, J. L. (2009). Qualitative or Quantitative Differences Between Asperger's Disorder and Autism? Historical Considerations. *Journal of Autism and Developmental Disorders*, *39*(11), 1560.

Satzger, W., Fessmann, H., & Engel, R. R. (2002). Liefern HAWIE-R, WST und MWT-B vergleichbare IQ-Werte? *Zeitschrift für Differentielle und Diagnostische Psychologie*, *23*(2), 159–170.

Schmidt, K.-H. & Metzler, P. (1992). *Wortschatztest: Wst*. Weinheim: Beltz.

Schneider, D., Slaughter, V. P., Bayliss, A. P., & Dux, P. E. (2013). A temporally sustained implicit theory of mind deficit in autism spectrum disorders. *Cognition*, *129*(2), 410–417.

Schumacher, P. B. & Baumann, S. (2010). Pitch Accent Type Affects the N400 During Referential Processing. *NeuroReport*, *21*(9), 618–622.

Schwarzschild, R. (1999). Givenness, AvoidF and Other Constraints on the Placement of Accent. *Natural Language Semantics*, *7*(2), 141–177.

Segal, O., Kaplan, D., Patael, S., & Kishon-Rabin, L. (2017). Comprehension of Narrow Focus by Adolescents in the Autism Spectrum. *Folia Phoniatrica et Logopaedica*, *69*(1-2), 67–77.

Selkirk, E. O. (1984). *Phonology and Syntax. The Relation between Sound and Structure.* Cambridge, MA: MIT Press.

Senju, A. (2013). Atypical development of spontaneous social cognition in autism spectrum disorders. *Brain and Development*, *35*(2), 96–101.

Sharda, M., Subhadra, T. P., Sahay, S., Nagaraja, C., Singh, L., Mishra, R., ... Singh, N. C. (2010). Sounds of melody - Pitch patterns of speech in autism. *Neuroscience Letters*, *478*(1), 42–45.

Shattuck-Hufnagel, S. & Turk, A. E. (1996). A Prosody Tutorial for Investigators of Auditory Sentence Processing. *Journal of Psycholinguistic Research*, *25*, 193–246.

Shport, I. A. & Redford, M. A. (2014). Lexical and phrasal prominence patterns in school-aged childrenś speech. *Journal of Child Language*, *41*(04), 890–912.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and Prosody Characteristics of Adolescents and Adults With High-Functioning Autism and Asperger Syndrome. *Journal of Speech, Language, and Hearing Research*, *4401*(October), 1097–1115.

Silverman, K. (1987). *The Structure and Processing of Fundamental Frequency Contours* (PhD thesis, University of Cambridge:).

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C. C., Price, P., ... Hirschberg, J. (1992). TOBI: A Standard for Labeling English Prosody. *Second International Conference on Spoken Language Processing*, 867–870.

Simmons, J. Q. & Baltaxe, C. (1975). Language patterns of adolescent autistics. *Journal of autism and childhood schizophrenia*, *5*(4), 333–351.

Singh, L. & Harrow, M. S. (2014). Influences of Semantic and Prosodic Cues on Word Repetition and Categorization in Autism. *Journal of Speech Language and Hearing Research*, *57*(5), 1764–1778.

Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosodic phrasing. *Laboratory Phonology*, *2*(1), 35–98.

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in autism spectrum conditions from voices and faces. *Autism*, *17*(1), 6–14.

't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach.* Cambridge: Cambridge University Press.

Tager-Flusberg, H. (2000). Language and understanding minds: Connections in autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. H. Cohen (Eds.), *Un-*

derstanding other minds: Perspectives from developmental cognitive neuroscience (Vol. 2, pp. 124–149). Oxford University Press.

Tager-Flusberg, H. (1996). Brief report: Current theory and research on language and communication in autism. *Journal of Autism and Developmental Disorders*, *26*(2).

Terken, J. & Hermes, D. (2000). The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and Experiment: Studies Presented to Gösta Bruce* (pp. 89–127). Dordrecht: Springer Netherlands.

Terken, J. & Hirschberg, J. (1994). Deaccentuation of Words Representing 'Given' Information: Effects of Persistence of Grammatical Function and Surface Position. *Language and Speech*, *37*(2), 125–145.

Thurber, C. & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders*, *23*(2), 309–322.

Trim, J. L. M. (1959). Major and minor tone-groups in English. *Le Maître Phonètique*, *112*, 26–29.

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

Vallduví, E. & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics*, *34*(3), 459–520.

van Krevelen, D. A. (1971). Early infantile autism and autistic psychopathy. *Journal of autism and childhood schizophrenia*, *1*(1), 82–86.

Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, *5*(2-3), 207–226.

Van Santen, J. P. H., Prud'hommeaux, E. T., Black, L. M., & Mitchell, M. (2010). Computational prosodic markers for autism. *Autism*, *14*(3), 215–236.

Vogel, I. & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of child language*, *29*(2), 225–50.

Volden, J., Mulcahy, R. F., & Holdgrafer, G. (1997). Pragmatic language disorder and perspective taking in autistic speakers. *Applied Psycholinguistics*, *18*(02), 181–198.

Warren, P. (2016). *Uptalk: The Phenomenon of Rising Intonation*. Cambridge: Cambridge University Press.

Weinreich, U. (1954). Stress and Word Structure in Yiddish. In *The Field of Yiddish: Studies in Yiddish Language, Folklore and Literature* (pp. 1–27). Linguistic Circle of New York.

Wells, B., Peppé, S., & Goulandris, N. (2004). Intonation development from five to thirteen. *Journal of child language*, *31*(4), 749–78.

Williamson, K. (1979). Sentence tone in some southern Nigerian languages. In E. Fischer-Jorgensen, J. Rischel, & N. Thorsen (Eds.), *Proceedings of the ninth international congress of phonetic sciences* (pp. 424–430).

Wing, L. (1981). Aspergers syndrome: A clinical account. *Psychological Medicine*, *11*(1), 115–129.

Wing, L. (1991). The relationship between Asperger's syndrome and Kanner's autism. In U. Frith (Ed.), *Autism and asperger syndrome* (pp. 93–121). Cambridge: Cambridge University Press.

Wing, L. (2000). Past and future of research on Asperger syndrome. In A. Klin, F. R. Volkmar, & S. S. Sparrow (Eds.), *Asperger syndrome* (pp. 418–432). New York, London: The Guildford Press.

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

Appendix A

Additional Material: Perception Experiment

A.1 Participant Information

Table A.1: Participant information of perception experiment: Control group.

Subj.	Sex	Age	Grown up	Living currently	Occupation
29	w	62	NW, Germany	NW, Germany	Journalistin
37	m	26	NW, Germany	BE, Germany	Student
69	m	56	NW, Germany	NW, Germany	KFZ-Sachverstndiger
71	w	61	NW, Germany	NW, Germany	Hausfrau
75	m	30	NW, Germany	NW, Germany	Student
85	m	26	NW, Germany	NW, Germany	Student
92	m	29	BW, Germany	BW, Germany	IT Specialist
96	w	52	NW, Germany	NW, Germany	Hausfrau
98	w	54	NW, Germany	Vaals, Niederlande	Lehrerin
99	m	27	NW, Germany	NW, Germany	Doktorand Ing.Wiss.
134	w	37	HB, Germany	HE, Germany	Projektmanagerin
138	m	31	NW, Germany	BE, Germany	Projektl. Produktmanagem.
148	m	25	BB, Germany	Saratov, Russland	Betriebswirt Logistik
158	m	29	SN, Germany	BE, Germany	Student
161	w	51	MV, Germany	BE, Germany	Angestellte
181	w	59	NW, Germany	NW, Germany	Vertriebsassistentin
186	m	44	RP, Germany	NW, Germany	Vertrieb Maschinenbau
189	m	29	BE, Germany	BE, Germany	Musikbusiness, angestellt
195	m	62	NW, Germany	BW, Germany	Akustiker
209	w	57	NW, Germany	NW, Germany	Rentnerin
211	w	21	NW, Germany	NW, Germany	Student
217	m	21	HE, Germany	NW, Germany	Student
220	m	21	NW, Germany	NW, Germany	Student
226	m	28	BW, Germany	NW, Germany	Student
231	m	29	NW, Germany	NW, Germany	Wiss. Mit.
234	w	30	HE, Germany	NW, Germany	Sprachtherapeutin
238	w	65	NW, Germany	NW, Germany	Sonderschullehrerin
242	w	56	NI, Germany	NW, Germany	Krankenschwester
248	w	59	SH, Germany	SH, Germany	k.A.
251	m	29	TH, Germany	BW, Germany	IT-Spezialist
253	m	31	NW, Germany	BE, Germany	Referent Bundesministerium
254	m	30	SH, Germany	BW, Germany	Software Consultant
256	m	29	NW, Germany	NW, Germany	Vertriebsassistent
258	m	32	NW, Germany	NW, Germany	Student
261	m	30	RP, Germany	HH, Germany	Unternehmensberater (IT)
262	m	31	NW, Germany	NW, Germany	Student
265	m	31	RP, Germany	RP, Germany	Ingenieur

Table A.2: Participant information of perception experiment: ASD group.

Subj.	Sex	Age	Grown up	Living currently	Occupation
55	w	51	NW, Germany	NW, Germany	Buchhalterin
56	w	49	HE, Germany	NW, Germany	PTA
62	m	44	NW, Germany	NW, Germany	Angestellter
65	m	36	NW, Germany	NW, Germany	k.A.
75	w	49	NW, Germany	NW, Germany	Projektingenieurin
80	m	53	NW, Germany	NW, Germany	Angestellter
86	m	43	NI, Germany	NI, Germany	Ingenieur
87	w	32	NW, Germany	NW, Germany	Hausfrau / Mutter
89	m	52	NI, Germany	NI, Germany	Verw.fachangestellter
91	m	24	NW, Germany	NW, Germany	k.A.
92	w	44	NW, Germany	NW, Germany	Schriftsetzerin
99	w	29	NW, Germany	NW, Germany	Studierende
103	m	28	NI, Germany	NI, Germany	Kchenhilfe
105	w	23	NW, Germany	NW, Germany	Studentin
107	m	49	NW, Germany	NW, Germany	Informatiker
108	m	44	RP, Germany	NW, Germany	Arbeitsvermittler
109	m	41	NW, Germany	NW, Germany	Manager
113	m	53	NW, Germany	NW, Germany	k.A.
116	m	50	ST, Germany	NW, Germany	Fachinformatiker SI
117	m	41	NW, Germany	NW, Germany	k.A.
119	m	52	NW, Germany	NW, Germany	k.A.
120	w	45	ST, Germany	RP, Germany	Sachbearbeiterin
121	m	27	NW, Germany	NW, Germany	Chemisch-technischer Ass.
122	m	51	HH, Germany	NW, Germany	SW-Entwickler
124	w	55	NW, Germany	NW, Germany	Verwaltungsangestellte
125	m	46	NW, Germany	NW, Germany	Lagerist
129	w	31	NW, Germany	NW, Germany	k.A.
132	w	48	NW, Germany	NW, Germany	Frhrentnerin
135	m	36	NW, Germany	NW, Germany	z.Z. arbeitslos
136	w	25	TH, Germany	NW, Germany	k.A.
138	w	43	NW, Germany	NW, Germany	archologische Zeichnerin
140	m	54	NW, Germany	NW, Germany	Dipl. Ing. Maschinenbau
145	m	48	NW, Germany	NW, Germany	IT-Systemadministrator
151	w	32	RP, Germany	NW, Germany	Studentin
153	m	54	NW, Germany	NW, Germany	IT-Branche
154	m	40	RP, Germany	RP, Germany	Softwareentwicklung
157	m	49	NW, Germany	HE, Germany	Fhrungskrftetrainer

A.2 Test Material

Table A.3: Target words and information about their corresponding log word frequency values according to the SUBTLEX corpus (Brysbaert et al., 2011) and their animacy (y = yes; n = no).

target word	**log frequency**	**animacy**
(Dr.) Bahber	0	y
Romana	0.6	y
Janina	0.78	y
(Dr.) Bieber	0.85	y
Rosine	0.9	n
Lawine	1.6	n
Ballade	1.77	n
Banane	2.08	n
Nina	2.68	y
Dame	3.25	y

Table A.4: Overview of speakers ('F' indicates female speakers, 'M' indicates male speakers) and target words in the selection of target sentences for the perception study.

speaker–word	**given**	**displaced**	**bridging**	**new**
H*	F05–*Dame*	F07–*Nina*	M02–*Nina*	M03–*Rosine*
!H*	M01–*Romana*	F02–*Nina*	F07–*Bahber*	M02–*Dame*
H+!H*	F05–*Janina*	F05–*Lawine*	M02–*Banane*	F06–*Rosine*
H+L*	F05–*Ballade*	M03–*Romana*	M01–*Lawine*	F04–*Dame*
L*	M02–*Romana*	F01–*Bahber*	F01–*Bahber*	F01–*Romana*
L*(PN)	F03–*Romana*	F03–*Ballade*	M03–*Ballade*	F01–*Dame*
Ø	F03–*Banane*	F02–*Rosine*	F02–*Lawine*	F01–*Bieber*

Appendix B

Additional Material: Production Experiment

B.1 Participant Information

Table B.1: Participant information of production experiment: Control group. The coding for subject includes gender information (F for female and M for male).

Subj.	Age	Grown up	Living currently	AQ	IQ (WST)
F02	27	Koeln, NW	Koeln, NW	22	122
F03	22	Loerrach, BW	Koeln, NW	10	114
F04	26	Brachbach, RP	Koeln, NW	12	107
F06	21	Moers, NW	Koeln, NW	10	99
F07	21	Muelheim/R., NW	Koeln, NW	20	110
M01	21	Unteriflingen, BW	Koeln, NW	12	110
F08	20	Koeln, NW	Koeln, NW	33	110
F11	19	Berlin	Koeln, NW	17	122
F13	19	Schoenecken, NW	Koeln, NW	11	99
M03	24	Gummersbach, NW	Gummersbach, NW	22	122
F14	26	Essen, NW	Koeln, NW	13	114
F15	20	Bruehl, NW	Koeln, NW	11	122
M05	22	Kerpen, NW	M.gladbach, NW	11	110
F16	25	Frechen, NW	Koeln, NW	24	110
F17	22	Burscheid, NW	Koeln, NW	14	101
M06	19	Kerpen, NW	Kerpen, NW	17	107

Table B.2: Participant information of production experiment: ASD group. The coding for subject includes gender information (F for female and M for male).

Subj.	Age	Grown up	Living currently	AQ	IQ (WST)
F01	25	Koeln, NW	Koeln, NW	39	95
M01	44	Emden, NI	Leer, NI	36	122
M02	46	M.gladbach, NW	M.gladbach, NW	40	139
M03	38	Greven, NW	Bonn, NW	45	122
F02	46	Hennef, NW	Hennef, NW	45	110
F03	46	Herford, NW	Bad Salzuflen, NW	45	118
M04	39	Koeln, NW	Koeln, NW	39	107
M05	51	Dortmund, NW	Schwerte, NW	41	118
M06	45	Koblenz, RP	Lahnstein, RP	46	107
F04	46	Koeln, NW	Eitorf, NW	45	101
M07	44	B. Gladbach, NW	B. Gladbach, NW	43	125
M08	39	Kruemmel, RP	Kruemmel, RP	42	143
M09	55	Duisburg, NW	Duisburg, NW	40	110
M10	51	Wuppertal, NW	Taunusstein, HE	40	125
M11	31	Bremen, HB	Bremen, HB	35	114
F05	34	Linz am Rhein, RP	Koeln, NW	41	114