Prosodic Phrasing and Syllable Prominence in Spoken Prose.

A Validated Coding Manual¹

ISABELLE FRANZ*

Max Planck Institute for Empirical Aesthetics Frankfurt

Goethe University Frankfurt

CHRISTINE A. KNOOP*

Max Planck Institute for Empirical Aesthetics Frankfurt

GERRIT KENTNER

Max Planck Institute for Empirical Aesthetics Frankfurt

Goethe University Frankfurt

SASCHA ROTHBART

Max Planck Institute for Empirical Aesthetics Frankfurt

VANESSA KEGEL

Max Planck Institute for Empirical Aesthetics Frankfurt

JULIA VASILIEVA

Max Planck Institute for Empirical Aesthetics Frankfurt

SANJA METHNER

Max Planck Institute for Empirical Aesthetics Frankfurt

MATHIAS SCHARINGER

Philipps-University Marburg

WINFRIED MENNINGHAUS

Max Planck Institute for Empirical Aesthetics Frankfurt

*Isabelle Franz and Christine A. Knoop contributed equally to this article and the underlying research and hence share first authorship.

¹ The authors would like to thank Marissa Gemma and Karen Henrich for helpful input in the early stages of the project, Muralikrishnan, Wolff Schlotz, Cornelius Abel, and Dominik Thiele, Antonia Walsch for invaluable technical and intellectual advice and support regarding the collection and processing of the audio data, Thomas Haider for POS tagging, and Julia Merrill and Kerstin Kipp for initiating contact with our professional speakers from the Hochschule für Musik und Darstellende Kunst, Stuttgart. Not least, we would like to thank the speakers themselves, Irene Baumann, Chantal Busse, Dominik Eisele, Elias Hartung, Paula Scheschonka, Ramon Schmid, Jonathan Springer, Patrick Suhm, Lina Syren, and Dorothee Wolfsberger, for providing us with audiobook-quality recordings.

Current systems for predicting prosodic prominence and boundaries in texts focus on syntax/semantic-based automatic decoding of sentences that need to be annotated syntactically (Atterer & Klein 2002; Windmann et al. 2011). However, to date, there is no phonetically validated replicable system for manually coding prosodic boundaries and syllable prominence in longer sentences or texts. Based on work in the fields of metrical phonology (Lieberman & Prince 1977), phrase formation (Hayes 1989) and existing pause coding systems (Gee and Grosjean 1983), we developed a manual for coding prosodic boundaries (with 6 degrees of juncture) and syllable prominence (8 degrees).

Three independent annotators applied the coding system to the beginning pages of four German novels and to four short stories (20 058 syllables, Fleiss kappa .82). For the phonetic validation, eight professional speakers read the excerpts of the novels aloud. We annotated the speech signal automatically with MAUS (Schiel 1999). Using PRAAT (Boersma & Weenink 2019), we extracted pitch, duration, and intensity for each syllable, as well as several phonetic parameters for pauses, and compared all measures obtained to the theoretically predicted levels of syllable prominence and prosodic boundary strength. The validation with the speech signal shows that our annotation system reliably predicts syllable prominence and prosodic boundaries. Since our annotation works with plain text, there are many potential applications of the coding system, covering research on prose rhythm, synthetic speech and (psycho)linguistic research on prosody.

KEYWORDS: prosodic annotation, prose rhythm, reading, text annotation, pauses, syllable prominence.

1. INTRODUCTION

It is a matter of conventional wisdom that prose texts are marked by rhythmical properties, i.e., patterns in prosodic prominence and prosodic boundaries, which may differ between writers, speakers, genres, as well as passages within the same text (e.g., Blohm et al., in press; Fant et al. 1991; Nord et al. 1990; Roberts 1996). This is especially true of literary prose, where rhythmical qualities contribute to the specific ‘sound’ of authors or individual texts, and hence support the emergent impression of a unique prosodic signature. A variety of systematic efforts from linguistics, classical studies, rhetoric, and literary studies have identified local rhythmical structures at the level of the phrase or clause that are, at least partly, realized by facultative selection and combination of lexical elements or syntactic constructions (e.g. Antilla 2016; Breiss & Hayes 2020; Franz et al. 2021; Lee & Gibbons 2007). To this day, however, there is no one convincing system that reliably captures rhythmical qualities in longer, more complex phrases or sentences.

In this article, we introduce a coding system that is capable of capturing rhythmical qualities of any kind of prose, from basic single sentences to complex literary texts. The system combines and extends a variety of extant approaches. We build on the fundamental assumption that prose rhythm arises from the concatenation of syllables with various graded levels of prominence, and from their grouping in nested prosodic constituents, and that these two aspects covary systematically. Guided by this assumption, we developed a phonetically validated, replicable system for the manual annotation of a) graded syllable prominence and b) the different levels of breaks marking prosodic boundaries that can be applied to any type of prose text. We chose to use a metrical grid, i.e., a system that represents multiple graded levels of syllable prominence (see, e.g.

Liberman & Prince 1977; Halle & Vergnaud 1987) and combined it with a detailed coding system for prosodic boundaries (see, e.g. Atterer 2002; Gee & Grosjean 1993).

In order to facilitate usability, we disregarded certain insights concerning the potentially complex recursive representation of prosodic structure (e.g. Ladd 1988; Wagner 2005; Kentner & Féry 2013) as well as discourse-semantic effects. We consider this an advantage rather than a disadvantage: First, annotations based on semantic information rely on the individual coder's interpretation and may therefore not be generalizable (see, for instance, early attempts by Marbe 1904; Kagarov 1928, and Lipsky 1907, who did not differentiate between semantic and other linguistic information). Second, concepts such as semantic focus or individual interpretation, while undoubtedly affecting prosodic prominence, can be considered part of the specialist knowledge that we did not want to posit as a coding requirement (for a similarly discourse-neutral annotation see Bachenko & Fitzpatrick 1990).

As quality criteria, we strove for replicability, broad and simple usability, accuracy, and relative brevity. While our system is deeply rooted in previous linguistic approaches, we sought to avoid any highly theoretical categories or vague definitions in the coding instructions themselves to make sure that they could be implemented by coders without specialist training or even a linguistic background. Our focus on manual over computational annotation is owed to the complexity of the texts for which this coding system is intended: We stipulated that a coding manual developed on the basis of (1) theoretical linguistic foundations, (2) the intuition of our first coders and (3) a phonetic validation as a control instance would make for the soundest system. Under this approach, a computational version is a future, additional step. Moreover, manual annotation of

syllable prominence and prosodic boundaries allows for broad usability since the texts do not need to be syntactically annotated (not even POS tagged) in advance.

The first part of this article offers a detailed overview of the linguistic evidence underlying our predictions. In support of our coding system, we then present the results of an empirical validation which has been performed on recordings of prose texts read aloud by professional speakers. In line with prior research, we used the measures of pitch, intensity and duration to capture prominence (see, e.g. Grice & Kügler 2021; Wagner et al. 2015), and the measures of silence in the phonetic signal and final lengthening (Petrone et al. 2017; Ferreira 1993; Zellner 1994) to capture phrase boundaries. In the appendix (I), we present our step-by-step manual of coding rules that can be readily implemented by coders with or without expert knowledge in linguistics.

1.1 *Materials*

We used a sample of eight literary texts of comparable length as sample texts for the development and the validation of our coding system (see Table 1). The first four are contemporary short stories that were originally chosen for a reading study; the others are the opening pages of famous German novels and novellas.

Author	Title	Text type	Number of words	Number of syllables	Year	Recordings
Brigitte Kronauer	<i>Das Ehepaar Dortwang</i> (<i>Mr and Mrs Dortwang</i>)	short story	1032	1942	1997	8
Peter Bichsel	<i>Der Mann mit dem Gedächtnis</i> (<i>The Man with the Memory</i>)	short story	1054	1654	1969	8
Christoph Ransmayr	<i>Blut</i> (<i>Blood</i>)	short story	1003	1836	2012	8
Martin Walser	<i>Selbstporträt als Kriminalroman</i> (<i>Self portrait in form of a crime novel</i>)	short story	2000	3262	1978	8
Johann Wolfgang Goethe	<i>Die Wahlverwandtschaften</i> (<i>Elective affinities</i>)	novel, extract	1544	2672	1809	8
Theodor Fontane	<i>Effi Briest</i> (<i>Effi Briest</i>)	novel, extract	1503	2599	1896	8
Heinrich von Kleist	<i>Die Marquise von O.</i> (<i>The Marquise of O.</i>)	novel, extract	1727	3080	1808	8
Franz Kafka	<i>Der Prozess</i> (<i>The Trial</i>)	novel, extract	1816	3013	1925	8

Table 1. Overview of the texts on which the system was developed.

1.2 Theoretical framework

1.2.1 Nested prosodic phrase structures

Syllables are not simply concatenated in a linear fashion. Instead, in line with the theory of prosodic phonology (Nespor & Vogel 2007), we assume that the prosodic representation of every sentence or utterance is hierarchically structured such that small prosodic domains are nested within larger ones: syllables are embedded in prosodic feet, which in turn constitute prosodic words; prosodic words combine to form prosodic phrases which in turn form intonational phrases either by themselves or with neighboring phrases (Selkirk 1984). In the prosodic hierarchy, feet typically consist of at least one stressed syllable, prosodic words of at least one lexical word, and phonological phrases of at least one syntactic head. The boundaries of intonational phrases usually match those

of syntactic phrases (see, e.g. Nespor & Vogel 1986; Selkirk 2011). This, of course, is a very simplified description which we will flesh out in more detail later.

Following this logic of prosodic grouping and its phonetic implementation in speech (see Krivokapic 2015, for a review), boundaries between higher-level units like intonational phrases (which roughly correspond to clauses, see, e.g., Selkirk 2011) are more noticeable in the phonetic signal than boundaries between lower-level units like phonological phrases or words (Féry & Schubö 2010; Kentner & Féry 2013; Krivokapić 2007).

This hierarchical conception of phrasing or prosodic grouping serves as the basic assumption for our annotation of prosodic boundaries. However, it does not account for the different levels of prosodic prominence that the syllables exhibit. For this, we drew on metrical grids.

1.2.2 *Metrical grids*

Sentential prominence is a matter of degree, i.e. some syllables are weaker and others stronger, with various gradations of prominence in between (Anttila et al. 2020). One established way of representing multiple graded levels of syllable prominence is the metrical grid (see, e.g., Liberman & Prince 1977; Halle & Vergnaud 1987), which represents a string of syllables as a sequence of cumulated beats¹ (see Figure 1). Usually, every beat assigned to a syllable corresponds to the head of some prosodic domain (syllable, foot, phonological word, phonological phrase etc., see Halle & Vergnaud 1987). As a result, syllables that serve as heads of phonological phrases accumulate more beats than those that only act as heads of a phonological word. In this way, metrical grids can account for a central principle in phonology, namely culminativity: each domain has an obligatory head which is more prominent than the rest of the domain (Selkirk 1984; Hayes

1989). As a consequence, the hierarchical structure leads to different strengths of prosodic boundaries and syllable prominence which in turn influence one another.

			x	Prosodic phrase
	x		x	Prosodic word
	x		x	Foot
x	x	x	x	Syllable
He	stole	The	vase	

Figure 1. Example of a metrical grid illustrating the gradual prominence structure in a neutral sentence: syllable prominence is captured by the height of the grid, i.e., the number of beats assigned to any one syllable.

For all their practical value, metrical grids were developed as abstract models for prosodic structure and are typically only applicable to relatively short linguistic units. By contrast, our ultimate goal was to capture rhythmical qualities in sentence structures of all degrees of complexity, including those of literary prose. Moreover, extant approaches that predict both syllable prominence and prosodic boundaries so far remain scarce (but see Ferreira 1993, for speech, and Lehrdahl 2001, for poetry). To capture both, we developed an extended grid structure and combined it with a coding system that predicts several degrees of prosodic boundaries.

However, under the assumption that prosodic groupings should be represented recursively (Wagner 2010), the annotation of graded prosodic boundaries in longer sentences – that is required for the prediction of pauses – is not trivial. Further, these difficulties extend to the annotation of syllable prominence: prosodic phrases as well as intonational phrases often coincide with syntactic units. However, restrictions on sentence and phrase length, speech tempo and discourse-semantic factors co-determine to what extent syntactic features actually govern prosodic parsing in speech (Selkirk 1984, 1995; Chomsky & Halle 1968; Nespor & Vogel 2007). For instance, according to Watson and Gibson (2004), prosodic boundary strength systematically depends on the

length of the units to the left and right of the boundary position. Additionally, there are limits to length and culminativity of phrases, which we will cover in more detail below.

While all coding systems for prosodic boundaries that we are aware of offer valuable information, we did not find a validated, replicable system for graded coding which a) works for prose texts with indeterminate sentence lengths, b) can be applied by non-specialists, and c) is combined with a coding system for syllable prominence.

1.3 Combining prosodic boundaries and syllable prominence

The notion that rhythmicity arises from prominence variations and from the distribution of pauses can be traced back to Cicero (trans. 2001), who suggested a close link between the rhythmicity of sentence- and passage-final cadences (the so-called *clausulae*) and the pauses (the so-called *caesurae*) that follow. This is intuitively convincing, but difficult to capture, most notably, because differences between individual prominences and breaks are graded rather than categorical, and the relation between the two is more intricate than Cicero and a long tradition of classical poetics were aware of. We found that the mutual influence of phrasing on prominence distribution and vice versa) proved to be the source of most of the unconvincing results in earlier annotation systems. For this reason, we developed a manual that allows us to reflect on, and in fact make use of, the systematic relation between the two in the annotation process itself.

1.3.1 Potential prominence targets – a new category for the process of the annotation

In accordance with Hayes (1989), Selkirk (1984), and Windmann et al. (2011, above others), we classify by part-of-speech the available lexical material that serves as head of a prosodic unit. This final marking of higher level prominences is implemented only *after* the annotation of prosodic boundaries; it accounts for the position of these prominences in the respective unit and the neighbouring parts-of-speech.

Additionally, our approach introduces a further markup phase *prior* to the annotation of prosodic boundaries, in which our annotators identify and label specific POS before the main part of the annotation. These marked POS play an important role in the process of phrasing and prominence assignment; we henceforth call them *potential prominence targets*. The category includes all nouns, proper names, adjectives, adverbs and full verbs (see Windmann et al. 2011), but also participles combined with auxiliary verbs (for evidence supporting this, see Röhr et al. 2020), as well as numerals (including quantifiers, see Widera et al. 1997) and verb particles without prominent neighbors (see Samlowski et al. 2012; for details, please consult the manual in the Appendix). Example (1) shows the potential phrasal accents in bold.

(1) Sie **gebe** ja **zu**, **sagte** die **Frau**, während ihr die **Tränen** über das **Gesicht** liefen, über

Mundwinkel und **Kinn** auf ihre **schwarze Bluse** **tropften** und **dort noch schwärzere**

Spuren hinterließen, sie **gebe** ja **zu**, dass ihr **Adi** ein **wilder junger Mensch** gewesen sei.

She admitted, the woman said, as tears ran down her face, dripping down the corners of her mouth and chin onto her black blouse, leaving even blacker marks there, she admitted that her Adi had been a wild young man.

We mark these potential prominence targets not just for assigning syllable prominence later, but also for assigning syllable prominence depending on neighbouring words. More importantly, the potential prominence targets serve the process of phrasing itself, as we use them to determine phrase length. For example, to fulfil the culminativity principle (see 1.2.2), every annotated unit needs at least one potential prominence target, which may become an actual phrasal accent in the later phases of the annotation.

Our multi-level annotation is hence guided by the understanding that prominence and boundary annotation systematically influence one another. In this way, the

hierarchical structure of prosodic units does not need to be actively identified by our annotators, but is implicitly included in the annotation.

1.3.2 *Transforming potential prominence targets into actual phrasal and sentential prominences*

Our extensive and detailed phrasing procedure (described in Section 2) allows for a fine-grained template that is the basis for the annotation of prominences above the word level. As we will outline in Sections 2.1 and 2.2, we distinguish four levels of prosodic boundaries and hence four different types of prosodic units. For each unit we annotate prominences that serve as heads of those units. We decided to distinguish two different types of prominence above the word level: for the two lower level units, we assign phrasal prominence, and for the two higher level units, we assign sentential prominence.

As discussed in Section 1.3.1 potential prominence targets are marked as those POS that have the potential to carry phrasal or sentential prominence, but may not end up carrying it depending on their position in the respective unit (for details see the manual in the appendix.).

Once the prosodic boundaries are annotated, *potential* prominence targets on nouns and adjectives/adverbs receive *actual* phrasal prominence if they are the rightmost element in the respective unit (following the Nuclear Stress Rule, Selkirk 1984). Potential prominence targets on verbs carry actual phrasal prominences if they are the rightmost element in the phrase and are not directly preceded by a noun or adjective/adverb in the same phrase (see the Predicate Deaccenting Rule, Schmerling 1976; Windmann et al. 2011). Further, we identified several syntactic conditions (see Appendix) in which a verb (although preceded by nouns or adverbs in the same phrase) carries phrasal prominence (cf. Vogel et al. 2015); to name but one example here, participles carry phrasal

prominence if they occur together with an auxiliary (cf. Féry & Herbst 2004; Röhr et al. 2020; see Appendix for details).

Additionally, to conform to the aforementioned principle of culminativity, we consider any rightmost POS as carrying actual phrasal prominence if the annotated unit (usually a small comma phrase) does not include a potential prominence target. The sentential prominences the annotation of which follows the same rules as the one for the phrasal prominences apply to the two higher levels of prosodic boundaries.

In this way, our system meets the objective of culminativity on several prosodic levels, and thus adheres to the hierarchical organization, as envisaged in previous work on prosodic phonology. In Example (2), the final phrasal prominence marks are in bold, sentential prominences are underlined. The boundary strengths 1–4 are annotated as (I – IIII).

(2) Sie gebe ja zu, III sagte die Frau, III während ihr die **Tränen** I über das Gesicht liefen, III über Mundwinkel und **Kinn** I auf ihre schwarze **Bluse** tropften I und **dort** I noch schwärzere **Spuren** hinterließen, III sie gebe ja zu, III dass ihr **Adi** I ein wilder junger Mensch gewesen sei.III

She admitted, the woman said, as tears ran down her face, dripping down the corners of her mouth and chin onto her black blouse, leaving even blacker marks there, she admitted that her Adi had been a wild young man.

In the following, we will detail the process of distributing these prosodic boundaries in the process of annotation.

2 THE ANNOTATION OF PROSODIC BOUNDARIES

Extant approaches regarding prosodic boundary prediction based on written texts simply assume phrase boundaries at each punctuation mark (e.g., Windmann et al. 2011). However, different types of punctuation marks, such as sentence-internal and sentence-

final punctuation, have different effects on the phonetic realization of phrase boundaries (Chafe 1988; Kalbertodt et al. 2015; Trouvain & Moebius 2018). Second, prosodic boundaries may also occur without punctuation. Chafe (1988) found that punctuation and prosodic boundaries do not always coincide in English prose; he specifically observed cases where long intervals between punctuation marks are divided into shorter prosodic intervals by the reader. Trouvain and Moebius (2018) also provided evidence that pauses can occur independently of punctuation. In the same vein, research on prosodic phonology (Nespor & Vogel 1989 [2007]) defines prosodic phrases as units ideally comprising two content words (Selkirk 2000); this implies that they are usually significantly shorter than the distance between punctuation marks in written texts. Complicating matters, prosodic phrase boundaries often reflect syntactic phrase boundaries, but need not be identical with these. In light of these findings, we chose to implement a rule system for prosodic boundary annotation that a) reflects graded levels of prosodic boundaries; and b) accounts for prosodic boundaries that occur with and without punctuation. We introduce our fundamental points regarding prosodic phrasing in the following.

2.1 Prosodic boundaries with punctuation

Overall, we distinguish four levels of prosodic boundaries. The three highest levels coincide with punctuation marks. As evidenced by prior research, sentence-final punctuation is accompanied by longer pauses than sentence-internal punctuation (Trouvain & Moebius 2018; Chafe 1988). Therefore, we assign the highest level of prosodic boundary strength (level 4, represented by ||||) to this first category (marked by a full stop, exclamation mark, question mark, or semicolon). Next in the punctuation hierarchy come commas and other sentence-internal punctuation marks (em dashes,

parentheses). Depending on the length of the preceding unit (Krivokapic 2010; Watson & Gipson 2004), these cases trigger different levels of prosodic boundary strength. Based on our experience with the sample texts, we determined the following: if preceded by at least three words, a comma or dash is followed by a level 3 (III) break, whereas we assign level 2 (II) breaks after units that are no longer than two words. Level 1 (I) breaks occur without punctuation (see Section 2.2.2).

2.2 *Prosodic boundaries without punctuation*

How can prosodic boundaries that do not coincide with punctuation be predicted from written text by linguistically naïve coders?² To solve this problem, we opted to turn the aforementioned approach to prosodic boundaries on its head.

2.2.1 *Linkage (prohibition of prosodic boundaries)*

Since phrasing *ex positivo*, i.e., the rule-based identification of phrasal units, could not be easily implemented without specialist coders, we used a phrasing *ex negativo*, i.e., we defined the circumstances under which a boundary was *not* allowed to occur. This approach has also partly been used in prior research (Atterer 2002; Atterer & Klein 2002; Atterer & Schulte im Walde 2004; Gee & Grosjean 1979; Grosjean et al. 1983; Selkirk 1984; Cooper & Paccia-Cooper 1980). However, these approaches require either an extensive amount of chunking rules or a syntactic annotation with hierarchical structuring or clause boundaries. This renders them unsuitable for easily learned manual annotation. As an alternative, we decided to formulate a limited number of rules which apply between two predefined POS and lead to phrasal cohesion or linkage between them. Under this approach, all that is required of the annotator is correct POS identification. In total, our manual contains seven such rules (see the coding manual in the Appendix [Table 6-7]).

Example 3 shows a sentence complete with phrasing according to the previously introduced steps, potential prominence targets (in bold) and blocked phrase boundaries; the resulting linkages are represented with (~) or (-).

(3) „Das wäre **neu**“, ||l **sagte K.**, || **sprang** aus ~ dem ~ **Bett** und **zog** – **rasch** seine ~ **Hosen** an.||l

“That would be new”, said K., jumped out of bed and quickly put on his trousers.

The example shows how significantly the linkages limit possible further phrase boundaries beyond punctuation. In particular, the linkages directly after prepositions (*aus ~ dem Bett*, ‘from ~ the bed’) and between determiner / possessive pronoun and neighbouring reference word (*dem ~ Bett* ‘the ~ bed’, *seine ~ Hose* ‘his ~ trousers’) are based on the tendency of function words to cliticize with prosodic heads (for English see, e.g., Hayes 1989; Bachenko & Fitzpatrick 1990; Gee & Grosjean 1993; for German see, e.g., Atterer 2004). The linkage between verb and adverb (*zog – rasch*, ‘put – quickly’) is supported by previous research on chunk formation (Schmid & Schulte im Walde 2004; Atterer 2004; see also Gee & Grosjean 1993, who provide an algorithm for English). Since both verbs and adverbs and both nouns and adjectives can carry phrasal prominence (Selkirk 1984; Windmann et al. 2011), these non-cliticized linkages may be dissolved later in the coding if certain conditions apply (see below; the linkages that we consider as weak and thus potentially dissolvable linkages described in our manual in the Appendix). Due to this strategy, we avoid word chains, e.g. exceedingly long phrases (see Siebenhaar et al. 2007) which often occur in recursive structures (multiple embeddings) or with multiple modifiers.

2.2.2 Placing prosodic boundaries without punctuation

Having established the linkages (prohibition of prosodic boundaries), we apply the topological model for German sentence structure (cf. Wöllstein 2014) to predict simple

phrase boundaries that do not coincide with punctuation (level 1, marked as |). The topological model distinguishes core parts of the sentence from peripheral ones (prefield and postfield). In canonical declarative sentences, the inflected verb marks the beginning of the core part, which is generally preceded by a syntactic constituent that constitutes the prefield. The boundary between prefield and inflected verb affords a level 1 break, provided the prefield constituent contains at least two words, one of which needs to be a potential prominence target (see Section 1.3.1; for constraints on prosodic boundaries after the prefield due to phonological weight in English see Grosjean et al. 1979). To illustrate this, consider the following example: a boundary after the prefield occurs in Example 4a, but not in Example 4b.

(4a) Unser Freund | **hat** den Verdacht, || dass wir ihn alle hassen.

Our friend suspects that we all hate him.

(4b) Er **hat** den Verdacht, || dass wir ihn alle hassen.

He suspects that we all hate him.

Level 1 boundaries may also occur between the sentence core and the postfield. There, they may be indicated by certain function words introducing verb phrases (see also Chafe 1988, and Bachenko & Fitzpatrick 1990, for English, and Trouvain & Moebius 2018, for German) or sentence-final prepositional phrases and/or extrapositions or afterthoughts (see also Chafe 1988; Bachenko & Fitzpatrick 1990). This group of function words includes conjunctions, comparative particles and prepositions in the postfield (underlined in the following examples):

(5) Er habe gerauft | und er habe getrunken.

He had scuffled and he had drunk.

(6) Dann erst dürfen wir aufhören | mit unserem Lob.

Only then may we stop with our praise.

(7) Weil er häufiger gelobt werden will | als alle anderen.

Because he wants to be praised more often than anyone else.

Whether or not a prosodic boundary is actually annotated again depends on the weight or complexity of the element in the postfield. We determined that the postfield needs to be at least four syllables in length and contain a potential prominence target.

Examples 8a and 8b show this for conjunctions that introduce a unit of an inflected verb with an argument (8a) and without one (8b) (see also Trouvain & Moebius 2018).

(8a) „Das wäre neu“, lll sagte K., ll sprang aus ~ dem ~ Bett | und zog – rasch seine ~ Hosen an. llll

„That would be new“, said K., jumped out of bed and quickly put on his trousers.

(8b) „Das wäre neu“, lll sagte K., ll sprang aus ~ dem ~ Bett und lachte. llll

“That would be new”, said K., jumped out of bed and laughed.

For prepositions and comparative particles, the following examples show differences regarding the length and the presence of a potential prominence target (in bold), which leads to either the occurrence of a prosodic boundary (a) or the lack thereof (b, c), respectively (see also Grosjean et al. 1979).

(9a) Ich ~ habe nichts **gesagt** | über ~ den ~ **Inspektor**. llll

I said nothing about the inspector.

(9b) Ich ~ habe nichts **gesagt** über ~ **Kim**. llll

I said nothing about Kim.

(9c) Ich ~ habe nichts **gesagt** über ~ ihn. llll

I said nothing about him.

2.3 Phrase formation: an algorithm

So far, we have distinguished four levels of prosodic boundaries depending on punctuation, number of words, the prefield, and certain function words as triggers, in the

case of the postfield; furthermore, we have defined contexts in which prosodic boundaries must *not* occur. Hence, the annotation system considers 1) strong cohesion (~); 2) weak cohesion (–); 3) the basic level of prosodic boundaries (Level 1, no punctuation mark: I); 4) the second level of prosodic boundaries (short comma phrase with two words at most: II); 5) the third level of prosodic boundaries (long comma phrase with three words at least: III); and 6) the fourth level of prosodic boundaries (sentence final punctuation: IIIL). Example 10 shows a longer sentence which has been annotated accordingly (potential prominence targets in bold).

(10) Sie **gebe** ja **zu**, III **sagte** die ~ **Frau**, III während ihr die ~ **Tränen** über ~ das ~ **Gesicht**

liefen, III über ~ **Mundwinkel** und **Kinn** auf ~ ihre ~ **schwarze** – **Bluse tropften** I und **dort noch** – **schwärzere** – **Spuren hinterließen**, III sie **gebe** ja **zu**, III dass ihr ~ **Adi** ein ~ **wilder** – **junger** – **Mensch** gewesen sei. IIIL

She admitted, the woman said, as tears ran down her face, dripping down the corners of her mouth and chin onto her black blouse, leaving even blacker marks there, she admitted that her Adi had been a wild young man.

Some of the phrases afford more prosodic boundaries due to sheer length. While most of these additional prosodic boundaries would probably be set intuitively by any native speaker, we decided, in the service of replicability, on the use of an algorithm for phrase formation. To this end, we modified an extant, relatively abstract algorithm to fit into our system (Wheeldon 2000: 257, following Hayes 1989)³.

Specifically, we posit, in line with the principle of culminativity, that every prosodic phrase has to contain a single prosodic head. Consequently, all words that were marked as potential prominence targets may become a prosodic head. Ideally, a simple phrase contains two such potential targets⁴ (Selkirk 2000).

The algorithm proceeds as follows: To obtain additional level 1 breaks, the annotated phrases are parsed from right to left, i.e., against the direction of reading. Once two potential prominence targets have been encountered, a phrase break is set at the next possible location to the left, i.e., at a word transition that does not constitute a linkage. In case a third prominence target is encountered before a neutral (non-linkage) transition is reached, this target is integrated into the current phrase. If a fourth prominence target is encountered before a neutral transition is reached, a phrase break is set at a position of a dissolvable linkage (see Section 2.2.1), such that the maximum of three prominence targets is maintained.

This direction of parsing may at first appear unintuitive, but is motivated by economy (for a similar notion, see Hayes 1995); we illustrate its usefulness for our purposes in the following examples. Example 11a below illustrates the phrase boundaries obtained when parsing proceeds from the left (in the direction of reading). Example 11b shows the resulting phrasing when parsed from the right. Parsing from left to right (11a) creates ‘orphaned’ elements, namely the verbs which should bear no phrasal accent when preceded by nouns, adjectives or adverbs, while phrasing from right to left (11b) leaves no such elements (the phrase boundaries which differ between example versions are marked with a “#”).

(11a) Sie **gebe** ja **zu**, **lll** **sagte** die ~ **Frau**, **lll** während ihr die ~ **Tränen** über ~ das ~ **Gesicht** #

liefen, **lll** über ~ **Mundwinkel** und **Kinn** | auf ~ ihre ~ **schwarze** – **Bluse** # **tropften** | und

dort noch – **schwärzere** – **Spuren** # **hinterließen**, | sie **gebe** ja **zu**, | dass ihr ~ **Adi** ein ~

wilder – **junger** – **Mensch** # gewesen sei. **llll**

(11b) Sie **gebe** ja **zu**, **lll** **sagte** die ~ **Frau**, **lll** während ihr die ~ **Tränen** # über ~ das ~ **Gesicht**

liefen, **lll** über ~ **Mundwinkel** und **Kinn** | auf ~ ihre ~ **schwarze** – **Bluse** **tropften** | und **dort**

noch – **schwärzere** – **Spuren** hinterließen, III sie gebe ja **zu**, III dass ihr ~ **Adi** # ein ~ **wilder**
– **junger** – **Mensch** gewesen sei. IIII

Overall, proceeding against the direction of reading leads to more intuitive results compared to parsing in the direction of reading. In particular, when we parse from the left, isolated verbs form their own phrases, appearing oddly brief compared to the neighbouring phrases; this is at odds with the preferred balance between phrase lengths (see below). Since parsing against the direction of reading required no additional steps for ‘orphaned verbs’ (see attachment rules in Atterer & Schulte im Walde 2004, or verb balancing/ adjacency rules in Bachenko & Fitzpatrick 1990) and yielded intuitively convincing results throughout our samples, we opted for the right-to-left annotation.

2.4 Length restrictions

All relevant systems of prosodic boundary annotation or pause prediction include length restrictions (see, among others, Gee & Grosjean 1993; Atterer 2002). Indeed, due to our linkage and phrasing rules, we found that in the complex literary texts we used, some of the resulting phrases were uncommonly long, as can be seen in Example 12a below. It is syntactically identical to Example 12b, but contains more syllables (13 versus 10 syllables in the last phrase of the sentence).

(12a) ...I sie gebe ja zu, III dass ihr ~ Adi I ein ~ wilder – junger – Casanova gewesen sei. IIII

... she admitted that her Adi had been a wild young Casanova.

(12b) ... I sie gebe ja zu, III dass ihr ~ Adi I ein ~ wilder – junger – Mensch gewesen sei. IIII

... she admitted that her Adi had been a wild young man.

Our phrase formation algorithm does not create a further boundary in the last phrase irrespective of length, due to the linkages between adjectives and noun. However, as detailed in Section 2.2.1, we distinguish non-dissolvable linkages (as created by clitization and represented as ~) from preliminary linkages (between particular types of

potential prominence targets as detailed in the manual in the Appendix; represented as –, which can be dissolved if they exceed certain length restrictions. Length restrictions, in turn, are operationalized based on the number of syllables in combination with the number of words in the respective phrase. Siebenhaar and colleagues (2004) have shown that speakers avoid phrases with more than 12 syllables, whereas Chafe (1988) found that speakers have a preference for phrases of five to seven words. In keeping with these results, we restrict phrase length for longer units of at least five words to a maximum of 12 syllables. In phrases with more than 12 syllables and at least five words, the preliminary linkages are dissolved, so that the phrase can be further divided.

The location of the additional prosodic boundary is based on findings regarding a preference for balanced length in adjacent phrases (Gee & Grosjean 1993). With the number of syllables as the reference value, and the goal to create two phrases of approximately equal length, the phrasing of Example 12a can be amended as follows (12c):

(12c) ...I sie gebe ja zu, III daß ihr ~ Adi I ein ~ wilder junger I Casanova gewesen sei. IIII

... she admitted that her Adi had been a wild young Casanova.

This rule can be easily applied in cases in which the result consists of two new phrases with at least one primary accent each and a similar syllable count. However, in some sentences several divisions are possible, with each not quite leading to a balance (Example 13):

(13) ...I sie gebe ja zu, III daß ihr X Adi I ein ~ uncharakteristisch wilder junger Mensch
gewesen sei. IIII

.... she admitted that her Adi had been an uncharacteristically wild young man.

For such unclear structures, we refer to Behaghel's (1909) Law of Increasing Terms (see also Hawkins 1994): the boundary is set such that each phrase contains at least one prominence target and the first phrase contains fewer syllables than the second.

Still, some sentences allow for different, intuitively equally acceptable phrasings, and hence a number of facultative phrase boundaries (see Atterer 2002). While our system cannot accommodate facultative phrase boundaries, it also does not contain linkages that would forbid such boundaries. Instead, we bypass the problem of different acceptable phrasing variants by first implementing the obligatory phrase boundaries (levels 2-4), then establishing a number of linkages, and subsequently generating what we consider a possible version (of level 1 phrase boundaries) as a kind of baseline from which individual renditions may possibly depart. Everything else (level 0) is covered by unmarked word boundaries; these constitute a category where prosodic boundaries are neither predicted nor prohibited.

This solution might at first be seen as a weakness of our system. However, it is a virtue insofar as it allows for a pragmatic way of handling an irresolvable issue: each annotation is a vote for one of potentially several possible versions of phrasing the same sentence. Our system gains in predictive value by differentiating between possible and impossible locations for phrasal boundaries, without claiming that there is always only one ideal annotation for all cases – which would falsely imply that there is but one single definitive way of realizing breaks when speaking any sentence.

3 THE ANNOTATION OF SYLLABLE PROMINENCE

We developed our annotation system for prominence at the level of the syllable and the word based on a module from Windmann and colleagues' (2011), the Bonn Open Synthesis System (BOSS). This module, which is based on previous work by Wagner

(2000), provides annotations of written texts for the production of synthetic speech in a text-to-speech system. The module assumes graded degrees of (perceived) syllable prominence. Prominence is automatically assigned to syllables based on an integrated lexicon that is enriched by part-of-speech and some syntactic information. Specifically, the authors distribute prominence by allocating so-called ‘beats’ to individual syllables according to the system specified in Table 2.

Step	Unit	Number of allocated beats
1	every syllable	1
2	every syllable whose nucleus is not a reduced vowel or a syllabic consonant	1
3	According to the part-of-speech of a word, beats are assigned to the constituent syllable bearing primary stress:	
3a	nouns, numerals, proper names	5
3b	adjectives, adverbs	4
3c	verbs, pronouns	3
3d	auxiliary verbs, particles	2
3e	all other parts-of-speech	1
4	phrase-final nouns, adjectives, and adverbs, as well as phrase-final verbs which do not follow nouns, adjectives, or adverbs	1
5	utterance-initial function words	1
6	every other syllable in a sequence of three or more syllables with 2 beats each	1

Table 2. Prominence prediction according to Windmann et al. (2011)

This system served as a rough blueprint for our prominence predictions. While it is, for the most part, easily applicable for native speakers performing a manual annotation, we found that it requires a number of modifications in order to produce intuitive results for our relatively complex literary material. We also adjusted it in cases where it conflicted with prior linguistic evidence regarding syllable prominence. Apart from the

potential prominence targets discussed above, our modifications pertain to three main categories: 1) parts-of-speech, 2) rhythmical processes (including effects of neighboring material on syllable prominence), and 3) the position of a word in a sentence or phrase. We outline our main changes in the following.

3.1 *Prominence at the level of the word*

3.1.1 *Adjusting prominence values from Windmann et al. (2011)*

For assigning prominence to the stressed syllables of content words, we fine-tuned and extended Windmann et al.'s (2011) hierarchy (nouns, proper names > adjectives / adverbs > full verbs; for details, please see the Manual in the Appendix).

1) Participles and verb particles: While derived from verbs, we assign the same prominence value to participles as to adjectives or adverbs (4 beats), with which they often co-distribute in syntax (for the perceived high prominence of German participles see Roehr et al. 2020, as well as Féry and Herbst 2004, who showed in a speech production experiment that German participles may receive phrasal accents). Further, separated verb particles may also receive 4 beats. However, this depends on the POS of neighboring words (see Section 2.1.2).

2) Pronouns: In contrast to Windmann et al. (2011), we assign 2 beats to demonstrative pronouns and 1 beat to other pronouns. This decision is based on Samlowski et al. (2014) who found in a reading production study that demonstratives are realized more prominently than other pronouns (see Widera et al. 1997, for perceptual evidence; for evidence that full verbs are more prominent than pronouns, see Vogel et al. 2015).

3) We do not limit word-level prominence assignment to primary lexical stress. Rather, for multisyllabic words and phrasal compounds we also assign secondary lexical stress (for phonetic correlates of secondary stress, see Plag et al. 2011). Secondary stress affords

one additional beat on the respective syllable. The conditions of secondary stress assignment are sketched in Section 3.1.2 below.

4) In accordance with chunking-rules proposed by Abney (1996), complex names (first and last names, or title and name) are considered as one phrasal compound and as such receive primary stress on the rightmost constituent; the left member of the phrasal compound receives secondary stress (see Table 3).

			x	beat level 7
			x	beat level 6
			x	beat level 5
			x	beat level 4
x			x	beat level 3
x		x	x	beat level 2
x	x	x	x	beat level 1
Pe	ter	Schmitt.		Syllable

Table 3. Stress distribution in proper names (in red).

3.1.2 The impact of the Rhythm Rule

In line with the principle of rhythmic alternation, Selkirk (1984), Kiparsky (1966) and Bohn et al. (2011, above others) showed that adjacent stressed syllables, so-called stress clashes, are avoided via stress reduction (the Rhythm Rule). Based on this observation, we assign secondary word stress only to syllables that are separated from the primary stress by at least one unstressed syllable (i.e., one syllable with less than three beats, see Table 4). Syllables adjacent to the primary stress are considered targets for stress reduction and therefore cannot be assigned secondary stress.

x				x			beat level 7
x				x			beat level 6
x				x			beat level 5
x				x			beat level 4
x				x			beat level 3
x		x		x	x		beat level 2
x	x	x	x	x	x	x	beat level 1
Vor	trags	a	bend	Vor	a	bend	syllable
<i>Evening of lectures</i>				<i>Eve</i>			

Table 4. Prominence distribution that avoids stress clash; secondary stress (in red) is applied only in the left panel.

We propose that alternation in prominence works not only for neighbouring syllables but also for neighbouring words. Thus, beat assignment is sensitive to the prominence of neighbouring words – at least for some function words, the acoustic prominence of which has been shown to be particularly flexible (Vogel et al. 2015). Consequently, separated verb particles receive four (rather than the usual two) beats if they do not have a prominent neighbouring word (see section 2.2). Quantifiers and numerals receive five beats if separated from nouns, adjectives or adverbs by at least one other word; otherwise, they receive two or one beat (cf. Widera et al. 1997). In Table 5 this difference is illustrated for *jeder*, ‘everybody’.

x									x	beat level 7
x									x	beat level 6
x				x					x	beat level 5
x		x		x					x	beat level 4
x		x		x				x	x	beat level 3
x		x		x				x		beat level 2
x	x	x	x	x			x	x	x	beat level 1
Je	der	muss	Schla	fen			Je	der	Mensch...	syllable
<i>Everyone must sleep</i>							<i>Every human...</i>			

Table 5. Impact of neighbouring words on prominence distribution (in red).

4 INTER-CODER AGREEMENT

We applied the coding system to eight different German literary texts (listed in Table 1). Eight annotators (five of them without specific linguistic training) worked on syllabified versions of the texts resulting in three independent annotations of every text⁵. As index of coder agreement, we calculated Fleiss’ kappa, first separately for beat assignment and for prosodic boundaries, and then for both marks together. This was done for each text and then for the total corpus. As seen in Table 6, the kappa values range from .69 to .89

for inter-coder agreement when beats and caesurae annotations were combined, confirming substantial inter-coder agreement.

Author (n syllables)	Fleiss' kappa (beats)	Fleiss' kappa (prosodic boundaries)	Fleiss' kappa (total)
Bichsel (1654)	.87	.85	.84
Fontane (2599)	.89	.89	.87
Goethe (2672)	.71	.78	.69
Kafka (3013)	.89	.91	.88
Kleist (3080)	.89	.94	.89
Kronauer (1942)	.82	.82	.79
Ransmayr (1836)	.76	.87	.75
Walser (3262)	.85	.92	.84
Total corpus (20058)	.84	.88	.82

Table 6. Fleiss' kappa values for inter-coder agreement of three independent annotations of beats and prosodic boundaries.

For each text, the conflicting cases were discussed and resolved between the annotators resulting in a final consensus coding. We used the consensus coding of the beginning excerpts of the four novels (Fontane, Goethe, Kafka, Kleist) to predict prosodic boundary strength and syllable prominence in the phonetic realization.

5 VALIDATION

5.1 Recordings

We evaluated recordings of eight professional speakers (four male, four female). Each speaker was seated in a sound-attenuated recording booth with a Neumann U87 Ai Studio microphone placed ~30 cm from their mouths, reading aloud four prose text samples (Fontane, Goethe, Kafka, Kleist). The excerpts contained roughly ~1500–1800 words each.

5.2 Data preparation

When slips of the tongue or disfluencies occurred during the recordings, speakers were asked to re-read the sentence or paragraph concerned. The faulty sentences were later erased, yet the original pauses preceding the affected sentences were preserved. The renditions cleaned this way comprised roughly 6h of speech with a total of >50,000 spoken words.

The cleaned renditions were automatically segmented for words and syllables using MAUS (Schiel et al., 1999). Using Praat, we extracted duration, minimum and maximum F0, and mean and maximum intensity values for each syllable. Where the F0 algorithm did not register F0 oscillations, the F0 values were set to 0.

All spoken syllables were aligned with the citation form syllables from the codings using an awk script. Mismatching syllables (citation form syllables without a spoken equivalent) were discarded. This mostly affected unstressed syllables, as, e.g., the first syllable in “Geschenk” (“present”, citation form [gə.ʃɛŋk] realized as monosyllabic [kʃɛŋk]) or, e.g., the last syllable of “haben” (“have”, citation form [ha:.bən] realized as monosyllabic [ha:m]). In these cases, the beat annotations of the unuttered syllables were erased, and the prosodic boundary value shifted to the last spoken syllable of the word.

5.3 Data analysis and results

As detailed above, the coding system distinguishes six degrees of phrasal cohesion or boundary strength for the transitions between words: linkage (-1), a neutral transition (0), and four degrees of prosodic boundary (level 1–4). Likelihood of a pause and pause duration is predicted to increase with every degree. This prediction is borne out, as shown in Table 7 and Figure 2. While a pause is registered at only 3% of linkages, 80% of the

sentence boundaries (level 4) afford a pause, with pause likelihood increasing monotonically along the phrasing scale, as predicted.

	-1 Linkage, cohesion	0 neutral	1 simple phrase break	2 small comma phrase	3 long comma phrase	4 Sentence boundary
N of transitions	20281	18248	4480	1216	5480	2992
N of Pauses	648	890	412	471	2487	2390
Likelihood of pause	3%	5%	9%	39%	45%	80%

Table 7. Likelihood of a pause at each degree of transition.

The left panel of Figure 2 shows the pause durations (rescaled values to square root-transformed y-axis) broken down by the six degrees of the phrasing scale. The figure shows pause durations for all word transitions for which a pause (phonetic silence of at least 20ms) was registered (with the mean duration as average marker): a monotonic increase in pause duration along the prosodic break hierarchy is clearly visible.

We employed mixed effects models to corroborate the empirical appearance. For pause likelihood, we fit a generalized linear mixed effects model with binomial link function that takes presence vs absence of a pause at every word-to-word transition as the binomial dependent variable and boundary degree or strength as the predictor term (fixed effects coded as five successive difference contrasts); Speaker and Author serve as grouping terms (random intercepts). The model output is tabulated in the first column of Table 8.

Since the pause duration data is clearly non-normally distributed and skewed, we fit a mixed effects Gamma regression with log link, but otherwise with the same fixed effects and random effects. Only data from word-to-word transitions that actually exhibit a pause was considered, i.e., this model uses a subset of the data (transitions without

pauses were discarded). The second column of Table 8 summarizes the coefficients of interest.

Regarding pause likelihood, significant successive difference contrasts between the six degrees of boundary strength confirm the stepwise monotonic increase in boundary strength along the predicted boundary strength scale. As for pause duration, all contrasts apart from the one between neutral transitions and level 1 breaks come out as significant.

Apart from pause likelihood and pause duration, we examined pre-boundary syllable duration as a potential cue to prosodic boundaries. To this end, we measured the duration of all word-final syllables for each degree of boundary strength (shown in the right panel of Figure 2). Visual inspection suggests an inverted u-shaped pattern with pre-boundary syllable duration increasing up to the level of small comma phrases (level 2 breaks) and then slightly decreasing towards sentence boundaries. A mixed effects Gamma regression (with log link) was fit with pre-boundary syllable duration as dependent variable, employing the same fixed effects and random effects as in the above models (3rd column in Table 8). The model confirms significant successive increases in syllable duration at the junctures between levels -1-0, 0-1, and 1-2. The subsequent decreases in syllable duration between levels 2 and 3, and level 3 and 4 are significant as well. Given that both pause likelihood and pause duration show substantial and significant increases along the scale of predicted boundary strength, the decrease in pre-boundary syllable duration between level 3 and level 4 breaks does not invalidate the boundary strength scale. Note that Ferreira (1993) similarly proposes certain windows during the process of speaking which might either be filled by syllable lengthening or pauses. This

phenomenon deserves further consideration; however, since discussing it in detail would go beyond the scope of this paper, we refrain from delving into the issue here.

Figure 2. Violin plots with pause durations in ms (left panel) and durations of the final syllables (right panel) broken down by boundary strength index. Average lines show the mean value for each distribution.

Successive differences	Pause likelihood (generalized linear mixed model with binomial link)				Pause duration (Gamma regression with log link; only transitions with pauses >20ms)				Pre-boundary syllable duration (Gamma regression with log link)			
	beta	SE	z	p	beta	SE	t	p	Beta	SE	t	p
Linkage–neutral	.42	.05	7.97	<.001	.13	.04	3.53	<.001	.084	.005	18.04	<.001
Neutral–level 1	.68	.06	10.85	<.001	.07	.05	1.62	0.106	.097	.008	12.85	<.001
Level 1–level 2	2.03	.08	25.29	<.001	.21	.05	4.06	<.001	.191	.015	13.02	<.001
Level 2–level 3	.22	.07	3.43	0.002	.14	.04	3.71	<.001	-.035	.014	-2.47	0.014
Level 3–level 4	1.53	.05	26.17	<.001	.90	.02	41.44	<.001	-.041	.01	-3.94	<.001

Table 8. Summary statistics for coefficients of interest (successive differences between boundary strength values) of three linear mixed effects models, computed in R using the lme4 package (Bates et al. 2015): 1) a generalized linear mixed model on pause likelihood (binomial link function), 2) a Gamma mixed effects regression on pause duration, and 3) a Gamma mixed effects regression on pre-boundary syllable duration. For models 2 and 3, the p values were estimated using the lmerTest package (Kuznetsova et al. 2017). Significant effects against predictions are shaded, non-significant effects are written in italics.

5.4 *The composite prominence scale - Beats*

The prominence scale employed in the coding system derives syllabic prominence mainly from part of speech and from (sub)lexical stress and phrasal accentuation. Additional beats are due to rhythmic adjustments (additional prominence marks for sentence-initial function words, application of the rhythm rule), as explained in detail in Section 3.

The beats assigned by all these sources are simply added up such that, together, the system distinguishes up to nine degrees of syllabic prominence, ranging from least prominent (1 beat, schwa syllable) to most prominent (nuclear accent on a noun, 9 beats). Analyses of the part of speech hierarchy, as well as (sub)lexical stress and phrasal accentuation on the phonetic signal are presented in the appendix.

In order to examine the predictive value of the nine-part prominence scale, we considered three widely used measures of syllabic prominence, namely syllable duration, pitch (in terms of F0 range, i.e., maximum F0 – minimum F0), and maximum intensity on each syllable. The resulting violin plots in Figure 3 show a general increase in the three dependent measures along the prominence scale. Apart from the last prominence degree, for every increase in predicted prominence, at least two of the three measures show an increase as well.

Figure 3. Violin plots showing Syllable duration (upper panel), F0 range (middle panel) and maximum intensity distributions (lower panel) broken down by predicted syllabic prominence on the nine-part prominence scale. Average lines show the mean value for each distribution.

To substantiate this impression, we fitted, for each of the three dependent variables, mixed effects models with the 9-degree prominence scale as ordinal predictor variable (i.e., eight successive differences contrasts); since the distributions of duration and F0 range values are heavily skewed, we fit mixed effects gamma regressions with log link function for these. Speaker and Author served as random intercepts. The outputs of the models (Table 9 displays the coefficients of interest) confirm, for every successive difference contrast apart from the last two, a significant increase in phonetic prominence in at least two of the three variables under scrutiny. The weaker phonetic expression of the two highest prominence degrees is likely due to a lack of phonetic expression of nuclear or sentence level accents. This finding is in line with research by Pierrehumbert (1979) who has shown that the prominence of nuclear accents is first and foremost a perceptual phenomenon and not directly rooted in the phonetic signal. Arguably, the perceived prominence of the nuclear accent is instead, in great part, due to its position (i.e., the last accent in the phrase).

To corroborate this interpretation, we subtracted all and only the nuclear accents from the tally of cumulative beats and re-fitted the above models with the resulting 8-degree prominence scale as ordinal predictor variable (Figure 4). This time, at least two of the three dependent measures show significant increases for all successive difference contrasts (see Table 10).

Figure 4. Violin plots showing Syllable duration (upper panel), F0 range (middle panel) and maximum intensity distributions (lower panel) broken down by predicted syllabic prominence on the eight-part prominence scale (nuclear accents discarded). Average lines show the mean value for each distribution.

In sum, all seven prominence contrasts exhibit differences in the predicted direction for at least two of the three variables examined (duration, F0 range, intensity). Moreover, occurrence of pauses and pause durations corroborate the predictions regarding the caesura or phrase boundaries.

Together, we take this as successful phonetic validation of the proposed ordinal six-degree scale of phrasal cohesion or caesurae, and of the likewise ordinal prominence scale. The final or ninth degree of prominence that is part of this coding system, namely nuclear accent, does not engender an additional increase in the phonetic measures under study. The perceptual prominence of it is nonetheless undisputed.

Successive differences	Syllable duration (Gamma regression with log link)				F0 range (Gamma regression with log link)				Maximum intensity			
	beta	SE	t	p	beta	SE	t	p	beta	SE	t	p
Beat 1 – Beat 2	.185	.004	43.39	<.001*	<i>-.001</i>	<i>.01</i>	<i>-0.10</i>	<i>.919</i>	.23	.043	5.31	<.001*
Beat 2 – Beat 3	<i>-.04</i>	<i>.005</i>	<i>-8.85</i>	<i><.001*</i>	.021	.01	2.05	.041*	.21	.046	4.61	<.001*
Beat 3 – Beat 4	.165	.006	25.56	<.001*	.055	.015	3.69	<.001*	1.01	.065	15.47	<.001*
Beat 4 – Beat 5	.1	.01	10.47	<.001*	.11	.022	5.05	<.001*	0.36	.096	3.8	<.001*
Beat 5 – Beat 6	<i>-.058</i>	<i>.009</i>	<i>-6.17</i>	<i><.001*</i>	.1	.021	4.91	<.001*	0.39	.094	4.14	<.001*
Beat 6 – Beat 7	.084	.007	11.21	<.001*	.04	.017	2.67	.007*	<i>-.16</i>	<i>.075</i>	<i>-2.13</i>	<i>.033*</i>
Beat 7 – Beat 8	.024	.008	3.17	0.0015*	<i>.03</i>	<i>.017</i>	<i>1.85</i>	<i>.06</i>	<i>-.23</i>	<i>.077</i>	<i>-2.91</i>	<i>.004*</i>
Beat 8 – Beat 9	<i>.005</i>	<i>.009</i>	<i>0.57</i>	<i>0.567</i>	<i>-.03</i>	<i>.02</i>	<i>-1.54</i>	<i>.123</i>	<i>-.27</i>	<i>.09</i>	<i>-3.02</i>	<i>.002*</i>

Table 9. Summary statistics for coefficients of interest (successive differences between nine degrees of syllabic prominence or “beats”) of three mixed effects models, computed in R using the lme4 package (Bates et al. 2015): 1) a mixed Gamma regression on syllable duration (with log link), 2) a mixed Gamma regression model on F0 range, and 3) a linear mixed effects model on maximum intensity. The p values were estimated using the lmerTest package (Kuznetsova et al. 2017). Significant effects against predictions are shaded, non-significant effects are printed in italics.

Successive differences	Syllable duration (Gamma regression with log link)				F0 range (Gamma regression with log link)				Maximum intensity			
	beta	SE	t	p	beta	SE	T	P	beta	SE	t	p
Beat 1 – Beat 2	.184	.004	43.08	<.001*	<i>-.004</i>	<i>.01</i>	<i>-0.41</i>	<i>.68</i>	.22	.043	5.2	<.001*
Beat 2 – Beat 3	<i>-.04</i>	<i>.005</i>	<i>-8.15</i>	<i><.001*</i>	.028	.01	2.66	.008*	.23	.046	4.9	<.001*
Beat 3 – Beat 4	.163	.006	25.27	<.001*	.051	.015	3.45	<.001*	.99	.065	15.31	<.001*
Beat 4 – Beat 5	.1	.01	10.47	<.001*	.11	.022	5.05	<.001*	.36	.096	3.8	<.001*
Beat 5 – Beat 6	<i>-.03</i>	<i>.009</i>	<i>-3.46</i>	<i><.001*</i>	.12	.021	5.81	<.001*	.16	.091	1.72	.086
Beat 6 – Beat 7	.06	.007	8.81	<.001*	.029	.015	1.87	.061	.05	.069	.71	.478
Beat 7 – Beat 8	.024	.007	3.24	0.0012*	.027	.016	1.67	.09	<i>-.16</i>	<i>.073</i>	<i>-2.25</i>	<i>.024*</i>

Table 10. Summary statistics for coefficients of interest (successive differences between eight degrees of syllabic prominence or “beats”; nuclear accents subtracted) of three mixed effects models, computed in R using the lme4 package (Bates et al. 2015): 1) a mixed Gamma regression on syllable duration (with log link), 2) a mixed Gamma regression model on F0 range, and 3) a linear mixed effects model on maximum intensity. The p values were estimated using the lmerTest package (Kuznetsova et al. 2017). Significant effects against predictions are shaded, non-significant effects are printed in italics.

6 LIMITATIONS

While our system captures a large part of the rhythmical qualities of prose texts, it still has a few limitations. For one, our determination to maintain a rulebook of relative brevity and easy applicability proved to be both a virtue and a source of imperfection. Specifically, in a few cases where our system differed from the intuitions of our annotating native speakers, the detected outliers were not accommodated by the creation of individual additional rules. Similarly, our decision to neglect discourse structure certainly leads to some locally unconvincing annotations that do not correspond to what a native speaker would realize. Nevertheless, the scarcity of these cases overall, as reflected by the results of our validation study, appears to justify our pragmatic approach: the impact of these limitations is outweighed by the overall success of our relatively brief and easily implemented coding system.

Lastly, the coding system we propose here could of course be complemented by including information about additional larger-scale units, such as the distribution of word, phrase, sentence, and paragraph lengths. While these measures certainly do not pertain to rhythmicity as directly as the caesuras and syllable prominences we record, they would still serve to round out the picture, allowing further insight into which features contribute to the rhythmicity of prose and how they interact.

7 CONCLUSION AND OUTLOOK

Our annotation system offers a way to predict quite reliably two key variables that shape the rhythm of spoken and written prose: syllable prominence and the distribution of prosodic boundaries which is inextricably linked to phrase length. We also managed to show and operationalize the systematic covariation between the two factors. While not the first attempt of this kind, it is original insofar as it allows for the annotation of prose

sentences of any length and complexity, including literary prose such as the texts on which it was developed. Reassuringly, it has proven to work very well in our validation study.

Within our own research context, the manual described in this paper is part of a larger project on the rhythm of literary prose, which in turn contributes to the greater field of research on prosodic particularities of literary language. To date, this field mainly includes research on poetry, such as on poetic speech melody (Menninghaus et al. 2018), on poetic rhyme and metre (e.g., Blohm et al. 2017; Menninghaus et al. 2017), as well as on sound-meaning interfaces in poetic language (e.g., Auracher et al. 2019; Kraxenberger et al. 2018). The present paper paves the way for additional insights into the reading of prose. Thus, especially with regard to literary prose, it is a far cry from the relatively simplistic, binary prominence models, that first tried to capture literary rhythms (see Marbe 1904; Kagarov 1928, and Lipsky 1907, for early intuitive attempts and, e.g., Best 2002; Kaßel 2002, and Knaus 2008, for more systematic and sophisticated versions).

In particular, our manual carries great potential to further the study of prose rhythm: central aspects of the rhythmical signatures of, and differences between, individual author styles, epochal styles, genres, and intratextual positions can now be systematically analyzed. It thus could make a direct contribution to older works from classical studies, rhetoric and literary studies, which tried to corroborate Cicero's (trans. 2001) claim that rhythmicity arises specifically in sentence- and passage-final cadences (e.g., Aili 1979; Aumont 1996; Bornecque 1907; de Groot 1921; Knapp 2015; Primmer 1968; Zielinski 1904, 1914). For modern languages, position-specific rhythmical trends have been identified in broad samples of English literature (e.g., Anttila and Heuser 2015, who report similar findings for Finnish literature), but extant approaches have not been

able to identify significant evidence for systematic regularity in individual positions (see also Saintsbury 1912; Borgeson et al. 2018).

Our annotation system also has the potential to further enlighten the relation between the rhythmically free forms of prose and strictly metered poetry (see Shengeli 1921, and Yarkho 1925, for earlier attempts to accomplish this for Russian) as well as the forms that lie in between the two, including phenomena like free verse or so-called ‘prose poems’. It could further be used to extend earlier research on interactions between the distribution of rhythmical units and word length, interval frequencies, genre, and affective content (for earlier research in these areas see, e.g., Gropp 1915; Kullmann 1909; Tomashevsky 1929; Unser 1906).

Other areas of potential research using this system include the systematic comparison of rhythmical realization between speakers in general, and between lay and professional speakers in particular. Aiming to differentiate between the stress or quantity distributions of the lexicon and the acoustic prose rhythm actually realized by readers, whether they read silently or aloud, previous research (Esser 1988) has evaluated reading performances and found significant differences between readers based on their levels of experience and proficiency. In addition, rhythm performance has been shown to be marked by the complexity of interactions between syllable rhythm and accent rhythm, the knowledge and foresight a reader needs to recognize and represent thematic focus, as well as the level of text interpretation through performance (Esser 2011). Our system could serve to add to and refine these results. In addition, it could be used to differentiate between rhythmical signatures of particular types of prose utterances, for example in political, journalistic, clerical, or judicial language, or in the language of advertisement.

A rather practical use of the system might lie in its applicability for (1) teaching and (2) speech therapy contexts: (1) to our knowledge, there is no tool to objectively evaluate oral reading capacities of children; predicting (non-)appropriate locations for pauses, the presented manual could serve as an instrument to evaluate the fluency with which children read. (2) The evaluation of prosodic units in spontaneous speech has been shown to be a fruitful source for diagnosing residual aphasia – a syndrome which is otherwise very difficult to capture (Jaacks et al. 2012). With our predictions for pauses and syllable prominence, analyzing patients’ reading behaviour might be a further step towards a reliable tool for the diagnosis of residual aphasia.

To get a better understanding of the relation between rhythmical realization in oral and silent reading, tracking the eye during reading might offer interesting insights into if and how the prominences and phrase boundaries predicted by our system influence patterns of eye movement. In terms of applicability, while our system was developed for manual annotation, its long-term use on larger corpora would of course benefit from full or at least partial automatization. Lastly, our system was developed specifically for German; an adaptation to other languages may be another desideratum for future research.

REFERENCES

- Aili, H. (1979). *The prose rhythm of Sallust and Livy*. Stockholm: Almqvist & Wiksell.
- Anttila, A. (2016). Phonological effects on syntactic variation. *Annual Review of Linguistics*, 2, 115-137.
- Anttila, A., & Heuser, R. (2015). Phonological and metrical variation across genres. *Proceedings of the 2015 annual meeting on phonology*, 1-12.
- Atterer, M., & Klein, E. (2002). Integrating linguistic and performance-based constraints for assigning phrase breaks. In: *Proceedings of the 19th international conference on Computational linguistics*, vol. 1 (pp. 1-7). Association for Computational Linguistics.
- Aumont, J. (1996). *Métrique et stylistique des clausules dans la prose latine: De Cicéron à Pline le Jeune et de César à Florus*. Paris: Champion.
- Auracher, J., Scharinger, M., & Menninghaus, W. (2019). Contiguity-based sound iconicity: The meaning of words resonates with phonetic properties of their immediate verbal contexts. *PLoS ONE*, 14(5), e0216930.
- Bachenko, J., & Fitzpatrick, E. (1990). A computational grammar of discourse-neutral prosodic phrasing in English. *Computational Linguistics*, 16(3), 155-170.
- Barry, W. J. (1995). Schwa vs. Schwa+/r/ in German. *Phonetica*, 52(3), 228-235.
- Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*, 25, 110.
- Berg, T. (2008). Emphatic stress shift in German. *Zeitschrift für Sprachwissenschaft*, 27(2), 165-187.
- Best, K.-H. (2002). The distribution of rhythmic units in German short prose. *Glottometrics*, 3, 136-142. Retrieved from <https://www.ram-verlag.eu/wp-content/uploads/2018/08/glo3abstract.pdf>
- Blohm, S., Versace, S., Methner, S., Wagner, V., Schlesewsky, M., & Menninghaus, W. (in press). Reading poetry and prose: Eye movements and acoustic evidence. *Discourse Processes*.
- Blohm, S., Menninghaus, W., & Schlesewsky, M. (2017). Sentence-level effects of literary genre: Behavioral and electrophysiological evidence. *Frontiers in Psychology*, 8, 1887.
- Borgeson, S., Anttila, A., Heuser, R., & Kiparsky, P. (2018). The rise and fall of antimetricity. Unpublished manuscript, Stanford University, Palo Alto, USA. Retrieved from <https://web.stanford.edu/~kiparsky/Papers/Antimetricity-Draft-2018-04-20.pdf>
- Bornecque, H. (1907). *Les clausules métriques latines*. Lille: Au siège de l'Université de Lille.

- Breen, M., Dilley, L. C., Kraemer, J., & Gibson, E. (2012). Inter-transcriber reliability for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguistics and Linguistic Theory*, 8(2), 277-312.
- Breiss, C., & Hayes, B. (2020). Phonological markedness effects in sentence formation. *Language*, 96(2), 338-370.
- Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication*, 5(4), 395-426.
- Cicero (2001). *De oratore* (E. W. Sutton & H. Rackham, Eds. and Trans.). Loeb Classical Library. Cambridge, MA: Harvard University Press. (Original work written 55 B.C.)
- Cinque, G. (1993). A null theory of phrase and compound stress. *Linguistic Inquiry*, 24(2), 239-297.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.
- Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11(1), 51-62.
- de Groot, A.W. (1921). *Der antike Prosarhythmus* [Prose rhythm in antiquity]. Groningen: J.B. Wolters.
- Esser, J. (1988). *Comparing reading and speaking intonation*. Amsterdam: Rodopi.
- Esser, J. (2011). *Rhythm in speech, prose, and verse: A linguistic description*. Berlin: Logos.
- Fant, G., Kruckenberg, A., & Nord, L. (1991). Stress patterns and rhythm in the reading of prose and poetry with analogies to music performance. In J. Sundberg, L. Nord & R. Carlson (Eds.), *Music, Language, Speech, and Brain* (pp. 380-407). Stockholm: MacMillan.
- Féry, C., & Herbst, L. (2004). German sentence accent revisited. *Interdisciplinary Studies on Information Structure*, 1, 43-75.
- Féry, C. (2011a). German sentence accents and embedded prosodic phrases. *Lingua*, 121(13), 1906-1922.
- Féry, C., Hörnig, R., & Pahaut, S. (2011b). Correlates of phrasing in French and German from an experiment with semi-spontaneous speech. *Intonational Phrasing in Romance and Germanic*, 11-41.
- Féry, C. (2015). Extraposition and prosodic monsters in German. In Frazier, L. & Gibson, E. (Eds.): *Explicit and Implicit Prosody in Sentence Processing. Studies in Honor of Janet Dean Fodor* (pp. 11-37). Cham: Springer.
- Féry, C. (2017). *Intonation and prosodic structure*. Cambridge University Press.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100(2), 233.

- Franz, I., Kentner, G. & Domahs, F. (2021). The impact of animacy and speech rhythm on the word order of conjuncts in German preschoolers and adults. *Laboratory Phonology* 12(1), 1-38.
- Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15(4), 411-458.
- Grice, M., & Kügler, F. (2021). Prosodic prominence—a cross-linguistic perspective. *Language and Speech*, 64(2), 253-260.
- Gropp, F. (1915). *Zur Ästhetik und statistischen Beschreibung des Prosarhythmus*. Würzburg: Königliche Universitätsdruckerei H. Stütz.
- Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11(1), 58-81.
- Halle, M., & Vergnaud, J. R. (1987). Stress and the cycle. *Linguistic Inquiry*, 18(1), 45-84.
- Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. The University of Chicago Press.
- Ito, J., & Mester, A. (2012). Recursive Prosodic Phrasing in Japanese. In Borowsky, T., Kawahara, S., Shinya, T., & Sugahara, M. *Prosody Matters: Essays in Honor of Elisabeth Selkirk* (pp. 280-303). Sheffield: Equinox Publishing Ltd.
- Jackendoff, R., & Lerdahl, F. (2006). The capacity for music: What is it, and what's special about it? *Cognition*, 100(1), 33-72.
- Jaacks, P., Hielscher-Fastabend, M., & Stenneken, P. (2012). Diagnosing residual aphasia using spontaneous speech analysis. *Aphasiology*, 26(7), 953-970.
- Keller, E., & Forst, M. (2006). Speech synthesis of dialectal variants as a method for research on prosody. *Topics in Dialectal Variation*, 145-162.
- Kagarov, E. G. (1928). O ritme prozaicheskoi rechi. *Doklady Akademii Nauk SSSR, Seria B*, 44–51.
- Kaßel, Anja (2002). *Zur Verteilung rhythmischer Einheiten in deutschen und englischen Texten* (Unpublished master's thesis). University of Göttingen, Göttingen.
- Knapp, R. (2013). *Der Prosarhythmus als Stilmittel: Eine Untersuchung an Ciceros De natura deorum mit statistischer Auswertung und Softwareentwicklung*. (Unpublished master's thesis). Retrieved from <http://eplus.uni-salzburg.at/obvusbhs/content/titleinfo/4375771>
- Knaus, M. (2008). Zur Verteilung rhythmischer Einheiten in russischer Prosa. *Glottometrics*, 16, 57–62. Retrieved from <https://www.ram-verlag.eu/wp-content/uploads/2018/08/g16zeit.pdf>
- Knowles, G., Wichmann, A., & Alderson, P. (1996). Working with speech: Perspectives on research into the Lancaster/ IBM Spoken English Corpus. Longman.

- Kompe, R., Batliner, A., Kiessling, A., Kilian, U., Niemann, H., Noth, E., & Regel-Brietzmann, P. (1994, April). Automatic classification of prosodically marked phrase boundaries in German. In: *Proceedings of ICASSP 94. IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2 (pp. II-173). Institute of Electrical and Electronics Engineers (IEEE).
- Kraxenberger, M., Menninghaus, W., Roth, A., & Scharinger, M. (2018). Prosody-based sound-emotion associations in poetry. *Frontiers in Psychology*, 9, 1284.
- Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, 35(2), 162-179.
- Krivokapić, J. (2010). Speech planning and phrase length. *Proceedings of the 5th Conference on Speech Prosody*.
- Krivokapić, J., & Byrd, D. (2012). Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics*, 40(3), 430-442.
- Kullmann, P. (1909). Statistische Untersuchungen zur Sprachpsychologie. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 54, 290-310.
- Lerdahl, F. (2001). The sounds of poetry viewed as music. *Annals of the New York Academy of Sciences*, 930(1), 337-354.
- Lee, M. W. & Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition*, 105(2), 446-456.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Massachusetts.
- Liberman, Mark; Prince, Alan (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249-336.
- Lipsky, A. (1907). *Rhythm as a distinguishing characteristic of prose style*. New York: The Science Press.
- Marbe, K. (1904). *Über den Rhythmus der Prosa*. Giessen: J. Ricker'sche Verlagsbuchhandlung.
- Menninghaus, W., Wagner, V., Knoop, C. A., & Scharinger, M. (2018). Poetic speech melody: A crucial link between music and language. *PLoS ONE*, 13(11), e0205980.
- Menninghaus, W., Wagner, V., Wassiliwizky, E., Jacobsen, T., & Knoop, C. A. (2017). The emotional and aesthetic powers of parallelistic diction. *Poetics*, 63, 47-59.
- Nord, L., Kruckenberg, A., & Fant, G. (1990). Some timing studies of prose, poetry and music. *Speech Communication*, 9(5), 477-483.
- Petrone, C., Truckenbrodt, H., Wellmann, C., Holzgreffe-Lang, J., Wartenburger, I., & Höhle, B. (2017). Prosodic boundary cues in German: Evidence from the production and perception of bracketed lists. *Journal of Phonetics*, 61, 71-92.
- Primmer, A. (1968). *Cicero numerosus: Studien zum antiken Prosarhythmus*. Vienna: Böhlau.

- Roberts, A. (1996). Rhythm in prose and the serial correlation of sentence lengths: a Joyce Cary case study. *Literary and Linguistic Computing*, 11(1), 33-39.
- Röhr, C. T., Baumann, S., Schumacher, P. B., & Grice, M. (2020). Perceptual prominence of accent types and the role of expectations. *Proceedings of the 10th International Conference on Speech Prosody 2020*, 366-370.
- Samlowski, B., Wagner, P., & Möbius, B. (2012). Disentangling lexical, morphological, syntactic and semantic influences on German prominence – Evidence from a production study. *Proceedings of Interspeech 2012*, 2406-2409.
- Saintsbury, G. (1912). *A history of English prose rhythm*. London: Macmillan.
- Schmerling, S. F. (1976). *Aspects of English sentence stress*. University of Texas Press.
- Selkirk, E. (1984). *Phonology and syntax: The relationship between sound and structure*. Cambridge, Mass.
- Selkirk, E. (1996). The prosodic structure of function words. In Demuth, K. & Morgan, J. (Eds.): *Signal to Syntax*. (pp. 199-226). Lawrence Erlbaum.
- Shengeli, G. A. (1921). *Traktat o russkom stikhe*. Moscow: Gos. izd-vo.
- Siebenhaar, B., Forst, M., & Keller, E. (2004). Prosody of Bernese and Zurich German. What the development of a dialectal speech synthesis system tells us about it. In Gilles, P. & Peters, J. (Eds.) *Regional Variation in Intonation* (pp. 219-238). Tübingen: Niemeyer.
- Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language*, 86(1), 142-164.
- Tomashevsky, B. (1929). Ritm prozy: Pikovaya dama. In *O stikhe* (pp. 254–318). Leningrad: Priboy.
- Truckenbrodt, H. (1999). On the relation between syntactic phrases and phonological phrases. *Linguistic Inquiry*, 30(2), 219-255.
- Truckenbrodt, H. (2012). Effects of indefinite pronouns and traces on verb stress in German. In Borowsky, T., Kawahara, S., Shinya, T., & Sugahara, M. *Prosody Matters: Essays in Honor of Elisabeth Selkirk* (pp. 487-513). Sheffield: Equinox Publishing Ltd.
- Trouvain, J. & Möbius, B. (2018). Zu Mustern der Pausengestaltung in natürlicher und synthetischer Lesesprache. *Proceedings of the 29th Conference Elektronische Sprachsignalverarbeitung (ESSV '18)*, Ulm, 334-341.
- Unser, H. (1906). *Über den Rhythmus der deutschen Prosa* [On the rhythm of German prose]. Heidelberg: Universitäts-Buchdruckerei J. Hörning.
- Vogel, R., van de Vijver, R., Kotz, S., Kutscher, A., & Wagner, P. (2015). Function Words in Rhythmic Optimisation. In van de Vijver, R. & Vogel, R. (Eds.): *Rhythm in Cognition and Grammar. A Germanic Perspective* (pp. 253-274). Berlin: De Gruyter.

- Vogel, I., Bunnell, H.T., & Hoskins, S. (1995). The phonology and phonetics of the Rhythm Rule,” In Connell, B. & Arvaniti, A. (Eds.) *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV* (pp. 111-127). Cambridge: Cambridge University Press.
- Wagner, P., & Fischenbeck, E. (2002). Stress perception and production in German stress clash environments. *Proceedings of Speech Prosody 2002*.
- Wagner, P., Origlia, A., Avezani, C., Christodoulides, G., Cutugno, F., d’Imperio, M., Mancebo, D., Fivela, B.G., Lacheret, A., Ludusan, B & Moniz, H. (2015). Different parts of the same elephant: A roadmap to disentangle and connect different perspectives on prosodic prominence. *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Wagner, P. (2000). Evaluating metrical phonology—a computational-empirical approach. In: *Proceedings of Konvens 2000 – Sprachkommunikation*, 161.
- Watson, D. G. (2002). *Intonational phrasing in language production and comprehension*. Doctoral dissertation, Massachusetts Institute of Technology.
- Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713-755.
- Widera, C., Portele, T., & Wolters, M. (1997). Prediction of word prominence. In: *Fifth European Conference on Speech Communication and Technology*.
- Wiese, R. (2001). The structure of the German vocabulary: edge marking of categories and functional considerations. *Linguistics*, 39(1), 95-116.
- Wöllstein, A. (2014). *Topologisches Satzmodell*. 2. rev. ed., Heidelberg: Winter (= Kurze Einführungen in die germanistische Linguistik, KEGLI 8).
- Yarkho, B. I. (1925). Granitsy nauchnogo literaturovedeniia. *Iskusstvo*, 2, 45–60.
- Zellner, B. (1994). Pauses and the temporal structure of speech. In: E. Keller (Ed.): *Fundamentals of Speech Synthesis and Speech Recognition*. (pp. 41-62). Chichester: John Wiley.
- Zielinski, T. (1904). *Das Clauselgesetz in Ciceros Reden: Grundzüge einer oratorischen Rhythmik*. Leipzig: Dieterich’sche Verlagsbuchhandlung.
- Zielinski, T. (1914). *Der constructive Rhythmus in Ciceros Reden: Der oratorischen Rhythmik zweiter Teil*. Dieterich’sche Verlagsbuchhandlung.

APPENDIX

I Manual for coding syllable prominence and prosodic boundaries

This manual contains step-by-step instructions for coding syllable prominence and prosodic boundaries in any kind of prose text. It serves to predict how speakers realize prominences and pauses.

As for syllable prominence, we not only code whether a syllable is stressed or unstressed, but also how strongly stressed the syllable is in relation to others. Hence, this relative syllable prominence reflects the strength of a syllable compared to the strength of the neighbouring syllables. Syllable prominence is symbolized by beats (each of which is visualized by an (x) in the paper). The stronger a syllable, the more beats are assigned.

As for prosodic boundaries, the annotation is aimed at predicting where a reader will pause, as well as the length of the break. Prosodic breaks, just like syllable prominences, can occur to different degrees. The present system accounts for six degrees of prosodic boundaries.

Prominences and pauses will be annotated in a four-part process. Our annotators will first complete rules 1 to 3, which pertain to syllable prominence (see **Part 1**). After that, they will code the distribution of breaks in six steps (see **Part 2**).

This section is dedicated to the annotation of different types of prosodic boundaries. Afterwards, in **Part 3** and **Part 4**, which again pertain to syllable prominence, rules 4-8 will be applied based on the annotated boundaries. These rules include information which is to be gleaned from the surroundings of the word, including syntactic structure.

Working with a coding mask is advisable if you work with longer texts. Below, you find an example sentence in a mask as we used it (**Fig. 1**).

While we previously visualized beats as (x), breaks as |, and linkages as (~) or (–), we here represent all of the above in numbers.

The first three lines of the input mask are created prior to coding. They contain the syllabified text, word boundaries and punctuation.

Lines 5 and 6 contain information on linkages and breaks determined during phrasing in **Part 2**.

Lines 7-16 follow the rules determined for prominence annotation in **Part 1**, **Part 3**, and **Part 4**. They are summed up in line 4.

The coding process begins in line 7.

<i>syllabified text</i>	1	Das	wä	re	neu	sag	te	K.	sprang	aus	dem	Bett	und	zog	rasch	sei	ne	Ho	sen	an
<i>punctuation (behind the syllable)</i>	2	»			«			,												.
<i>word boundaries</i>	3	1		1	1		1	1	1	1	1	1	1	1	1		1		1	1
<i>overall number of beats</i>	4	4	4	1	8	5	1	8	5	3	2	8	2	5	6	3	1	9	1	4
<i>boundary blocking (Part 2)</i>	5	1	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0
<i>boundary degree (Part 2)</i>	6	0	0	0	3	0	0	2	0	0	0	1	0	0	0	0	0	0	0	4
<i>beats, rule 1 (Part 1)</i>	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<i>beats, rule 2 (Part 1)</i>	8	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	0	1	0	1
<i>beats, rule 3a (Part 1)</i>	9	0	0	0	4	3	0	5	3	0	0	5	0	3	4	0	0	5	0	0
<i>beats, rule 3b (Part 1)</i>	10	1	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	2
<i>beats, rule 3c (Part 1)</i>	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>beats, rule 4 (Part 3)</i>	12	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0
<i>beats, rule 5 (Part 3)</i>	13	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
<i>beats, rule 6 (Part 4)</i>	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>beats, rule 7 (Part 4)</i>	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1. Example of a coding mask for one sentence. The leftmost row explains what information is contained in each line.

PART 1: Syllable Prominence on the Word Level

The following rules allow for a replicable way of coding syllable prominence regardless of sentence and text length or text type. Linked to the rules for annotating phrasing, this coding system can predict where readers are likely to change their fundamental voice frequency, adjust the volume, or prolong syllables. The annotator should follow the manual closely. The parts marked in colour and bold in the examples always pertain to the step they are set to explain. Rules 1 to 3 are based on information which can mostly be inferred directly from the syllable or the word itself. If a coder feels uncertain, the phonetic transcription, word stress and

syllabification as well as part-of-speech classification (POS) can be looked up in a dictionary (Duden). When there are several possibilities, the more common one is preferred. In some cases, the surrounding words might help in determining the part-of-speech classification. Coders should always work with the spelling and the related word boundaries used in the text, regardless of the current rules of German orthography. E.g., *mithilfe*, ‘with the help of’ (old spelling), is classified as an adverb, whereas *mit Hilfe*, ‘with the help of’ (new spelling) is classified as preposition and noun.

Rule	Description	Example
1	Default beat on every syllable: visualized in numbers.	<p>+1 +1 +1 +1 +1</p> <p><i>Je</i> <i>der</i> <i>muss</i> <i>schla</i> <i>fen.</i></p> <p>Everybody must sleep</p> <p>,Everybody needs to sleep.’</p>
2	<p>Stressability: every syllable by default receives a default beat if its vowel is not a schwa and if it is not reduced.</p> <p>Schwa sounds in German are those vowels that require the least tension in the mouth area. We distinguish between e-schwa (as in <i>Birne</i> [ˈbɪʁnə], or <i>Bedienung</i> [bəˈdiːnʊŋ]), and a-schwa (as in <i>Jeder</i> [ˈjeːdə]).</p> <p>Syllables in which the schwa is the only vowel are called schwa syllables, they remain unstressed. (The e as part of the determiner <i>der</i>, ‘the’, is not considered a schwa.)</p>	<p>+1 +1 +1</p> <p>1 1 1 1 1</p> <p><i>Je</i> <i>der</i> <i>muss</i> <i>schla</i> <i>fen.</i></p> <p>Everybody must sleep</p> <p>,Everybody needs to sleep.’</p> <p>+1</p> <p>1 1</p> <p><i>A</i> <i>bend</i></p> <p>evening</p> <p>‘evening’</p>

Table 1. Part 1 of the Manual (Prominence Annotation, Rules 1-2)

Rule	Description	Example
3a	<p>Primary stress: potential phrasal prominences (3a-accent) Syllables bearing primary stress receive a different number of extra beats depending on part-of-speech classification (unlisted POS do not receive extra beats):</p> <p>5 beats:</p> <ul style="list-style-type: none"> nouns proper names (complex names, including title, form of address or job title, receive only one 3a accent, which falls onto the last word. The other words that are part of the name receive one beat, see rule 3c.) numerals (unless they are directly neighbouring nouns, proper names, adjectives or adverbs). quantifiers are treated as numerals (for a list of quantifiers see Addendum 1) <p>4 beats:</p> <ul style="list-style-type: none"> adjectives (participles without an auxiliary verb are treated as adverbs) adverb (verb particles are treated as adverbs unless they are directly neighbouring verbs, proper names, adjectives or adverbs) participles with auxiliary verb <p>3 beats:</p> <ul style="list-style-type: none"> full verbs 	<p>+5 +3 +5</p> <p>2 2 2 2 1 2</p> <p>Han na ruft Pe ter Schmitt.</p> <p>Hanna calls Peter Schmitt</p> <p>‘Hanna calls Peter Schmitt.’</p> <p>+5 +3</p> <p>2 1 2 2 1</p> <p>Je der muss schla fen.</p> <p>Everybody must sleep</p> <p>‘Everybody needs to sleep.’</p>
3b	<p>Primary stress: further prominences Syllables bearing primary stress receive a different number of extra beats depending on part-of-speech classification (unlisted POS do not receive extra beats):</p> <p>2 beats</p> <ul style="list-style-type: none"> auxiliary verbs (<i>sein</i>, ‘to be’, <i>werden</i>, ‘to become’ and <i>haben</i>, ‘to have’, always receive 2 beats) particles (<i>nicht</i>, ‘not’, is treated as a negating particle) verb particles directly neighbouring nouns, proper names, adjectives, or adverbs demonstratives <p>1beat</p> <ul style="list-style-type: none"> prepositions, indefinite articles pronouns (all kinds except of demonstrative pronouns) quantifiers directly neighbouring nouns, proper names, adjectives, or adverbs 	<p>+2</p> <p>7 1 2 5 1</p> <p>Je der muss schla fen.</p> <p>Everybody must sleep</p> <p>‘Everybody needs to sleep.’</p> <p>+1</p> <p>2 1 7</p> <p>Je der Mensch...</p> <p>Every human...</p> <p>‘Every human...’</p>

Table 2. Part 1 of the Manual (Prominence Annotation, Rules 3a-b)

Rule	Description	Example
3c	<p>Secondary stress Syllables bearing secondary stress (the second strongest syllable in a multi-syllable word) receive one extra beat depending on part-of-speech classification (unlisted POS do not receive extra beats):</p> <p><i>1 beat</i></p> <ul style="list-style-type: none"> • nouns (compounds) • proper names (compounds and complex proper names) • numerals, adjectives, adverbs, full verbs <p>Please note that not all words have secondary stress/ a secondary accent. Here, secondary stress is only annotated if at least one syllable with no more than 2 beats lies between primary and secondary accent (e.g., <i>Vórtragsàbend</i>, ‘evening of recitals’, bears secondary stress, whereas <i>Vóràbend</i>, ‘eve’, does not).</p>	<p style="text-align: right; color: red;">+1</p> <p>7 2 2 1 <i>Vor trags a bend</i> evening of recitals ‘evening of recitals’</p> <p style="text-align: right; color: red;">+0</p> <p>7 2 1 <i>Vor a bend</i> eve ‘eve’</p> <p style="text-align: right; color: red;">+1</p> <p>7 2 5 2 1 7 <i>Han na ruft Pe ter Schmitt.</i> Hanna calls Peter Schmitt ‘Hanna calls Peter Schmitt.’</p>

Table 3. Part 1 of the Manual (Prominence Annotation, Rule 3c)

PART 2: Phrasing (Annotation of Prosodic Boundaries)

Whereas in the previous sections, beats per syllable were annotated in order to code the respective syllable prominence, this section is dedicated to the annotation of different types of prosodic boundaries between words. Thus, it is aimed at predicting where a reader will pause, as well as the length of the break. Prosodic breaks, just like syllable prominences, can occur to different degrees.

The present system accounts for six degrees of prosodic boundaries, increasing from unlikely locations for breaks, symbolized by (~) or (–), neutral locations (word boundaries), shortest (symbolized by |) to longest breaks (symbolized by ||||). Breaks || – |||| are coded based on punctuation marks and number of words (**Step 1**). The

shortest type of break (|) is more complex and will be explained in **Steps 2-7** of the phrasing manual. Phrase boundaries form the smallest breaks, i.e. degree (|), and as such they occur without punctuation marks.

In the following, the rules for phrasing will be presented. In order to obtain correct results, these rules should be followed step by step in chronological order. A phrase boundary once set will not be removed. In the examples, potential phrasal prominences (see 3a accents in **Part 1**) are marked in bold. They will serve as candidates for the annotation of phrasal and sentence accents in **Part 3**.

<i>Step</i>	<i>Description</i>	<i>Example</i>
1	<p>Annotation of breaks – </p> <ul style="list-style-type: none"> • behind full stops / semicola / cola / exclamation marks / question marks • behind all other punctuation marks if they are preceded by at least three words without punctuation mark • behind all other punctuation marks if they are preceded by no more than two words without punctuation mark. (For opening quotation marks or brackets, the break is placed before the punctuation mark.) 	<p><i>Hanna, Mara Jörg und Jens werden von Peter gerufen, obwohl er heiser ist. </i> Hanna Mara Jörg and Jens are by Peter called although he hoarse is ‘Hanna, Mara, and Jörg are being called by Peter, although he is hoarse.’</p> <p><i>Er staunte (mit offenem Mund) und ging weg. </i> He wondered with open mouth and went away ‘He stood amazed (his mouth agape) and then left.’</p>
1a	<p>The breaks are not added up.</p> <p>This means that in the case of several punctuation marks in a row, only the strongest will be counted. (Several dots marking an ellipsis as well as single dots marking abbreviations will be ignored.)</p>	<p><i>„Das wäre neu!“, sagte K.</i> That would (be) new said K ‘That would be new!’ said K.’</p>
1b	<p>Relation to the potential phrasal prominences (Rule 3a in Part 1)</p> <p>Through the annotation of breaks, smaller or larger units are formed. All of these units should have at least one 3a accent (in bold, cf. rule 3a in Part 1). If this is not the case after the assignment of breaks, the respective unit will be assigned a 3a accent (three beats) – if the unit is formed of several elements, the 3a accent will be assigned to the element farthest to the right (color highlighted in red in the example). (If the element already has a 3b accent, this accent is eliminated.)</p>	<p><i>Die Tiere krabbelten auf die Stühle, auf sie, auf ihn und auf den Tisch. </i> The animals crawled on the chairs on her on him and on the table ‘The animals crawled on the chairs, on her, on him and on the table.’</p>

Table 4. Part 2 of the Manual (Phrasing, Step 1)

Step	Description	Example
2	<p>Annotation of first-degree breaks () and the prefield</p> <ul style="list-style-type: none"> behind the prefield if it consists of at least two elements and if at least one of these elements has a 3a accent (in the examples in bold). <p>A note regarding the prefield: The Topological Field Model explains the word order in German sentences. The sentence is segmented into five successive fields – for our purpose, however, the distinction between prefield and left bracket is sufficient.</p> <p>In German main clauses, the position called “left sentence bracket” is filled by the finite verb (in the examples above, the finite verbs are “hat”/”hatte” and “weint”). All elements <u>before</u> the finite verb form the prefield.</p> <p>In the examples above, this would be <i>a. Unser Freund</i>, <i>b. Er</i>, <i>c. Schon lange</i>, ‘For a long time already’, <i>d. Lange</i>, ‘For a long time’, <i>e. Das Kind</i>, ‘the child’, and <i>f. Peter</i>.</p> <p>However, a phrase boundary is only placed behind the prefield of a., c. and e., because the rule above states the precondition that the prefield should consist of at least two elements, one of them having a 3a accent (in bold).</p>	<p><i>a. Unser Freund hat den Verdacht, dass wir ihn alle hassen.</i> Our friend has the suspicion that we him all hate</p> <p>‘Our friend has the suspicion that we all hate him.’</p> <p><i>b. Er hat den Verdacht, dass wir ihn alle hassen.</i> He has the suspicion that we him all hate</p> <p>‘He has the suspicion that we all hate him.’</p> <p><i>c. Schon lange hat unser Freund den Verdacht, dass wir ihn alle hassen.</i> Already long has our friend the suspicion that we him all hate</p> <p><i>d. Lange hat unser Freund den Verdacht, dass wir ihn alle hassen.</i> Long has our friend the suspicion that we him all hate</p> <p>‘For a long time (already), our friend has had the suspicion that we all hate him.’</p> <p><i>e. Das Kind weint.</i> The child cries</p> <p>,The child cries.’</p> <p><i>f. Peter weint.</i> Peter cries</p> <p>‘Peter cries.’</p>

Table 5. Part 2 of the Manual (Phrasing, Step 2)

Step	Description	Example
3	Blocking of phrase boundaries, symbolized by (~) and (–) Blockings are placed in order to mark positions where no phrase boundaries will be allowed in the subsequent steps. Blockings marked by (~) are not removable once set whereas those marked by (–) can be removed during the further process of phrasing.	
3a	(~) Complex names <ul style="list-style-type: none"> Names consisting of several words may not be separated by a phrase boundary. 	<i>Johann ~ Wolfgang ~ Goethe.</i> Johann Wolfgang Goethe ‘Johann Wolfgang Goethe’
3b	(~) Prepositions <ul style="list-style-type: none"> Phrase boundaries may not be placed directly after prepositions. (This rule does not apply to postpositions.) 	<i>Die unter ~ der Wucht der Schüsse <i>splitternd</i> aufsprang.</i> Which under the force theGEN shotsGEN splintering open burst ‘Which splintered and burst open under the force of the shots.’
3c	(~) Determiners, possessive pronouns and demonstrative pronouns <ul style="list-style-type: none"> No separation of determiners / possessive pronouns / demonstrative pronouns from a directly neighbouring reference word. 	<i>Der ~ Peter mag sein ~ Haus. Der ~ Peter mag dieses ~ Haus.</i> The Peter likes his house. The Peter likes this house ‘Peter likes his house. Peter likes this house.’
3d	(~) Reflexive pronouns, object pronouns, personal pronouns & verb particles <ul style="list-style-type: none"> No separation of two directly neighbouring pronouns. No separation of a verb and a directly neighbouring pronoun. No separation of a full verb and a directly neighbouring verb particle. 	<i>Gestern hat er ~ sich ~ <i>rasiert</i>. Die ~ Hanna <i>trägt</i> ~ ihn.</i> Yesterday has he himself shaved. The Hanna carries him ‘Yesterday, he shaved.’ ,Hanna carries him.’ <i>Die ~ Hand <i>schwoll</i> ~ an und <i>schmerzte</i>.</i> The hand swelled up and hurt ‘The hand swelled up and hurt.’
3e	(~) Numerals <ul style="list-style-type: none"> No separation of numerals (see Prominence-Annotation) and directly neighbouring nouns. 	<i>Jedes ~ Kind, alle ~ Menschen, <i>zwanzig</i> ~ Tiere.</i> Every child all humans twenty animals ‘Every child, all people, twenty animals.’

Table 6. Part 2 of the Manual (Phrasing, Step 3a-e)

<i>Step</i>	<i>Description</i>	<i>Example</i>
3f	<p>(–) Genitive constructions</p> <ul style="list-style-type: none"> No separation of genitives and their left-hand reference word <p>Please note: This rule applies only if the elements of the genitive construction <u>do not have any modifiers</u>.</p>	<p>a. <i>Weil er aber in ~ der ~ Aufregung – dieses ~ Morgens ...</i> Because he but in the excitement thisGEN morningGEN</p> <p>‘But because, in the excitement of this morning...’</p> <p>b. <i>...in ~ der ~ Aufregung dieses ~ <u>wundervollen</u> Morgens...</i> in the excitement thisGEN wonderful morningGEN</p> <p>‘...in the excitement of this wonderful morning...’</p> <p>c. <i>...wegen der ~ Charakterlosigkeit und Minderwertigkeit des ~ Inspektors.</i> due theGEN character lack and inferiority theGEN inspectorGEN</p> <p>‘...due to the inspector’s lack of character and his inferiority.’</p>
3g	<p>(–) No separation of directly neighbouring words of the following word classes (to be applied chronologically)</p> <ul style="list-style-type: none"> adjective – noun, (no linkage of noun – adjective) adverb – full verb, full verb – adverb, adjective – adverb, adverb – adjective, adverb – adverb, adjective – adjective <p>Note: Adjectives used with an adverbial function are counted as adverbs.</p>	<p><i>Die unter ~ der ~ Wucht der ~ immer – noch – fallenden – Schüsse splitternd – aufsprang.</i> Which under theGEN force theGEN still falling shots splintering burst open</p> <p>‘Which splintered and burst open under the force of the gunshots still ringing out.’</p>

Table 7. Part 2 of the Manual (Phrasing, Step 3f-g)

Step	Description	Example
4	Annotation of first-degree breaks triggered by function words	
4a	<p>Clause-initial conjunctions</p> <ul style="list-style-type: none"> before conjunctions that introduce a new verb phrase with corresponding arguments (finite verb + arguments). <p>This rule applies only if there is at least one more element between the conjunction and the next punctuation mark.</p>	<p>...auf ~ ihre ~ schwarze – Bluse tropften und dort – noch – schwärzere – Spuren hinterließen...</p> <p>on her black blouse dripped and there even blacker traces left</p> <p>‘...dripped onto her black blouse and left even blacker traces there...’</p>
4b	<p>Comparative particles "als" / "wie"</p> <ul style="list-style-type: none"> before the comparative particles "als" and "wie" if they follow a verb or a verb particle and are followed by at least four syllables (before the next punctuation mark). 	<p><i>Dass er häufiger gelobt sein will als alle anderen.</i></p> <p>That he more often praised be wants than all others</p> <p><i>Dass er häufiger als alle anderen gelobt sein will.</i></p> <p>That he more often than all others praised be wants</p> <p>‘That he wants to be praised more often than anyone else.’</p>
4c	<p>Prepositions</p> <ul style="list-style-type: none"> between participles / infinitives / verb particles (on the one hand) and prepositions (on the other) if the preposition refers to (introduces) a word with a 3a accent (in the examples in bold). 	<p><i>Dann erst dürfen wir aufhören mit unserem Lob.</i></p> <p>Then only may we stop with our praise</p> <p>‘Only then are we allowed to stop with our praise.’</p> <p><i>Ich habe nichts gesagt über den Inspektor. Ich habe nichts gesagt über ihn.</i></p> <p>I have nothing said about the inspector I have nothing said about him</p> <p>‘I didn't say anything about the inspector / him.’</p> <p><i>Steigt aus, lässt einen allein zurück in einer Welt ohne jeden Reiz.</i></p> <p>Backs out leaves one alone behind in a world without any stimulus</p> <p>‘Backs out, leaves you behind alone in a world without any charm.’</p>

Table 8. Part 2 of the Manual (Phrasing, Step 4)

Step	Description	Example
5	<p>Annotation of first-degree breaks (Active phrasing)</p> <ul style="list-style-type: none"> Procedure: Breaks can be placed in positions that have not been blocked by (~) or (–) an in a previous step. Phrasing is carried out per sentence from right to left; every phrase has to carry at least one 3a accent. Elements without a 3a accent are integrated into the phrase to the right as long as they are not tied to the left phrase by a blocking. Number of 3a accents (in the examples in bold): Ideally, exactly two elements with a 3a accent are placed within a phrase. In the case of adjectives or adverbs that are not tied to other material by blockings, the ideal number of 3a accents is three. (Due to blockings, however, phrases with more than three 3a accents may emerge.) 	<p><i>Sie gebe ja zu, sagte die ~ Frau, während ihr die ~ Tränen über ~ das ~ Gesicht liefen, über ~ Mundwinkel</i> She admits said the woman while her the tears over the face ran over mouth corners</p> <p><i>und Kinn auf ~ ihre ~ schwarze – Bluse tropften und dort noch – schwärzere – Spuren hinterließen, sie</i> and chin on her black blouse dripped and there even blacker traces left she</p> <p><i>gebe ja zu, dass ihr ~ Adi ein ~ wilder – junger – Mensch gewesen sei.</i> admits that her Adi a wild young human had been</p> <p>‘She admitted, said the woman, while tears ran down her face, dripped over the corners of her mouth and chin onto her black blouse, and left even blacker traces there, she admitted that her Adi had been a wild young man.’</p> <p><i>Die unter ~ der ~ Wucht – der ~ Schüsse splitternd – aufsprang.</i> Which under the force the shots splintering burst open ‘Which splintered and burst open under the force of the shots.’</p> <p><i>Die unter ~ der ~ Wucht der ~ immer – noch – fallenden – Schüsse splitternd – aufsprang.</i> Which under the force theGEN still fallingGEN shots splintering burst open ‘Which splintered and burst open under the force of the gunshots still ringing out.’</p> <p><i>...auf ~ den ~ Steinplatten – des ~ Kirchenbodens blutige – Tritte...</i> on the stone slabs theGEN church floor bloody footprints ‘...bloody footprints on the stone slabs of the church floor...’</p>

Table 9. Part 2 of the Manual (Phrasing, Step 5)

Step	Description	Example
6	Annotation of first-degree breaks (Length restrictions) In this final step of phrasing, unnaturally long phrases are divided according to 6a-b.	
6a	Number of syllables / words: Phrases with more than 12 syllables and at least 5 words are segmented in such a way that each of the two new phrases has at least one 3a accent. If a segmentation would result in a phrase without 3a accents, the segmentation is not carried out. <ul style="list-style-type: none"> ○ Annulment of blockings: The (–) blockings of breaks carried out in steps 3f and 3g are annulled if segmentation is not otherwise possible. 	<p><i>Dass einer andauernd – gelobt sein will für etwas.</i> That one always praised be wants for something</p> <p>Correction: <i>Dass einer andauernd gelobt sein will für etwas.</i> That one always praised be wants for something</p> <p>‘That someone wants to be praised for something all the time.’</p>
6b	Balance: If there is more than one possibility of segmentation (because the phrase has more than two 3a accents that are not tied together by blockings), the segmentation is carried out in a way that leads to two phrases the length of which (in number of syllables) is as equal as possible. <ul style="list-style-type: none"> ○ Short before long: If there are several possibilities of balancing, the rule "short before long" applies (segmentation in such a way that the left phrase is shorter than the right). 	<p><i>Sie gebe ja zu, dass ihr X Adi ein~wilder~junger– Casanova gewesen sei. </i> She admits that her Adi a wild young Casanova had been</p> <p>Correction: <i>Sie gebe ja zu, dass ihr ~ Adi ein~wilder junger Casanova gewesen sei. </i> She admits that her Adi a wild young Casanova had been</p> <p>‘She admits that her Adi had been a wild young Casanova.’</p>

Table 10. Part 2 of the Manual (Phrasing, Step 6)

PART 3: Syllable Prominence Based on Phrasing (Phrasal and Sentence Accents)

The different levels of boundaries that are now annotated in the text allow for dividing it into phrases, clauses, and sentences. These units receive different types of accents that are, again, annotated as beats. For greater simplicity, these beats are only marked in capitals (rule 4) and in bold (rule 5) in the following examples. Please

note that rule 4 and 5 have the same wording. However, rule 4 (phrasal accents) refers to units divided by all kinds of boundaries (|-|||), whereas rule 5 (sentence accents) exclusively refers to units divided by ||| and |||| boundaries. Hence, a sentence accent is always added onto a phrasal accent

Rule	Description	Examples
4-5	<ul style="list-style-type: none"> Phrasal accents (rule 4): pertain to all units limited by a boundary of any degree (-). In the examples, phrasal accents are capitalized. Sentence accents (rule 5): pertain to the material limited by break types and , and added to phrasal accents. In the examples, sentence accents are in bold. <p><i>1 beat on the primary accent of the last word in a unit, if it belongs to the following POS:</i></p> <ol style="list-style-type: none"> Nouns/ proper names/ numerals Adjectives/ adverbs Full verbs (including participles) which come together with an auxiliary verb (need not be neighbouring or part of the same phrase) Full verbs used as intransitively and without any further words other than the subject. All other full verbs unless they directly follow an adjective, adverb, or noun within the same phrase Full verbs in the infinitive which directly precede a full verb Full verbs in the infinitive with an inserted <i>zu</i>, ‘to’ Elements bearing potential phrasal prominence (3a accent) 	<ol style="list-style-type: none"> <i>Die kleine HANna ruft PETER. </i> The little Hanna calls Peter ‘Little Hanna calls Peter.’ <i>Die kleine HANna ruft Peter LAUT. </i> The little Hanna calls Peter loudly ‘Little Hanna calls Peter loudly.’ <i>Die kleine HANna wird von Peter geRUFen. </i> The little Hanna is by Peter called ‘Little Hanna is being called by Peter.’ <i>Peter LACHT. Peter lacht LAUT. </i> Peter laughs Peter laughs loudly ‘Peter laughs.’ ‘Peter laughs loudly.’ <i>Weil er sie ABholt, ...</i> Because he her picks up ‘Because he picks her up, ...’ <i>er will sie TANzen sehen. </i> he wants her dancing see ‘he wants to see her dancing.’ <i>um langsam FESTzustellen, ...</i> to slowly find out ,to find out slowly, ...’ <i>..., was wir SIND, </i> what we are ... ‘what we are,’ ...

Table 11. Part 3 of the Manual (Prominence Annotation, Rule 4-5)

PART 4: Final Touches on Syllable Prominence

Rule	Description	Example
6	Utterance-initial function words receive 1 beat <ul style="list-style-type: none"> We define function words as any word that is neither noun nor full verb, adjective, or adverb. We define utterance as a unit between two breaks of the strongest degree, i.e. . 	+1 2 6 <i>Er</i> <i>lacht.</i> He laughs ‘He laughs.’
7	RhythmRule <ul style="list-style-type: none"> If three or more syllables with precisely 2 beats (per syllable) follow each other within a phrase, every other syllable receives 1 beat. 	 7 2 +1 2 8 <i>Fahr</i> <i>rad</i> <i>und</i> <i>Re</i> <i>gal</i> bike and shelf ‘bike and shelf’

Table 12. Part 4 of the Manual (Prominence Annotation, Rule 6-7)

Addendum 1: List of quantifiers (German)

- eins/eines, zwei, drei...
- erste(r), zweite(r), dritte(r)....
- doppelt/zweimal/zweifach, dreimal/dreifach...
- zu zweit, zu dritt, zu viert...
- erstens, zweitens, drittens...
- and indefinite numerals: wie einige(s), beide(s), manche(s), verschiedene(s), mehrere(s), viel(es), zahlreiche(s), zahllose(s), unzählige(s), etliche(s), [ein] paar, einzeln(e/s), vereinzelte(s), wenige(s), gesamte, etwas (im Sinne von „ein bisschen“), ein bisschen, weitere(s), sonstige(s), übrige(s), ganze (im Sinne von „alle“: „die ganzen Sachen“), alle(s), sämtliche(s), jede(r), nichts, keine(r)/keins, niemand

II Additional information concerning the validation

The prominence scale employed in this coding system derives syllabic prominence from two main sources and their interplay: (sub)lexical stress and phrasal prominence as subsumed in the hierarchy in (1) and part of speech of the prominent syllable as subsumed in the hierarchy in (2). However, since we included various additional factors such as neighbouring POS and speech rhythm, these sources provide only a rough approximation.

As for lexical stress and phrasal accent, we distinguish, for every syllable, six degrees of stress and accent (seven with the potential prominence targets, here included in syllables bearing primary stress), in line with extant work on German phonology:

(1) unstressable schwa-syllables >> unstressed syllables with full vowels >> syllables bearing secondary stress >> syllables bearing primary stress >> syllables bearing phrasal accent >> syllables bearing nuclear or sentence accent.

Moreover, previous work predicts differences in the degree of prominence on syllables bearing primary stress, depending on the part of speech of the word. In line with this research, the present coding system is based on the predictions of the following hierarchy (with several additions as outlined in the manual).

(2) Nouns/proper names >> adjectives/adverbs >> full verbs >> auxiliary verbs, particles, demonstrative pronouns >> prepositions, indefinite articles, other pronouns.

In the following, we examine the effect of a) stress/accent level and b) POS on three phonetic variables that conjointly determine syllabic prominence, namely duration, pitch (F0 range on the syllable), and maximum intensity of the syllable. To this end, we fit a suite of mixed effects models with Author and Speaker as grouping terms (random intercepts), and 1) stress/accent level and 2) position in the POS-hierarchy as ordinal predictor variables.

Figure 1 shows the phonetic expression of the stress/accent hierarchy in terms of duration (upper panel), F0 range (middle panel), and maximum intensity (lower panel) of the respective syllables. Figure 2 shows the phonetic expression of the part of speech hierarchy. In the latter case, only the syllables with primary stress that lack phrasal or sentence accent entered the analysis; in this way, we avoid that the tendency for nouns to bear phrasal/sentence accent confounds the part-of-speech effect. The mixed effects models (Gamma

mixed effects regressions in the case of duration and F0 range data; linear mixed model in the case of intensity) are tabulated in Table 1.

While associated with longer syllable duration, nuclear accents exhibit lower F0 range and lower maximum intensity compared to phrasal accents, against predictions of the stress/accent hierarchy. The distinction between unstressed syllables with full vowels and syllables exhibiting secondary stress is clearly expressed in terms of duration (longer duration for secondary stress), but not in terms of F0 range (significant effect against predictions) or intensity (no significant difference). Otherwise, the predictions of the stress hierarchy hold with significant differences in the predicted directions in at least two of the three dependent variables.

As for the part of speech hierarchy on the word level, the predictions largely hold as well. One exception concerns the difference in duration between full verbs and adjectives/adverbs, with full verbs having longer stressed syllables. However, please note that the Manual includes several additions concerning verbs and their phrasal prominence as detailed in Appendix 1 (Table 11). The other dependent measures (F0 range, intensity) turn out as predicted.

Figure 1. Phonetic expression of the stress/accent hierarchy in terms of duration (upper panel), F0 range (middle panel), and maximum intensity (lower panel) of the respective syllables.

Figure 2. Phonetic expression of POS hierarchy in terms of duration (upper panel), F0 range (middle panel), and maximum intensity (lower panel) of the respective syllables.

Successive differences	Syllable duration (Gamma regression with log link)				F0 range (Gamma regression with log link)				Maximum intensity			
Stress/Accent hierarchy	Beta	SE	t	p	beta	SE	t	p	beta	SE	t	p
Schwa – unstressed	.19	.004	42.98	<.001*	<i>-.001</i>	<i>.01</i>	<i>-0.11</i>	<i>.91</i>	.23	.043	5.33	<.001*
Unstressed – 2dry	.14	.008	17.06	<.001*	-.08	.02	-4.67	<.001*	<i>-.11</i>	<i>.083</i>	<i>-1.295</i>	<i>.195</i>
2dry – primary	-.08	.008	-9.53	<.001*	.22	.018	12.46	<.001*	1.12	.08	13.95	<.001*
Primary – phrasal	.19	.006	31.21	<.001*	.25	.014	17.85	<.001*	1.36	.062	21.85	<.001*
Phrasal – nuclear	.018	.007	2.47	.014*	-.04	.016	-2.55	.011*	-1.19	.074	-16.16	<.001*
POS hierarchy												
PRON – AUX	.2	.007	29.52	<.001*	.06	.016	3.9	<.001*	.83	.069	11.94	<.001*
AUX – Verbs	.11	.009	11.59	<.001*	.07	.023	3.09	<.001*	.37	.098	3.77	<.001*
Verbs – ADJ/ADV	-.08	.009	-8.17	<.001*	.09	.023	3.94	<.001*	.27	.097	2.75	.006
ADJ/ADV – Nouns	.09	.009	10.58	<.001*	.04	.022	1.87	0.06	.17	.096	1.75	.08

Table 1. Upper panel: summary statistics for coefficients of interest (successive differences between six degrees of stress/accent) of three mixed effects models, computed in *R* using the *lme4* package (Bates et al. 2015): 1) a mixed Gamma regression on syllable duration (with log link), 2) a mixed Gamma regression model on F0 range, and 3) a linear mixed effects model on maximum intensity. *Lower panel:* The same models with the POS-hierarchy as predictor variable (successive differences between five POS-classes). The p values were estimated using the *lmerTest* package (Kuznetsova et al. 2017). Significant effects against predictions are shaded, non-significant effects are printed in italics.

FOOTNOTES

¹ Following Windmann et al. (2011, above others) we use the term ‘beat’ as an auxiliary unit which can be accumulated with other beats and thus serves to describe the degree of syllable prominence (as opposed to the ‘beat’ in the rhythm of music). In this understanding, a syllable with one beat is less prominent than one with two or more beats, etc.

² Previous work in prosodic phonology assumes a type of phrase formation in which syntactic heads become prosodic heads around which the rest of the phrase is gathered, as long as the syntactic heads belong to particular word types and have no prominent neighbors (Selkirk 1984; Nespor & Vogel 1986). In longer and recursive structures, however, this account is not always easy to trace. A pragmatic solution is the creation of clearly defined chunks that roughly resembles phrase formation (e.g., Abney 1996). A chunk cannot be internally separated by a prosodic boundary without changes to its semantic content (Bachenko & Fitzpatrick 1990). An extant complex chunk grammar for German with more than 4500 rules for automatic annotation (Schmid & Schulte im Walde 2000) could not be sufficiently reduced to serve for our manual annotation.

³ Put both briefly and very generally, Hayes’ (1989) algorithm defines how certain syntactic categories form phonological phrases (depending on POS and complexity). To make the task easier on our annotators and leave less room for error, we replaced these syntactic categories with our category of potential prominence targets (see also Wheeldon 2000).

⁴ Minimally, a simple phrase (closed off by a level 1 break) must contain one such target. The maximal number is two – three if an adverb or adjective is involved, see chunking algorithms in Atterer (2002).

⁵ Please note that each one of the eight annotators coded only some of the texts, resulting in three independent annotations for each text.