

# Multi-modal Visualization and Search for Text and Prosody Annotations

Markus Gärtner Katrin Schweitzer Kerstin Eckart Jonas Kuhn

Institute for Natural Language Processing  
University of Stuttgart

{markus.gaertner,kati,eckartkn,kuhn}@ims.uni-stuttgart.de

## Abstract

We present ICARUS for intonation, an interactive tool to browse and search automatically derived descriptions of fundamental frequency contours. It offers access to tonal features in combination with other annotation layers like part-of-speech, syntax or coreference and visualizes them in a highly customizable graphical interface with various playback functions. The built-in search allows multi-level queries, the construction of which can be done graphically or textually, and includes the ability to search  $F_0$  contours based on various similarity measures.

## 1 Introduction

In this paper we present *ICARUS for intonation*, a new module for the query and visualization tool ICARUS by Gärtner et al. (2013).<sup>1</sup>

So far, ICARUS included modules for the handling of dependency treebanks (Gärtner et al., 2013) and coreference data (Gärtner et al., 2014), thus supporting typical annotation layers from the processing of written data. However, the graphical query builder and the intuitive example-based search could prove just as expedient for other types of data, such as speech corpora, transcribed and annotated for sub word features. This also allows combined research on speech and text data, e.g. the analysis of different tonal realizations of a certain syntactic structure.

ICARUS for intonation allows to import syllable-based prosodic features into ICARUS, which can then be visualized and queried either

individually or in a combined search with e.g. syntactic features or coreference information. The latter targets several user groups: speech data experts can adjust fine-grained settings on pitch accent shapes in their queries and can easily add constraints on part-of-speech or syntax information, while an expert user of dependency treebanks can get a simple visualization of the intonation contour of a sentence.

Furthermore ICARUS focuses on automatic annotations to allow for search on large data sets. Thus ICARUS for intonation's main features for prosodic search are based on PaIntE, a parametric intonation model (Möhler, 1998; Möhler, 2001). So far, most data in intonation research is manually annotated, which is a very time consuming task: the time for annotating speech data is many times higher than the real time of the audio recording. For example the Tones and Break Indices (ToBI) system for American English (Beckman and Hirschberg, 1999) takes experienced annotators about 100-200 times the real time (Syrdal et al., 2001). While manual annotations for pitch accents and prosodic phrase boundaries can also be imported, our main goal with this module is to provide intonation researchers with a customizable tool to conduct thorough studies on very large sets of only automatically annotated speech data.

In Sections 2 and 3 we introduce the PaIntE model and describe the current input format for the data importer. Section 4 demonstrates several visualization functionalities, and Section 5 discusses the search facilities, including dependency and intonation as well as coreference and intonation queries. After discussing some related work in Section 6 we conclude in Section 7.

## 2 The PaIntE Model

The PaIntE model (Möhler, 1998; Möhler, 2001) approximates a peak in the  $F_0$  contour by employing a model function operating on a 3-syllable

<sup>1</sup>ICARUS for intonation 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>

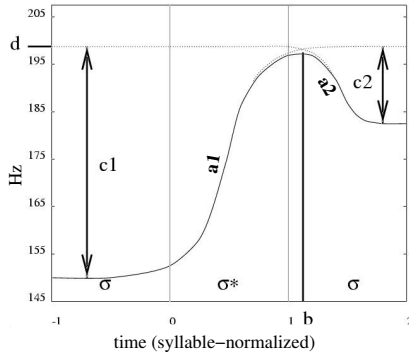


Figure 1: The PaIntE model function and its parameters. Figure adapted from (Möhler, 2001).

window. There are 6 free parameters in the function term which are set by the model so that the actual  $F_0$  shape is fit best. They are linguistically meaningful: parameter  $b$  locates the peak within the 3-syllable window, parameter  $d$  encodes its absolute height. The remaining parameters specify the steepness and amplitude of the rise before, and the fall after the peak (parameters  $a1$  and  $a2$  for the steepness and  $c1/c2$  for the amplitude).

Figure 1 illustrates the function. It displays the syllable for which the parametrization is carried out ( $\sigma^*$ ) and its immediate neighbors. The x-axis indicates time (normalized for syllable duration, the current syllable spans from 0 to 1) and the y-axis displays the fundamental frequency in Hertz. The PaIntE model has been used for the modeling of different languages, e.g. Norwegian, Italian, German and English (Cosi et al., 2002; Kelly and Schweitzer, in press; Schweitzer et al., 2015).

### 3 Data Representation

ICARUS for intonation ships with reader implementations for two very different formats. One is an extended version of the format used for the 2011 and 2012 CoNLL shared tasks (Pradhan et al., 2011; Pradhan et al., 2012) with a number of additional columns to accommodate features for the syllable level. This format stores all annotations corresponding to a word token in one line and packs syllable features into a list separated by pipe-characters ('|'). To address syllable centric data like the typical output of speech processing systems, a second flexible tabular format was specified where each line of text corresponds to a single syllable and a global header describes the content of all columns and how to read and map them to the internal data model of ICARUS.



Figure 2: PaIntE Editor currently displaying 2 curves and their respective parameters. The lower section shows saved and named prototypes.

To enable audio playback functionality ICARUS for intonation requires access to the appropriate sound files. In both formats described above, special properties define the name of a sound file to be used for playback. Timestamp values on various levels (syllable, word, sentence or document) point to the respective section in the audio data, which currently is required to be in the Waveform Audio File Format (\*.wav files).

## 4 Visualization

Since the ICARUS for intonation module is build on the data model used for corpora with coreference annotations in ICARUS, existing visualizations for coreference data can be used. However, they make no use of syllable level features and do not provide playback functionality. Therefore a couple of new visualizations have been implemented, adding visual information about PaIntE curves in several levels of granularity.

### 4.1 PaIntE Editor

To get familiar with the visualization of PaIntE parameters the PaIntE Editor (Figure 2) offers users with little or no knowledge about PaIntE a starting point to directly see the impact of changes to certain parameters. In this editor the user can define multiple PaIntE curves either from scratch or by importing them from real examples in a corpus. Changes to individual parameters can be applied via sliders or input fields and are displayed in real-time. Additionally a persistent storage of PaIntE curves is provided where the user can save parameter sets that are of interest to him along with a description and identifier, the latter of which can be used when searching (see Section 5).

### 4.2 Curve Preview

For all visualizations dealing with PaIntE curves ICARUS for intonation provides a compact “pre-

view” on the sentence level (lower parts of Figures 3 and 4b). Instead of drawing the full curves for all syllables, only syllables in which a peak was found (based on the peak’s timing encoded in the PaIntE parameter  $b$ ) are displayed. The visualization of the curve then only uses the amplitudes of rise and fall and the absolute height of the peak ( $c1$ ,  $c2$  and  $d$ ). Since the user can freely customize the filter window for the peak this curve preview offers a fast way to spot interesting parts of the  $F_0$  contour when exploring data manually.

### 4.3 Document Outline

Figure 3 shows parts of the main entry point for manual exploration in ICARUS for intonation. Having selected a section of the corpus the user wishes to inspect (with sentences grouped into documents in the left section of the figure) he then gets a detailed outline of the contents of that document using one of several available presentation modes. The default visualization for data holding PaIntE annotations arranges the document’s content one sentence per line, making use of the above mentioned curve preview to provide the user with a very compact overview of an entire document. For each sentence a detail panel can be unfolded which renders the complete PaIntE curves above the preview area. Several aspects of the visualization are highly customizable (like the number of words to show detailed curves for) and the user can select the content of the detail panel by moving a slider through the sentence.

An important feature of the Document Outline is the fine-grained playback functionality. The user is free to play a variety of sections of the sound data linked to the document currently being displayed. Speaker buttons at the left border play predefined parts of the sound data like sentences or the current content of a detail panel. By clicking on individual word or syllable labels in the detail panel the playback can be selected even finer.

### 4.4 Sentence Outline

When only single sentences are visualized, ICARUS for intonation displays a more detailed outline showing the PaIntE curves for all syllables in the sentence grouped by the surrounding words. In Figure 4b part of a sentence is visualized in this way (the screenshot also contains visual highlighting as its content is the result of a search).

In contrast to the more condensed document outline, this visualization offers a great deal more

space for additional information on the syllable level to be displayed. As for playback functionality it offers granularity similar to the document outline, allowing the user to play the entire sentence or restrict it to individual words or syllables.

### 4.5 Label Patterns

Both formats currently read by ICARUS for intonation can contain more information on the syllable and word level than can be presented to the user without overloading the visualization. Therefore the two visualizations described above make heavy use of so called *label patterns* to produce the actual text displayed at various locations. A label pattern is a string describing a format according to which a certain text should be created. Expressions of the form “{<level>:<property>}” define where information extracted from the visualized data should be inserted. The <level> specifies the level of data to query ( $\{syl, word, sent, doc\}$  for the syllable, word, sentence and document levels). For example the default pattern “{word:form}\n{word:pos}”, used in the Document Outline (see Section 4.3) to display the text for a sentence, extracts the surface form and part-of-speech tag for a word and places them below each other as shown in Figure 3. The user can freely define the default patterns for a number of locations as well as change the patterns used for the active visualization on the fly. Besides directly extracting data and displaying it as text, patterns offer additional options that define how to convert e.g. numerical values into strings or how to post process or aggregate generated texts. However, going into details of the pattern engine is beyond the scope of this paper.

## 5 Search

ICARUS for intonation augments both the coreference and dependency search facilities already available in ICARUS by adding access to various syllable features and implementing multiple specialized search constraints based on the PaIntE model. For example the user can search for predefined  $F_0$  contours (rise, fall, rise-fall or unaccented) based on customizable criteria or use one of several similarity measures available, like Euclidean distance or cosine similarity.

Sets of PaIntE parameters can either be defined explicitly by listing all values or by referencing a previously saved prototype from the PaIntE Editor

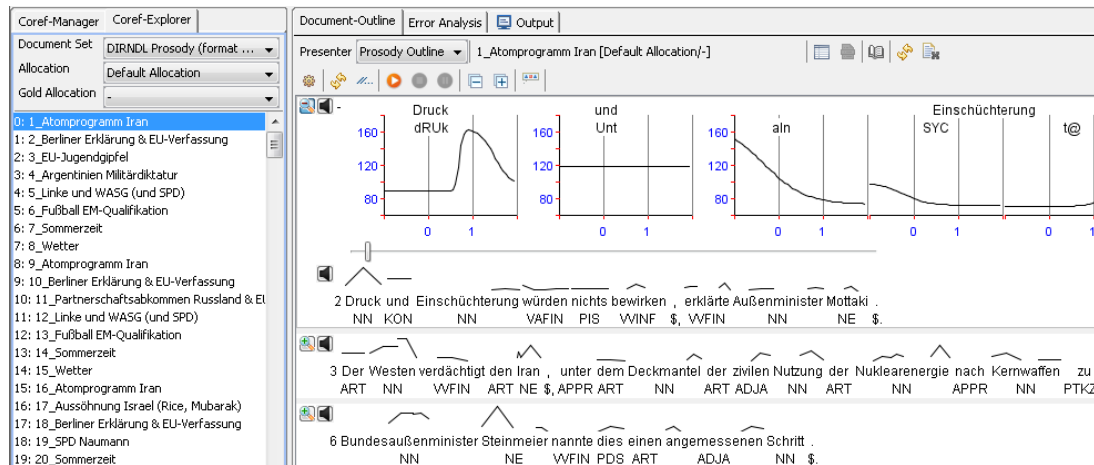


Figure 3: Visualization of the first few sentences in a document with preview curves painted above the raw text outline. The top sentence has its detail panel unfolded, showing PaIntE curves for all syllables of a selected number of words.

by name (see Section 4.1). The ICARUS search engine allows queries to be created either graphically (by creating nodes and attaching constraints to them) or textually via a simple query language (Gärtner et al., 2013).

The following two sections outline some example use cases that combine prosodic features with structural information on different layers for analysis and Section 5.3 shows some of the similarity measures used for searching. Example data in those sections is taken from the DIRNDL corpus (Eckart et al., 2012) with coreference information (Björkelund et al., 2014) and some added features.

### 5.1 Syntax and Intonation

As part of a recent study (Riester and Piontek, in press) adjective-noun sequences from the DIRNDL corpus have been analyzed based on their tonal realization. Of interest in this study concerning *relative givenness* (Wagner, 2006) were those adjective-noun sequences where the adjective was tonally more prominent than the adjacent noun. An example of how to find them is shown in Figure 4. The query (Figure 4a) will match adjectives (ADJA) adjacent to a following noun (NN) which must not have another dependent that is either a modifying noun or name (NE). The results are presented to the user using the detailed Sentence Outline (Figure 4b) from Section 4.4.

### 5.2 Coreference and Intonation

Besides finding exact matches in a data set the search engine in ICARUS can be used to analyze value distributions for an annotation. Using the

query in Figure 5a the search engine is asked to look for mentions the size of up to 2 words that are not the root of a coreference chain. The special *grouping operator* <\*> results in the creation of a frequency list (Figure 5b) over the Boolean *tonal prominence* property (which purely relies on the peak excursion with a customizable threshold) of the head word of each mention that was found based on the above constraints. By clicking on one of the entries in this list the user will then be presented with all the instances that contributed to the respective frequency for further exploration.

### 5.3 Similarity Search

The continuous nature of the PaIntE parameters makes using absolute values to search for curve forms very impractical. Therefore ICARUS for intonation provides a collection of similarity measures and other constraints that can be used to find syllables with PaIntE curves similar to a given prototype. Most of them are customizable by the user and investigation and refinement of the available similarity measures is subject of ongoing work.

Figure 6 shows an example of using cosine similarity to find instances in the data set that are similar to a defined prototype curve. In this case the first syllable of the accented word “Steinmeier” was found to be of interest and saved in the PaIntE editor with the identifier `prototype_stein`. The query `[painteAngle$"$prototype_stein"<="5.0"]` then looks for PaIntE curves which do not differ from the prototype by more than 5 degrees.

When using PaIntE curves as part of a search

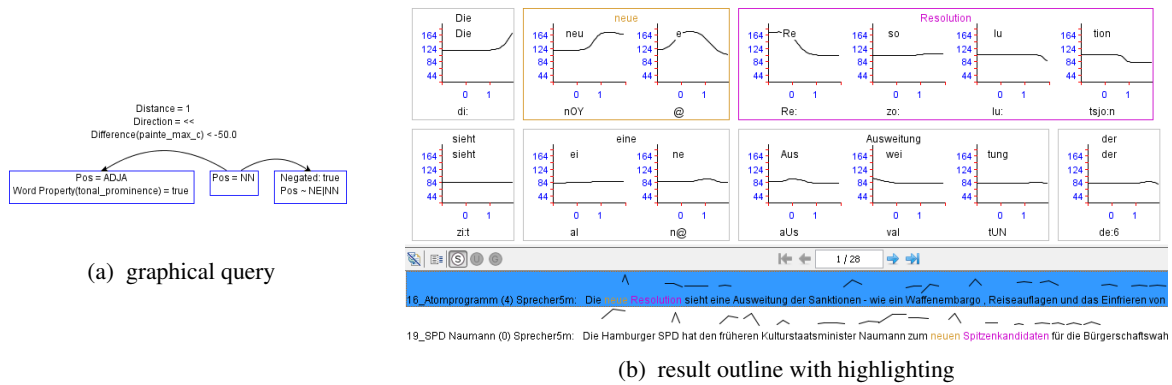


Figure 4: Example search query combining syntax and intonation constraints and an excerpt of the corresponding result outline.

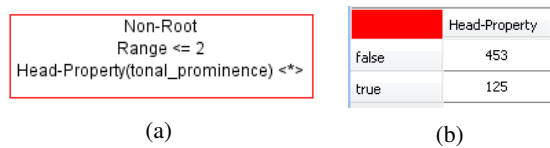


Figure 5: Simple search combining coreference and intonation features. It is meant to investigate the distribution of “tonally prominent” mentions that are *given* (already introduced) in a discourse.

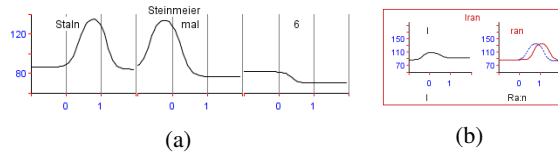


Figure 6: Prototype of a PaIntE curve as found in the data and an example result of a search using cosine similarity.

constraint the corresponding result visualization will render those curves when highlighting result instances as can be seen on the first peak (dashed blue curve) in Figures 6b. This provides the user with accurate information on how “visually close” a match is towards the used constraints.

## 6 Related Work

A number of well established tools exist for visualization of text corpora annotated with dependency or coreference, many of which have been discussed in other ICARUS related papers (Gärtner et al., 2013; Gärtner et al., 2014). In terms of search functionality those tools offer a broad range of complexity, ranging from string-searching on surface forms<sup>2</sup> up to queries on multi-level anno-

<sup>2</sup><http://code.google.com/p/whatswrong/>

tations (Zeldes et al., 2009; Pajas and Štěpánek, 2009). However, they do not support a dedicated search and visualization for prosodic syllable level annotations. Tools like ELAN (Wittenburg et al., 2006) provide an interface for adding (flat) annotations to multi-modal corpora, but focus on audio and video data. More importantly, ICARUS for intonation is so far the first tool using the PaIntE model for  $F_0$  contour visualizations, a task previously worked around via general curve plotting tools like R<sup>3</sup> and also is first to provide a collection of search constraints dedicated to PaIntE curves.

Eckart et al. (2010) describe a database that serves as a generic query tool for multiple annotation layers. It allows to take annotations of tonal features into account and has also been tested with the DIRNDL corpus. However, this database has been designed as an expert system, e.g. for internal use in projects that create annotations. It does not provide any visualization or query functions besides basic SQL queries and no sound playback.

The focus on preprocessed or completely annotated data in ICARUS distinguishes it from typical tools in the domain of Spoken Document Retrieval (SDR) or Spoken Term Detection (STD). These use automatic speech recognition and information retrieval technologies in order to prepare and process audio data (Garofolo et al., 2000).

## 7 Conclusion

We presented ICARUS for intonation, a flexible visualization and search tool for multi-modal (text and speech) data. The tool augments existing visualization and search features of ICARUS to handle prosodic annotations and introduces a collec-

<sup>3</sup><http://www.r-project.org>

tion of novel visualizations and search functionalities. In addition to the highly customizable visualizations it allows for a very fine-grained playback of speech data for displayed sections of a corpus directly from within the graphical user interface. The built-in search engine lets the user combine prosodic constraints with constraints of other annotation layers like syntax or coreference, thereby supporting complex search queries, and it features aggregated result views. Being based on the ICARUS platform’s plugin-engine, the module can be extended to cover additional data formats.

## Acknowledgments

This work was funded by the German Federal Ministry of Education and Research (BMBF) via CLARIN-D, No. 01UG1120F and the German Research Foundation (DFG) via the SFB 732, project INF.

## References

- Mary Beckman and Julia Hirschberg. 1999. The ToBI Annotation Conventions. [http://www.ling.ohio-state.edu/~tobi/ame\\_tobi/annotation\\_conventions.html](http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html).
- Anders Björkelund, Kerstin Eckart, Arndt Riester, Nadja Schauffler, and Katrin Schweitzer. 2014. The Extended DIRNDL Corpus as a Resource for Coreference and Bridging Resolution. In *LREC*.
- P. Cosi, C. Avesani, F. Tesser, R. Gretter, and F. Piansi. 2002. A modified “PaIntE” model for Italian TTS. In *Speech Synthesis, 2002. Proceedings of 2002 IEEE Workshop on*, pages 131 – 134.
- Kerstin Eckart, Kurt Eberle, and Ulrich Heid. 2010. An Infrastructure for More Reliable Corpus Analysis. In *LREC: Workshop on Web Services and Processing Pipelines in HLT*, pages 8–14, Valletta.
- Kerstin Eckart, Arndt Riester, and Katrin Schweitzer. 2012. A Discourse Information Radio News Database for Linguistic Analysis. In Christian Chiarcos, Sebastian Nordhoff, and Sebastian Hellmann, editors, *Linked Data in Linguistics*, pages 65–75. Springer, Heidelberg.
- John S. Garofolo, Cedric G. P. Auzanne, and Ellen M. Voorhees. 2000. The TREC Spoken Document Retrieval Track: A Success Story. In *Text Retrieval Conference (TREC) 8*, pages 16–19.
- Markus Gärtner, Gregor Thiele, Wolfgang Seeker, Anders Björkelund, and Jonas Kuhn. 2013. ICARUS – An Extensible Graphical Search Tool for Dependency Treebanks. In *ACL: System Demonstrations*, pages 55–60, Sofia, Bulgaria.
- Markus Gärtner, Anders Björkelund, Gregor Thiele, Wolfgang Seeker, and Jonas Kuhn. 2014. Visualization, Search, and Error Analysis for Coreference Annotations. In *ACL: System Demonstrations*, pages 7–12, Baltimore, Maryland.
- Niamh Kelly and Katrin Schweitzer. in press. Examining Lexical Tonal Contrast in Norwegian Using Intonation Modelling. In *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, UK.
- Gregor Möhler. 1998. Describing intonation with a parametric model. In *ICSLP*, volume 7, pages 2851–2854.
- Gregor Möhler. 2001. Improvements of the PaIntE model for F<sub>0</sub> parametrization. Technical report, Institute of Natural Language Processing, University of Stuttgart. Draft version.
- Petr Pajas and Jan Štěpánek. 2009. System for Querying Syntactically Annotated Corpora. In *ACL-IJCNLP: Software Demonstrations*, pages 33–36, Suntec, Singapore.
- Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes. In *CoNLL: Shared Task*, pages 1–27, Portland, Oregon, USA.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *EMNLP-CoNLL: Shared Task*, pages 1–40, Jeju Island, Korea.
- Arndt Riester and Jörn Piontek. in press. Anarchy in the NP. When new nouns get deaccented and given nouns don’t. *Lingua*.
- Katrin Schweitzer, Michael Walsh, Sasha Calhoun, Hinrich Schütze, Bernd Möbius, Antje Schweitzer, and Grzegorz Dogil. 2015. Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation. *Speech Communication*, 66(0):65–81.
- Ann K. Syrdal, Julia Hirschberg, Julie McGory, and Mary Beckman. 2001. Automatic ToBI Prediction and Alignment to Speed Manual Labeling of Prosody. *Speech Commun.*, 33(1-2):135–151.
- Michael Wagner. 2006. Givenness and Locality. In Masayuki Gibson and Jonathan Howell, editors, *Proceedings of SALT XVI*, pages 295–312.
- P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. 2006. ELAN: a Professional Framework for Multimodality Research. In *LREC*.
- Amir Zeldes, Julia Ritz, Anke Lüdeling, and Christian Chiarcos. 2009. ANNIS: A Search Tool for Multi-Layer Annotated Corpora. In *Proceedings of Corpus Linguistics*.