Embarking on PoLaR Explorations:
A Framework for Intonational Annotation and Analysis

Byron Ahn,
Nanette Veilleux,
Stefanie Shattuck-Hufnagel,
and Alejna Brugos

draft
November 2022

Contents

List of Figures

List of Tables

Chapter 1

PoLaR Beyond these Current Guidelines

Up to this point, PoLaR labels have been (purposefully) defined in a way that allows them to be used with a variety of phonological analyses. For example, the PrStr tier labels for prominences and boundaries are not defined according to any particular phonological inventory, but rather encode distinctions that are widely adopted by a variety of theorists and that apply widely across languages and language varieties.

While this reduction in theoretical commitments is useful for ensuring that labels produced by annotators working in a variety of phonological frameworks can be compared, PoLaR labellers may also want to annotate more detail and insight, as informed by other frameworks or theories. PoLaR labelling systematically places all prosodic annotations in one place, which is useful for later analysis, and its extensibility provides the option of adding new labelling options, while ensuring that these additional labels can be easily ignored by other labellers who prefer to do so, and also (critically) by PoLaR-related scripts.

To allow PoLaR users to do this, while preserving PoLaR as a single prosodic annotation system whose labels can be compared across labelling groups, we offer two means of extending PoLaR. First, users can add "tags" to existing PoLaR labelling tiers, that take on a form like #X, where X is any parameter of interest, and the # symbol identifies the label as a non-core PoLaR annotation. Second, a user can add additional tiers where appropriate. In the remainder of this section, these two methods of extending PoLaR will be illustrated with examples from several different domains. The illustrative labels and tiers introduced in this section are not to be taken as a core aspect of PoLaR. None of them are intended as necessary for PoLaR annotation.

1.1 Tagging a Deeper Analysis (Expert-Level Labellers Only)

For certain goals and datasets, it may be useful to systematically label more specific analyses. To allow this flexibility, while also remaining consistent to enable comparison across labellers, an 'analysis tag' —introduced with a '#' delimiter— can be added to an existing PoLaR labelling tier.

Tag	Significance
#X	The PoLaR label to which this tag has been added is being analyzed as having property X (where X is a variable).

Table 1.1: A proposal for a means to encode additional grammatical analyses. (Note that these #X tags can be stacked, to create a PrStr label such as]#X#Y#Z, indicating a perceived boundary that has characteristics X, Y and Z.)

For example, a labeller who would like to distinguish nuclear stress prominences can augment the PrStr * label to create *#NS. Below we discuss several particular cases where we anticipate that such 'analysis tags' may be useful.

1.1.1 Categories of Prosodic Phrasing

English has been argued to make reference the following elements prosodic structure, in its intonational phonology:

A Prosodic Hierarchy of English (cf. Nespor & Vogel 1986, Shattuck-Hufnagel & Turk 1996, Féry 2017)
syllable (σ) < foot (Σ / Ft) < word (ω) < phonological phrase (φ) / intermediate phrase (ip) < intonational phrase (ι / IP) < Utterance (υ / Utt)

A labeller who adheres to this (or some other) prosodic hierarchy may wish to keep track of particular prosodic boundaries in PoLaR labels, instead of the perceptually (and not phonologically) defined labels like] (or [,]], and [[). This can be achieved with analysis tags like #IP, to create PrStr annotations like]#IP, to indicate the right edge of an intonation phrase.

These sorts of tags may be especially useful for languages that make use of more than two phrasal levels of prosodic hierarchy (e.g., a labeller of Korean may wish to use]#AP,]#ip, and]#IP; see Jun 2007), or for languages/varieties/data whose prosodic hierarchy is less clear but for which more than two hierarchical elements can be identified. A labeller of such data where the phonological analysis is still uncertain may wish to use labels like]#A,]#B,]#C, etc., to remain agnostic about particular analyses.

Two examples are given below. The first is an example from MAE_ToBI training materials (se also Beckman et al. 2005). In this Advanced PoLaR annotation, each] on the PrStr tier is annotated with MAE_ToBI's numerical labels for constituent categories in the prosodic hierarchy: #2, #3, and #4.
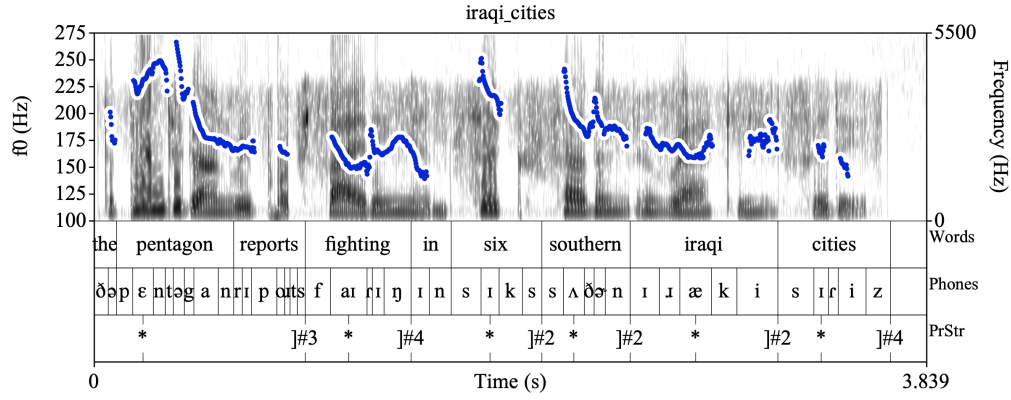
Figure 1.1: `iraqi_cities`, with analysis tags showing the MAE_ToBI break indices on the phrase boundary labels.

The second example comes from Korean. In this example, each] on the PrStr tier is marked with one of the boundary types in the analysis proposed in Jun 2007, including]#AP,]#ip, and]#IP.
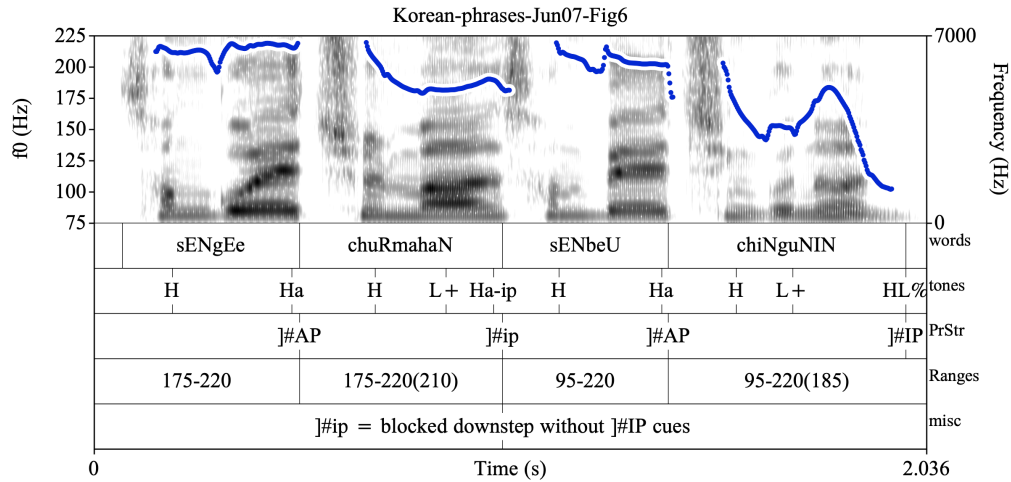
Figure 1.2: `Korean-phrases-Jun07-Fig6`, with analysis tags showing the proposed addition of ip boundary types to K_ToBI, as in Jun 2007 Fig.6, with PoLaR labels added.

1.1.2 Categories of Phrase-Level Prominence

The literature on prosodic prominence above the word level has recognized that different strengths of prominence may be linked to different constituents in the prosodic hierarchy. For example, some traditions (e.g., Liberman & Prince 1977, Halle & Vergnaud 1987, *et seqq.*) state as a principle that in each domain of prosodic hierarchy, the grammar identifies a uniquely prominent element (i.e., "the head"), which often manifests as acoustically/perceptually stronger than the other elements. Such analyses predict that each prosodic phrase has a perceptually/ grammatically "most prominent" element, in addition to the other "prominent" elements. (This approach is directly related to ideas of "nuclear stress" and "nuclear pitch accent", as well as

3

concepts of "focus stress" and "focus pitch accent"; we return to issues of focus in the next section.)

Given such claims, an advanced labeller may find it useful to identify as many levels of prosodic strength as there are prosodic domains in the prosodic hierarchy. To do so, one can add analysis tags to * labels, to create PrStr labels like *#ɩ prominences for the strongest * in an intonation phrase (ɩ), *#φ prominences for the strongest prominence in a phonological phrase (φ), etc., alongside a plain * for elements that are "prominent" without being most prominent in any particular domain. An example is given below:
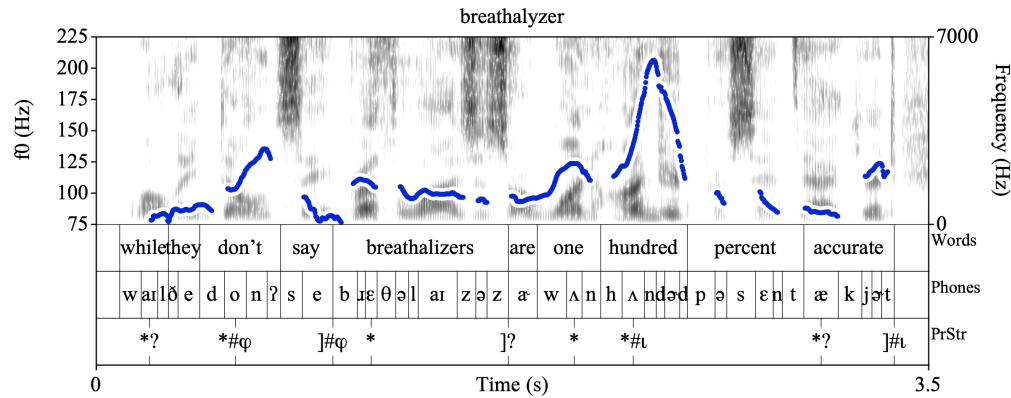

Figure 1.3: `breathalyzer_BURNC`, showing the first half of the recording, with analysis tags identifying which prominences are strongest in a domain, and which domain each is strongest for.

In this example, the word "*don't*" carries the strongest prominence in the phonological phrase "*while they don't*", but "*hundred*" carries the strongest prominence in the intonation phrase "*while they don't say breathalyzers are one hundred percent accurate*". This sort of grammatical tagging reflects a particular theoretical point of view, and may be subject to a fair amount of disagreement. For this reason, such tags are not a core part of PoLaR. At the same time, PoLaR provides a means for systematically labelling these distinctions, for those who are interested (and who can provide a theoretical stance that allows them to label the proposed distinctions systematically).

The second half of this sound file illustrates a potential distinction between phonological phrases (φ) and intonational phrases (ɩ), as well as the most prominent stress in each of these phrasal domains. Labels that reflect this distinction are displayed in Figure 1.4.
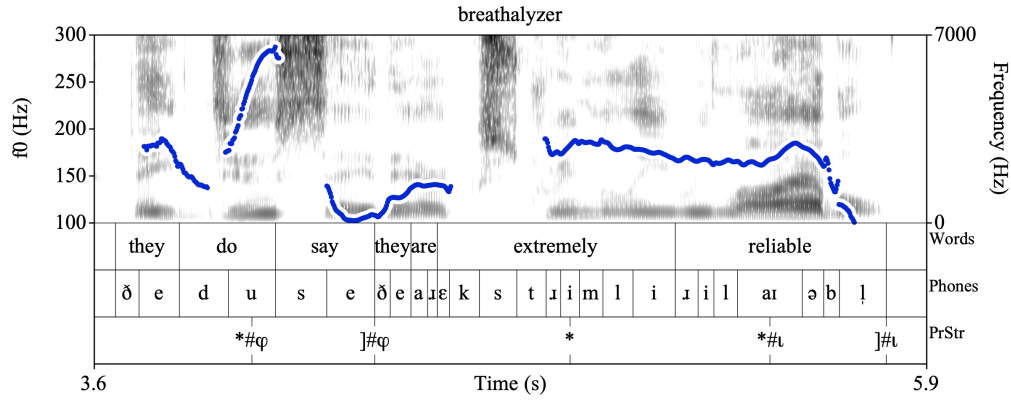
4

Figure 1.4: `breathalyzer_BURNC`: This is the second half of this recording; the labeller analyzes the first phrase boundary as lower on the hierarchy than the boundary at the end.

1.1.3 Contrastive Focus Prosody

In terms of semantic interpretation, contrastive focus is often treated as a grammatical primitive, resulting in the calculation of semantic alternatives (see, e.g., Rooth 1992, Truckenbrodt 1995, Wagner 2005, Kratzer & Selkirk 2020). In terms of prosody, a variety of languages have been argued to mark contrastively focused elements with particular prosodic forms. For example, contrastive focus may manifest as a particularly strong prominence, which may have a particular intonational form (Pierrehumbert & Hirschberg 1990 for American English, Baltazani & Jun 1999 for Greek, etc.) or which may interact in particular ways with surrounding prosodic structure (Ishihara 2007 for Japanese, Kanerva 1990 for Chichewa, etc.). Taking this idea into account, labellers may wish to systematically label PrStr events with the optional #CF tag. An English example is provided below.
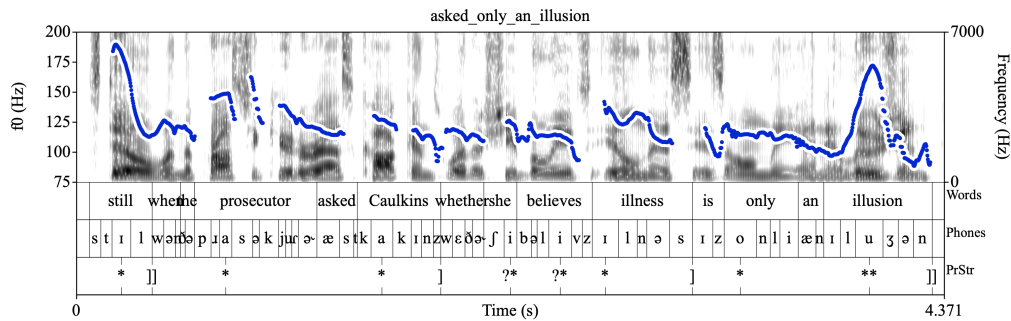
Figure 1.5: The labeller has added the 'analysis tag' #CF, to indicate that the prominence on "illusion" expones the grammatical category of contrastive focus.

Contrastive focus is analyzed as having a grammatical relationship with prominence (and not with phrase boundaries) in English, as the *#CF label notates in Figure 1.5 above. On the other hand, contrastive focus in other languages is (also) grammatically related to phrase boundaries (e.g., contrastively focused constituents in Korean are marked with an additional phrase boundary; Jeon & Nolan 2017); in such languages, the #CF tag should (also) be added to the appropriate] or [label.

1.1.4 Summary of Some Proposed Analysis Tags

We have reviewed some examples of potential #X t tags as a way of annotating grammatical analyses made by the labeller. Some of the types of tags discussed here are listed in Table 1.3.

Label	Phonological Object	Time-Aligned with
]#X	Phrase Boundary for a Phrase of Type "X"	A prosodic phrase ends at this point; the type of prosodic phrase is specified by "X" (e.g.,]#IP)
[#X	Phrase Boundary for a Phrase of Type "X"	A prosodic phrase begins at this point; the type of prosodic phrase is specified by "X" (e.g., [#IP)
*#X	Strongest Prominence in a Domain	The constituent this occurs in is the most prominent in the prosodic domain specified by "X" (e.g., *#IP)
*#CF	Contrastive Focus Prominence	The prominence marks Contrastive Focus meaning

Table 1.3: A summary of some of the analysis tags described in this section, as expert-level PrStr annotation.

The goal of including these examples here is to show how an annotator can use tags to systematically notate their theoretical analyses. PoLaR annotators are strongly encouraged to devise their own tags; when doing so, they should explicitly define the grammatical features that each tag marks, and provide guidelines on when it should (or should not) be used. Analysis tags thus provide PoLaR with flexibility, so that it can be used for systematic purposes not foreseen by these guidelines.

1.2   Adding a New Tier

PoLaR is meant to be extensible and new Tiers may be added to address particular research questions. New tiers can always be added without serious consequence for the existing tiers and their labels, because the tiers are designed to, by default, isolate particular characteristics of the speech signal. These Tiers can be shared with the community by sending supporting material to the authors for inclusion on polarlabels.com. This repository can also be referenced in publications if useful.

There are several issues, both theoretical and practical, to consider when developing a new Tier. First, one should consider whether the new Tier is necessary or whether additional labels on existing tiers would suffice. Second, labels should be unique, memorable and consistent. Finally, labels should be machine readable.

Supporting materials should include:

1. a README.txt text file that includes
   a) the date and version number (e.g. 1.0)
   b) a brief description of the motivation, structure and use of the new Tier's labels. This README can also include labelling guidelines.
   c) bibliography and reference materials for the phenomenon to be labelled

d) a summary table of the labels (formatted like Table XX and others in this guide)

2. sound and Textgrid file(s) for a labelled example (or several, to illustrate contrastive phenomena),

3. any scripts that would be helpful in labelling.

1.2.1 New Tier Candidate: Labelling Individual Cues to Prosodic Structure

At the heart of PoLaR's approach to labelling prosody is a focus on cues in the intonational domain that involve f0 patterns. But as has often been noted in this monograph, speakers use many other cues to signal both prominence and grouping, and to signal other aspects of meaning such as emotion, attitude, and stance. As noted earlier, the range of cues in the speaker's toolbox includes patterns of duration, amplitude and voice quality, in addition to f0, and it has been argued that changes in the cues to the distinctive features of the phonemes of a word can also serve as cues to prosodic structure. It is thus of considerable interest to some researchers to understand how speakers (and listeners) use the different individual cues available to them. Brugos et al. (2018), for example, have shown that a location that is marked by more cues (e.g. f0 movement at the end of a word, f0 reset at the beginning of a word, amplitude changes, final creak at the end of a phrase, irregular pitch periods at the onset of a phrase, pause and final lengthening) is more likely to be labelled as a boundary by ToBI labellers. They have also argued that different listeners may weight each cue type differently in perception. PoLaR's flexible extensibility facilitates investigation of such ideas, by allowing labellers to develop tiers and conventions for annotating individual cues separately.

Taking the broader view, the expansion of labelling systems to include individual acoustic cues is also occurring in the domain of cues to the distinctive features that define natural classes of phonemes. For example, Stevens (2002) proposed that listeners attend to Landmark cues (i.e. moments of abrupt change in the spectrum of the wave form, associated with narrowings and widenings of the vocal tract) and other acoustic cues to feature contrasts in a spoken utterance. Relatedly, Stevens & Keyser (1989) and Keyser & Stevens (2001) proposed that speakers can enhance a contrast either phonologically, by adding an enhancing feature (e.g. [+round] for /ʃ/ in American English), or phonetically, by adding an enhancing cue (e.g. vowel shortening before a voiceless coda in American English). (See Clements 2009 for further discussion.) In this sense, PoLaR offers the opportunity not only to extend prosodic labelling to individual cues, but also to integrate prosodic cue labelling with distinctive feature cue labelling in the segmental domain. This integration would provide a powerful tool for quantifying the interaction between prosodic structure and systematic variation in segmental feature cues, and investigating the principles and constraints that govern this interaction.

1.2.2 New Tier Candidate: Analyzing Range Changes

The selection of pitch range(s) for an utterance has been associated with a variety of linguistic functions, communicative/emotional signalling, and paralinguistic phenomena. In the Basic and Advanced Ranges chapters above, we have described and given examples for a number of phenomena in which range changes are frequently observed. While we have offered guidance about how to determine where to put range boundaries and choose range values, this guidance is somewhat agnostic about the reasons why range changes occur. In other words, the Ranges tier labels capture what the values of the ranges are, but do not specify why range changes occur.

Much like the way in which the Advanced Points tier labels associate/assign individual turning points to PrStr tier objects (prominences and boundaries), we envision that researchers may want to link range changes to aspects of prosodic structure, or even discourse or syntactic structure. We encourage researchers to develop annotation schemes for such links that are consistent with PoLaR conventions. Such annotations might include, for example, elements like associating a Ranges tier annotation with a PrStr element (i.e. *, [, or]) as well as additional expansions. While future developments may incorporate additional labels in the Ranges tier itself, we encourage researchers to consider making use of a new tier for such 'linking labels'. As in any expansion to PoLaR, it is critical that the researchers proposing the expansion provide a clear set of guidelines, to explain and describe such new tiers and/or labels.

Conventions laid out in § **?REF?** for the Advanced Points tier labelling (namely, association between PrStr objects and f0 points) incorporate insights based on theories accounting for intonational movements that are well-established in the intonational phonology literature. Associations between pitch ranges (sometimes called registers) and prosodic structure are less firmly established in the same literature. That being said, the flexibility of PoLaR's tier-based annotation system encourages the use of annotations to explore hypotheses about associations between ranges and aspects of prosodic structure, facilitating potential future studies in this domain.

Examples of changes to ranges that may be governed by prosodic structure and discourse structure:

- Range maximum is conditioned by a compression relating to a sequence of prominences (i.e. frequently considered a "downstepped" High pitch accent)
    - This may include cross-phrase pitch-accent scaling relations (Féry & Truckenbrodt 2005), an issue that is in need of study, since current AM theory (and the ToBI annotation system) provides for downstep only within an intonational phrase.
- Range maximum is conditioned by an expansion relating to a sequence of prominences (i.e. sometimes considered an "upstepped" High pitch accent)
- Range (max and/or min) is conditioned by phrase boundaries or higher level grouping, including:
    - parentheticals
    - embedded registers (Truckenbrodt 2002; d'Imperio & Michelas 2010)
    - long-distance range dependencies (Ladd 1988)
    - larger discourse structure (including what has been called "paragraph intonation") (Lehiste 1975, Sluijter & Terken 1993, Wichmann 2014)

As our understanding of such phenomena deepens, and testable hypotheses emerge, PoLaR's ability to expand by adding new tiers and connecting information across tiers will prove useful.

Bibliography

Baltazani, Mary & Jun, Sun-Ah. 1999. Focus and topic intonation in Greek. In *The proceedings of 14th international congress of phonetic sciences (ICPhS)*. 1305–1308.

Beckman, Mary E. & Hirschberg, Julia & Shattuck-Hufnagel, Stefanie. 2005. The original ToBI system and the evolution of the ToBI framework. In Jun, Sun-Ah (ed.), *Prosodic typology: The phonology of intonation and phrasing*, 9–54. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199249633.003.0002

Brugos, Alejna & Breen, Mara & Veilleux, Nanette & Barnes, Jonathan & Shattuck-Hufnagel, Stefanie. 2018. Cue-based annotation and analysis of prosodic boundary events. In *Speech Prosody 2018*, 245–249. doi:10.21437/SpeechProsody.2018-50

Clements, George Nickerson. 2009. The role of features in phonological inventories. In Raimy, Eric & Cairns, Charles (eds.), *Contemporary views on architecture and representations in phonology*, 19–68. MIT Press.

d'Imperio, Mariapaola & Michelas, Amandine. 2010. Embedded register levels and prosodic phrasing in French. In *Speech prosody*. 4.

Féry, Caroline. 2017. *Intonation and prosodic structure*. Cambridge, UK: Cambridge University Press.

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

Halle, Morris & Vergnaud, Jean Roger. 1987. Stress and the cycle. *Linguistic Inquiry* 18(1). 45–84.

Ishihara, Shinichiro. 2007. Major phrase, focus intonation, multiple spell-out (MaP, FI, MSO). *The Linguistic Review* 24(2-3). 137–167.

Jeon, Hae-Sung & Nolan, Francis. 2017. Prosodic marking of narrow focus in Seoul Korean. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1). 2. doi:10.5334/labphon.48

Jun, Sun-Ah. 2007. The intermediate phrase in Korean: Evidence from sentence processing. In Gussenhoven, Carlos & Riad, Tomas (eds.), *Tones and tunes*, vol. 2, 143–167. Mouton de Gruyter.

Kanerva, Jonni M. 1990. Focusing on phonological phrases in Chichewa. In Inkelas, Sharon & Zec, Draga (eds.), *The phonology-syntax connection*, 145–162. Chicago: University of Chicago Press.

Keyser, Samuel Jay & Stevens, Kenneth N. 2001. Enhancement revisited. In Kenstowicz, Michael (ed.), *Ken hale: A life in language*, chap. 8, 271–291. Cambridge, MA: MIT Press.

Kratzer, Angelika & Selkirk, Elisabeth. 2020. Deconstructing information structure. *Glossa* 5(1). 113. doi:10.5334/gjgl.968

Ladd, D. Robert. 1988. Declination "reset" and the hierarchical organization of utterances. *The Journal of the Acoustical Society of America* 84(2). 530. doi:10.1121/1.396830

Lehiste, Ilse. 1975. The phonetic structure of paragraphs. In *Structure and process in speech perception*, 195–206. Springer.

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

Nespor, Marina & Vogel, Irene. 1986. *Prosodic phonology*. Dordrecht, Holland ; Riverton, N.J.: Foris.

Pierrehumbert, Janet & Hirschberg, Julia. 1990. The meaning of intonational contours in the interpretation of discourse. In Cohen, Philip R. & Morgan, Jerry L. & Pollack, Martha E. (eds.), *Intentions in communication*, 271–311. Cambridge, MA: MIT Press.

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

Shattuck-Hufnagel, Stefanie & Turk, Alice E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research* 25(2). 193–247.

Sluijter, Agaath MC & Terken, Jacques MB. 1993. Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica* 50(3). 180–188.

Stevens, Kenneth N. 2002. Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* 111. 1872–1891. doi:10.1121/1.1458026

Stevens, Kenneth N. & Keyser, Samuel Jay. 1989. Primary features and their enhancement in consonants. *Language* 65(1). 81–106.

Truckenbrodt, Hubert. 1995. *Phonological phrases: Their relation to syntax, focus, and prominence*: MIT dissertation.

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

Wagner, Michael. 2005. *Prosody and recursion*: MIT dissertation.

Wichmann, Anne. 2014. *Intonation in text and discourse: Beginnings, middles and ends*. Routledge.