Visualisation in Speech Corpora: Maps and Waves in the Glossa System

Michał Kosek, Anders Nøklestad, Joel Priestley, Kristin Hagen and Janne Bondi Johannessen
The Text Laboratory, Department of Linguistics and Scandinavian Studies
University of Oslo
Oslo, Norway
{michalkk,noklesta,joeljp,kristiha,jannebj}@iln.uio.no

Abstract

We present the Glossa web-based system for corpus search and results handling, focussing on two modes of visualisation implemented in the system. First, we describe the use of maps to show the geographical distribution of search results and its utility for exploring dialectal variation and discovering new isoglosses. Secondly, we present a functionality for speech visualisation, yielding dynamically generated representations of spectrograms, pitch and formants. The analyses are accompanied by the ability to replay selected parts of the waveform, as well as export and compare maximum, minimum and average values of the parameters for different selections. Among other things, this can be used to explore in more detail the set of spoken variants revealed by the geographical map view.

1 Introduction

The current availability of large corpora presents a new challenge for corpus research: how to process hundreds of millions or even billions of tokens, extract relevant information and transform it into a shape that can be used to explore specific hypotheses about language. In addition, the emergence of extensive collections of audio- and video-recorded speech, accompanied by transcriptions as well as metadata about the speakers such as their age, sex and geographical location, presents us with the challenge of how to make these additional sources of linguistic data readily available to language researchers. To help with these tasks, the Glossa corpus search system has been developed. The present article discusses the visualisation possibilities that Glossa provides for speech corpora, which are particularly useful for research within phonetics, phonology and dialectology.

The most important role of a speech corpus is to give access to data that are otherwise difficult to obtain. Without a corpus, researchers either need to trawl through a large amount of pre-existing speech recordings, listening for words or patterns they are interested in, or elicit such patterns from speakers. The former option is very time-consuming, and the latter requires setting up an experiment, which may introduce a bias, since the experimental setup would be designed with specific research questions in mind. Speech corpora do not necessarily solve the observer's paradox (Labov, 1972). They do, however, provide speech data that are not affected by a specific research question. Furthermore, in some cases participants are not even aware that their speech will inform linguistic research. BigBrother[1] is one such corpus.

However, mere access to speech data is not sufficient. With the large amount of material currently available for many languages, techniques for extracting and visualising potentially interesting patterns in the data have become more important than ever. It is both important to have a way to group a large number of search results in a meaningful way, and to have the possibility to "zoom in" and analyse detailed features of the speech. If we take dialect research as an example, the corpus user may be interested in the relation between pronunciation and the place where the speakers live. The first of the presented features of Glossa groups the search results according to the phonetic transcription and geographical location and visualises them on a map. The user may then want to perform a more detailed analysis of pronunciation samples from some specific places shown on the map. Another feature of Glossa may then be used: the sound visualisation, which gives quick access to the spectrogram, waveform, pitch and formant

[1] `http://www.tekstlab.uio.no/nota/bigbrother/english.html`

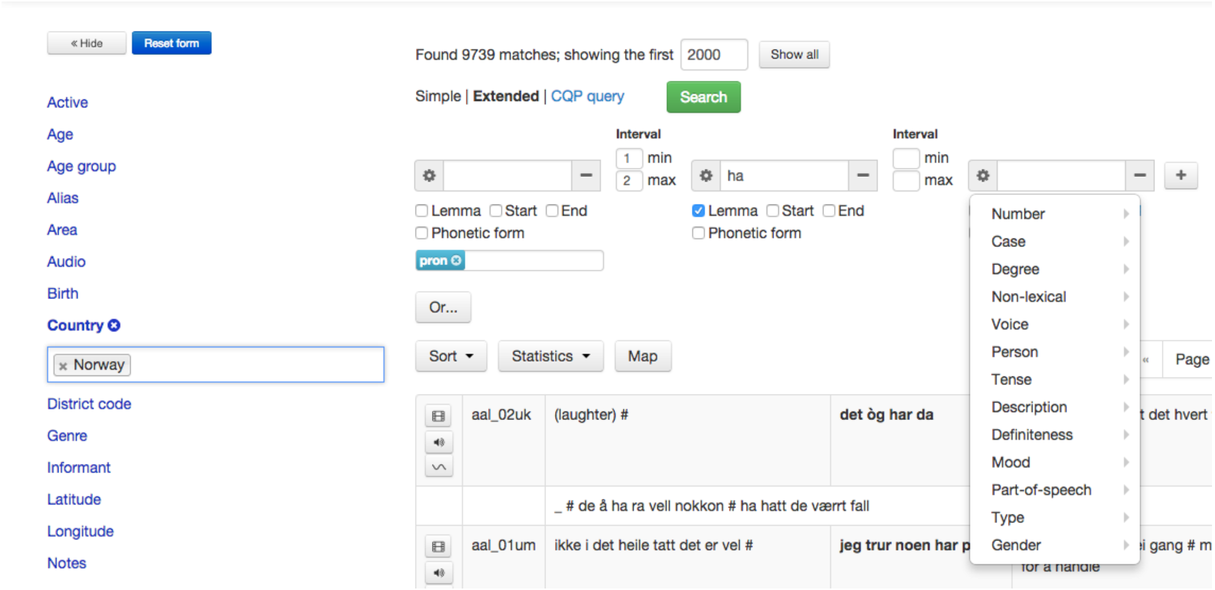

Figure 1: Advanced search in Glossa. We have searched in Norwegian dialects for a pronoun and the lemma *ha* separated by 1 or 2 tokens, and are in the process of selecting grammatical features for a third search term.

plots of the speech.

The article is structured as follows. Section 2 introduces general information about Glossa, and in particular, features related to speech corpora. The focus of the rest of the article is on visualisation possibilities of speech corpora in Glossa: geographical visualisation is presented in section 3, and speech visualisation in section 4. Section 5 describes technical details related to the implementation of these features. Then, section 6 discusses research applications that can benefit from these types of visualisation. Finally, section 7 discusses possible improvements of the features that may make research even more effective.

2 The Glossa Corpus Search System

2.1 Features of Glossa

Glossa is a web application that provides powerful methods for corpus search and result visualisation combined with a strong focus on user-friendliness. It allows a user to search monolingual and multilingual (parallel) text and speech corpora annotated with grammatical analyses or other types of token information. By selecting a set of metadata values (such as the author and publisher of a written text or the location and age of a speaker), the user can limit the search to a certain subcorpus.

There are three alternative search interfaces, ranging from maximum ease of use to maximally powerful queries:

a) a Google-like search box for simple token or phrase searches,

b) a set of text inputs, checkboxes and drop-down menus for more complex, grammatically specified searches (see Figure 1),

c) a search box for queries that are directly passed to the underlying search engine.

Results are presented as KWIC concordances, with the additional possibility to generate frequency distributions for tokens, lemmas or parts of speech. Glossa is easily installed on servers or laptops via Docker (see section 5). Alternatively, the source code can be freely downloaded from GitHub[2] under a very permissive open-source licence (MIT).

Out of the box, Glossa comes with support for corpora encoded with the IMS Open Corpus Workbench (Christ, 1994; Evert and Hardie, 2011)[3], which supports up to 2.1 billion tokens per corpus. However, Glossa was built from the ground up to be easily extended with support for different search engines and corresponding search and result views, and there is already an optional module for searching corpora on remote servers

[2] `https://github.com/textlab/glossa`
[3] `http://cwb.sourceforge.net`

using the Federated Content Search protocol defined by the European CLARIN infrastructure[4].

Glossa provides a simple admin interface which allows a Corpus Workbench (CWB) corpus to be created by uploading a zip file containing CWB indexes and potentially also tab separated value files with metadata as well as audio and video files if applicable. Glossa itself does not provide functionality to create these input resources; however we are currently working on a corpus processing pipeline for creating corpora from XML or plain text files, including TEI[5] format for written corpora and ELAN[6] format for speech corpora.

It should be noted that Glossa is not the only web-based corpus search system available; some examples of powerful alternatives are CQPweb[7], Corpuscle[8], Korp[9], and SketchEngine[10]. What sets Glossa apart from these is a unique combination of characteristics:

- the functionality for audio analysis and display of geographical distribution described in this paper,
- support for parallel queries in multilingual corpora,
- a strong focus on ease of use for non-technical users,
- ease of installation (particularly through its Docker distribution),
- extensibility with respect to different search engines and database systems,
- it is freely available without charge.

2.2 Speech Corpora in Glossa

With speech corpora, search results can be linked to audio and video clips that are accompanied by an auto-cue display showing each transcribed utterance as it is spoken. The utterances may have several different transcriptions. For example, in the Nordic Dialect Corpus (Johannessen et al., 2009), they are transcribed into the standard orthography, and to a simplified phonetic transcription, which shows how the word was actually pronounced in a particular utterance. The phonetic search feature allows looking for utterances where a word is pronounced in a particular way.

If the speakers in a corpus were recorded at different geographical locations (such as in a dialect corpus) and geographical coordinates are provided for these locations, search results can also be visualised as plots on a geographical map. Furthermore, if audio recordings are available, each search result can be analysed in an interface that implements the most important functionality found in desktop applications for sound analysis such as Praat (Boersma and Weenink, 2001). The rest of this paper will focus on the latter two functionalities: geographical maps and sound analysis.

3 Geographical Visualisation in Glossa

Corpus linguistic investigation commonly draws on analytic and communicative techniques taken from other fields. Dialectologists interested in regional variation have long turned to manually rendered maps to represent linguistic features. Perhaps the earliest example of such work is *Der Deutsche Sprachatlas*, carried out in the first half of the twentieth century and more recently digitised as a result of the *Digitaler Wenker-Atlas* (DiWA) project (Schmidt et al., 2001). However, as late as 2005, the lack of automation was still a concern (Labov et al., 2005, 41–42). Work has since been done to automatically render linguistic data in geographical maps, for example *Dynamic Syntactic Atlas of the Dutch dialects* (Barbiers and others, 2006, DynaSAND). The following section details one simple approach to achieving the type of digital linguistic topography heralded by Labov.

Geographical location is an essential metadata component of any speech corpus, and can be utilised in search visualisations. The Google Maps Embed API[11] has proven useful in extending the functionality of Glossa to this end. While originally incorporated into Glossa as a way of providing a metadata overview for corpus queries, it soon became evident that the spatial distribution of data reveals interesting patterns, particularly for corpora comprising multiple layers of transcription. The Norwegian component of the Nordic Dialect Corpus (Johannessen et al., 2009) is one such example; it is transcribed using a simplified phonetic

[4] http://clarin.eu
[5] http://www.tei-c.org/index.xml
[6] https://tla.mpi.nl/tools/tla-tools/elan
[7] https://cqpweb.lancs.ac.uk
[8] http://clarino.uib.no/korpuskel/page
[9] http://spraakbanken.gu.se/eng/korp-info
[10] http://www.sketchengine.co.uk
[11] https://developers.google.com/maps/web

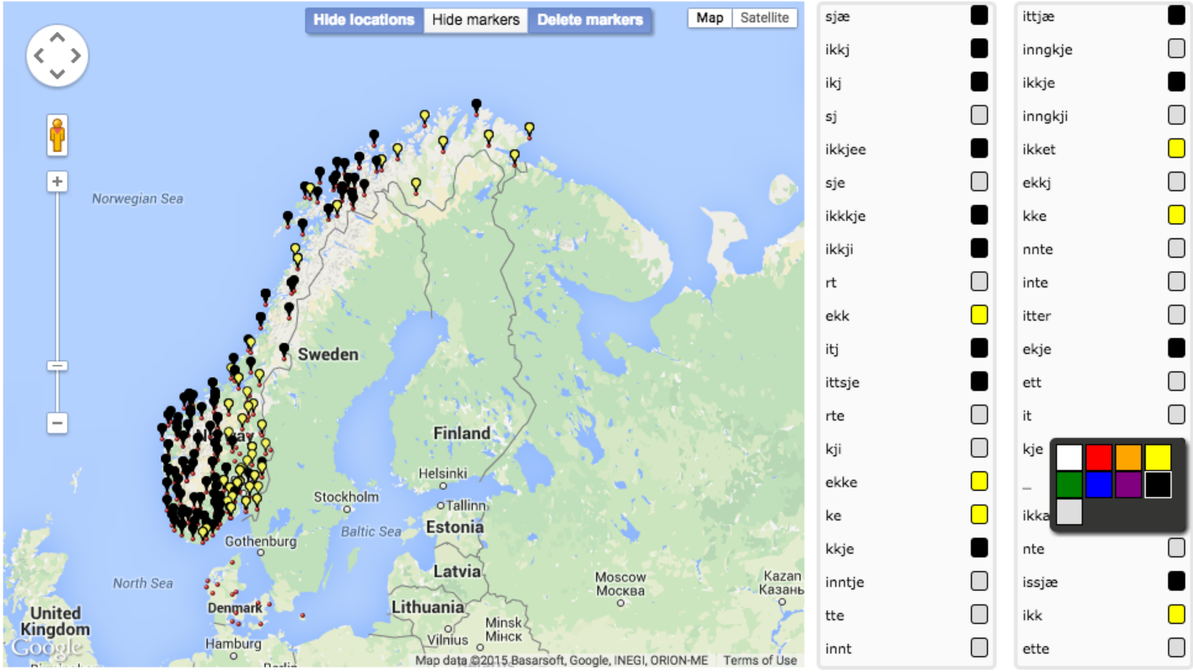

Figure 2: Geographical visualisation in Glossa. We can see the geographical distribution of dialectal variations of *ikke* 'not' in Norway: yellow = velar plosive; black = fricative/affricate, non-nasal

system as its initial layer[12], with an automatically transliterated orthographic transcription providing a subsequent layer. This dual-tiered approach preserves the rich dialectal variation while ensuring searchability. Simple orthographic searches may yield dozens of distinct dialectal variants. Using the Embed API, Glossa takes advantage of this system of transcription by compiling the phonetic variants along with their corresponding geographical coordinates into a data structure and plotting the locations onto a Google map. Colour palettes are provided, giving the user the option of colour coding specified variants. The resulting clustering enables users to easily scan the distribution of linguistic features. For corpora of a sufficient size and geographical distribution, isoglosses are readily visible (see Figure 2). The size of the corpus, in terms of tokens per informant, will determine which linguistic features will be available for examination; less frequent features requiring more data. In the example, there is an average of about 4000 tokens per informant, with 564 informants spread amongst 163 locations.

[12]The transcription standard is described here: `http://www.tekstlab.uio.no/nota/scandiasyn/Transkripsjonsrettleiing%20for%20ScanDiaSyn.pdf` (in Norwegian). It is a coarse-grained transcription standard based on the Oslo Norwegian pronunciation of the alphabet (Papazian and Helleland, 2005).

4 Speech Visualisation in Glossa

4.1 Background

Sound visualisation applications are an important part of the phonetician's toolkit; they provide information that does not depend on the subjective perception of the sound signal and can be presented and referred to in a written form. Among various sound analysis programs, Praat (Boersma and Weenink, 2001) is specifically designed to visualise and extract parameters of speech and therefore it is a standard tool within phonetics. Among its many features, the most significant ones include: visualisation of the waveform and the spectrogram, pitch and formant analysis.

Speech corpora not only enable easy access to a large amount of spoken utterances, but also allow the user to restrict the search according to a range of variables, based on corpus annotations and metadata. For instance, phonological annotations allow the user to find all words that share a particular pronunciation, part-of-speech tagging can be used to differentiate between some of the homonyms, and metadata may be used to specify sex and dialect of the speaker.

It would therefore be natural to use Praat to analyse search results from speech corpora. Unfortunately, it is often difficult or even impossible.

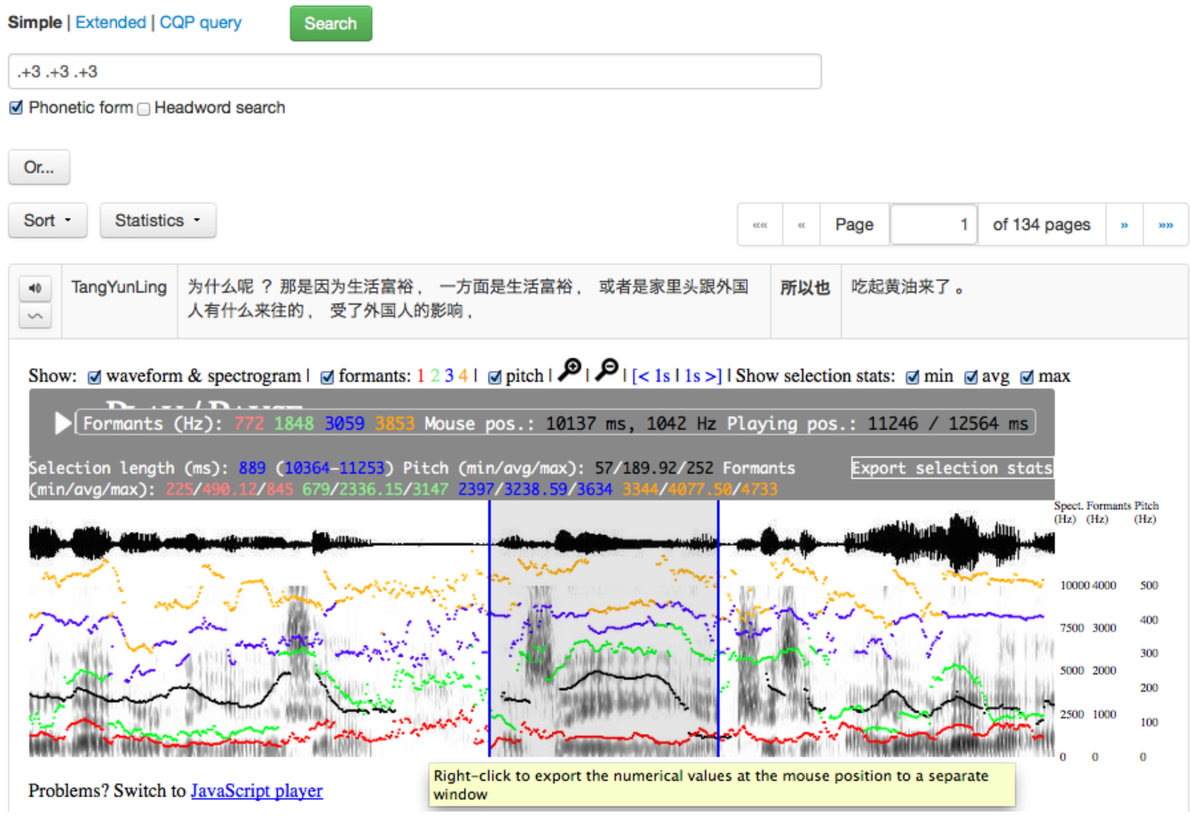

Figure 3: Sound visualisation in Glossa. We have searched a Mandarin Chinese corpus for utterances with three consecutive third tones. The pitch plot (black) shows that in the presented search result only the last syllable is pronounced with the actual low third tone. The figure also shows waveform, spectrogram and formant plots, and the user interface that provides functionalities described in the article.

Praat operates only on local files saved on the hard drive. Due to copyright restrictions, speech corpora do not commonly allow sound files to be directly downloaded. Most often the search results are only available via streaming and can only be listened to through a web browser.

If the download links are available, Praat might be sufficient for research that requires deep analysis of a relatively small amount of samples. There are, however, many situations in which the use of Praat is suboptimal. For instance, in the exploratory phase of the research, when one tries to formulate a hypothesis, it is often beneficial to conduct a lot of different searches according to different criteria and perform a quick analysis of the results. In such cases, downloading each result separately and repeatedly switching between separate applications is ineffective.

4.2 Features of the Current Visualisation System

The online speech visualisation in Glossa contains the most important features of Praat and solves the above-mentioned problems. The sound visualisation is accessible through an icon next to each result presenting KWIC concordances. An overview of the user interface is presented in Figure 3.

The following plots are available: the waveform in the upper part, the spectrogram in the lower part, and four formants and pitch overlaid over the spectrogram. The plots can be turned on or off individually according to user preference. Just like in Praat, the sound can be played and an animated vertical line shows the current playing position in real time.

The user can select a part of the spectrogram that may, for example, correspond to one sound or syllable, listen to that part, and zoom in to see the features of the sound in more detail. As the selection is made, the numerical values of the parameters are displayed: duration and statistics for the selected period, namely maximal, minimal and average value of the pitch and formants. The time and frequency of any point on the spectrogram can be accessed via a mouse hover-over. Right-clicking allows the numerical values to be exported to a separate window. Also the statistics of the selection can be exported.

The visualisation is fully integrated with the search functionality of Glossa. For instance, in a corpus with speaker metadata, part-of-speech annotation and phonetic transcription, the user may specify a particular word in its orthographic and/or phonetic transcription, its part of speech, and restrict the search to utterances from speakers who meet particular criteria.

5 Technical Details

Even though Praat has batch processing facilities, it does not provide any application programming interface (API) that would make integration into a larger system possible. Therefore we used the Snack Sound Toolkit[13] instead, which contains Tcl/Tk and Python bindings.

The visualisation is generated by a Python daemon using Snack, which communicates with the main Glossa daemon, written in Ruby on Rails, through a simple protocol. Such an architecture is motivated by several factors. Snack does not have Ruby bindings, so calling it directly from the main Glossa daemon is not possible. Moreover, having this functionality in a separate daemon allows Glossa to be more responsive: other functionalities of Glossa are not blocked during the generation of the spectrograms, nor is it necessary to create a separate process for every request. Additionally, initialisation of Snack and other related code requires several seconds. In our architecture this needs to be done only once, after starting the daemon, and does not affect the time of the generation of the spectrogram.

The speed of the spectrogram generation depends on the size of the segments[14]. The daemon can generate spectrograms for sound chunks of any length. Since Snack needs to load the whole file into memory in order to generate the spectrogram, and the generation of some of the plots (e.g. the formants) may be time-consuming for longer files, we decided to split sound files into one-minute chunks, generate the plots for each of them, and stitch them together in a way that makes the result practically indistinguishable from a plot generated directly from a larger segment. This solution makes visualisation of large sound files possible without using too many resources. That said, if the chunks are too long, the system becomes less usable for the researchers, as they need to spend time looking for words or sounds they are interested in. In our corpora, the segments are generally not longer than a few dozens of seconds. The plot for a typical, 10-second segment is generated on our server in less than 4 seconds, which allows high interactivity. If a corpus is divided into larger segments, the user will need to wait a bit longer, but our solution is still faster than downloading the file and opening it in Praat: the plot for a 38-second segment is generated in less than 9 seconds, and for a 94-second segment in less than 21 seconds.

The interactive audio player that allows any part of the displayed sound to be played was written in JavaScript and makes use of the API provided by SoundManager 2[15]. Our visualisation tool may use Adobe Flash to play sounds, or use a purely JavaScript-based alternative when Flash is not available.

In order to minimise installation problems and support reproducible research (Chamberlain and Schommer, 2014), Glossa is released as a Docker[16] image, which contains all the required dependencies. Such packaging guarantees that it will give exactly the same results on the same data, avoiding problems caused by different software versions. Moreover, although the system makes use of several programming languages and libraries, due to Docker packaging, it is cross-platform and installable with just a few clicks.

In order to render the maps, Glossa communicates with the Google Maps Embed API using a JSON object. The colour-coding widget is borrowed from the JQuery API, and shares the same JSON object. The reason for choosing Google's service was simply practicality. At the time of implementation, Google's API was deemed most well developed and its interface most familiar to potential users. However, most of the development involved was agnostic with regards to which service was chosen, meaning any future move to another service will require little adaptation. It is worth noting here that using an open-source map service, such as OpenStreetMap, would enable hosting the maps together with the Glossa in-

[13] `http://www.speech.kth.se/snack`

[14] A segment is a piece of transcription synchronised with the audio, with a defined start and end time.

[15] `http://www.schillmania.com/projects/soundmanager2`

[16] `http://www.docker.com`

stallation, thus enabling offline viewing.

6 Research Applications

As mentioned in subsection 4.1, the visualisation features of Glossa may be useful in the exploratory phase of the research. For example, one may start by searching for a word in the Nordic Dialect Corpus and see the geographical distribution of different pronunciation variants. The phonetic search feature, mentioned in subsection 2.2, may then be used to restrict the search to one of the variants presented by the geographical visualisation. The phonetic transcription used in the corpus is coarse-grained, which means that there may be subtler differences within a particular variant. The sound visualisation allows the user to find out whether there are differences within the chosen pronunciation variant. For instance, one may measure duration of the vowels or the voice onset time of the consonants. The spectrogram analysis may improve accuracy in differentiating relatively similar phones (such as alveolar tap and trill), compared to a situation where only audio is available. The tool is designed to work interactively, and one may easily repeat the procedure with different words and different features in order to produce a hypothesis.

The map visualisation has already had a significant impact on the Scandinavian linguistic research community. Researchers from the projects NorDiaSyn and NorDiaCorp have written more than sixty papers on various syntactic phenomena in the North Germanic languages, in which maps from the Nordic Dialect Corpus have played a crucial role. These papers have been published in a new online open access journal which requires its papers to use empirical data from, *inter alia*, the Nordic Dialect Corpus and maps generated from it: *Nordic Atlas of Language Structures Journal* (Johannessen and Vangsnes, 2014).

Visualisation in speech corpora is also useful for more reliable data gathering. For example, stress is a feature of spoken Mandarin Chinese that received relatively little attention. One may want to use a Mandarin speech corpus, such as MAID[17], to investigate the patterns of its occurrence. The problem, however, is that stress is not marked in the Chinese writing system, and *Hanyu Pinyin*, the official phonetic transcription system for Mandarin, only distinguishes unstressed syllables that completely lose their underlying tone. Therefore, without a sound recording it is impossible to distinguish unstressed syllables that still retain their tone from stressed syllables. Another feature of Mandarin, present in the Beijing dialect, is the retroflex suffixation. This suffix is often omitted in speech transcription, and therefore an actual recording is again more reliable than transcription alone. But even with a recording, researchers often have no other choice than to rely on their subjective evaluation of whether stress or a suffix is present in a particular syllable. The sound visualisation makes it possible to refer to objective features of the waveform, spectrogram and formants, such as F3 decrease in case of the retroflex suffixation (Lee, 2005), instead of researchers' subjective perception.

Those who work with specific groups of speakers, for example children or people with pronunciation difficulties, may take advantage of features of Glossa, even if corpora covering their target group are not available. For example, Norwegian children may tend to produce epenthetic vowels in word-initial consonant clusters. Researching this feature requires a control group of adult Norwegian speakers, and this is where data from the Norwegian Speech Corpus[18] or the Nordic Dialect Corpus[19] may be useful. The sound visualisation makes it possible to quickly find out whether adults produce such epenthetic vowels and to measure their duration.

One of the easiest features to analyse in the sound visualisation is the pitch contour, which may give valuable information, especially in the case of tonal languages. For example, one may analyse the 3rd tone sandhi in Mandarin Chinese: the patterns occurring when there are two or more subsequent syllables with the 3rd tone. When the syllables occur within a prosodic foot, all but the last one change to the 2nd tone (Shih, 1997). In other cases the change is not obligatory, but may occur. The pitch plot is a useful tool for verification of whether the tone sandhi actually occurs and analyse patterns of its occurrence. The visualisation is even more useful for investigating effects of tone coarticulation – while the tone sandhi is categorical, the coarticulation effect changes the pitch in different degrees, depending on the sit-

[17]Mandarin Audio Idiolect Dictionary, a dictionary and corpus of Beijing Mandarin: `http://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/maid`

[18]`http://tekstlab.uio.no/nota/oslo/english`

[19]`http://tekstlab.uio.no/nota/scandiasyn`

uation (Zhang and Liu, 2011). Figure 3 shows how one may use Glossa to look for particular tone combinations, visualise and analyse the pitch contour. Average pitch values over specified periods of time can then be exported, which allows for the investigation of degrees of tone changes in natural speech, depending on the adjacent tonal context.

No formal evaluation of the presented visualisation features has been performed. However, the fact that many researchers use the corpora served by Glossa on a daily basis, and publish research papers based on them, is a sign that they have succeeded in satisfying the user groups. 16,142 individual searches were performed by more than 220 different users between 20. December 2014 and 19. April 2015 (i.e. 135 searches per day on average). The many papers in the *Nordic Atlas of Language Structures Journal* provide yet more evidence. The quantitative results from Google Scholar are also worth mentioning: 141 scholarly publications refer to "Nordic Dialect Corpus", which is just one of the many corpora using Glossa, and another 24 refer to its Norwegian name "Nordisk dialektkorpus". These are high numbers for a corpus that was only ready for use in 2011.

7 Future Work

The detection of isoglosses discussed still leaves the job of charting them. One interesting line of future development would be attempting to perform this task automatically. The application of k-means clustering and computing convex hulls (Wiedenbeck and La Touche, 2008) would be one such avenue.

The currently available sound visualisation is a multi-purpose tool that provides a wide range of acoustic data about the speech signal. It reduces the time required to visualise the features of the utterances that are of interest. However, corpora should not only give access to specific examples, but also provide useful statistics that allow generalisations to be drawn from the search results. In this case, the efficiency of research would be increased even more if the corpus search tool could directly provide statistics relating to the search results, for example average voice onset time, vowel length or formant values within a vowel. This would, however, require synchronisation of speech at the level of words and/or syllables.

In the speech corpora currently available in Glossa, the sound and the text are synchronised at the utterance level. There are, however, no technical problems with producing corpora that are synchronised at the word level. When such corpora become available, Glossa may be extended with an API that allows algorithms that find and/or measure specific phonetic details to be applied, such as the automatic measurement of voice onset time (Sonderegger and Keshet, 2012) or the detection of retroflex suffixation (Zhang et al., 2014).

8 Conclusion

This paper has discussed the visualisation possibilities of the Glossa corpus search system. The main focus was on the features available for speech corpora: geographical visualisation and speech visualisation. Geographical visualisation makes it possible to display pronunciation variants of the search results on a map and use colour-coding to cluster them into larger groups. The phonetic search feature allows specific pronunciation variants in the corpus to be found. Each search result in a speech corpus can be visualised in the built-in tool for audio analysis. The user may select its part and plot its parameters or export their values. These visualisations may be used to explore the data and formulate a research hypothesis, verify the existence of particular phonetic features, and easily analyse various parameters of speech.

Acknowledgements

We would like to thank Professor Christoph Harbsmeier for his suggestions and input on the visualisation of speech, and Pernille Hansen and Anders Vaa for their ideas on the use of the speech features now included in Glossa. Further, we are grateful to those who have provided speech content, and to the transcribers who have facilitated the searchability of our corpora. The two projects NorDiaSyn (financed by the Research Council of Norway) and NorDiaCorp (financed by Nordforsk) were responsible for many of the recordings and the transcriptions used in the present speech corpora.

This work was partly supported by the Research Council of Norway through its Centres of Excellence funding scheme, project number 223265, and partly through the Research Council's infrastructure project CLARINO.

References

Sjef Barbiers et al. 2006. *Dynamic Syntactic Atlas of the Dutch dialects (DynaSAND)*. Meertens Institute, Amsterdam. `http://www.meertens.knaw.nl/sand`.

Paul Boersma and David Weenink. 2001. Praat, a system for doing phonetics by computer. *Glot International*, 5(9/10):341–345.

Ryan Chamberlain and Jennifer Schommer. 2014. Using Docker to support reproducible research. Technical report, Invenshure, LLC. `http://dx.doi.org/10.6084/m9.figshare.1101910`.

Oliver Christ. 1994. A modular and flexible architecture for an integrated corpus query system. In *Proceedings of the 3rd International Conference on Computational Lexicography (COMPLEX)*, pages 22–32, Budapest.

Stefan Evert and Andrew Hardie. 2011. Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. In *Proceedings of the Corpus Linguistics 2011 conference*, Birmingham. University of Birmingham.

Janne Bondi Johannessen and Øystein Alexander Vangsnes. 2014. *Nordic Atlas of Language Structures Journal*. Department of Linguistics and Scandinavian Studies, University of Oslo. `http://www.tekstlab.uio.no/nals/`.

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.

William Labov, Sharon Ash, and Charles Boberg. 2005. *The atlas of North American English: Phonetics, phonology and sound change*. Walter de Gruyter.

William Labov. 1972. *Sociolinguistic patterns*. Number 4 in Conduct and Communication. University of Pennsylvania Press.

Wai-Sum Lee. 2005. A phonetic study of the "er-hua" rimes in Beijing Mandarin. In *Ninth European Conference on Speech Communication and Technology*.

Eric Papazian and Botolv Helleland. 2005. *Norsk talemål*. Hyskoleforlaget, Kristiansand.

Jürgen Erich Schmidt, Joachim Herrgen, Tanja Giessler, Alfred Lameli, Alexandra Lenz, Karl-Heinz Müller, Wolfgang Näser, Jost Nickel, Roland Kehrein, Christoph Purschke, et al. 2001. *Digitaler Wenker-Atlas*. Forschungszentrum Deutscher Sprachatlas, Marburg. `http://www.diwa.info`.

Chilin Shih. 1997. Mandarin third tone sandhi and prosodic structure. *Studies in Chinese Phonology*, 20:81–123.

Morgan Sonderegger and Joseph Keshet. 2012. Automatic measurement of voice onset time using discriminative structured prediction. *The Journal of the Acoustical Society of America*, 132(6):3965–3979.

Bryce Wiedenbeck and Kit La Touche. 2008. Drawing isoglosses algorithmically. In *Class of 2008 Senior Conference on Computational Geometry*, page 22.

Jie Zhang and Jiang Liu. 2011. Tone sandhi and tonal coarticulation in Tianjin Chinese. *Phonetica*, 68:161–191.

Long Zhang, Haifeng Li, Lin Ma, and Jianhua Wang. 2014. Automatic detection and evaluation of Erhua in the Putonghua proficiency test. *Chinese Journal of Acoustics*, 1:83–96.