Spokes – a search and exploration service for conversational corpus data

Piotr Pęzik
University of Lodz
Corpus & Computational Linguistics Laboratory
pezik@uni.lodz.pl

Abstract

Spokes is an online service for conversational corpus data search and exploration, currently developed as part of CLARIN-PL – the Polish CLARIN infrastructure. This paper describes the data sets currently available through Spokes, the architecture of the service and the data and metadata search functionality it provides to its users. We also introduce some of the more experimental features which have been developed to facilitate more advanced research on multimodal conversational corpora.

1 Introduction

Open-access speech corpora and speech databases are still rare and undersized, if at all available, for most languages. Many such resources have been made available simply as collections of annotated transcription and media files, which can be downloaded and processed by their prospective users, e.g. (Du Bois et al., 2000), (Coleman et al., 2012). For other corpora, dedicated, web-based tools have been developed, which make it possible not only to browse selected transcriptions and play the associated recordings, but also to search and retrieve text spans matching corpus queries, cf. (Johannessen et al., 2009), (Gasch, 2010), (Freitas and Santos, 2008). These tools vary widely with respect to their general functionality, query syntax and the range of supported export formats. For example, while some online tools only support basic searching for exact strings occurring in the transcriptions (Douglas, 2003), others make it possible to directly search and display time-aligned phonetic transcriptions[1].

The PELCRA Conversational Corpus (PELCRA CC) contains over 2.2 million words of casual Polish spoken data collected since 1999 in a number of research projects (Waliński and Pęzik, 2007), (Pęzik, 2012). The most recent set of samples was acquired and added to the corpus in the CLARIN-PL project. In contrast to other speech databases and spoken corpora available for Polish, the PELCRA CC includes mostly transcriptions of *in vivo* recordings of casual conversations, many of which were taken surreptitiously in everyday situations by trained acquisition agents. Although prior and ex post facto permissions were granted by the recorded speakers to process and distribute the transcriptions for research purposes, many of the speakers did not realise their conversations were being recorded at the exact time of recording[2]. This in turn makes this corpus particularly useful for casual spoken discourse studies as well as for the development of formal models of casual Polish speech (Pęzik, 2012).

Although parts of the PELCRA CC corpus have previously been released in raw source formats under open-source licenses, its full research potential has remained dormant for many potential users such as linguists and spoken discourse analysts from domains other than linguistics. This was mainly due to the technical difficulties related to exploring large quantities of casual conversational data. Many researchers simply lack the technical expertise needed to process XML-encoded transcription files in order to extract relevant samples of texts. Also, due to their sheer size, the sound files available for the transcriptions have proved problematic for users who need to identify and analyse large sets of audio samples.

[1] See, for example, the Spock system (http://spock.iltec.pt/) developed for the CORP-ORAL corpus (Freitas and Santos, 2008).

[2] By contrast, the term *in vitro* speech corpora can be used to describe corpora which contain mainly data from scheduled interviews arranged specifically for the purpose of corpus acquisition

To address the need for a centralized, user-friendly tool which would make this data more useful and more available to both technical and non-technical users, we have developed Spokes – a web-based service providing search and analysis functionality with GUI and programmatic access. Once the first version of Spokes for Polish conversational data was released, we proceeded to develop an experimental version of the service for the spoken component of the British National Corpus (BNC) in order to test those search features which require phone-level time-alignment of spoken data. Both of these versions of Spokes are discussed in this paper.

2 Annotation

As mentioned above, Spokes has so far been used for two PELCRA CC and spoken BNC data. The general nature of speaker and conversation metadata as well as the linguistic annotations is similar in both corpora and they are stored and searched using similar backend models. From the point of view of Spokes development, the most important difference between these two corpora is related to the level of phonetic annotation available as it has significant implications on how the data is searched and accessed.

2.1 PELCRA CC

The original PELCRA CC recordings were transcribed orthographically, anonymized and aligned manually at the level of utterances with ELAN (Wittenburg et al., 2006). In addition to basic demographic metadata about the conversations (such as place of recording, date, register) and speakers (such as age, sex, education), the transcriptions were automatically part-of-speech tagged and lemmatized. Using the manual time-alignments, it was also possible to extract the corresponding fragments of the recordings, process them and add pitch pattern annotation to individual utterance spans. The pitch properties for the data shown in Listing 1 were extracted with Praat (Boersma, 2002) and they include observed strength, intensity and frequency values for each time point.

Listing 1: Automatic pitch annotation in PELCRA CC.

```
<audio-segment id="Ekz6a">
   <pch s="0.778" i="0.171"
      t="0.230">164.648</pch>
   <pch s="0.899" i="0.150"
      t="0.240">164.273</pch>
   <pch s="0.915" i="0.135"
      t="0.250">164.214</pch>
  <pch s="0.936" i="0.176"
      t="0.260">163.977</pch>
  <pch s="0.960" i="0.199"
      t="0.270">163.405</pch>
  <pch s="0.934" i="0.203"
      t="0.280">161.971</pch>
  ...
</audio-segment>
```

Since the quality of some of the audio recordings is poor due to the conditions in which they were taken, each conversation was additionally rated and annotated for its overall acoustic quality. The ELAN-annotated transcriptions are transferred to a relational database for management and further processing. The audio recordings are stored in wav files, which are currently approximately 69 gigabytes in size, and accessed on demand by a file retrieval mechanism which is separate from the two other backends.

2.2 Spoken BNC

The spoken component of the BNC contains both text metadata such as source, text types and classification codes as well as speaker metadata including sex, social class, occupation and dialect codes. As a result of a joint project of the British Library Sound Archive and the Oxford University Phonetics Laboratory, most of the original recordings from the spoken component of the BNC were recently digitized from cassette tapes and made available with the time-aligned transcriptions (Coleman et al., 2012). In order to test the flexibility of the solution described in this paper we have transferred this release of spoken BNC data to a separate instance of Spokes. In contrast to the PELCRA CC data, which is only manually time-aligned at the level of utterances, the alignment available in the BNC relates individual phone units to time offsets. The phonetic transcription is available in the SAMPA (Wells and others, 1997) and IPA alphabets. Needless to say, this level of alignment opens up the possibility of supporting more sophisticated phonetic queries against this data, some of which are discussed below.

3 Architecture

A basic overview of the current Spokes architecture is presented in Fig. 1. The three main tiers of Spokes are the search and storage backends, the REST API service and the Web application. We believe that this separation has several advantages. First of all, all these three modules are separated and they can be distributed and hosted independently. Secondly, access to the backend modules is always mediated by the REST API which means that the backends do not need to be directly exposed to third party users. Finally, because the dedicated Spokes web application uses the same REST API service which is available for other (programmatic) clients, the web application developers were the first users to thoroughly test the API. In the process of developing the web application, a number of new API methods of serving and accessing the data were developed and made publicly available.

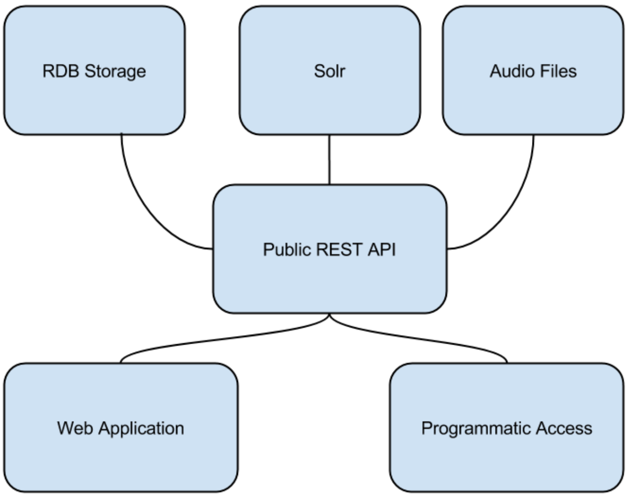


Figure 1: An overview of the Spokes architecture.

3.1 Backends

The three backend modules used in Spokes serve distinct purposes. The main function of the relational database backend[3] is to store and manage the data in a highly normalized model. It is also used for relatively standard data retrieval operations which are well supported in the SQL syntax, such as joins and aggregations. The Solr backend is used to provide very fast full text search capabilities and aggregated views of data matched by corpus queries. The main Solr core used in Spokes contains a flat index of all

[3] We are currently using PostgreSQL 9.2 for the RDB backend.

utterances, which in Solr parlance, are called "documents". In some cases, separate backend modules are used in tandem for different phases of a complex query. For example, while Solr is always used to fetch concordance spans, the RDB backend is used to fetch pitch data for the matching utterances which are then visualised in the front-end web application (see Fig. 5)[4] .

3.1.1 RDB

The XML-encoded datasets were transferred and normalized in the the RDB backend. For example, in the BNC data model, utterances, word tokens and phone segments are stored in separate tables in order to represent the information available in the original annotation. As a result, with a relatively simple SQL query, word and utterance durations can be aggregated from the durations of their constituent phones. The RDB storage has proved to be highly suitable for data management tasks such as batch updates and versioning or validation routines including detecting duplicated, missing and erroneous data values which are less straightforward to detect in the original XML-format of the transcriptions. Additionally, the RDB backend is used to directly support those search functions of Spokes which involve complex metadata aggregation and joining.

3.1.2 Solr

Apache Solr, which is based on Apache Lucene, is a general-purpose search technology known for its maturity, speed, scalability and advanced full-text search capabilities. As explained below, with some modifications to its standard index structure and query syntax, Solr can also be used to support positional queries on part-of-speech tagged corpus data. By customizing this technology, we took advantage of its highly performant "faceting" functionality which opens up interesting possibilities of query-based corpus metadata aggregation.

3.2 REST API

In addition to the Spokes web-application specific methods, the REST API exposes full metadata and partial data access methods. All of its methods and resources are documented using mashery/iodocs [5], which makes it convenient for API users to learn and test them interactively.

4 Search and exploration

4.1 Metadata browsing

Basic data retrieval and metadata browsing are the most obvious functionalities of Spokes. As shown in Fig. 2, users can browse and filter full transcription metadata. The PELCRA CC transcriptions can also be viewed, played, and downloaded with full utterance-level metadata about the individual speakers who took part in the conversations. Similar browsers are available for word frequency and formulaic sequence lists extracted from the two corpora.

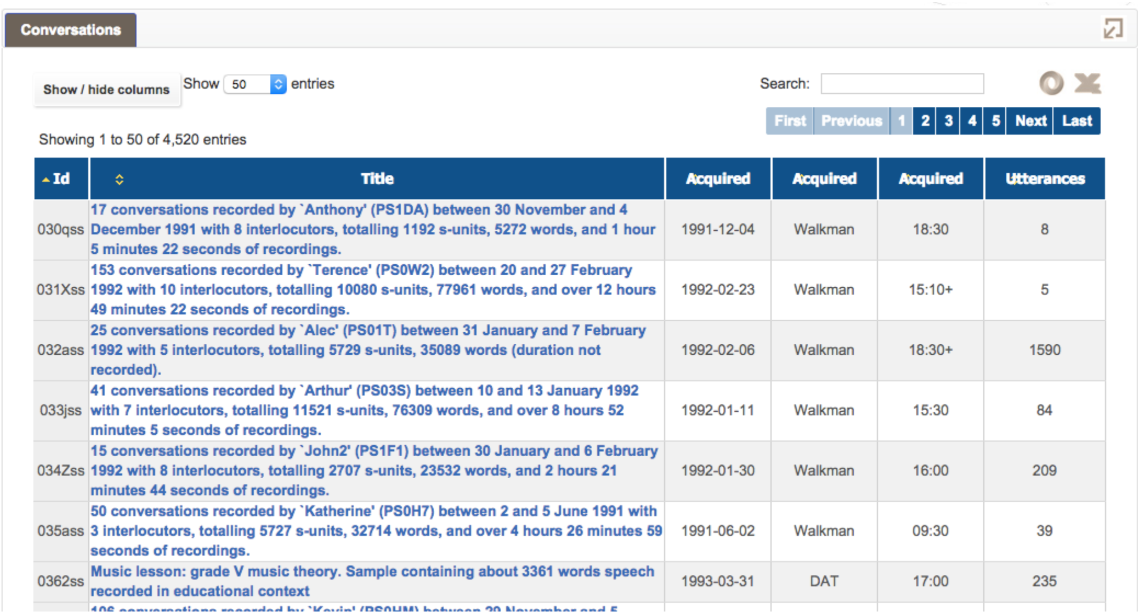
4.2 Corpus query syntax

In order to support positional concordance queries for annotated token sequences, we developed a dedicated text analyser and a query syntax for the Solr backend. The query syntax, called SlopeQ, is illustrated in Table 1. Apart from simple queries for surface and lemmatised terms, it supports regular expressions, part-of-speech terms, variant negation and slop-factor (proximity) operators.

The two proximity operators supported by Spokes can be particularly useful when searching for spans of word tokens which are often discontinued by discourse markers and hesitation devices in casual conversational data. For example, the query `(an|a way of <pos=v.+>)=2` will match sequences such as *Yeah, I mean there could be* ***a way of sort of coming*** *together.* It should be noted that regular expressions are only matched against single tokens which are listed in the inverted index of the corpus. Although it is possible to specify that a token matching a regular expression such as `<lemma=.*>` is required at a certain position, it is more convenient to use the slop factor syntax in order to match loosely

[4]All of the visualisations provided by Spokes can be downloaded in bitmap (PNG, JPEG) and vector (PDF and SVG) formats.

[5]See https://github.com/mashery/iodocs.



Id	Title	Acquired	Acquired	Acquired	Utterances
030qss	17 conversations recorded by `Anthony' (PS1DA) between 30 November and 4 December 1991 with 8 interlocutors, totalling 1192 s-units, 5272 words, and 1 hour 5 minutes 22 seconds of recordings.	1991-12-04	Walkman	18:30	8
031Xss	153 conversations recorded by `Terence' (PS0W2) between 20 and 27 February 1992 with 10 interlocutors, totalling 10080 s-units, 77961 words, and over 12 hours 49 minutes 22 seconds of recordings.	1992-02-23	Walkman	15:10+	5
032ass	25 conversations recorded by `Alec' (PS01T) between 31 January and 7 February 1992 with 5 interlocutors, totalling 5729 s-units, 35089 words (duration not recorded).	1992-02-06	Walkman	18:30+	1590
033jss	41 conversations recorded by `Arthur' (PS03S) between 10 and 13 January 1992 with 7 interlocutors, totalling 11521 s-units, 76309 words, and over 8 hours 52 minutes 5 seconds of recordings.	1992-01-11	Walkman	15:30	84
034Zss	15 conversations recorded by `John2' (PS1F1) between 30 January and 6 February 1992 with 8 interlocutors, totalling 2707 s-units, 23532 words, and 2 hours 21 minutes 44 seconds of recordings.	1992-01-30	Walkman	16:00	209
035ass	50 conversations recorded by `Katherine' (PS0H7) between 2 and 5 June 1991 with 3 interlocutors, totalling 5727 s-units, 32714 words, and over 4 hours 26 minutes 59 seconds of recordings.	1991-06-02	Walkman	09:30	39
0362ss	Music lesson: grade V music theory. Sample containing about 3361 words speech recorded in educational context	1993-03-31	DAT	17:00	235

Figure 2: A transcription metadata browser.

defined phrases which contain two or more obligatory tokens. Additionally, the slop factor queries are noticeably faster than their regular expression equivalents in large corpora.

4.3 Metadata query syntax

Spokes allows users to run metadata queries which are formulated in the Solr Extended DisMax syntax[8]. Currently these queries are always appended as filters to an obligatory concordance query. For example, the following DisMax filter: `speaker_age:[0 TO 10]` can be appended to the SlopeQ query "mamo" (Pol. "mom" in the vocative case) in order to make sure that only concordances of this word found in utterances of speakers up to 10 years old are returned. Some of these filtering criteria can also be set using the graphical controls of the web application user interface.

4.4 Concordance grouping

Another search option of Spokes based on Solr is the ability to group concordances matched by a SlopeQ query by one or more metadata field values. For example, it is possible to define the maximum number of concordance results per speaker or text identifier. This in turn may serve as a basic way of sampling the results of queries which are likely to match many spans per conversation. The range of the matching concordances can be specified as well. It is thus possible to sample and group results from different sections of the corpus. It is also possible to implement hierarchical grouping of concordances, which would make it possible to specify the maximum size of samples matching a conjunction of metadata field values. For example, the maximum number of utterances with a unique combination of speaker and text identifiers or any other metadata field value stored in the Solr backend could be specified to be fetched in the concordances matching a query.

4.5 Facets

One of the Solr-based features of Spokes is its facet generation mechanism. For any concordance query, the Solr backend automatically runs a full aggregative query which collects counts of distinct metadata values found in all the matching documents. In other words, even if the user chooses to fetch, say, only 20 matching concordance spans from the index, a full report about the number of matching results found

[6] The pipe operator is always interpreted as a token boundary.
[7] The negated variant is marked with "!" and subtracted from the set of alternatives specified on the same position.
[8] See https://wiki.apache.org/solr/ExtendedDisMax.

#	Query	Returns text spans containing
1	mamo	A single surface token
2	wiesz co	A sequence of surface tokens
3	<lemma=palić>	All variants of a single lemma token
4	<lemma=mieć> <lemma=szansa>	A sequence of immediately adjacent lemma tokens
5	słuchaj <lemma=ja>	A sequence of surface and lemma tokens
6	tu|tutaj	Variants separated by the pipe operator[6]
7	<lemma=facet>|<lemma=koleś>	Lemma variants
8	bardzo|strasznie dużo	Surface token variants in a sequence
9	(ta kobita)=1	Sequence separated by zero or one unspecified tokens
10	(<lemma=jechać> tam)=2	A lemma and a surface token separated by up to 2 tokens
11	(<lemma=jechać> tam)∼2	As above except that the tokens may occur in any order
12	(<lemma=dać> do zrozumienia)=2	3 obligatory tokens separated by up to 2 unspecified ones
13	<lemma=prosić>|!proszę	Any form of *prosić* except for *proszę*[7]
14	t.* bab.*	Tokens matching regular expressions
15	szykow.+|przygotow.+	Variants with regular expressions
16	<lemma=p.+biec>	Lemmas matching a regular expression
17	<tag=subst:pl:.*>	Any plural noun
18	<tag=subst:.+:inst:.+>	Any noun in the instrumental case
19	<lemma=zdać pos=verb:sg:.*>	Singular forms of the verb "zdać"
20	<tag=adj:.+> <lemma=temat>	Sequences of adjectives preceding the noun "temat"
21	(<lemma=słuchać> <tag=.*:gen:.*>)=1	Lemma followed by any genitives with a slop factor

Table 1: SlopeQ 2 query syntax

in the different sections of the corpus is returned. Fig. 3 shows an example facet-based report for the concordance query *aye*. The available facets are listed in the left panel and they can be visualised in the middle panel as piechart graphs or bar plots. The one hundred most frequent values of each facet are also listed in the right panel, where they can be selected and deselected as filtering criteria.

4.6 Collocations

Another feature of Spokes which is aimed at helping users digest large sets of concordance results is the positional collocation extraction module[9]. The module can be used to aggregate a frequency list of the most frequent tokens co-occuring with the spans matched by any SlopeQ query. Fig. 4 shows an example concordance query for all inflections of the adjective *dobry* which has been transformed into a collocation query. Users need to specify the maximum number of contexts from which collocates will be extracted as well as the allowed part-of-speech tags and positions of potential collocates. The resulting list of frequent collocates contains 35 combinations which occurred four or more times in the concordance results. Each of the positionally defined collocates in this list is presented as a hyperlink to its full concordance. It is possible to extract potential collocates from sets of up to 100 000 matching spans in a single query.

4.7 Data export

Although we have tried to make the web application as powerful and easy to use as possible, it is nevertheless possible to envisage non-technical users who will still want to download thousands of results in order to further process and analyse them. For example, some researchers may want to filter and annotate all the instances of a specific linguistic feature which may be difficult to describe with the query syntax supported by Spokes. To address this need, we make it possible to download up to 100 000 results per

[9]The term "positional collocation extraction" is used here to refer to extraction methods which only rely on aggregating co-occurrences of words in a predefined window rather than explicitly encoded syntactic relations between them, cf. (Evert, 2004).

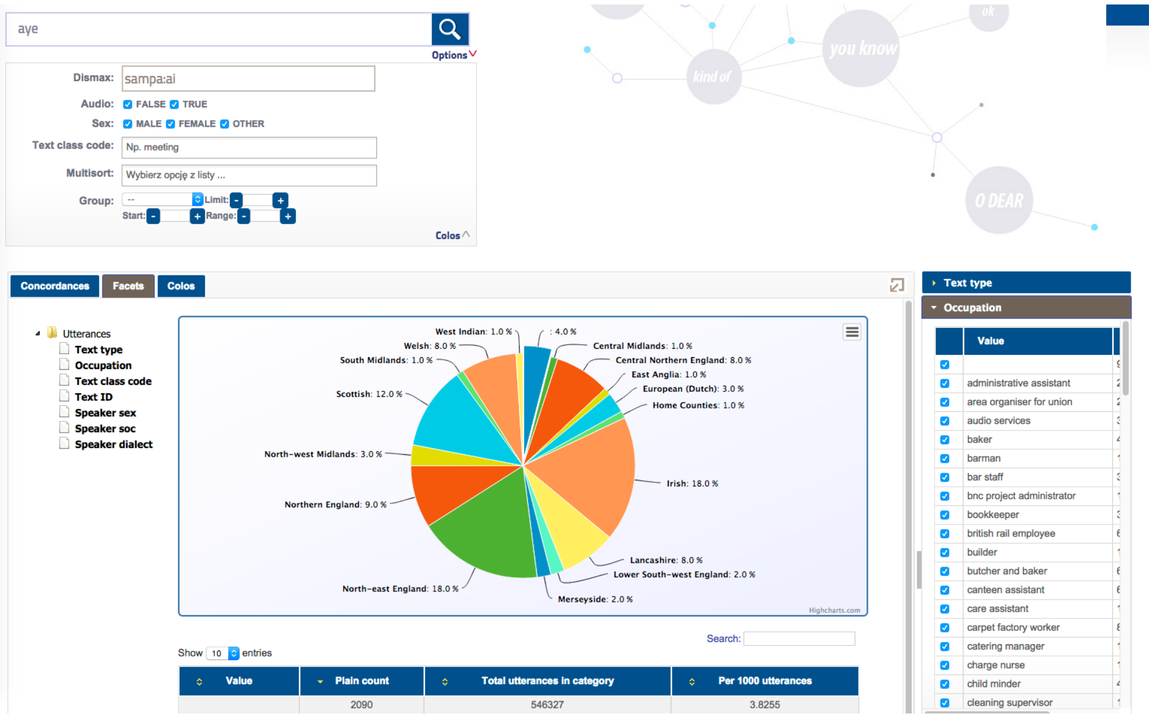


Figure 3: Interactive search facets.

single query in the form of an MS Excel spreadsheet (which can also be processed in OpenOffice). Apart from metadata-annotated listings of concordance lines, such spreadsheets also contain complete sets of facets extracted for a given query.

4.8 Searching phonetic annotation

The spoken BNC data indexed in Spokes can be searched using SAMPA- and IPA-encoded query terms which correspond to the phonetic transcriptions of word tokens. Such queries can be combined by means of logical operators with conditions specified for other index fields. For example, the SlopeQ query `row` can be combined with the Sampa query `r@U` or the equivalent IPA query rəʊ to return spans where the corresponding pronunciation of the word *row* was recognized.

Users can display pitch annotations (such as the f0 values shown in Listing 1) for any concordance query, which returns spans aligned with the time offsets of the utterances. This is illustrated in Fig. 5, which shows a pitch contour for an utterance matching a concordance query. Thanks to the availability of phone and word level time alignment in the spoken BNC data, it is possible to generate similar contours for the exact spans (and not just utterances) matching concordance queries. This functionality is also an example of how the separate backend modules are combined to serve different types of data.

5 Experimental features

One of the experimental features of Spokes, which may be particularly useful in spoken discourse analysis makes it possible to carry out on-the-fly analyses of the duration of spans matching users' corpus queries. For example, users who type in a corpus query matching the word "right", which happens to be both polysemous as a lexical item and multifunctional as a discourse marking device in conversational English. may want to order the resulting concordances by their observed duration whenever such time-alignment is available for a transcribed span. The purpose of generating such a ranking would be to check whether certain meanings or functions of "right" are marked by longer or shorter durations. For instance, one working hypothesis here could be that instances of "right" as a turn-opening discourse marker may be characterized by significantly higher average or median duration values than instances of "right" as an adjective pre-modifing heads of noun phrases.

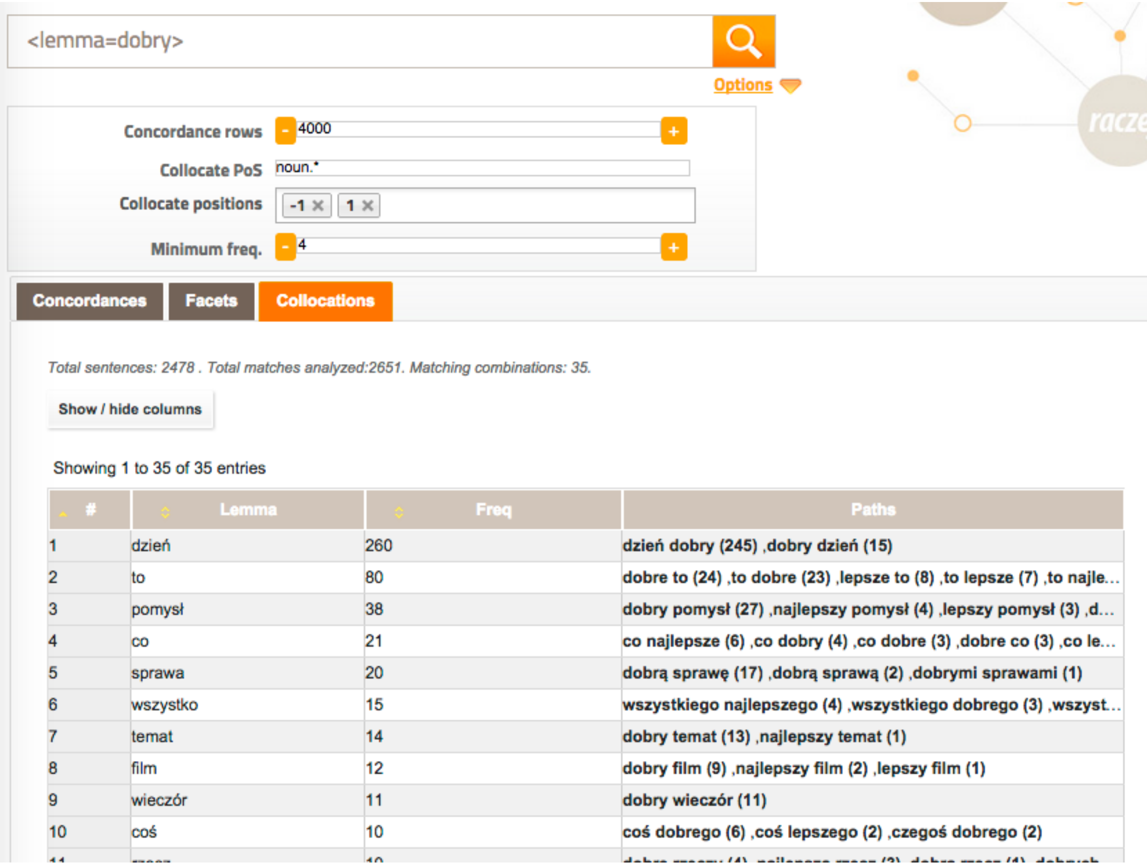


Figure 4: Collocation extraction in Spokes.

This type of analysis is made possible by combining information available in two backend modules. Table 2 shows a simplified representation of word token values stored in the RDB backend. The last column in this view shows the duration of a given word derived from the relative offsets shown in columns 3 and 4 extracted from the original BNC transcriptions. The lists of concordances retrieved from the Solr index for any corpus query contains the same identifiers of the word tokens occurring in the matching spans. It is therefore possible to use those identifiers to join the database records and sort or pass them to an aggregate function.

The result of this operation is not only a duration-sorted list of concordances, but also a summary of descriptive statistics for the sample of concordance spans retrieved. For example, Fig. 6 shows a standard box plot for the 6433 time-aligned instances of the word "right" found in the BNC data indexed by Spokes. The median duration of "right" is 190 ms with a mean of 216.4 ms and standard deviation of 132.8. The box plot reveals number of outliers with the maximum value of 2.2 seconds, which could be rather long but genuine instances of the word or simply misalignments. A similar analysis can be carried out for any concordance results matching multiword unit spans which can be specified in the SlopeQ syntax.

6 Challenges and planned developments

As described above, the analysis of the distribution of duration as a prosodic feature is fairly straightforward to implement. A much more challenging extension of Spokes which we are currently developing is motivated by the need to enable automatic identification and classification of pitch patterns of concordance spans. There is a considerable amount of research into recurrent discourse-functional lexical sequences which seem to exhibit regular prosodic characteristics such as "stereotyped" intonation contours (Bolinger, 1986). One hypothesis formulated in such studies is that "intonation conveys information about the intentional as well as the attentional structure of discourse" (Hirschberg and Pier-

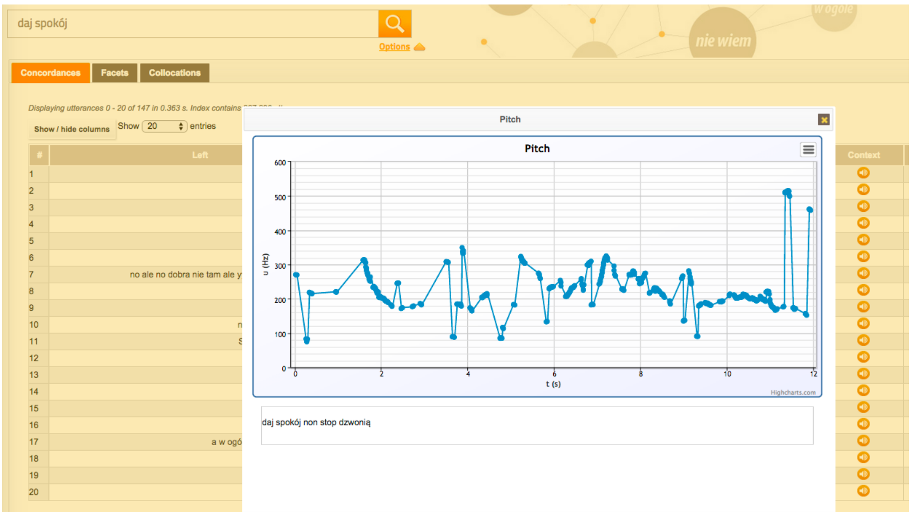


Figure 5: Pitch annotations for utterances and matched concordance spans.

Id	Word	Start	Stop	Duration
45799700	right	538673	539013	340
46153393	right	811603	811733	130
49168719	right	1999793	2000043	250
46154674	right	2296643	2296823	180
45802240	right	2651043	2651133	90
47388854	right	372829	373159	330
46155247	right	2306773	2306863	90
47388909	right	375643	375733	90
45802580	right	2659473	2659613	140
45803157	right	195935	196295	360

Table 2: Time offsets for retrieved concordance spans.

rehumbert, 1986). For example, one of the functions of “hello” is to express surprise or irritation at what has just been said or done. This function, as opposed to its “greeting” function, seems to be prosodically stereotyped in that it is marked by a rising pitch contour of this word.

We are currently investigating the possibility of using time-aligned conversational data to facilitate such analysis in Spokes. As already mentioned, pitch annotations for every utterance indexed in the Solr backend are stored and can be retrieved from the RDB backend. In the case of the BNC data, they can also be mapped to word tokens matched in concordance spans. This in turn makes it possible to extract and analyze thousands of automatically recognized pitch patterns for any concordance query. Such results could simply be presented to the user as shown in Fig. 5 for further inspection. Users of Praat may actually prefer to download the audio snippet provided by Spokes and perform their analysis offline.

At the same time, it may also be possible to use different methods of automatic detection and categorization of prosodic events (**?**) to provide an automatic classification of pitch contours observed in the concordance spans retrieved for a query. The technique we have experimented with so far involves producing a distance matrix based on dynamic time warping similarity values (cf. (Müller, 2007), which is computed for all pairs of pitch contours observed in the concordance spans. The resulting distance matrix is then used to produce a dendrogram showing clusters of concordance spans with “similar” pitch contours.

This feature of Spokes is still highly experimental and it requires careful validation. The choice of a reliable measure of comparing and clustering highly volatile prosodic signals remains a challenge. Also,

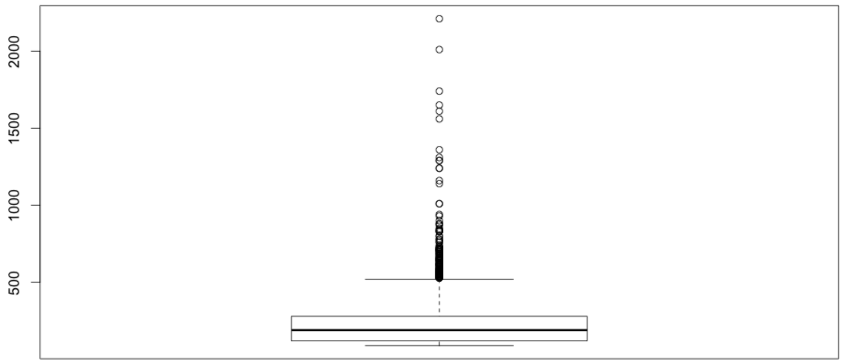


Figure 6: A box plot generated for the 6433 time-aligned instances of “right” found in the spoken BNC data.

the poor audio quality of many of the original recordings also makes it difficult to compare pitch contours for different instances of the same word forms.

7 Availability

Current versions of the Spokes for PELCRA CC and Spokes for BNC web application services are publicly available at `http://spokes.clarin-pl.eu`[10] and `http://pelcra.clarin-pl.eu/SpokesBNC`. The help pages of these applications provide up-to-date links to the REST API. The entire Spokes data will also be available through a Federated Content Search endpoint as part of the CLARIN-PL resource center.

Acknowledgements

The work described in this paper has been financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education. The relational database backend and the REST service was implemented mainly by Łukasz Dróżdż. Paweł Wilk and Paweł Kowalczyk are the main developers of the Web application. Finally, we extend thanks to Danijel Koržinek for sharing his expertise in developing speech signal analysis modules used in Spokes.

References

[Boersma2002] Paul Boersma. 2002. Praat, a system for doing phonetics by computer. *Glot international*, 5(9/10):341–345.

[Bolinger1986] Dwight Bolinger. 1986. *Intonation and its parts: Melody in spoken English*. Stanford University Press.

[Coleman et al.2012] John Coleman, Ladan Baghai-Ravary, John Pybus, and Sergio Grau. 2012. Audio BNC: the audio edition of the Spoken British National Corpus.

[Douglas2003] Fiona M Douglas. 2003. The scottish corpus of texts and speech: Problems of corpus design. *Literary and linguistic computing*, 18(1):23–37.

[Du Bois et al.2000] John W. Du Bois, Wallace L. Chafe, Charles Meyer, Sandra A. Thompson, Robert Englebretson, and Nii Martey. 2000. Santa Barbara corpus of spoken American English.

[10]See also `http://hdl.handle.net/11321/47`.

[Evert2004] Stefan Evert. 2004. *The statistics of word cooccurrences*. Ph.D. thesis, PhD Dissertation, Stuttgart University.

[Freitas and Santos2008] Tiago Freitas and Fabíola Santos. 2008. Corp-oral: Spontaneous speech corpus for european portuguese. In *Proceedings of LREC*.

[Gasch2010] Joachim Gasch. 2010. Dgd 2.0: A web-based navigation platform for the visualization, presentation and retrieval of german speech corpora. *Sprache und Datenverarbeitung*, 34(1):27–38.

[Hirschberg and Pierrehumbert1986] Julia Hirschberg and Janet Pierrehumbert. 1986. The intonational structuring of discourse. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, pages 136–144. Association for Computational Linguistics.

[Johannessen et al.2009] Janne Bondi Johannessen, Joel Priestley, Kristin Hagen, Tor Anders Åfarli, and Øystein Alexander Vangsnes. 2009. The nordic dialect corpus-an advanced research tool. In *Proceedings of the 17th Nordic conference of computational linguistics NODALIDA 2009. NEALT proceedings series*, volume 4, pages 73–80.

[Müller2007] Meinard Müller. 2007. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84.

[Pęzik2012] Piotr Pęzik. 2012. Język mówiony w NKJP. In Adam Przepiórkowski, Mirosław Bańko, Rafał Górski, and Barbara Lewandowska-Tomaszczyk, editors, *Narodowy Korpus Języka Polskiego*, pages 37–47. Wydawnictwo Naukowe PWN, Warszawa.

[Waliński and Pęzik2007] Jacek Waliński and Piotr Pęzik. 2007. Web access interface to the PELCRA referential corpus of polish. pages 65–86. Lang.

[Wells and others1997] John C Wells et al. 1997. Sampa computer readable phonetic alphabet. *Handbook of standards and resources for spoken language systems*, 4.

[Wittenburg et al.2006] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes. 2006. Elan: a professional framework for multimodality research. In *Proceedings of LREC*, volume 2006.