Analysing automatic descriptions of intonation with ICARUS

Katrin Schweitzer, Markus Gärtner, Arndt Riester,
Ina Rösiger, Kerstin Eckart, Jonas Kuhn, Grzegorz Dogil

Institute for Natural Language Processing, University of Stuttgart, Germany
<firstname>.<lastname>@ims.uni-stuttgart.de

Abstract

We present ICARUS for intonation – a graphical tool which allows to access automatically derived F_0 features in an intuitive and user-friendly way. It can be used for data exploration and search. Tonal features can be accessed together with information from other linguistic levels; this is exemplified in two search queries where we combine the search for a specific tonal contour with a) coreference annotations and b) automatically derived syntactic annotations. Thereby we demonstrate how ICARUS for intonation bridges the gap between manual/semi-automatic analysis of relatively small, manually annotated data sets and automatic analysis of larger corpora with automatically derived features.

Index Terms: Intonation, parametric F_0 analysis, multi-level annotations, syntax interface, semantic interface

1. Introduction

Research on intonation is often based on manual annotations for pitch accents and intonation boundaries. Manual prosodic annotation requires trained annotators and is a very time consuming task. For instance, the time needed for labeling speech data according to the Tones and Break Indices (ToBI) system for American English [1] takes experienced annotators about 100-200 times the real time [2]. Even though for some research questions it might suffice to not look at all tiers that the ToBI system provides – for instance, research focusing on pitch accents might only need pitch accent type and placement annotations – manual annotations are still very time consuming: even the time needed for accent types, will still be many times longer than the length of the respective speech sample.

On the other hand, in the past years research centering around the automatic annotation of intonation or automatic analysis of the fundamental frequency (F_0) contour has grown, furthering the processing of large data sets (e.g.[3, 4, 5, 6]). However, direct access to the data to look at specific instances is often not straight-forward and the output of intonation models can lack an obvious, intuitive representation of the tonal contour which makes it hard to interpret. Moreover, often the tonal information cannot be accessed together with state-of-the-art text based tools.

In this paper we present a methodology employing a new module for the ICARUS platform.[1] This tool, ICARUS for intonation, bridges the gap between manual or semi-automatic analyses of relatively small data sets with (manually annotated) intonation labels and large-scale analyses of automatically derived parameters describing the tonal contour. ICARUS (**I**nteractive platform for **C**orpus **A**nalysis and **R**esearch tools **U**niversity of **S**tuttgart) was developed as a search tool for dependency treebanks [7], and has been extended to be used for automatic error mining [8] and for coreference research [9]. The intonation module for ICARUS offers direct access to the output of a parametric intonation model, here: PaIntE [10, 11], visualization of the numeric parameters of the model (including the possibility to modify them with the changes being visualized) and various possibilities of instance-based search. The user can search based on the shape of the tonal contour or the configuration of the linguistic context (POS tags, dependency parses, coreference annotations). Ranges of PaIntE values can be specified as search criteria, and built-in (but customizable) definitions of F_0 shapes (e.g. rising/falling) can be employed. The instances found can be directly visualized and exported, and the sound file can be played in various granularities, enabling the user to redefine their search criteria and deepen their understanding of the data. Moreover, various similarity measures can be utilized to find instances that have a similar shape.

We will first outline the characteristics of the PaIntE model and the specifics of ICARUS for intonation, before we turn towards two real-world examples: we outline search queries where ICARUS for intonation adds to a comprehensive understanding of the research question. The first example investigates intonation in connection with coreference resolution, that is, a comparison between the tonal realization of expressions that are given, and such that are new to the discourse is carried out (section 4). The second example demonstrates how automatic syntactic annotations of large data sets can be incorporated and how specific tonal configurations (based on the automatically derived tonal parameters) can be searched in conjunction with search criteria that are based on syntax (section 5).

2. Automatic analysis of intonation: PaIntE

The PaIntE model employs a function term to approximate a peak in the F_0 contour, comprising 6 parameters which are set by the model so that the resulting curve fits the actual F_0 shape best. The parameters are linguistically meaningful: they specify the steepness of the rise before, and the fall after the peak (parameter *a1* and *a2*, respectively), the temporal alignment of the peak (*b*), the amplitude of the rise / fall (*c1/c2*) and the absolute peak height *(d)*. Figure 1 illustrates the parameters which are calculated over a span of 3 syllables: the one for which the parametrization is currently carried out ($\sigma*$) and its immediate neighbors. The x-axis indicates time (normalized for syllable duration, i.e. the current syllable spans from 0 to 1) and the y-axis displays the fundamental frequency in Hertz.

[1] ICARUS is written in Java and is therefore platform independent. It is open source (under GNU GPL) and we provide both sources and binaries for download on http://www.ims.uni-stuttgart.de/data/icarus.html

Figure 1: *The PaIntE model function and its parameters. Figure adapted from [11].*

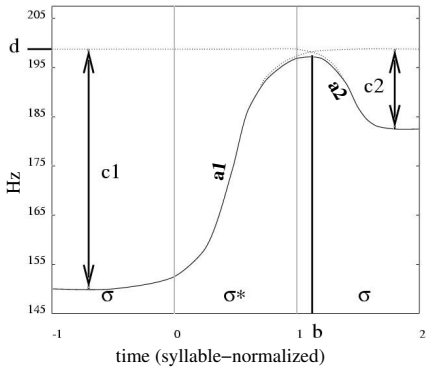
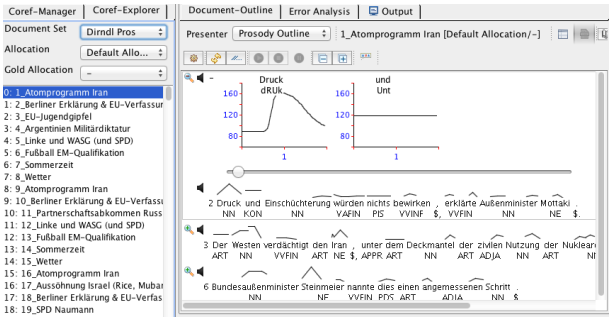


Figure 2: *The* Prosody Outline *of ICARUS.*



3. ICARUS for intonation

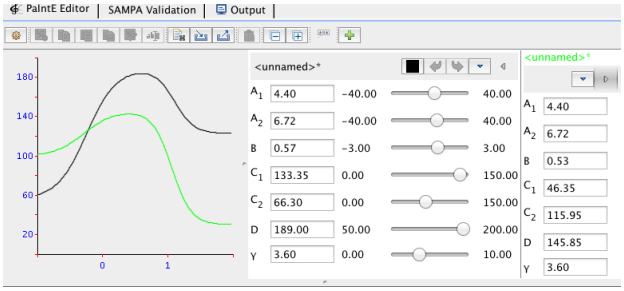
ICARUS is set up to read in PaIntE parameters along with other information about the linguistic context available for a particular data set. Here we will use the DIRNDL data set [12] with coreference information [13] and some additional features in a tabular format. Note that the DIRNDL corpus actually provides manually labeled prosody information – however we will not use this information in order to a) allow a purely phonetic exploration of other linguistic levels without any additional information from phonology or other linguistic assumptions about intonation, and to b) provide an idea how (larger) corpora without such information can be explored.

3.1. Data exploration

ICARUS for intonation offers various ways of visualizing the tonal contour of an utterance. For a first overview it provides a "compact mode" in which the utterance is displayed with a number of customizable features e.g. word form and POS tag, and a simplified visualization of the PaIntE function for all syllables on which a peak was found (according to PaIntE parameter *b*, encoding the peak's timing) . This visualization only uses the amplitudes of rise and fall (*c1*/*c2*) and the absolute peak height (*d*). The view can then be expanded so that the complete PaIntE curve can be seen for a (user-specified) number of syllables (see Figure 2).

There are various ways of playing the corresponding sound files, ranging from playing the whole corpus or utterance (using the play buttons) to a syllable-wise audio playback (by clicking on the respective labels, e.g. for syllables or words).

Figure 3: *The PaIntE editor.*



3.2. Search

Search in ICARUS offers a vast variety of features; describing all of them is beyond the scope of this paper, therefore we will concentrate on some key features here. Generally, for searching an audio corpus, the user can combine search criteria from all linguistic levels provided with the corpus. So for our example corpus, DIRNDL, we can search for syntactic configurations, POS sequences, coreference structures, phonetic features and, of course, PaIntE parameters. The search can be defined graphically, arranging nodes with feature values and (e.g. syntactic) dependencies between the nodes, or in a text-based manner. Graphically defined searches can also be exported to text files and be imported (and graphically displayed) again in future studies. In ICARUS one can search for specific PaIntE values or ranges (e.g. all syllables with a peak excursion greater than 50Hz). Users can also define criteria for specific shapes, e.g. a rising F_0 contour needs to have a peak in the following syllable (PaIntE parameter $b>1$ and <2) and the rise needs to be greater than the fall (users can define in Hz how big the delta *c1*-*c2* is) and the peak excursion (here *c1* has to be above a user-defined Hertz value). We also defined a feature *tonal prominence* with which we tag words where any syllable has a peak that exceeds a (customizable) Hertz value (the default is 50Hz).

Users can define search criteria based on numerical comparisons (e.g. greater than, equals, less than), and for categorical values the instances can be searched via a match with one of the possible values, or the levels can be displayed along with their frequency distribution, using a grouping operator (cf. section 4).

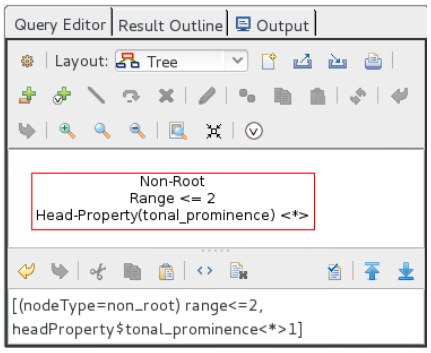
The results in ICARUS are displayed in a sentence based overview which the user can flip through. The tonal realization of the results is presented as a preview which can be expanded to a detailed view and, of course, results can be inspected also auditorily, using the play features (cf. section 3.1).

The results can be exported in a flexible tabular format which can be specified by the user and can be used as input in further processing/analyzing steps, e.g. for statistical analyses.

3.3. The PaIntE editor

The PaIntE editor provides users with no or little knowledge about PaIntE with the possibility to directly see the impact of changes in the PaIntE parameters. The user can define or modify one or more PaIntE curves here by changing the parameters with a slider. Changes are displayed in real-time. Different tonal contours can be overlayed, saved (along with a description) and used in search. For instance, a prototype of tonal contour can be created and similar shapes can be searched for, choosing from a number of different similarity measures, e.g. cosine similarity and Euclidean distance. Figure 3 shows the PaIntE editor interface with two different peak shapes.

Figure 4: *The search query editor in ICARUS. Top window shows the graphical representation of the search, bottom window shows the search as text.*



4. The tonal features of coreference

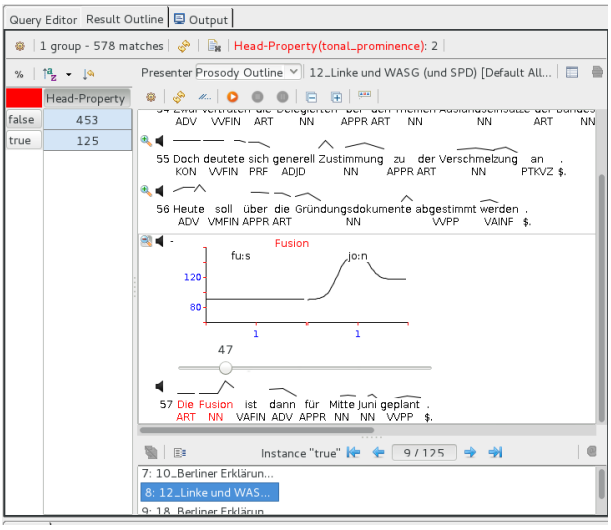
In the following we present a case in which ICARUS for intonation is used to test and explore potential features for automatic coreference resolution.

Coreference resolution and intonation In the area of coreference resolution, i.e. identifying expressions which refer to entities already mentioned in the discourse, phonetic features have, to our knowledge, not been taken into account. In an ongoing research project, we are examining if and how phonetic parameters improve automatic coreference resolution. There is evidence that entities that have been previously mentioned in the discourse (*given* items) are often deaccented [14, 15, 16]. However, deaccentuation can be overruled by contrast [17], i.e. complex interactions are at play, potentially influencing the tonal realisation of givenness in different directions. Therefore, looking at phonetically derived features of intonation might improve automatic coreference resolution. To this end, we investigate whether the automatically derived PaIntE values are distributed differently between expressions that are coreferent to an already mentioned entity, and such that are not.

Search query The search query in ICARUS for intonation is straightforward. To define expressions that are not new to the discourse, we employ DIRNDL's coreference annotations: root nodes in the coreference chain are new, all other nodes in the chain are given. Using the feature *tonal prominence* together with the grouping operator (cf. section 3.2) we acquire a frequency distribution of cases, where given expressions are "tonally prominent" (note: the definition relies purely on peak excursion, we kept the default value of 50Hz for this query). For the sake of simplification, we restricted the search to expressions with one or two words, here. Figure 4 shows the graphical and the textual representation of the query.

Search results Figure 5 displays the result in ICARUS. The top left corner shows the frequency table displaying how often given expressions are realized with *tonal prominence*. By clicking on the table, the user can look at (and listen to) both sets of results in preview or detailed view (cf. section 3.1). In the example, the expression "die Fusion" (the fusion) in sentence 57 is coreferent with "die Verschmelzung" (the merger) in sentence 55 and is a case where *tonal prominence* according to the

Figure 5: *Coreference and intonation in ICARUS.*



default definition is true. As can be seen in the frequency table, cases where expressions that are coreferent with another expression in the discourse, are realized more often without *tonal prominence* (according to our definition). The result indicates that automatically derived pitch excursion is a feature that can be helpful in automatic coreference resolution. Modifying the definition of *tonal prominence*, i.e. the pitch excursion, will provide further inside into the parameter distribution for given (or new) expressions, so that in a step-wise process, we gain a better understanding of the data before applying machine learning algorithms.

5. The intonation of adjective-noun sequences

A recent study [18] analyzed adjective-noun sequences from the DIRNDL corpus with respect to their tonal realization. The study examines the *relative givenness* assumption [19], which claims that, in adjective-noun combinations, deaccentuation of the noun does not depend on the givenness of the noun but on the salience of an alternative adjective-noun sequence.

Adjective noun sequences and their intonation The main interest of the study was to compare adjective and noun with respect to which one is more prominent. The researchers employed the manual prosodic annotations to the DIRNDL corpus to compare the prominence of adjective and noun. They manually examined the context of each match with respect to the availability of salient alternatives. In some cases the researchers disagreed with the manual prosodic annotations which lead them to disregard some of the data. That is, even though inter annotator agreement for prosodic annotations is in general reasonable (87% for placement, 51% for type [20]), some subjectivity remains. Analyzing the intonation of such sequences phonetically can add to the overall picture and provide an entirely objective measure. Moreover, while in [18] the distinction between pre-nuclear and nuclear pitch accents is utilized to distinguish between levels of prominence, an acoustic analysis can reveal finer details of differing tonal prominence. Searching the data with ICARUS for intonation allows seeing the matches

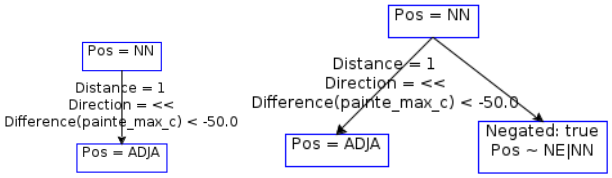


Figure 6: *The search query for ADJ-NN sequences, where the the adjective's tonal contour has a peak which is at least 50Hz higher than the one of the noun. The distance between the two nodes in the dependency syntax structure may only be one edge (Distance=1), i.e. they are directly adjacent and the adjective precedes the noun (Direction = ≪). The second search (Fig 6b) refines the search by excluding matches where the noun is additionally modified by another noun (NN) or a proper name (NE).*

in their context, listen to them and then refine the search criteria in a step-wise manner.

Search query There are various ways to tackle the area of relative prominence of the adjective and the noun in ICARUS for intonation. We will present one query here which can be seen as a starting point to examine the intonation of these sequences in an iterative way, as well as a follow-up query, demonstrating how one can refine queries, based on the previous matches. The main interest in [18] lay in cases where the adjective was more prominent than the noun, i.e. a marked intonation was used (usually, the noun would be expected to receive an accent, being the head of the phrase). Here we will present how such cases can be searched for, only using automatically derived features, and how the results can be accessed in their context, along with a visualization and play-back of the respective sound files. To look at two nodes in comparison, ICARUS allows to apply relative search constraints, using the (automatically acquired) dependency syntax structure of the data. Hence, we can search for an attributive adjective (ADJA) which is a dependent of a noun (NN). In phrase-structure grammar terms, this would be similar to an ADJA embedded in an NP. We apply a search constraint to the edge which connects these two nodes. The constraint compares the maximal value of PaIntE parameter *c1* and *c2* (i.e. the excursion of the peak) of one word with the one of the other. The user can then define how much these two maximal values should differ. In our sample query we set the difference to 50Hz, so a considerable difference in the acoustic tonal prominence of these two words is required. Figure 6a shows the graphical representation of the search.

Search results Figure 7 shows the results. Again, the user can get an overview of all matches (bottom window), see and hear them in their context (middle window) and get a detailed overview (top window). The first match "neue Resolution" (new resolution) is displayed in detail. As can be seen, the adjective "neue" has indeed a peak that has a greater excursion than the noun "Resolution". While this is true for the majority of the matches, a few matches, e.g. the the second match "früheren Kulturstaatsminister Naumann" (former minister for culture Naumann), are undesired matches: the noun is followed by the proper name (NE) "Naumann". Here, the NE is more prominent than the ADJA. The increased prominence is

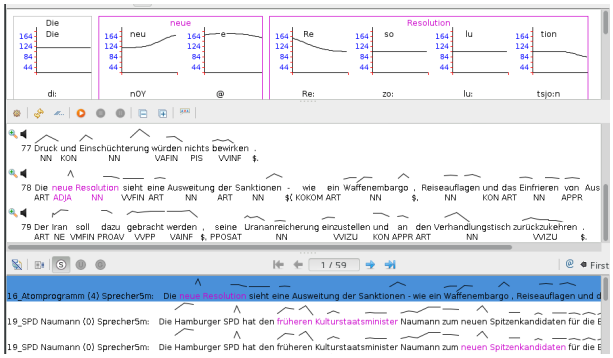


Figure 7: *Search result for adjective noun combinations in DIRNDL.*

reflected in a considerable rise on the word "Naumann", which is already visible in the overview of the matches (Fig. 7 lower window) which provides a preview of the PaIntE parameters (on top of the orthographic representation of the match).[2] That is, taken together, the phrase "Kulturstaatsminister Naumann" does not match the criterion of having a peak with 50Hz more excursion than the adjective. Examining the other matches reveals more false positives of that kind, e.g. "der iranische Außenminister Mottaki" (the Iranian minister for foreign affairs Mottaki) as well as cases where the modifier is another NN, e.g. "das umstrittene Wort Verfassung" (the controversial word constitution). To ensure that the matches in our search are not modified by another noun or name, as was the case here, we will refine the search query.

Refined search query Modifiers of the noun are marked in the dependency structure as dependants of the NN. Thus, we add a restriction to the previous search query, excluding cases where the NN has a dependant that is either NE (a modifying name) or NN (a modifying noun). The graphical representation of the search is displayed in Figure 6b. The matches are effectively improved. These results can now be refined further (e.g. by examining the effect of smaller differences in tonal prominence), looked at manually (as in [18]), or exported for statistical or other machine-based analyses.

6. Conclusion

We presented ICARUS for intonation, a tool that allows for a conjoint analysis of intonation, here represented as parametrized peaks in the F_0 contour, with different linguistic levels. We described two search queries where we looked at intonation and coreference, and at intonation and syntax, respectively and where we demonstrated how the features that ICARUS for intonation comprises can help directly accessing automatically derived tonal parameters and thereby gaining a deeper understanding of the data in context. Moreover we hope that the easy access and visualization of tonal parameters together with other annotation layers will foster interdisciplinary research on data from speech corpora. Future work includes the incorporation of more annotation layers as well as an investigation and refinement of the similarity measures to gain insight into the relation of acoustic and perceptual similarity.

[2] Interestingly enough, the intonation of "Naumann" is a "tonal slip-of-the-tongue" with the second syllable (which canonically does not bear the word stress) being pitch accented.

7. References

[1] M. Beckman and J. Hirschberg, "The ToBI annotation conventions," http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html, 1999.

[2] A. K. Syrdal, J. Hirschberg, J. McGory, and M. Beckman, "Automatic ToBI Prediction and Alignment to Speed Manual Labeling of Prosody," *Speech Commun.*, vol. 33, no. 1-2, pp. 135–151, Jan. 2001. [Online]. Available: http://dx.doi.org/10.1016/S0167-6393(00)00073-X

[3] V. K. R. Sridhar, A. Nenkova, S. Narayanan, and D. Jurafsky, "Detecting Prominence in Conversational Speech: Pitch Accent, Givenness and Focus ," in *Proceedings of Speech Prosody (SP-2008)*, Campinas, Brazil, 2008, pp. 453–456.

[4] R. Fernandez and B. Ramabhadran, "Discriminative training and unsupervised adaptation for labeling prosodic events with limited training data," in *Proceedings of Interspeech 2010 (Makuhari, Japan)*, 2010, pp. 1429–1432.

[5] A. Rosenberg, "AuToBI - a tool for automatic ToBI annotation," in *Proceedings of Interspeech 2010 (Makuhari, Japan)*, 2010, pp. 146–149.

[6] A. Schweitzer and B. Möbius, "Experiments on automatic prosodic labeling," in *Proceedings of Interspeech 2009 (Brighton, UK)*, 2009, pp. 2515–2518.

[7] M. Gärtner, G. Thiele, W. Seeker, A. Björkelund, and J. Kuhn, "Icarus – an extensible graphical search tool for dependency treebanks," in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Sofia, Bulgaria: Association for Computational Linguistics, August 2013, pp. 55–60. [Online]. Available: http://www.aclweb.org/anthology/P13-4010

[8] G. Thiele, W. Seeker, M. Gärtner, A. Björkelund, and J. Kuhn, "A graphical interface for automatic error mining in corpora," in *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg, Sweden: Association for Computational Linguistics, April 2014, pp. 57–60. [Online]. Available: http://www.aclweb.org/anthology/E14-2015

[9] M. Gärtner, A. Björkelund, G. Thiele, W. Seeker, and J. Kuhn, "Visualization, search, and error analysis for coreference annotations," in *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. Baltimore, Maryland: Association for Computational Linguistics, June 2014, pp. 7–12. [Online]. Available: http://www.aclweb.org/anthology/P14-5002

[10] G. Möhler, "Describing intonation with a parametric model," in *Proceedings of the International Conference on Spoken Language Processing*, vol. 7, 1998, pp. 2851–2854.

[11] ——, "Improvements of the PaIntE model for F_0 parametrization," Institute of Natural Language Processing, University of Stuttgart, Tech. Rep., 2001, draft version.

[12] K. Eckart, A. Riester, and K. Schweitzer, "A discourse information radio news database for linguistic analysis," in *Linked Data in Linguistics. Representing and Connecting Language Data and Language Metadata*, C. Chiarcos, S. Nordhoff, and S. Hellmann, Eds. Heidelberg: Springer, 2012, pp. 65–75.

[13] A. Björkelund, K. Eckart, A. Riester, N. Schauffler, and K. Schweitzer, "The extended dirndl corpus as a resource for coreference and bridging resolution," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, may 2014.

[14] D. Ladd, *Intonational Phonology*, 2nd ed. Cambridge, UK: Cambridge University Press, 2008.

[15] D. Büring, "Intonation, semantics and information structure," in *The Oxford Handbook of Linguistic Interfaces*, G. Ramchand and C. Reiss, Eds. Oxford University Press, 2007.

[16] S. Baumann and A. Riester, "Coreference, lexical givenness and prosody in German," *Lingua*, vol. 136, pp. 16–37, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S002438411300168X

[17] M. Rooth, "A theory of focus interpretation," *Natural Language Semantics*, vol. 1, pp. 75–116, 1992.

[18] A. Riester and J. Piontek, "Anarchy in the NP. When new nouns get deaccented and given nouns dont," *Lingua*, no. 0, pp. –, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0024384115000522

[19] M. Wagner, "Givenness and Locality," in *Proceedings of SALT XVI*, M. Gibson and J. Howell, Eds. CLC Publications, 2006, pp. 295–312.

[20] M. Grice, M. Reyelt, R. Benzmüller, J. Mayer, and A. Batliner, "Consistency in transcription and labelling of German intonation with GToBI," in *Proceedings of ICSLP*, 1996, pp. 1716–1719.