Methods XVII | Mainz 2022

JG|U

Book of Abstracts

Deutsche Forschungsgemeinschaft

JOHANNES GUTENBERG UNIVERSITÄT MAINZ

DE GRUYTER MOUTON

Representative Natalie Fecher will be present Mon-Fri, find her during breaks or contact her @ Natalie.Fecher@degruyter.com.

Thank you

to

our sponsors!

JOHN BENJAMINS PUBLISHING COMPANY

Universitätsverlag WINTER Heidelberg

EDINBURGH University Press

See representative Dagmar Konetzka at the Winter book table!

Deep acoustic representations for clustering Dutch dialect pronunciations

Martijn Bartelds and Martijn Wieling

University of Groningen
The Netherlands

{m.bartelds, m.b.wieling}@rug.nl

Generally, dialectometric methods have focused on computing the string edit distance to compare phonetic transcriptions of speech samples (e.g., Nerbonne et al., 1999). The process of manually transcribing speech samples is, however, time-consuming and labor-intensive (Hakkani-Tür et al., 2002; Novotney and Callison-Burch, 2010). Another limitation of this procedure is that transcriptions can not fully capture all acoustic details of human speech, since often a limited set of transcription symbols is used (Liberman, 2018).

Alternatively, deep acoustic models automatically learn linguistic information on the basis of large audio corpora by taking the complete audio signal into account. While these acoustic models are primarily developed for automatic speech recognition (e.g., `wav2vec 2.0`; Baevski et al., 2020, `XLSR-53`; Conneau et al., 2020), they might learn information that is useful for other tasks. In this study, we therefore investigate whether we can use deep acoustic models to develop an acoustic distance measure for investigating differences between Dutch dialect pronunciations. Specifically, we use a deep acoustic `wav2vec 2.0` model pre-trained and fine-tuned on Dutch, and a multilingual `XLSR-53` model fine-tuned on Dutch. In addition, we compare the deep acoustic models to the adjusted Levenshtein distance algorithm of Wieling et al. (2012).

We extract data from the Goeman-Taeldeman-Van Reenen-Project (Goeman and Taeldeman, 1996), which contains audio recordings and phonetic transcriptions of hundreds of words for 613 dialect varieties in the Netherlands and Flanders. The metadata with time stamps to segment the recordings into words was only available for a small subset of the data. We therefore use recordings and phonetic transcriptions for 106 Netherlandic dialect varieties for which we have dialect pronunciations recordings of the same 10 words.

We extract acoustic representations from the deep acoustic models, and compare representations of the same word for every pair of locations using dynamic time warping (Senin, 2008). Similarly, the adjusted Levenshtein distance algorithm is used to compare the phonetic transcriptions. We average word-based distances between two locations to obtain a single pronunciation distance score, and do this for each pair of locations. The resulting distance matrices are subsequently clustered[1] into four groups and compared (using the spatially-sensitive `CDistance` score of Coen et al., 2010) to a gold standard clustering, distinguishing the three officially recognized regional (minority) languages spoken in the Netherlands (i.e. Frisian, Low Saxon, and Limburgish) and Dutch.

[1]While there are many clustering algorithms, we only included the ones available in *Gabmap* (Nerbonne et al., 2011), and for each approach we selected the clustering algorithm resulting in the highest cophenetic correlation coefficient (Sokal and Rohlf, 1962).

Using this approach, we obtain `CDistance` scores (lower is better) of 0.34 (Dutch `wav2vec 2.0` using WPGMA (Weighted Pair Group Method using Arithmetic averages) clustering), 0.20 (fine-tuned `XLSR-53` using complete link clustering), and 0.46 (adjusted Levenshtein distance using UPGMA (Unweighted Pair Group Method using Arithmetic averages) clustering). Consequently, we find that the fine-tuned `XLSR-53` model can be most effectively used to distinguish between language varieties in the Netherlands.

Combined with earlier work of Bartelds et al. (2021), which showed that these deep acoustic models were also superior in distinguishing accented speech, our results suggest that our approach is a suitable alternative to dialectometric analysis requiring (time-consuming) phonetic transcriptions. Importantly, our analysis appears to be effective even when only few audio samples are available.

References

Baevski, A., Zhou, Y., Mohamed, A., and Auli, M. (2020). wav2vec 2.0: A framework for self-supervised learning of speech representations. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. F., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12449–12460. Curran Associates, Inc.

Bartelds, M., de Vries, W., Sanal, F., Richter, C., Liberman, M., and Wieling, M. (2021). Neural Representations for Modeling Variation in Speech. *arXiv preprint arXiv:2011.12649*.

Coen, M. H., Ansari, M. H., and Fillmore, N. (2010). Comparing Clusterings in Space. In *Proc. of ICML*.

Conneau, A., Baevski, A., Collobert, R., Mohamed, A., and Auli, M. (2020). Unsupervised Cross-lingual Representation Learning for Speech Recognition. *arXiv preprint arXiv:2006.13979*.

Goeman, A. and Taeldeman, J. (1996). Fonologie en morfologie van de Nederlandse dialecten; een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval*, 48:38–59.

Hakkani-Tür, D., Riccardi, G., and Gorin, A. (2002). Active learning for automatic speech recognition. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages IV–3904. IEEE.

Liberman, M. (2018). Towards progress in theories of language sound structure. In Brentari, D. and Lee, J. L., editors, *Shaping phonology*. University of Chicago Press.

Nerbonne, J., Heeringa, W., and Kleiweg, P. (1999). Edit distance and dialect proximity. *Time Warps, String Edits and Macromolecules: The theory and practice of sequence comparison*, 15.

Nerbonne, J., Rinke, C., Charlotte, G., Peter, K., and Therese, L. (2011). Gabmap - a web application for dialectology. *Dialectologia: revista electrònica*, pages 65–89.

Novotney, S. and Callison-Burch, C. (2010). Cheap, fast and good enough: Automatic speech recognition with non-expert transcription. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 207–215. Association for Computational Linguistics.

Senin, P. (2008). Dynamic Time Warping Algorithm Review. *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, 855(1-23):40.

Sokal, R. R. and Rohlf, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *Taxon*, 11(2):33–40.

Wieling, M., Margaretha, E., and Nerbonne, J. (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, 40(2):307–314.

Extracting Non-Standard Varieties via the Twitter API: The Case of AAE

Kimberley Baxter

The present paper examines methodology in the use of Twitter in the corpus-based analysis of African American English (AAE) syntax. Widespread use of AAE on archived social media posts creates a living database of timed, dated, and geotagged utterances from which corpora may be built. Twitter's API allows access to their full database of tweets, which is a much larger and more accessible dataset than its large social media contemporaries.

There are well-documented challenges in analyzing non-standard varieties of English commonly used in social media (Plank, 2016), and AAE is no different. Efforts to normalize non-standard varieties often do not allow for the extraction or analysis of non-standard language as it occurs among speakers. In the case of AAE, standard methods of extracting data via Twitter's API are insufficient, lacking the specifications necessary to isolate certain parts of speech exclusive to AAE, and differentiate them from similar lexical items in Mainstream American English (MAE). Despite habitual be and copula/auxiliary be having two separate uses, the word "be" and its conjugations look and sound exactly the same, as seen below in examples (1) and (2).

(1) He be running. (Meaning: He tends to run. He usually runs. He is not necessarily running right now.)

(2) He is running. (He is currently running right now.)

This results in a high number of false positives and data which is rendered unusable due to the sheer size of the dataset, which may number in the millions of tweets, thus rendering the manual elimination of false positives unfeasible. This project ultimately aims to produce an alternative, syntax-based method which allows the user to eliminate a great deal of the aforementioned false positives by coding the syntactic constraints of this part of speech, thus allowing for the extraction of non-standard varieties, which would otherwise be inaccessible due to the comparative lack of specialized part of speech taggers designed for this task. While the initial focus is on Twitter, this tool will ultimately combine a range of methodological approaches and hopes to foster collaboration between researchers.

Infinitival Ain't in African American English

Kimberley Baxter and Jonathan Stevenson

The present paper documents the first stage in the creation of an atlas of African American English (AAE) syntax, charting the relative use of the uninflected form of ain't (infinitival ain't - inf.aint) (1) against didn't (2) in a large Twitter corpus spanning three years from 2012-2015.

(1) I ain't see that coming (inf.aint)
(2) I didn't see that coming (inf.didnt)

The semantic near-equivalence of (1) and (2) allows us (with caveats) to consider the two forms as variants of a single variable (Labov et al., 1968; Wolfram and Schilling-Estes, 2016; Fisher, 2018) whereby the relative frequency of inf.aint may be measured against inf.didnt to provide a reliable index of use across and between places. Meanwhile, the scale of the data available via Twitter's API allows for unprecedented resolution at the level of small towns and suburbs.

In line with previous studies that use Twitter data for dialect research (Eisenstein, 2013; Jones, 2015; Stevenson, 2016; Willis, 2020; Strelluf, 2019, 2020), results in many cases appear to follow established dialect 'faultlines' (Eisenstein, 2013, p.1) whilst also highlighting particular hotspots of use. Variation appears to correlate with African American population distribution, supporting the claim that inf.aint is indeed associated with AAE.

The distribution also supports the notion that inf.aint is a relatively recent development in AAE, innovated amongst populations that travelled to the Northern industrial areas of the US during the Great Migration. This is shown in notably higher pockets of use in those regions in the Twitter corpus as well as substantial variation between places. Further, the finding that inf.aint was originally rare in early AAE in the Southern US (Kautzsch, 2000), is partially supported, with on average lower rates on Twitter in many parts of the South Atlantic region that otherwise have high rates of AAE use, such as Atlanta, GA. However, the atlas highlights some localised areas of use that buck this trend as well as sharp boundaries that warrant more detailed investigation. Results show that inf.aint is extremely widespread, with some areas such as Gary, Indiana and Augusta, Georgia exhibiting up to 40% usage in comparison with inf.didnt.
Whilst there is semantic equivalence between inf.aint and inf.didnt, equivalence between other AAE forms and their Mainstream American English (MAE) counterparts is only partial and elusive. This is true for forms such as habitual be, and perfective done). The inf.aint data, then, can provide a metric against which the frequency of otherwise less trackable —but more stable— structures may be compared.

The project ultimately aims to produce a functioning syntactic atlas which can serve as a rich re-source for further investigation by the academic community. Whilst the initial focus is on Twitter, the atlas will ultimately combine a range of methodological approaches and hopes to foster collaboration between researchers.

(In)coherence across the linguistic architecture: change in Swabian across the lifespan

Karen V. Beaman
Eberhard Karls Universität Tübingen

Ever more studies show that adult vernaculars are malleable, influenced by shifting cultural contexts, prominent social, psychological or cognitive adjustments, exceptional historical incidents, and decisive life-changing events (e.g., Beaman and Buchstaller 2021; Wagner and Buchstaller 2018). But do all variables change in the same way and under the same conditions for all speakers? Are there notable deviations at different levels of the linguistic architecture across the lifespan?

To explore these questions, this research targets two speech communities of Swabian, an upper German dialect belonging to the Alemannic family. The real-time panel study comprises 20 native Swabian speakers first recorded in 1982 and re-recorded 35-years later in 2017. The hypothesis of this study predicts that greater coherence is found with phonological variables than with morphosyntactic ones. This may also be attributed to the expectation that morphosyntactic variables are more salient (and hence more stigmatized), making them more susceptible to change (e.g., Naro 1981), while phonological variables are more frequent (and hence more entrenched), making them less vulnerable to change (e.g. Bybee 2002). In addition, Chambers (1995:51) claims that "grammatical variables tend to mark social stratification more sharply [than phonological ones] so that it is probably safe to say that most … function as class markers." To investigate these claims, this study follows Buchstaller, Krause-Lerche, and Mechler (2021) in analyzing three common sociolinguistic heuristics – inventory change, frequency change, and constraint-based change for 10 phonological and 10 morphosyntactic variables to assess how lifespan change may differ across the linguistic architecture.

The results show that, while lifespan change generally follows community change, there are important individual patterns that diverge from the norm: some speakers change more quickly, some more slowly, some not at all, and some move in reverse of the change (e.g., Beaman 2020; Sankoff 2006). Overall, the metrics from the three sociolinguistic heuristics concur, although they signal crucial deviances across the architecture and with specific speakers – variances which can be explained with reference to immense societal change occurring in Germany, changing norms of prestige and stigma, and linguistic marketplace effects (Bourdieu and Boltanski 1975).

References

Beaman, Karen V. 2020. "Coherence in Real- and Apparent-Time: A Sociolinguistic Variationist Investigation of Language Change in Swabia. Unpublished PhD Thesis." Queen Mary University of London.

Beaman, Karen V., and Isabelle Buchstaller. 2021. *Language Variation and Language Change Across the Lifespan: Theoretical and Empirical Perspectives from Panel Studies*. New York: Routledge.

Bourdieu, Pierre, and Luc Boltanski. 1975. "Le Fétichisme de La Langue." *Actes de la recherche en sciences sociales* 1(4): 2–32.

Buchstaller, Isabelle, Anne Krause-Lerche, and Johanna Mechler. 2021. "Exploring the Effect of

Linguistic Architecture and Heuristic Method in Panel Analysis." In *Language Variation and Change across the Lifespan*, eds. Karen V. Beaman and Isabelle Buchstaller. New York: Routledge, 185–208.

Bybee, Joan L. 2002. "Phonological Evidence for Exemplar Storage of Multiword Sequences." *Studies in Second Language Acquisistion* 24(2): 215–21.

Chambers, J. K. 1995. *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Cambridge: Blackwell Publishing.

Naro, Anthony J. 1981. "The Social and Structural Dimensions of a Syntactic Change." *Language* 57(1): 63–98.

Sankoff, Gillian. 2006. "Age: Apparent Time and Real Time." In *Encyclopedia of Language and Linguistics. 2nd Edition*, ed. Keith Brown. Elsevier Science Ltd., 110–16.

Wagner, Suzanne Evans, and Isabelle Buchstaller. 2018. *Panel Studies of Variation and Change*. eds. Suzanne Evans Wagner and Isabelle Buchstaller. New York: Routledge.

A computational approach to detecting the envelope of variation

Isaac L. Bleaman and Rhea Kommerell (University of California, Berkeley)

Variationist sociolinguistic methodology is grounded in the principle of accountability (Labov 1972:72; Tagliamonte 2006:12–3), which compels the researcher to enumerate all the contexts in which a variable occurs or fails to occur (where one variant is used categorically or where the choice is neutralized). The process of defining the envelope of variation and determining which tokens "count" for analysis is notoriously time- and labor-intensive (Labov 1978:6). Moreover, although the variationist enterprise rejects the use of grammaticality/acceptability intuitions as data (Bayley 2013:89), researchers routinely rely on such intuitions when *selecting* data—especially in studies of morphosyntactic, lexical, and discourse variables.

In this paper, we demonstrate the usability of pre-trained computational language models to automatically identify tokens of sociolinguistic variables in raw text. We focus on two English-language variables from different linguistic domains: intensifier choice (lexical; e.g., *he is* {*very*, *really*, *so*} *cute*) and complementizer selection (grammatical; e.g., *they thought* {*that*, Ø} *she understood*). These variables exemplify different challenges for automatically detecting the envelope of variation: Intensifier variants are one-word strings, but basic search techniques cannot distinguish intensifier from non-intensifier usages (e.g., exclusions such as *she's the* {*very*, **really*, **so*} *person I had in mind*). Complementizer selection involves one variant that is overt and another that is phonetically null; the overt variant also appears in non-complementizer contexts (e.g., determiner or relativizer *that*), and the null variant necessarily eludes most search methods.

We employed BERT (Devlin et al. 2019) to train classifiers to predict whether sentences in raw text fall within or beyond the envelope of variation for each variable. The classifiers were trained and evaluated using manually annotated data. We adapted the dataset from Tagliamonte & Roberts's (2005) study of intensifiers in episodes of the American sitcom *Friends* to compile a list of sentences containing the words *very*, *really*, or *so* in both intensifier and non-intensifier contexts. We used the Penn Treebank to obtain sentences containing an overt complementizer, a null complementizer, or no complementizer. For each variable, classifier models were trained on random samples of different sizes in order to compare their performance; for complementizers, separate classifiers were trained for the overt and null variants (though these were combined during evaluation).

Our findings show that computational language models, like BERT, can dramatically reduce the burden of combing through raw language data in search of tokens of a variable—including when the surface forms are highly polysemous or phonetically null. Very little hand-annotated training data is required to achieve relatively high accuracy. Precision is somewhat lower than recall, but this is not crucial for our methodological purposes because it is much easier to remove false positives than it is to recover false negatives. Furthermore, by manually inspecting the sentences that receive high scores (indicating prototypical examples of the variable), low scores (likely exclusions), and intermediate scores around 0.5 (tricky edge cases), researchers can identify patterns that should be written into the description of the variable context for further study.

References

Bayley, Robert. 2013. The quantitative paradigm. In J. K. Chambers & Natalie Schilling (eds.), *The handbook of language variation and change*, chap. 4, 85–107. Malden, MA: Wiley-Blackwell.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, & Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL 2019*, 4171–4186. Minneapolis: Association for Computational Linguistics. doi:10.18653/v1/N19-1423.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 1978. Where does the linguistic variable stop? A response to Beatriz Lavandera. In Richard Bauman & Joel Sherzer (eds.), *Working papers in sociolinguistics*, vol. 44, 1–17. Austin, TX: Southwest Educational Development Laboratory. https://eric.ed.gov/?id=ED157378.

Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.

Tagliamonte, Sali & Chris Roberts. 2005. So weird; so cool; so innovative: The use of intensifiers in the television series *Friends*. *American Speech* 80. 280–300. doi:10.1215/00031283-80-3-280.

Genre coherence and distinctiveness in the International Corpus of English: A quantitative approach

Axel Bohmann, Albert-Ludwigs-Universität Freiburg

This paper introduces an innovative method for exploring relationships among sub-groups in a corpus of linguistic data. The specific focus is on the coherence and distinctiveness of text categories in ten national sub-corpora of the International Corpus of English (ICE) project (Greenbaum & Nelson 1996), with the aim of proposing a typology of genres. ICE corpora comprise four spoken and eight written text categories, for each of which the following relationships are explored:

- How linguistically distinct is this text category from the other corpus texts on the whole?
- Within the text category, how distinct are texts representing different varieties of English?
- For text categories with more fine-grained sub-distinctions, how distinct are the sub-categories from each other?

In order to quantify linguistic overlap/distinction, the study relies on the ten dimensions of variation developed in Bohmann (2019). These express general textual properties and have been constructed empirically on the basis of co-variances among 276 individual linguistic variables. Relationships between groups of texts in this ten-dimensional space are expressed via the Bhattacharyya coefficient (Bhattacharyya 1943), a measure of the overlap between two multivariate distributions. For instance, in response to the first research question above, it is possible to calculate the Bhattacharyya coefficient for overlap between all spontaneous conversation (S1A) text samples and all remaining text samples.

Addressing the three research questions above by means of the Bhattacharyya coefficient results in a categorization of corpus text types into three broad classes:

- Highly coherent text categories that are clearly distinct from other categories and show moderate regional distinction: Creative writing, Private dialogue.
- Less coherent categories that show relatively strong cross-varietal distinctiveness: Non-professional writing, Correspondence, Reportage, Instructional writing, Persuasive writing.
- Less coherent categories with lower degrees of cross-varietal distinctiveness: Public dialogue, Unscripted monologue, Scripted Monologue, Academic writing, Popular informational writing.

The findings are relevant both at a theoretical and a methodological level. In terms of the former, the importance of genre in mediating linguistic variation has long been recognized (e.g. Hundt & Mair 1999), but the relationship among genres themselves has received less attention. The typology offered above may help corpus compilers establish and empirically verify appropriate levels of granularity in their sampling frame; it may also help researchers identify the kinds of text in which regional divergences may be expected, the points of overlap between genres that enable the gradual spread of a feature from one to the other, etc.

At the methodological level, use of the Bhattacharyya coefficient is not limited to specific corpus linguistic questions. It can express overlap between any two groups along any number of dimensions, such as vowel formant measurements or frequencies of specific lexical items. As such, it is useful for studies of group relationships in sociolinguistics and dialectology more generally. Its ability to incorporate multiple dimensions of variation at once is promising for holistic perspectives on linguistic distinctiveness in the spirit of dialectometry (e.g Szmrecsanyi 2013).

References

Bhattacharyya, A. 1943. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*. 35: 99–109.

Bohmann, Axel. 2019. *Variation in English Worldwide: Registers and Global Varieties.* Cambridge: Cambridge University Press.

Greenbaum, Sidney & Gerald Nelson. 1996. The International Corpus of English (ICE) project. *World Englishes* 15(1). 3–15.

Hundt, Marianne & Christian Mair. 1999. “Agile” and “uptight” genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics* 4(2). 221–242.

Szmrecsanyi, Benedikt. 2013. *Grammatical Variation in British English Dialects: A Study in Corpus-Based Dialectometry.* Cambridge: Cambridge University Press.

How old are the moun[ʔə]ns in Utah? Utahn "*t*-dropping" over time and across the lifespan

A widespread folk-linguistic belief among speakers of Utah English is that "*t*-dropping" (that is, the glottalization of /t/ before syllabic nasals) distinguishes Utah English from other American Englishes. Eddington & Savage (2012) found that present-day Utah English is somewhat different, in that a widespread realization is an oral release of the glottal stop (e.g., *mountain* produced as [mɑʊnʔən], not [mɑʊnʔn̩]). This study pushes back in time by investigating same variable in the production of speakers in an archive of recordings made between 1940 and 2010 that were not originally collected for sociolinguistic analysis, but that have been successfully used for that purpose (Stanley & Renwick 2016, Bowie 2021, among others).

To gain insight into the development of this variable, two parallel studies were conducted using separate samples of Utah English speakers from this archive: a panel study of 10 speakers (born 1876–1928), and a trend study of 26 speakers (13 recorded in 1940 and 13 in 2010) who were demographically similar at the time of recording. Both studies found a decrease in the realization of /t/ before syllabic nasals as [tʰ] and a corresponding increase in both realization of the sound as [ʔ] and deletion. However, the glottal realization was almost never followed by an oral release (i.e., into [ən] rather than [n̩]), lending credence to Eddington & Savage's findings that that is a recent innovation. Further, there was relatively little intraindividual variation in the panel study, reflecting a difference between this variable and several of those reported earlier by Bowie (2011, 2015, 2019, 2021). However, unlike those this variable is socially salient, leading to the possibility that social awareness of linguistic variables can be reflected in lifespan variation and change, supporting the conclusions of Sankoff & Blondeau (2007).

References

Bowie, David. 2011. Aging and sociolinguistic variation. In Anna Duszak & Urszula Okulska (eds.), *Language, culture, and the dynamics of age*, 29–51. Berlin, Germany: Mouton de Gruyter.

Bowie, David. 2015. Phonological variation in real time: Patterns of adult linguistic stability and change. In Annette Gerstenberg & Anja Voeste (eds.), *Language development: The lifespan perspective*, 39–58. Amsterdam, The Netherlands: John Benjamins.

Bowie, David. 2019. Individual variation in the development of the Western Vowel System of Utah. *Linguistics Vanguard* 5. Article 20180020.

Bowie, David. 2021. Individual and group trajectories across adulthood in a sample of Utah English speakers. In Karen V. Beaman & Isabelle Buchstaller (eds.), *Language variation and language change across the lifespan: Theoretical and empirical perspectives from panel studies*, 101–118. New York, New York: Routledge.

Eddington, David & Matthew Savage. 2012. Where are the moun[ʔə]ns in Utah? *American Speech* 87. 336–349.

Sankoff, Gillian & Hélène Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83. 560–588.

Stanley, Joseph A. & Margaret E. L. Renwick. 2016. Phonetic shift /ɔr/ phonemic change? American English mergers over 40 years. Poster presented at LabPhon 15, Ithaca, New York.

Đorđe Božović
University of Belgrade, Faculty of Philology
djordje.bozovic@fil.bg.ac.rs

On dialect polyphony in the Central South Slavic continuum

This paper discusses some of the methodological issues related to constructing the research paradigm and conducting research into the actual state of (multi)dialectal variation in one area of the Central South Slavic dialect continuum. With the Ijekavian Neo-Štokavian dialect of south-western Serbia as a case study, the paper brings an analysis of the evolution of a traditional dialect in contact with the Ekavian-based standard language, in order to point out a range of methodological issues that arise in the study of present-day (dia)lect variation and the relationship between the standard language and non-standard dialects. Although typologically interesting, with the Ijekavian-based traditional dialect and the Ekavian-based superstrate in now over a century-and-a-half-long contact, the resulting dialect variation and change in this zone has remained largely underdescribed—as the traditional dialectology has aimed chiefly for 'vernacular authenticity' (Ilić 2016) and therefore largely ignored innovations in dialects brought by modernization.

Following Petrović's (2016) 'dialect as performace' approach, by examining the use and distribution of dialect features with different degrees of perceptual salience in dialogues situated in different socio-pragmatic contexts, the paper points out the importance of speakers' meta-linguistic awareness and identity negotiation strategies for the processes of traditional dialect features retention and change. This creates polyphony-like effects, in which speakers develop different accomodation strategies for different in-group and out-group interlocutors in different situational contexts, by making various choices from the inventory of traditional and more innovative dialect features, based on their perceptual salience.

References

Ilić, Marija (2016). Vernakularna autentičnost u srpskoj dijalektologiji: kritičko preispitivanje. In: R. Žugić (ed.), *Dijalekti srpskoga jezika: istraživanja, nastava, književnost*, vol. 1. Leskovac: Leskovački kulturni centar – Vranje: Pedagoški fakultet, 519–529.

Petrović, Tanja (2016). Dijalekat kao performans. In: R. Žugić (ed.), *Dijalekti srpskoga jezika: istraživanja, nastava, književnost*, vol. 1. Leskovac: Leskovački kulturni centar – Vranje: Pedagoški fakultet, 531–538.

Problems of Space and Times and Linguistics. Concepts and technical aspects of (digital) locating and placing linguistic data on a map – using the example of SyHD, DiÖ and LIÖ.

There are many problems regarding space and time or space-time in different research fields but at least the studies on dialects have neglected a lot of these discussions. This is remarkable considering that dialects and their exploration are not only inherently spatial but also temporal. Problems occur when trying to view data on a map and they become greater when comparing different maps – e.g. maps of different regions or maps showing data about different time periods. Linguists doing research on Variation and Change have to deal with concepts of time and space and strictly define the kind of space or time they are dealing with. This is also true for modelling linguistic databases (e.g. Dimitriadis 2009) or project mapping tools. Especially the socio-political dimension is often ignored by linguistic data mapping – even if it is a crucial aspect of (socio-)linguistic and dialect analyses.
From a digital perspective, considering the world wide web (or "cyber space") space and time raises challenges which may seem to be trivial but are crucial for modern variationist linguistic projects to be solved effectively. For reasons of research practice, studies on linguistic variation are limited to a specific space and they have to be achieved in a certain period of time. Nontheless, linguists are interested in a depicting linguistic variation as complete as possible, (see Lenz (2019)) e.g. in a country, in a linguistic area or over a long period of time. Practically spoken, the aimed linguistic data has to be gathered by various linguists or linguistic projects and brought together via one unifying point of referencein order to enable different projects to combine their data in one system (e.g. Breuer/Seltmann (2018)).
The talk will address different concepts of space & time with a focus on the technical implications of these concepts and therefore on digital linguistics (e.g. Lenz (2019)). It will show different approaches for mapping linguistic data from three different projects: SyHD (Syntax of Hessian Dialects, see Fleischer et al (2017) and SyHD-online (2016)), DiÖ (German in Austria, see. Budin et al and DiÖ (2019)) and LIÖ ("Lexicographic Informationsystem Austria", see. VaWaDiÖ (2019). These projects work with different a) (semantic) database models, b) kinds of places (on an administrative level and as a database representation) and c) different time periods. Furthermore, these projects provide the discussion with different d) sustainability and interoperability strategies for the long-time preservation of the data and e) different linguistic entities and annotations. In this combination, the projects are the perfect starting point for a discussion on the technical

implementation of linguistic databases and maps – and how to combine them across projects.

Bibliography

- Breuer, Ludwig Maximilian / Seltmann, Melanie E.-H. (2018): Sprachdaten(banken) – Aufbereitung und Visualisierung am Beispiel von SyHD und DiÖ. In: Börner, Ingo / Straub, Wolfgang / Zolels, Christian (eds.): Germanistik digital: Digital Humanities in der Sprach- und Literaturwissenschaft. Wien: Facultas UTB. 135–152.
- Budin, Gerhard, Stephan Elspaß & Alexandra N. Lenz. 2019. "The Research Project ‚German in Austria'. Variation – Contact – Perception". In: Bülow, Lars, Kathrin Fischer & Kristina Herbert (eds.). Dimensions of linguistic space: variation – multilingualism – conceptualisation / Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung. Wien: Peter Lang, 7–35.
- Dimitriadis, Alexis/Musgrave, Simon (2009): Designing linguistic databases: A primer for linguists. In: Everaert, Martin/Musgrave, Simon/Dimitriadis, Alexis (Eds.): The Use of Databases in Cross-Linguistic Studies. Berlin: de Gruyter Mouton, S. 13–75.
- DiÖ (2019): https://dioe.at/ [28.08.2019]
- Fleischer, Jürg/Lenz, Alexandra N./Weiß, Helmut (2017): SyHD-atlas. Konzipiert von Ludwig M. Breuer. Unter Mitarbeit von Katrin Kuhmichel, Stephanie Leser-Cronau, Johanna Schwalm und Thomas Strobel. Marburg/Wien/ Frankfurt.
- Lenz, Alexandra N. (2019). Digitale Sprachwissenschaft. Herausforderungen und Perspektiven. In: Akademie im Dialog 15. Wien: Österreichische Akademie der Wissenschaften. 5–16.
- SyHD (2016): SyHD-online: http://www.syhd.info/
- VaWaDiÖ (2019): https://vawadioe.acdh.oeaw.ac.at/projekte/lioe/ [28.08.2019]

Sociolinguistic variation in a non-native variety of Swiss German: Romansh migrants in the city of Berne

Andrin Büchler (University of Berne)

This paper is concerned with the sociolinguistic effects of Swiss intra-national migration involving speakers of the minority language Romansh. Since Switzerland is officially a quadrilingual country characterised by "territorial multilingualism" (cf. Riehl 2014: 64), people often need to acquire an additional language when moving to another region within the country. This also holds true for speakers of Romansh. However, given the strong economic and social prevalence of German-speaking Switzerland, Romansh speakers generally acquire the Grison variety of Swiss German as an L2 already during adolescence. If they move to German-speaking Switzerland (e.g., to attend university), they are confronted with yet other regional varieties of Swiss German, which leads to an intense dialect contact situation.

The Swiss German variety spoken by Romansh speakers as L2 has only recently gained some scholarly attention (cf. Eckhardt 2021). Previously, studies of Swiss German have focused more on regional rather than social variation and hence have excluded non-native speakers, such as Romansh speakers (cf. SDS; Glaser 2021). Furthermore, non-mobile speakers have been favoured. So, to some degree, Swiss dialect studies have adhered to traditional methods of dialectology and hence have only partially investigated the social processes underlying linguistic variation.

The present paper adopts variationist sociolinguistic methods to analyse long-term accommodation involving mobile, non-native speakers of Swiss German. Specifically, this research shows how variationist methods can better explain accommodation processes present in the Swiss German L2-variety of Romansh speakers who have migrated from their rural villages in Grisons to the city of Berne. I present data demonstrating internal as well as external factors to predict speakers' level of accommodation.

The sample consists of sociolinguistic interviews of 40 tertiary-educated Romansh speakers, aged between 20 and 40, who have migrated to Berne. The variationist analysis is based on three phonetic-phonological variables, Germanic word-initial (k), non-Germanic word-initial (k) and word-final (ə), for which the Grison variety of Swiss German has typical local variants. Long-term accommodation means that speakers level out these typical Grison features and approach variants common in most Swiss Midland varieties (that is, the adoption of supralocal variants). Results suggest a high degree of inter-speaker variability. This is no surprise given each speaker's distinct history of acquisition and contact to different varieties of Swiss German. However, variation is not random but constrained by internal factors such as the phonetic environment as well as a number of social factors (e.g., language biography, geographical orientation, network structure, school attended, etc.) which help to explain speakers' varying degrees of accommodation towards more supralocal variants.

References

Eckhardt, Oscar (2021): *Alemannisch in der Rumantschia: Die Alemannischen Dialekte im Romanischen Sprachraum von Trin, Ilanz, Trun und Scuol*. Stuttgart: Steiner. (= ZDL Beihefte 183).

Glaser, Elvira (eds.) (2021): *Syntaktischer Atlas der deutschen Schweiz (SADS)*. Vol. 1–2. Tübingen: Narr Francke Attempto.

Riehl, Claudia M. (2014): *Sprachkontaktforschung: Eine Einführung*. 3rd ed. Tübingen: Narr.

Sprachatlas der deutschen Schweiz (SDS) (1939–1997): Initiated by Heinrich Baumgartner and Rudolf Hotzenköcherle; in collaboration with Konrad Lobeck, Robert Schläpfer, Rudolf Trüb and Paul Zinsli, ed. by Rudolf Hotzenköcherle. Collected works. Bern/Basel: Francke.

Investigating dialect syntax in Austria. Empirical findings and methodological considerations

Lars Bülow (University of Vienna) | Philip C. Vergeiner (University of Salzburg) | Stephan Elspaß (University of Salzburg)

Dialect syntax is a thriving field of research in modern dialectology. Relatively little work, however, has been conducted on the syntax of dialects in Austria where traditional dialectology focussed almost completely on phonetics/phonology and to some extent on morphology. The few studies available are either pilot studies (cf. e.g. Lenz, Ahlers & Werner 2014; Breuer & Bülow 2019), not based on comprehensive and systematically built corpora (e.g. Patocka 1997), or studies with a regional focus (Scheutz 2005; Breuer 2016; Breuer 2021). The scarcity of empirical studies can chiefly be attributed to methodological difficulties that researchers are confronted with when investigating dialect syntax: syntactic features are considered to be distributed less locally and less systematically in comparison to phonetic/phonological or morphological features (cf. Kortmann 2010; König, Elspaß & Möller 2019: 163;). Additionally, not only are syntactic variables less frequent in natural data but also assumed to be less salient to speakers. Therefore, traditional methods of dialectological research – observations and questionnaire-based surveys – seem to be less suitable for inquiring syntactical phenomena (cf. Fleischer, Kasper & Lenz 2012; Glaser 2014 for discussion).

This paper will present findings from an ongoing project on selected syntactic features of traditional Austrian dialects. In comparing the results on different syntactic phenomena, we want to focus on two main aspects: (1) We will re-examine some popular claims on the nature of syntactic data, e.g. on the large-scale spatial distribution of syntactic variants and their supposedly 'unsystematic' patterns (in terms of spatial and social variation). (2) On a methodological note, we will discuss methodical benefits and limitations of different data types and methods. In particular, we will focus on the influence of standard norms on spoken and written data and the influence of word order in stimulus sentences in both spoken and written translation tasks.

To account for these aspects, we chose eight syntactic phenomena (based e.g. on Fleischer, Kasper & Lenz 2017, Lenz 2019): (a) article use before mass nouns, proper names and indefinite plural nouns, (b) adnominal possessive constructions, (c) indefinite-partitive pronouns, (d) subjunctive II (analytic vs. periphrastic), (e) negative concord, (f) complementizer agreement, (g) relative pronouns and (h) comparative constructions. The analyses on these phenomena are based on a comprehensive survey on the dialects of 163

speakers (balanced for age and gender) in 40 localities throughout Austria. The survey consists of different questionnaire-based methods of data collection, using translation tasks, rating tasks, and cloze tasks.

References

Breuer, Ludwig Maximilian (2016): Computers & Coffee: Computergestützte Sprachproduktionstests zur syntaktischen Variation des „unbestimmten Artikels vor Massennomen" im „Wienerischen". In: Helmut Kowar (ed.): Jahrbuch des Phonogrammarchivs. Wien: Verlag der Österreichischen Akademie der Wissenschaften, 86-111.

Breuer, Ludwig Maximilian (2021): 'Wienerisch' vertikal. Theorie und Methoden zur stadtsprachlichen syntaktischen Variation am Beispiel einer empirischen Untersuchung in Wien. Doctoral Thesis: University of Vienna.

Breuer, Ludwig; Bülow, Lars (2019): Experimental approaches in the realm of language variation – How Language Production Tests can help us to better understand language variation. In: Bülow, Lars et al. (eds.): Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptualisierung. Wien: Lang, 251-269.

Fleischer, Jürg; Kasper, Simon; Lenz, Alexandra N. (2012): Die Erhebung syntaktischer Phänomene durch die indirekte Methode: Ergebnisse und Erfahrungen aus dem Forschungsprojekt „Syntax hessischer Dialekte" (SyHD). In: Zeitschrift für Dialektologie und Linguistik 79/1, 2-42.

Fleischer, Jürg; Lenz, Alexandra N.; Weiß Helmut (2017): SyHD-atlas. Marburg, Wien, Frankfurt am Main: dx.doi.org/10.17192/es2017.0003.

Glaser, Elvira (2014): Wandel und Variation in der Morphosyntax der schweizerdeutschen Dialekte. In: Taal en Tongval 66/1. 21-64.

König, Werner; Elspaß, Stephan; Möller, Robert (2019): *dtv-Atlas Deutsche Sprache*. 19th rev. ed. München: dtv.

Kortmann, Bernd (2010): Areal variation in syntax. In: Auer, Peter; Schmidt, Jürgen Erich (eds.), Language and Space. An International Handbook of Linguistic Variation. Vol. 1: Theories and Methods. Berlin: Mouton de Gruyter, 837-864.

Lenz, Alexandra N.; Ahlers, Timo; Werner, Martina (2014): Zur Dynamik bairischer Dialektsyntax – eine Pilotstudie. In: Zeitschrift für Dialektologie und Linguistik 81/1, 1-33.

Lenz, Alexandra N. (2019): Bairisch und Alemannisch in Österreich. In: Herrgen, Joachim; Schmidt Jürgen Erich (eds.): Sprache und Raum. Ein internationales Handbuch der Sprachvariation. Volume 4: Deutsch. Berlin, Boston: de Gruyter, 318–363.

Patocka, Franz (1997): Satzgliedstellung in den bairischen Dialekten Österreichs. Frankfurt/M., Berlin, Bern, New York, Paris, Wien: Lang.

Scheutz, Hannes (2005): Perspektiven einer neuen Dialekt-Syntax. In: Eggers, Eckehard; Schmidt, Jürgen Erich; Stellmacher, Dieter (eds.): Moderne Dialekte – neue Dialektologie. Stuttgart: Steiner, 291-312.

Intra-speaker (in-)stability and varietal coherence across the lifespan: Findings from a real-time panel study in Austria

Lars Bülow and Dominik Wallner
University of Vienna and University of Salzburg

The varietal spectrum in the Bavarian parts of Austria is complex: it is supposed that traditional dialects and the standard language have been mutually influencing each other for decades, leading to a dialect/standard-continuum in a diaglossic language situation (Auer 2005). This assumption, however, is often based on impression or even speculation as current and up-to-date investigations into varietal coherence of Bavarian dialects in Austria are rare (exceptions comprise the pioneering work of Martin 1996 and Scheutz 1999 as well as Vergeiner 2019). Until now it remains unclear to what extent supposed dialect and standard features cluster together and build a 'unified whole' (Guy & Hinskens 2016). It is also not clear – not solely in the Austrian context – to what extent coherence changes over an individual's lifespan, i.e., whether aging is accompanied by a lectal focusing or diffusion. To explore both intra-speaker (in-)stability and lectal coherence across the lifespan we conducted a real-time panel study consisting of twelve speakers from Ulrichsberg (Upper Austria). Each of these twelve speakers were interviewed in 1975/6 and in 2018/19 in two situations: a formal interview and an informal conversation.

First, the data are analysed via a variable rule analysis, selecting six frequent, phonetic variables. We found differences in the distribution of the features in different settings and points in time (Vergeiner et al. submitted). In sum, the analysis reveals an increase of dialect features for each individual (retrograde-change) even though apparent-time as well as real-time trend studies indicate dialect loss in the Bavarian speaking parts of Austria. To reduce the complex dimensionality of the data, a factor analysis was computed, which identified latent regularities in the co-occurrence of linguistic variants. Crucially, this analysis indicated that there are coherent patterns in variation (Pickl 2013).

References

Auer, P. (2005). Europe's Sociolinguistic Unity, or: A Typology of European Dialect / Standard Constellations. In: N. Delbecque, J. van der Auwera & D. Geeraerts (eds.): *Perspectives on Variation*. Berlin: de Gruyter, 7–42.

Guy, G. & Hinskens, F. (2016). Linguistic coherence: Systems, repertoires and speech communities. *Lingua* 172-173, pp. 1-9.

Martin, V. (1996). Modelle der Umgangssprache. Überlegungen zum theoretischen Status eines linguistischen Begriffs am Beispiel des Wiener Deutsch. In: *Zeitschrift für Dialektologie und Linguistik* 63 (2), pp. 129-156.

Scheutz, H. (1999). Umgangssprache als Ergebnis von Konvergenz- und Divergenzprozessen zwischen Dialekt und Standardsprache. In: T. Stehl (ed.): *Dialektgenerationen, Dialektfunktionen, Sprachwandel*. Tübingen: Narr, pp. 105-131.

Pickl, S. (2013). *Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben*. Stuttgart: Steiner.

Vergeiner, P. C. (2019). *Kookkurrenz - Kovariation - Kontrast. Formen und Funktionen individueller Dialekt-/ Standardvariation in universitären Beratungsgesprächen*. Wien: Lang.

Vergeiner, P. C., Wallner, D., Bülow, L. & Scheutz, H. [submitted]. Redialektalisierung und Alter. Ergebnisse einer *real-time*-Studie zum Age-grading in Ulrichsberg. In H. Christen, B. Ganswindt, J. Herrgen & J. E. Schmidt (eds.). Conference volume of *6. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD),* 3.-15. September 2018.

Sociolinguistic variation in Kiezdeutsch and Namdeutsch

Oliver Bunk, Antje Sauermann, Fynn Raphael Dobler, and Heike Wiese
(Humboldt University Berlin)

This paper investigates sociolinguistic variation in multilingual speech communities, focusing on Kiezdeutsch and Namdeutsch, two German varieties spoken in distinct contact settings. Traditional sociolinguistics primarily focused on variables such as age, gender, and register (that this, formal vs. informal communicative contexts) in analysing variation. However, language contact is also known to boost linguistic dynamics, leading to new variants and varieties. Therefore, we target its impact on linguistic variation, and we do so at different levels: (a) speakers: multilingualism, i.e., speakers from multilingual vs. monolingual families, (b) languages: minority/heritage vs. majority language, (c) societal macro context: monolingualvs. multilingual habitus. Kiezdeutsch and Namdeutsch both emerged in multilingual settings, but differ with respect to (b) and (c) above: Kiezdeutsch is spoken in urban Germany, with German as the majority language and a monolingual habitus at the societal level, while Namdeutsch is spoken as a heritage language in the multilingual context of Namibia. We compare Kiezdeutsch and Namdeutsch data with that in monolingual settings, in order to evaluate the relevance of (a), and with heritage German in the US, – that is, German as a minority language in an environment with a societal monolingual habitus – in order to tease apartthe impact of (b) and (c). As our empirical basis, we use the Kiezdeutsch-corpus (KiDKo, Wieseat al. 2010ff), the DNam corpus of German in Namibia (Wiese et al. 2017, Zimmer et al. 2020),and the RUEG corpus of register-differentiated productions from mono- and bilingual speakers in Germany and the US (Wiese et al. 2020).

We investigate these factors and their possible interaction with traditional sociolinguistic variables by looking at the distribution of modal particles (MPs). We target the distribution of two MPs, "eben" and "halt" that are near- synonyms, both signalling evidentiality, and have been observed to change their distribution in ongoing developments. While "halt" is traditionally associated with southern varieties of German, and "eben" with northern ones, "halt" is currently gradually replacing "eben" in northern Germany (Elspaß 2005), suggesting an influence of sociolinguistic factors.

We compare the distribution of MPs in general and "eben" and "halt" in particular over registers, speaker groups, and language contact settings. We investigate whetherthe dynamics in the use of "eben" and "halt" that we can observe in Germany are reflected in contact varieties and whether these dynamics interact with sociolinguistic factors. Preliminaryfindings indicate that language contact factors, i.e., (a) – (c) above, play a role in the distribution of the two MPs, and that these factors interact with age and (in-)formality.

References

Bross, Fabian. 2012. "German modal particles and the common ground." *Helikon. A Multidisciplinary Online Journal* 2, 182-209.

Döring, Sophia. 2016. *Modal particles, discourse structure and common ground management: Theoretical and empirical aspects*. Doctoral Dissertation. Humboldt University Berlin.

Elspaß, Stephan. 2005. "Zum Wandel im Gebrauch regionalsprachlicher Lexik. Ergebnisse einer Neuerhebung." *Zeitschrift für Dialektologie und Linguistik*, 1-51.

Kokovidis, Alexandra. 2015. *Gendered discourse in German chatroom conversations: the use of modal particles by young adults*. Doctoral Dissertation. Boston University.

Thurmair, Maria. 1989. *Modalpartikeln und ihre Kombinationen.* Berlin, New York: Niemeyer.

Wiese, Heike, Rehbein, Ines, Schalowski, Sören, Freywald, Ulrike & Katharina Mayr. 2010ff. *KiDKo - Ein Korpus spontaner Unterhaltungen unter Jugendlichen im multiethnischen und monoethnischen urbanen Raum.*

Wiese, Heike, Simon, Horst, Zimmer, Christian & Kathleen Schumann. 2017. "German in Namibia: A vital speech community and its multilingual dynamics". In Péter Maitz & Craig A. Volker (eds.), *Language Contact in the German Colonies*. 221-245.

Wiese, Heike, Alexiadou, Artemis, Allen, Shanley, Bunk, Oliver, Gagarina, Natalia, Iefremenko, Kateryna, Jahns, Esther, Klotz, Martin, Krause, Thomas, Labrenz, Annika, Lüdeling, Anke, Martynova, Maria, Neuhaus, Katrin, Pashkova, Tatiana, Rizou, Vicky, Rosemarie, Tracy, Schroeder, Christoph, Szucsich, Luka, Tsehaye, Wintai, Zerbian, Sabine & Yulia Zuban. 2020. "RUEG Corpus (0.3.0) [Data set]. Zenodo". https://doi.org/10.5281/zenodo.3765218.

Zimmer, Christian, Wiese, Heike, Simon, Horst J., Zappen-Thomson, Marianne, Bracke, Yannic, Stuhl, Britta & Thomas Schmidt. 2020. "Das Korpus Deutsch in Namibia (DNam): Eine Ressource für die Kontakt- Variations- und Soziolinguistik". *Deutsche Sprache* 3: 210–232.

M. Emma Butterworth 05/10/2022

Methods in Dialectology XVII: https://methodsxvii.uni-mainz.de/calls-circulars/

Et puis, well: Discourse marker use in Ontarian Laurentian French

English and French have coexisted in what is now Ontario, Canada since the 18th century. This proximity between Ontarian-Laurentian French (hereafter OLF) and English has influenced OLF's lexicon, phonology, and semantic innovations (Mougeon 2004). A substantial amount of literature and recorded corpora exist investigating this influence (Mougeon 2004, Golembeski 1998, inter alia). However, comparing discourse marker (DM) use in majority versus minority French communities has not yet been studied. Expanding on existing research, I investigate a recent shift in the use of English and French DMs in OLF conversation in French-majority Hearst and French-minority Windsor.

OLF varieties developed in isolation from other LF varieties (e.g. Quebec, Manitoba), under ever-increasing English contact. This led to widespread lexical borrowing, semantic calquing, and increasing concerns about OLF's vitality (Golembeski 1998).

Recently, DMs have diverged from Standard French (1) to include English borrowings and calques adopted from English into French conversation. The result is the incorporation of English DMs such as *well*, illustrated in (2), and the semantic shift in the use and placement of French DMs *comme* (like) and *juste* (just), illustrated in (3) :

1. et puis les familles, **tsé**, sont plus, **disons**, françaises
   and then the families, ***y'know****, are more,* ***say****, French*

2. euh, **well** dans l'hiver j'aime faire du ski
   uh, ***well*** *in winter I like to ski*

3. Oui, c'est **juste comme** à cinq minutes de marche
   Yes, it's ***just like*** *a five-minute walk*

Sankoff et al. (1997) define DMs as informal lexical items acquired through interaction and linked to speaker fluency, linguistic security and speech-community integration. OLF speakers have altered their use of DMs, either reducing overall use or preferring English. This in turn indicates greater impacts of language contact and potential loss of speaker fluency (Chambers & Lapierre 2011).

Following the approach of Sankoff et al. (1997), I investigate DMs including *yeah*, *well*, *ok*, *t'sais / tsé* (y'know), *disons* (say), *puis* (then/and), and *donc* (so). I also include the DMs *comme* (like) and *juste* (just) as their meanings and positions mirror those of English *like* and *just* (Canac-Marquis & Walker 2016).

I draw on data from the *Phonologie du francais contemporain* corpus (PFC; Durand et. al 2002), including ten interviews respectively from the French-majority town of Hearst (Poiré & Tennant 2016), and the French-minority city of Windsor (Poiré et al. 2010). I explore the linguistic factors of overall DM use, overall English word use, and the DM's syntactic environment. Social factors include speaker's age, gender, and location (Canac-Marquis & Walker 2016).

Initial findings indicate a generational shift in the Windsor corpus, with older generations using French DMs (1), middle and younger generations incorporating English DMs (2), and younger generations decreasing overall DM use while increasing the use of direct translations *comme* and *juste* (3). How these findings compare to the majority context in Hearst sheds light on the influences of English on OLF, the current situation of the variety, and its future.

References

Canac-Marquis, Réjean and Douglas C. Walker. 2016. Variation in Canada: Effects of language contact in rural francophone Alberta. In *Varieties of Spoken French*, ed. Sylvain Detey, Jacques Durand, Bernard Laks, and Chantal Lyche, 463-477. Oxford University Press.

Chambers, J.K. and André Lapierre. 2011. Dialect Variants in the Bilingual Belt. In *Le Français en Contact: Hommages au Raymond Mougeon*, ed. France Martineau and Terry Nadasdi, 35-50. Les Presses de l'Université Laval.

Durand, Jacques, Bernard Laks & Chantal Lyche (2002). La phonologie du français contemporain: usages, variétés et structure. In C. Pusch & W. Raible (eds), *Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics – Corpora and Spoken Language.* Tübingen : Gunter Narr Verlag, 93-106. PDF

Golembeski, Dan. 1998. French Language Maintenance in Ontario, Canada: A sociolinguistic portrait of the community of Hearst. Doctoral dissertation, Indiana University.

Mougeon, Raymond. 2004. Perspective sociolinguistique sure le français en Ontario. In *Variation et francophonie*, ed. Aiden Coveney, Marie-Anne Hintze, Carol Sanders, 155-190. France: L'Harmattan.

Poiré, François and Jeff Tennant. 2016. French in Ontario: A speaker from Hearst. In *Varieties of Spoken French*, ed. Sylvain Detey, Jacques Durand, Bernard Laks, and Chantal Lyche, 290-299. Oxford University Press.

Poiré, François, Svetlana Kaminskaia, and Rémi Tremblay. 2010. Conséquences du contact avec l'anglais sur la réalisation de la liaison et du schwa en français de Windsor, Canada. In *Actes du XXVe Congrès International de Linguistique et de Philologie Romanes*, 365-374.

Sankoff, Gillian, Pierrette Thibault, Naomi Nagy, Hélène Blondeau, Marie-Odile Fonollosa, and Lucie Gagnon. 1997. Variation in the use of discourse markers in a language contact situation. *Variation and Change*, (9): 191-217. Cambridge University Press.

Evaluating Voices of Groningen

An interactive web-based approach to collecting Low Saxon dialect data

Raoul Buurke, Nanna Hilton, Martijn Wieling
University of Groningen

The *Stemmen van Grunnen* ('Voices of Groningen') web application (available at https://woordwaark.nl/stemmen) is aimed at collecting dialect data about the Low Saxon dialects within the Netherlands. It is based on *Stimmen van Fryslân* ('Voices of Friesland'; see Hilton, 2019) and similar applications (Leemann et al., 2016) in which the dialect of a speaker is located geographically on the basis of their selected dialectal variants for several words. Our application moreover asks the participant to record their own variants, which allows for investigating finer-grained dialect differences than when only selected variants are used. Advantageously, the acoustic recordings should no longer need manual transcription, as – in theory – the chosen provided transcription can be used.

Each participant is first asked to select the dialectal variant closest to their pronunciation for 10 Dutch words. The available variants were determined on the basis of the Goeman-Taeldeman-Van Reenen project (GTRP; Taeldeman & Goeman, 1996). Each variant is presented visually (in an intuitive spelling and IPA) and acoustically (if desired), after which participants are asked to record their own dialectal variant. Finally, the predicted geographical location of the speaker's dialect is shown on the basis of a GTRP-based decision tree.

Currently, over 1900 speakers have participated (i.e., approximately 19000 recordings). Before 2021, most data were collected from elderly speakers in the province of Groningen. Stemmen was included in a large-scale questionnaire covering a larger area in 2021, which doubled the amount of data.

The recordings of 377 words (10% of the available data at the time) were manually transcribed for evaluation purposes. Subsequently, we calculated the Levenshtein distance (Levenshtein, 1966) between these transcriptions and the transcriptions of the selected variant. The Levenshtein distance is a popular approach in dialectometry to quantify the difference between pronunciations (e.g., Kessler, 1995, Heeringa, 2004, and Wieling, 2012). The average Levenshtein distance was 0.5 (SD = 0.4), whereas the averaged normalized (over alignment length) Levenshtein distance was 0.09 (SD = 0.06), indicating a generally small difference between the actual pronunciation and the selected variant.

To assess whether participants chose the variant closest to their pronunciation, we ranked each Levenshtein distance between the option per word per participant and their pronunciation. We then normalized the ranks between 0 and 1 (with 0 representing the best possible choices and 1 the worst ones). In 16% of all cases, participants produced a form not present on the list. In a minority of these cases (28%) participants selected a non-optimal variant. In the 84% of cases when the pronounced variant was on the list only 10% of the time a non-optimal variant was selected. The feasibility of our approach is reflected by the close-to-optimal average normalized rank of 0.04 (SD = 0.05).

In sum, our approach with Stemmen van Grunnen seems suitable to obtain dialectal recordings together with (automatic) transcriptions. This opens the door to use both transcription-based dialectometric techniques, but also acoustic-based techniques to quantify pronunciation differences.

References

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance*. PhD thesis, University of Groningen.

Hilton, N. (2019). Smartphone sociolinguistics in minority language areas: 'Stimmen'. Published: Paper presented at International Conference on Language Variation in Europe (ICLaVE) 10, Leeuwarden.

Kessler, B. (1995). Computational dialectology in Irish Gaelic. In Abney, S. P. & Hinrichs, E. W. (Eds.), *Proceedings of the Seventh Conference of the European Chapter of the Association for Computational Linguistics*, (pp. 60–66). Morgan Kaufmann Publishers Inc.

Leemann, A., Kolly, M.-J., Purves, R., Britain, D., & Glaser, E. (2016). Crowdsourcing language change with smartphone applications. *PLOS ONE, 11*(1), e0143060.

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Taeldeman, J. & Goeman, A. (1996). Fonologie en morfologie van de Nederlandse dialecten: Een nieuwe materiaalverzameling en twee nieuwe atlasprojecten. *Taal en Tongval, 48*, 38–59.

Wieling, M. (2012). *A quantitative approach to social and geographical dialect variation*. PhD thesis, University of Groningen.

A Multidimensional Approach to Investigating Accent Attitudes in Britain

Amanda Cardoso, Erez Levon, Devyani Sharma, Dominic Watt, Christian Ilbury, Yang Ye

Language attitude research has long recognised that individuals can maintain conflicting explicit and implicit attitudes to the same attitude object (e.g., an accent or a variety), which researchers can access via direct or indirect methods, respectively. A recent critical summary of research on the topic, Rosseel and Grondelaers (2019), proposes moving beyond binary divisions and toward a more multi-faceted consideration of listeners' attitudinal reactions to language. In this paper, we take up this suggestion and present results from a multidimensional exploration of how listeners evaluate patterns of accent variation in 21st-century Britain. Drawing together approaches from linguistics, social psychology and labour market economics, we investigate how accent attitudes relate to subjective judgments of candidates in job interviews and, consequently, to social mobility in the UK more generally.

We report results from 5 studies, each targeting a different dimension of accent evaluation:

1) In Study 1, 826 members of the UK general public evaluate 38 English accent labels (e.g., Cockney, Liverpool) in terms of prestige and pleasantness (Giles 1970; Bishop et al. 2005). This study allows us to examine explicit attitudes, and so get a picture of the ideological landscape of accent variation in Britain. Results are remarkably consistent with prior research, demonstrating the existence of an enduring hierarchy of accents in the UK.
2) Study 2 asked 1106 UK listeners to evaluate native speakers of 5 English accents, who were "candidates" for entry-level positions in a major law firm. This study allows us to obtain more implicit attitudes related to an accent's perceived professionalism. Results show that differences in evaluation across accents are mitigated when using a more indirect approach, and are significantly conditioned by listener factors such as age, region and motivation to control a prejudiced response.
3) Study 3 asked 61 lawyers and professional recruiters to evaluate the same stimuli as in Study 2. By replicating the study with lawyers/recruiters, we further explore listener background and context as factors that moderate attitudinal outcomes. Results show an even stronger mitigation of accent attitude effects, with lawyers showing no bias across accents at all.
4) Study 4 examines real-time evaluations of accents among 160 UK listeners, allowing us to explore the relationship between evaluative endpoints and online attitudinal processing. Results show that standard versus non-standard accents engender distinct real-time response trajectories, demonstrating the existence of a deeper and more implicit form of accent bias.
5) Finally, Study 5 asked 80 UK listeners to rate 10 speakers of 5 UK accents for how "strong" their accent is and how "professional" they sound. Results show judgments of accent strength are a better predictor of professionalism ratings than accent, indicating that the target of evaluations may be prevalence of (non-standard) features rather than "accent".

Together, the 5 studies present us with a holistic perspective on attitudes to contemporary UK accents. In the talk, we discuss the ramifications of our findings for current theories of language attitudes, and for the role of accent as an impediment to social mobility.

References

Bishop, H., N. Coupland and P. Garrett (2005) Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia* 37: 131-154.

Giles, H. (1970) Evaluative reactions to accents. *Educational Review* 22: 211-227.

Rosseel, L. & Grondelaers, S. (2019) Implicitness and experimental methods in language variation research. *Linguistics Vanguard*, 5(s1): 1-6. Retrieved 30 Nov. 2019, from doi:10.1515/lingvan-2018-0005

PP Complements in Vernacular Spanish: Constituency Tests and Syntactic Variation of the *complemento de régimen* (CR)

Vanessa Casanova (Université de Liège / Universiteit Gent)

As described in Modern Spanish (Cano 1999; de Santiago 2007; RAE/ASALE 2009), the PP known as Verbal Government Complement – henceforth, *complemento de régimen* (CR) – differs from other obliques in that it expresses a participant of the argument structure (1a)[1], just as direct and indirect objects are the formal expression of core arguments (1b-c).

(1) a. Creo que tuiter me odia, o ***se burla de mí***$_{CR}$ (Twitter, Mexico, 10/17/19). < *burlarse de alguien* [to make fun of somebody]
b. Lo mejor es ir sonriendo por la calle. Así la gente piensa que ***tramas algo***$_{DO}$ [Twitter, Spain, 07/18/13). < tramar algo [to plot something]
c. *Achacaban* ***a la bruja***$_{IO}$ las epidemias reinantes y otros desaguisados (CdE, Cuba). < *achacar algo a alguien* [to blame something on somebody]

Despite being defined as core complements (Cano 1999), syntactic criteria are not sufficient alone to identify these PP arguments in a systematic way. The plethora of tests that have been proposed and contested (Alarcos 1968; Bosque 1983; Rojo 1990; Martínez 1986; Gutiérrez Araus 1987; Gutiérrez Ordóñez 1994; Serradilla-Castaño 1997-1998; Casanova 2021) rely on a precondition: argumenthood, a semantic feature that can only be assessed by following lexicographic criteria (RAE/ASALE 2009).
The identification of CRs is even more challenging when confronted to vernacular data. Common constituency tests – pronoun substitution of DPs (2a), non-adverbial coordination (2b), non-passivization (2c), ecuanditional structures (2d), indexical non-deletion (2e-f), etc. – are used to identify CRs in normative Spanish, but when applied to vernacular varieties the results may diverge.

(2) a. Si la Palabra de Dios se supone que sea la infalible palabra de Dios mismo, ¿cómo se supone que ***la confiemos*** [*General Spanish: confiemos EN ella*] lo suficiente como para vivir nuestras vidas 100 % de acuerdo a ella […]? (CdE, Paraguay).
b. La sociedad mundial se forma y obliga a personas provenientes del sur a ir a ***vivir al norte y recíprocamente*** [*vivir a algún lugar ≠ vivir de algún modo*] (CdE, República Dominicana).
c. Un tipo con el físico de Lincecum ***es dudado*** por las cosas que puede hacer. Un tipo con el físico de Kershaw es excusado por las cosas que no puede hacer [**dudar a alguien > alguien ser dudado*] (CdE, Venezuela).
d. Mira si me voy a esmerar en hacer sentir bien a una flaca que la disfruto cualquiera. ***Si me esmero es en una flaca que sea solo mia*** [*si en alguien me esmero es en una flaca*] (Twitter, Argentina, 03/18/19).
e. Para la Mesa, el borrador no señala ***qué consisten esas dos circunstancias*** [*consistir EN algo*] (CdE, El Salvador).
f. Mis amigos me kieren por lo que soy, pero la gente que no conosco ***se burla mio*** [*burlarse DE alguien*] porque soy diferente a ellos (CdE, Argentina).

[1] Examples in (1) and (2) are taken from *Corpus del Español: Web/Dialectos* and from geolocalized tweets.

This study is an attempt to answer the following research questions: 1) How can a dialectal corpus help revisiting current criteria for CR identification? 2) Which tests pass and which ones appear to fail? 3) Is syntactic variation related to diatopic factors as well? Which grammatical features are relevant to describe these verb alternations? For this purpose, I have tested a selection of PPs representing the various syntactic schemes of the CR. In this work I claim that 'test failures', rather than questioning the categorial status of CRs, reflect the ongoing changes in Spanish, in congruence with historical processes of [±transitive] alternations from Medieval Spanish to present day (Cano 1977-1978; Rivas 2004). These findings (cf. Casanova 2021), in the light of dialectal syntax, call for a more nuanced description of PP arguments, which appear to be placed on a *continuum* of core/adverbial complementation (Hagège 2010).

References

Alarcos Llorach, Emilio. 1968. "Verbo transitivo, verbo intransitivo y estructura del predicado". *Archivum* 16: 5-17.

Alarcos Llorach, Emilio. 1986. "Prólogo" to H. Martínez García, *El suplemento en español*. Madrid: Gredos, 3-5.

Bosque, Ignacio. 1983. "Dos notas sobre el concepto "suplemento" en la gramática funcional". *Dicenda* 2: 147-156.

Cano Aguilar, Rafael. 1977-1978. "Cambios en la construcción de los verbos en castellano medieval". *Archivum* XXVII-XXVIII: 335-379.

Cano Aguilar, Rafael. 1999. "Los complementos de régimen verbal" in Ignacio Bosque y Violeta Demonte (Coords.), *Gramática descriptiva de la lengua española*, Vol. 2. Madrid: Espasa Calpe, 1807-1854.

Casanova, Vanessa. 2021. *El complemento de régimen preposicional: construcción y distribución en español actual*. Tesis doctoral. Universidad de Montreal.

Demonte, Violeta. 1991. "La realización sintáctica de los argumentos: el caso de los verbos preposicionales" in *Detrás de la palabra: estudios de gramática del español*. Madrid: Alianza Editorial, 69-115.

De Santiago Guervós, Javier. 2007. *El complemento (de régimen) preposicional*. Madrid: Arco/Libros.

Gutiérrez Araus, M.ª Luz. 1987. "Sobre la transitividad preposicional en español". *Verba* 14: 367-381.

Gutiérrez Ordóñez, Salvador. 1994. "Estructuras ecuandicionales" in Violeta Demonte (Ed.), *Gramática del español*. México D.F.: El Colegio de México, 363-383 [recogido en S. Gutiérrez Ordóñez. 1997. *La oración y sus funciones.* Madrid: Arco/Libros, 549-576].

Hagège, Claude. 2010. *Adpositions: Function-marking in Human Languages*. Oxford University Press.

Martínez García, Hortensia. 1986. *El suplemento en español*. Madrid: Gredos.

Porto Dapena, José-Álvaro. 2014[1992]. *Complementos argumentales del verbo: directo, indirecto, suplemento y agente*. Madrid: Arco/Libros.

Real Academia Española y Asociación de Academias de la Lengua Española (RAE/ASALE). 2009. *Nueva gramática de la lengua española*. Madrid: Espasa.

Rivas, Elena 2000-2001. "Régimen directo y régimen preposicional en castellano medieval". *Philologica canariensia* 6-7: 117-144.

Rojo, Guillermo. 1990. "Sobre los complementos adverbiales" in VV. AA. *Homenaje al Profesor Francisco Marsá. Jornadas de Filología*. Barcelona: Universidad de Barcelona, 153-171.

Serradilla Castaño, Ana M.ª. 1997-1998. "El complemento de régimen preposicional: criterios para su identificación". *CAUCE. Revista de Filología y Didáctica*, *21-22*, 1017-1051.

A proto-Ryukyuan Database: an aggregating model of dialectal lexical data

CELIK Kenan (Nnationa Institute for Japanese Language and Linguistics)
NAKAZAWA Kohei (Shinshu University)
ASO Reiko (Meio University)

In the investigation into the history of the Japonic languages, the reconstruction of proto-Ryukyuan (pRk), i.e. the most recent common ancestor of all the languages belonging to the Ryukyuan branch of the Japonic languages, is a necessary step. Detailed reconstructions of Proto-Ryukyuan have already been proposed (Hattori 1978-79, Thorpe 1986), but they predate the publication of major lexical works on several Ryukyuan doculects, among others Maeara (2011), Tomihama (2013), Kajiku (2020), Tokuyama, Celik (2020). More recent research on pRk like Pellard (2009, 2015) does integrate more up-to-date data, but even since the publication of this research, the availablity of lexical data on Ryukyuan doculects has grown at a dizzying pace.

Needless to say, any proposed reconstruction of pRk should be tested against new data and be updated accordingly. The unprecedented growth in lexical data on Ryukyuan dialects opens up the possibility of a robust testing of the models proposed so far, but, at the same time, it raises the non-trivial problem of how to integrate this uninterrupted flow of new dialectal data into the reconstruction. A solution to this problem would be a data management system that is able to continuously integrate the latest data into the existing model of pRk.

A pioneering answer to this problem is Igarashi's "Japonic classified vocabulary" (Igarashi 2016), a periodically updated database of reconstructed pRk Forms, containing in its lastest version (v.7 Igarashi 2019) more than 1800 pRk words. However, this database suffers from several flaws. First, it does not make public the dialectal data on which the reconstructions are based. Another major flaw lies in the approach adopted, which could be described as "extractive data-building". Lexical data is extracted from each source into the database without keeping a co-reference with the data in the source. This design, in which the link with the primary source is lost, not only limits research possibilities, but, more importantly, it also cannot cope with dynamic lexical data building, in which the primary source is regularly updated.

In the face of this, we constructed a proto-Ryukyuan database (approximately 6000 cognate sets) adopting a radically different approach, which could be described as "aggregate data-building". Namely, we devised an ID system enabling us to link cognates across different lexical data sets, whatever their data structure. Since the link with all primary sources is kept, not only does this design efficiently cope with dynamic lexical building, but it also vouches for more flexible data building. For example, since we link cognates to a common cognate set and not to a proto-form, there is no need to decide whether a particular cognate set goes back to the proto-language or not. This approach also

ensures transparency in the reconstruction as we can refer to all the dialectal data with a unique ID. Lastly, this design also opens up many other research possibilities than the reconstruction of pRk, like for instance the mapping of isoglosses or the exploration of the geographic factors of word variation.

References

Hattori, Shirō. 1978–1979. Nihon sogo ni tsuite 1–22. *Gengo* 7(1)–7(3), 7(6)–8(12).

Igarashi, Yōsuke. 2016. Akusento-gatano taiōni motoduite nichiryū-sogowo saiken suru tameno goi-risuto "nichiryūgo-ruibetu-goi". *Nihongo-gakkai 2016 nendo shunki taikai yokōshū*, 233-238.

Igarashi, Yōsuke. 2019. Japonic Classified Vocabulary (v.7, 2019/5/17) electronic data (last accessed 2021/8/18).

Kajiku, Shin'ichi. 2020. *Hatoma hōgen jiten*. Tokyo: The National Institute for Japanese Language and Linguistics.

Maeara, Tōru. 2011. *Taketomi hōgen jiten*. Ishigaki: Nanzansha.

Pellard, Thomas. *Ōgami: Éléments de description d'un parler du Sud des Ryūkyū*. Linguistique. Ecole des Hautes Etudes en Sciences Sociales (EHESS) PhD thesis.

Pellard, Thomas. 2015. The linguistic archeology of the Ryukyu Islands. In *Handbook of the Ryukyuan languages: History, structure, and use*, eds., Patrick Heinrich, Shinsho Miyara, Michinori Shimoji, 13–37. Berlin: De Gruyter Mouton.

Thorpe, Maner L. 1983. *Ryūkyūan language history*. University of Southern California PhD thesis.

Tokuyama, Shun'ei, Celik, Kenan. 2020. *Minami Ryūkyū Miyakogo Tarama hōgen jiten*. Tokyo: The National Institute for Japanese Language and Linguistics.

Tomihama, Sadakichi. 2013. *Miyako Irabu hōgen jiten*. Okinawa: The Okinawa Times.

Yes-No Questions in Acadian French

Philip Comeau, UQAM; Ruth King, York University; Carmen L. LeBlanc, Carleton University

This study involves comparison of Atlantic Canada Acadian French varieties which differ in their sociodemographic histories. The data come from archival and sociolinguistic corpora recordings for four communities: Baie Sainte-Marie, which has a long history of dialect isolation; Chéticamp and Iles-de-la-Madeleine, which have had close historical ties; and L'Anse-à-Canards, which saw relatively late 19th century settlement involving Acadian settlers with origins in Chéticamp and Iles-de-la-Madeleine along with a northwestern France founder group.

The history of French yes-no questions has involved no fewer than seven variants, with those illustrated below robustly present in our Acadian varieties. 1 involves pronominal inversion (P-INV); 2 rising intonation (INT); and 3 the *ti* question particle (TI).

1 Voudrais- tu une tasse de café? (AC-03)
want.COND.2SG you.SG INDF.F.SG cup of coffee
'Would you like a cup of coffee?'

2 Vos garçons sont venus? (CH-02)
your.PL sons be.PRS.3PL came
'Your sons have come?'

3 Le puit est ti dehors icitte? (BSM-13)
DEF.M.SG well be.PRS.3SG Q outside here
'Is the well outside?'

P-INV and INT can be traced to Old French while TI is the result of 16th century reanalysis of the third-person singular subject pronoun *i*(*l*) and a preceding epenthetic [t] in complex inversion constructions (Brunot & Bruneau 1969; Foulet 1921).

Of a total of 1362 tokens extracted for analysis, P-INV is limited to the second person, arguably its last bastion in spoken French. The overall results contrast with Québec French in that further reanalysis of the Q particle ([ti] > [ty], i.e., TU) has not taken place (cf. Léard 1996) and TI is indeed compatible with negation as in 4 (cf. Vinet 2000).

4 Elle était ti pas rose? (MA-13)
she be.IMP.3SG Q NEG pink
'Wasn't it pink?'

A number of linguistic and social variables were tested in a series of mixed-effects Rbrul analyses for each community. No social variables were selected as significant. The main findings for the linguistic variables are as follows. TI is found throughout the verbal paradigm only for L'Anse-à-Canards, reminiscent of the northern Metropolitan French pattern of the late 19th century (Renchon 1967). For the remaining communities, TI is strongly favoured in third-person contexts but entirely absent from second-person contexts. INT is favoured in negative yes-no questions for Chéticamp, Iles-de-la-Madeleine and L'Anse-à-Canards but not to the extent that has been found for Québec French, where the effect is categorical (Elsig 2009). However, no polarity effect is found for Baine Sainte-Marie. To account for this case of interdialectal variation, we consider the semantico-pragmatic content of negative questions, specifically whether they are information-seeking or confirmatory, i.e., either affirmative or negative confirmation-seeking (Borillo 1979, Ladd 1981). Where there were sufficient negative tokens, our analysis revealed a relationship between TI and negative questions which are affirmative confirmation-seeking (clearly seen for Baie Sainte-Marie and to a somewhat lesser extent, L'Anse-à-Canards). This suggests a dialectal difference between Québec French and finely grained variation in Acadian French along

with a distinction involving subtype of negative question, a finding not reported in earlier studies of French varieties.

References

Borillo, Andrée. 1979. La négation et l'orientation de la demande de confirmation. *Langue française* 44. 27-41.

Brunot, Ferdinand, and Charles Bruneau. 1949. *Précis de grammaire historique de la langue française*. 3 ed. Paris, France: Masson.

Elsig, Martin. 2009. *Grammatical variation across space and time: The French interrogative system*. Amsterdam: Benjamins.

Foulet, Lucien. 1921. "Comment ont évolué les formes de l'interrogation." *Romania* 47: 243-348.

Ladd, D. Robert. 1981. A first look at the semantics and pragmatics of negative questions and tag questions. *Proceedings of Chicago Linguistic Society* 17. 164-171.

Léard, Jean-Marcel. 1996. "*TI /TU*, *EST-CE QUE*, *QU'EST-CE QUE*, *CE QUE*, *HÉ QUE*, *DON* : des particules de modalisation en français?" *Revue québécoise de linguistique* 24 (2):107-123.

Renchon, Hector. 1967. *Études de syntaxe descriptive, 2, La syntaxe de l'interrogation*. Brussels: Académie royale de langue et de littérature françaises.

Vinet, Marie-Thérèse. 2000. "La polarité pos/nég, -tu et les questions oui/non." *Revue québécoise de linguistique théorique et appliquée* 28 (1):137-149.

Is the sociolinguistic interview a useful tool for gathering data on everyday linguistic behavior?

Sociolinguistics is built on the assumption that the sociolinguistic interview, the primary mechanism for gathering data for the study of language variation and change, produces speech reflecting the everyday linguistic behavior of individuals in a defined population. In fact, this data-gathering technique is so common that many papers published in the field simply state that data were gathered through "sociolinguistic interviews" or "sociolinguistic interviews in the Labovian tradition" without any further elaboration. However, such statements essentially obfuscate rather than clarify the research methods used. "Sociolinguistic interviews" actually comprise a wide range of approaches to gathering data, and as we show in this paper, these approaches can have substantial consequences on the kind of data that emerges in a study.

The idea of sociolinguistic interview dates from the 1960s, when Labov (1966) provided a detailed analysis of the effects of the interview situation on the elicitation of natural speech and offered a number of techniques for producing casual speech. He used a variety of approaches (reading passages, minimal pairs) to extend the formal end of the stylistic continuum and developed a number of techniques to create situations where more casual speech occurs, including interruptions by third parties, danger of death questions, and interviews with peer groups. When researchers indicate that they did standard sociolinguistic interviews with no further elaboration, it is not clear exactly which of these techniques they have used or exactly what comprises the interviews. However, what happens in a sociolinguistic interview can have a dramatic impact on the data that emerges from that interview. In fact, in light of the research reported on in this paper, it may be problematic to assume that data from sociolinguistic interviews always represents the everyday linguistic behavior of those being interviewed.

Our data for exploring the impact of various approaches to sociolinguistic interviews comes from fieldwork in Springville, Texas, which includes both interviews using most of the techniques developed by Labov (along with a number of other techniques) and also individual, peer group, and site study interviews. A quantitative analysis of the occurrence of zero forms (copula and present 3rd singular) in the English of 67 African American residents shows that interview type affects the use of zero for some individuals but not for others, and that even the same individuals do not always use zero copula and zero 3rd singular consistently across interview contexts. Moreover, for some, although not all people interviewed, subsequent interviews provide substantially higher rates of zero. Finally, even when we account for factors such as interview context and familiarity, the total number of tokens, often a consequence of the number or length of interviews, has an impact on results. Since we cannot predict either the differing effects of various interview contexts or of multiple interviews among individuals, or which features will be affected, for the sociolinguistic interview to be a useful tool, speakers should be interviewed multiple times in a variety of contexts to ensure that results are not colored by the type of "sociolinguistic interview" used.

References

Labov, William. 1966. *The Social Stratification of English in New York City*. Washington, D.C.: Center for Applied Linguistics.

Reglindis De Ridder
Institutionen för slaviska och baltiska språk, finska,
nederländska och tyska
Avdelningen för nederländska

Abstract for "Borders, Dialects and Standard Varieties" session

The usage of Belgian Dutch in translated and non-translated audiovisual children's content and its sociolinguistic implications

Dutch as it is used in Belgium (henceforth *Belgian Dutch**) has taken a different course from *Netherlandic Dutch***, despite massive language planning efforts in the second half of the 20th Century to prevent this (Jaspers and Van Hoof 2013). For a long time, Belgian Dutch was considered a deviation from 'proper' Dutch. The turn of the century, however, marked the official recognition of the Dutch language area as a pluricentric language area with two equal national varieties in Europe by the Dutch language planning body (Nederlandse Taalunie 2003). Nevertheless, Netherlandic Dutch is still the dominant variety in translated (audiovisual) fiction. This paper discusses the results of a sociolinguistic analysis of 290 children's programmes focussing on the usage of Belgian and Netherlandic Dutch (De Ridder 2020a) with a reception study into parents' opinion as regards their children's exposure to both varieties of Dutch (idem 2020b). It highlights differences between local and imported programmes and calls for further sociolinguistic research into the language used in different children's media and how it may affect language development in children and language attitude. Children's television has been criticized for its lack of diversity, yet, linguistically, children's media may also be out of touch with reality.

* In English, often referred to as 'Flemish', however, the official term used by the Nederlandse Taalunie is 'Belgisch Nederlands'/'Belgian Dutch'.
** In Dutch, colloquially referred to as 'Hollands', however, the official term used by the *Nederlandse Taalunie* is 'Nederlands Nederlands'/'Netherlandic Dutch'.

References

De Ridder, Reglindis (2020). Dutch national varieties in contact and in conflict. In: R. Muhr, J. A. Mas Castells & J. Rueter (eds.) *European Pluricentric Languages in Contact and Conflict*. Wien: Peter Lang Verlag, 65-80.

De Ridder, Reglindis (2021). Linguistic diversity in audiovisual media for children in Belgium and Austria. In: R. Muhr and J. Thomas (eds.) *Pluricentric Theory beyond Dominance and Non-dominance*. Graz/Berlin: PCL press, 121-136. Available from https://pcl-press.org/publications/pluricentric-theory-beyond-dominance-and-non-dominance-a-critical-view/ (Last accessed on 26 September 2021).

Nederlandse Taalunie. (2003). Rapport Variatie in het Nederlands: eenheid in verscheidenheid [Report on the variation within Dutch: unity in diversity]. Available from https://taalunie.org/publicaties/94/variatie-in-het-nederlands-eenheid-in-verscheidenheid (Last accessed on 26 September 2021).

Van Hoof, Sarah & Jaspers, Jürgen (2013). Hyperstandardisation in Flanders. Extreme enregisterment and its aftermath. *Pragmatics*, 23(2), pp.331-359.

Towards a new lexicographical Infrastructure for the Dutch Dialects: the Database of Southern Dutch Dialects project.

De Tier, Veronique; Depuydt, Katrien; De Does, Jesse (Dutch Language Institute)
Chambers, Sally; Vandenberghe, Roxane; Hellebaut, Lien; Van Keymeulen, Jacques (Ghent University)

The Southern Dutch dialect area is described in four separate dictionaries, available both in print and online. The Brabantic, Limburgian and Flemish dialect dictionaries were set up as parallel onomasiological dictionaries whereas the Zeelandic dictionary was ordered alphabetically. The idea behind a common data model was to provide a comprehensive overview of the entire dialect area covered by these dictionaries. Through the Dictionary of the Southern Dutch Dialects (DSDD) project, the three dictionaries were brought together in one portal, realising the first phase of a new infrastructure for Dutch dialects. This paper will discuss: how these three conceptually similar dictionaries were brought together, what the challenges to harmonise these three dialects datasets were and how this enabled the integrated dataset to be made accessible both via a user application with cartographic tools, and an API.

The data model
Even though the dictionaries used the same data model it became clear that an overarching concept layer was required to deal with the problem of similar, but not necessarily equivalent concepts (e.g. the choice for either a concept “frog”, or two concepts “frog” and “green frog”) and the corresponding heteronomy (i.e. all dialect words for a single concept).

Data format
The source data was received in a range of different formats, e.g. database extracts from Oracle and FileMaker or as OCR (XML), and then stored in a relational database (PostgreSQL).

Data quality
Some of the original material had been OCR’d and semi-automatically/manually corrected, which had resulted in poorly structured data. It was therefore unfortunately necessary to leave out some of the data. However, it was made sure that the data ingestion method allows for future updates.

Data curation and enrichment
Before aligning the concepts some curation and enrichment was necessary to avoid inconsistencies, for instance, the differences: a) between the ‘Dutchification’ of keywords and lexical variants, b) in the assignment of lexical variants to keywords, c) in spelling, etc. had to be resolved.

DSDD concept layer and linking
In the pilot phase, 1500 concepts from a number of thematic dictionary volumes were selected to explore different methods for aligning the data. For each theme a list of overarching concepts was compiled. Lex’it, a rapid database application development platform for linguistic data, developed at the Dutch Language Institute

(INT), was used to do the linking. When the names of the concepts were identical, concepts were linked semi-automatically. However, when they were not, strategies such as keyword overlap, searching in concept definitions etc. were used for linking. Later on other concepts were integrated and linked. Now, the database consists of 29.000 concepts.

Future work

The dataset is now accessible via a user application with cartographic tools and an API. In 2022/2023, the database will be extended with additional semasiological Dutch dialect dictionary data, such as the *Zeelandic Dictionary,* until ultimately, the dialect data in the lexicographical infrastructure covers the entire Dutch language area.

References

De Vriend, F. & L. Boves, H. van den Heuvel, R. van Hout, J. Kruijsen & J. Swanenberg, Jos. (2006). A unified structure for dutch dialect dictionary data. in: *Proceedings of The fifth international conference on Language Resources and Evaluation* (LREC 2006).

De Vriend, F. (2012), *Tools for Computational Analyses of Dialect Geography Data*. PhD Radboud University Nijmegen.

Van Keymeulen, J. (2004), Trefwoorden en lexicale varianten in de grote regionale woordenboeken van het zuidelijke Nederlands (WBD, WLD, WVD). In: De Caluwe J, G. De Schutter, M. Devos en J. Van Keymeulen*, Taeldeman, man van de taal, schatbewaarder van de taal.* Vakgroep Nederlandse Taalkunde UGent – Academia Press, Gent (2004); 897-908.

Van Keymeulen, J., V. De Tier, R. Vandenberghe & S. Chambers (2019), The dictionary of the Southern Dutch Dialects (DSDD): designing a virtual research environment for digital lexicological research. in: *Dialectologia. Special issue*, 8 (2019), 93-115.

Van den Heuvel, H, E. Sanders en N. van der Sijs (2016), Curation of Dutch Dialect Dictionaries In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC 16).

Van Hout, R, N. van der Sijs, E. Komen en H. van den Heuvel (2018), A fast and flexible web interface for dialect research in the Low Countries. In: *Proceedings of The fifth international conference on Language Resources and Evaluation* (LREC 18).

Modelling standard varieties: epistemological considerations, “fail-safes”, and the peculiar case of German

Stefan Dollinger (UBC Vancouver)

This paper puts at its centre the German “Pluricentricity Debate” (Dollinger 2019), which explores the question whether the pluricentric view of languages is still adequate today. This debate is important, as recent critics have re-introduced the counter term of “pluri-areality” (Scheuringer 1996) and German dialectology has seen the branding of pluricentricity as an outdated model that is hampered by national limitations (e.g. Elspaß and Niehaus 2014, Herrgen 2015, Langer 2021, Koppensteiner & Lenz 2021). The pluricentric perspective of German – one language, several national standards – is, in German linguistics, now questioned more than at any point since Clyne’s (1984) landmark publication.

The debate affords the opportunity to inquire how German – and any other codified language – should be modelled in the 21st century and allows conclusions about gaps in English, Dutch and other varieties of comparable social use. To that purpose, a comparative view is taken in this meta study that contrasts the sociolinguistic situations, linguistic behaviours, attitudes and perceptions in German with other Germanic varieties. Although philology-specific concepts do have their place, it will be shown that “pluri-areality” represents no such case, leaving pluricentricity as the most appropriate theory to date, a concept that abides by the epistemological principle of hypothesis testing (Popper 1966).

Clyne, Michael G. 1984. *Language and Society in the German-speaking Countries*. New York : Cambridge University Press.

Dollinger, Stefan. 2019. *The Pluricentricity Debate: On Austrian German and other Germanic Standard Varieties*. London: Routledge

Elspaß, Stephan and Konstantin Niehaus. 2014. The standardization of a modern pluriareal language. Concepts and corpus designs for German and beyond. *Orð og tunga* 16: 47–67.

Herrgen, Joachim. 2015. Entnationalisierung des Standards. Eine perzeptionslinguistische Untersuchung zur deutschen Standardsprache in Deutschland, Österreich und der Schweiz. In: Lenz & Glauninger (2015), 139-164.

Koppensteiner, Wolfgang & Alexandra Lenz. 2021. Standard(s) aus der Perspektive von „Nicht-LinguistInnen“ in Österreich. In Hoffmeister et al. (2021), 391-416.

Langer, Nils. 2021. Review of Dollinger 2019, with rejoinder. *Zeitschrift für Rezensionen zur Germanistischen Sprachwissenschaft*. Online First. https://www.academia.edu/45577124/

Popper, Karl R. 1966. *Logik der Forschung*. Zweite, erw. Auflage. Tübingen: Mohr.

Democratising dialect fieldwork and research: desirable, doable, or doomed?

Fiona Douglas and Rosemary Hall

Despite widespread criticism (Chambers and Trudgill 1998, Wells 1978), the Survey of English Dialects (1950-61) is still considered one of the benchmarks of English dialect study, and its findings are an invaluable resource for modern dialectologists (Jansen et al 2020, Britain 2009). Its emphasis on 'traditional dialect, genuine and old' (Orton 1960: 332), its preference for non-mobile older rural males ('NORMS') and its focus on rural life and communities are deeply unfashionable nowadays. A long, onerous oral survey designed to elicit individual words, pronunciations and grammatical forms rather than spontaneous connected speech, administered by trained fieldworkers, and transcribed in real time might be dismissed as methodologically outdated and problematic. The policy of interviewing a few selected individuals in each locality and then aggregating their responses to form one village return raises questions of representativeness. And yet its influence persists.

Seventy years on, the Dialect and Heritage Project is a National Lottery Heritage-funded initiative with interconnected research, public engagement, and impact goals including (but not limited to) the digitisation of original SED materials and the gathering of new present-day dialect data. Its partnership with five folk life museums spread across England offers interesting opportunities for new, collaborative data collection methods, including:

- A distributed fieldwork model in which volunteers are trained to conduct oral history/dialect interviews in their communities
- Simultaneous data collection and public engagement activities at 'dialect roadshow' events using a specially designed pop-up dialect kit
- Dialect reminiscence sessions (held both virtually and in-person)
- A new dialect survey
- Reconnecting with original SED informants via their descendants

The project aims to investigate:

- Similarities and differences between dialect use past and present, and across different areas of the country
- The role of family intergenerational relationships in dialect maintenance/transmission ('inherited' words)
- 'Adopted' dialect words that people take on as they move around the country, develop new friendships and relationships
- The link between dialect, heritage and identity

Its distributed fieldwork approach, equipping and empowering members of the public to become fieldworkers, transcribers, and project ambassadors, is designed to embed knowledge and ownership within local communities and to create a lasting legacy — but it is methodologically risky. By involving the public as co-researchers, co-creators and co-custodians of knowledge, and by conducting research alongside public engagement activities open to all, we lose some of the control. And we must do our research in ways that are enabling, meaningful and fulfilling for our co-creators. This paper summarises the project's successes and challenges thus far, previews early results from the new present-day dialect research, and asks whether such methodological innovations are doomed. It argues that democratising research brings its own challenges but also rich and unexpected rewards.

References

Britain, D. 2009. "One foot in the grave? Dialect death, dialect contact, and dialect birth in England". *International Journal of the Sociology of Language* 196-197: 121-155.

Jansen, S. et al. 2020. Sussex by the sea: A descriptive analysis of dialect variation in the South East of England based on English Dialect App data. *English Today* 36/3: 31–39.

Orton, H. S. 1960. An English dialect survey: Linguistic atlas of England. Orbis: Bulletin international de documentation linguistique, IX/2: 331–348.

Convergence and divergence of dialects in a border area: evidence from Basque syntax

Maitena Duhalde de Serra
Sergio Monforte
University of the Basque Country (UPV/EHU)

Keywords: Basque dialects; border contact; syntax.

Classical dialectology has demonstrated that political-administrative and linguistic borders do not always have to coincide. In this paper we look into neighbouring Basque varieties, spoken on both sides of the political border, namely the guipuscoan and navarrese (Spain) and labourdin (France) dialects following the first classification of Basque dialects (Bonaparte, 1863). For that purpose, we take three grammar properties into account: a) the dummy verb *egin* 'to do' and the morpheme *ba-* for emphasis; b) the particles *ahal* and *al*; and, c) the non-adjacency of the *wh*-word and finite verb in embedded contexts.

Our corpus is composed not only of texts from the second part of the XXth century but also of interviews carried out according to the standard criteria of dialectology (Chambers & Trudgill [1998] 2004).

The first characteristic deals with the syntactic means to mark emphasis on the verb, namely the dummy *egin* ('do') mainly used in western dialects and the morpheme *ba-* particular of eastern dialects (cf. Euskaltzaindia 2016). Data show that both strategies arise in the Labourdin coast:

(1) Autsi ezta iñen.
break not.AUX do.FUT
'It won't break.'

(2) Baut uste senarra eztuen bate ongi.
CL.AUX think husband.ABS not.have at.all well
'I do think that her husband wasn't well at all.'

The second property concerns the particle *ahal* used in all dialects and the particle *al* attested only in Guipuscoan (Euskaltzaindia 1987; de Rijk 2008). Historically *ahal* − also pronounced as *al* − has been productive in the three areas not only in declarative contexts but

also in biased questions. However, its current function as a question particle in Guipuscoan has led to separate interpretations of the same utterance:

(3) Etorriko al aiz?
come.FUT PART AUX
'You're coming?!' (Labourdin)
'Are you coming? (Guipuscoan)

The third property looks into indirect questions. It is an acknowledged fact that the *wh*-word and the finite verb must be adjacent not only in matrix questions but also in embedded ones (Hualde & Ortiz de Urbina 2003); nevertheless, eastern varieties have developed a more flexible behaviour in indirect questions since other constituents can occur between both them (Monforte 2020).

(4) Ez dakit non ote kazeta utzi dudan.
not know where PART journal.ABS leave AUX.C
'I don't know where I might have left the journal.'

In conclusion, considering the properties mentioned above, we see that the border can be impermeable. The first property suggests that there is a linguistic continuum in the cross-border area covered by the guipuscoan, labourdin and navarrese dialects. As for the second characteristic, although the particle *ahal* is still in use in the labourdin dialect, the grammaticalization of *ahal* into *al* arisen in the guipuscoan dialect has not taken this step forward in the former. Finally, other innovations have not spread across the border such as the third property. Indeed, chronology accounts for the different grade of diffusion since the first property is longer attested than the second and third ones.

References

Bonaparte, L.-L. 1863. *Carte des sept provinces basques montrant la délimitation actuelle de l'euscara et sa division en dialectes, sous-dialectes et variétés*, Londres: Stanford's Geographical Establishment.

Chambers J. K. & P. Trudgill [1998] 2004. *Dialectology, second edition*, Cambridge: Cambridge University Press.

Euskaltzaindia (Royal Academy of the Basque language). 1987. *Euskal gramatika. Lehen urratsak II* Bilbo: Euskaltzaindia.

Euskaltzaindia (Royal Academy of the Basque language). 2016. *Euskararen Herri Hizkeren Atlasa. VII, lexikoa, izen morfologia, sintaxia.* Bilbo: Euskaltzaindia.

Hualde, J. I. & J. Ortiz de Urbina 2003. *A grammar of Basque.* Berlin/ Boston: de Gruyter Mouton.

Monforte, S. 2020. Syntactic analyses of discourse particles through the microvariation of Basque *ote*. *Glossa: a journal of general linguistics* 5(1): 126. 1–29. DOI: https://doi.org/10.5334/gjgl.1179

Moutard, N. 1975-1976. Étude Phonologique sur les Dialectes Basques, *Fontes Linguae Vasconum* 19, 5-42; 20, 141-189; 22, 9-54.

de Rijk, R. 2008. *Standard Basque: A progressive grammar* (Current Studies in Linguistics). Cambridge: MIT Press.

Quantifying potential: Non-canonical word order through a variationist perspective

Mercedes Durham
Cardiff University

The principle of accountability (Labov 1972), whereby all instances of a variable and not just those of interest are analysed, is rightly one of the main tenets of variationist sociolinguistics. In practice, however, the full envelope of variation can difficult to circumscribe, especially for morphosyntactic and discourse-pragmatic features. What is the best procedure to follow in cases when every sentence could potentially contain an overt variant, but generally has an unrealised form? How do we handle situations where separate but partly related features are all potential variants? It is of course possible to focus on the functions of a single variant rather than what could be there, but this is not suitable in cases where the ultimate aim is to compare rates of use across social categories and across different varieties.

This paper aims to suggest ways to resolve these issues by examining five types of non-canonical word order (Birner and Ward 1998).

(1) Left dislocation: Chester, he comes over several times a year twice
(2) Right dislocation: cos I remember we used to be able to buy it from Shaws, this hoop.
(3) Focus Fronting: Early sixties it started, yeah.
(4) Inversion: all they ever speak, really, is Welsh.
(5) Clefting: Oh golly, I was in hospital for- I think it was about a couple of weeks, I think, they kept me in.

Non-canonical word order is a good test case as certain variables are perceived to be more frequent in some varieties than others (e.g. right dislocation in the North of England (Durham 2011) and Wales (Penhallurick 2007), fronting in Yiddish English (Prince 1981)), but without a clear method to compare rates across varieties it is difficult to confirm whether these perceptions are accurate, as well as whether findings related to age, sex and other social factors in one variety are unique or shared across varieties.

Previous researchers have dealt with the issues in various ways: some have coded every sentence (but on a restricted data set), some have done their analysis using the numbers of overt tokens per 1000 or 10000 words, yet others have focused on the functions of the variants.

By examining the tokens of all five variables (as well as the unrealised forms) in a half a million word corpus of interviews from Cardiff stratified by age and sex, this paper will compare the various methods and offer suggestions of how best to deal with such types data in language variation and change research, as well as demonstrate which methods are most suited to cross-variety comparisons. It will also discuss why analyses of the interaction between perceived and actual frequency of features can further our understanding of sociolinguistic processes more generally.

References

Birner, Betty J. and Gregory L. Ward. (1998). *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Durham, Mercedes (2011). Right dislocation in Northern England: frequency and use perception meets reality. *English World-Wide* 32 (3): 257 279.

Labov, William (1972). *Sociolinguistic Patterns*. University of Pennsylvania Press, Philadelphia.

Penhallurick, Robert. (2007) English in Wales. In David Britain (ed.) *Language in the British Isles.* pp. 152-170. Cambridge: Cambridge University Press.

Prince, Ellen F. (1981). Topicalization, Focus-Movement, and Yiddish-Movement: A pragmatic differentiation. Berkeley Linguistics Society 7:249-64.

Mapping change in colloquial German in real and apparent time

Stephan Elspaß (Salzburg) & Robert Möller (Liège)

Colloquial vernaculars of German have been the subject of the *Wortatlas der deutschen Umgangssprachen* (WDU) ('Word atlas of colloquial German', cf. Eichhoff 1977–2000) and the long-term project *Atlas zur Alltagssprache* (AdA) ('Atlas of everyday language', cf. Elspaß & Möller 2003ff.). Such colloquial vernaculars can be defined as

> registers and variants in everyday communication, i.e. in the social and functional domains of private life, of spontaneous speech among friends, relatives, acquaintances, or in informal situations among people from the same place who are not necessarily close to each other, e.g. in the local corner shop (cf. Möller & Elspaß 2019: 760)

As both the WDU and the AdA essentially build on similar survey methods (questionnaires, responses assigned to ca. 500 cities and towns in the German-speaking countries), as both use point-symbol maps which account for maximally two variants per locality, and as they cover a time span of almost fifty years, the data from both linguistic atlases can be compared with respect to real time change (cf. for Elspaß 2005 for an early pilot study). Moreover, as the AdA uses data from almost 2,000 up to 20,000 informants per survey round, the sheer amount of data from different age groups can be used for apparent time studies.

In our presentation, we will present methods of mapping real time change, based on data from WDU and AdA (cf. also Leemann, Derungs & Elspaß 2019 for a comparison of data from WDU and Leemann et al. 2018), as well as apparent time change, based on data from AdA. Since the AdA informants are regularly asked how long they have been living at the respective locality and whether their parents were raises there, the AdA data also allow for the investigation of change pertaining to factors such as the informants' degree of mobility and their family roots in the respective localities. Thus, it will also be explored – in a tentative approach – how such factors can be mapped.

References

Eichhoff, Jürgen. 1977 ff. *Wortatlas der deutschen Umgangssprachen.* Bd. I/II: Bern [1977/78]; Bd. III: München et al. [1993]; Bd. IV: Bern/München [2000]: K.G. Saur.

Elspaß, Stephan. 2005. Zum Wandel im Gebrauch regionalsprachlicher Lexik. Ergebnisse einer Neuerhebung. In: *Zeitschrift für Dialektologie und Linguistik*, 1–51.

Elspaß, Stephan & Robert Möller. 2003 ff. *Atlas zur deutschen Alltagssprache (AdA)*. URL: <www.atlas-alltagssprache.de>, accessed 15.05.2022.

Leemann, Adrian, Curdin Derungs & Stephan Elspaß. 2019. Analyzing linguistic variation and change using gamification web apps: the case of German-speaking Europe. In: *PlosOne*. 14(12): e0225399.

Leemann, Adrian, Stephan Elspaß, Robert Möller & Timo Grossenbacher. 2018. *Grüezi, Moin, Servus! Wie wir wo sprechen.* Reinbek bei Hamburg: Rowohlt.

Möller, Robert & Stephan Elspaß. 2019. Die rezente Dynamik im arealsprachlichen Lexikon. In: Herrgen, Joachim & Jürgen Erich Schmidt (eds.): *Language and Space. An International Handbook of Linguistic Variation.* Vol. 4: *German*. In collaboration with Hanna Fischer and Brigitte Ganswindt. Berlin, Boston: De Gruyter Mouton, 758–783.

Dialect Variation as Geography

Sheila Embleton, Dorin Uritescu† and Eric Wheeler

embleton@yorku.ca; eric.wheeler@sympatico.ca

York University, Toronto, CANADA

It is well known that dialect variation correlates with many factors, including the sociolinguistic ones of age, gender and social class (e.g. Chambers 2009). But the obvious (and one might say, traditional) explanation for dialect variation is geography: the further apart people are, the more their languages can vary (for example, cf. Nerbonne, et al. 2007). How true is this?

Theory-driven thinking would suggest that the more difficult it is for two groups to communicate, the more possibility there is for variation between the groups (cf. Wheeler 2007). That suggests that distance between communities, both direct and as a function of travel time or travel distance would correlate with dialect variation. A first test of this idea, using Romanian data (Embleton, Uritescu and Wheeler 2012, 2017), demonstrated a correlation with r-squared of about 0.80. There were some small improvements when travel distance or travel time was substituted for direct distance "as the crow flies". Clearly, in this case, geography is an important but not exclusive correlate of dialect variation.

The question arises of whether or not the scale of distance is important. The Romanian area examined is about 250 km long, whereas the distance from Beijing to Hong Kong is about 2000 km. It is to be expected that there is "dialect variation" between the two Chinese locations, but does it accord with the distance?

At the other extreme, we have data from the Mambila region of Nigeria and Cameroon, where the distances are well under 100 km, but perhaps the cultural and social distances are stronger, and they override the simple geographic factors.

Using consistent approaches to the measures of geography and dialect variation, we report the quantitative correlation in these three situations (Northwest Romania, Chinese, Mambila), and discuss some of the issues that impact the answers. While positive results in any study are always welcome, other results can illustrate the limitations of a valid hypothesis. Yes, geography does correlate with dialect variation, but not always in the same degree, and for good reasons, such as the appropriate choice of linguistic data, varying geographic factors, and over-riding social conditions.

References

Chambers, J.K. 2009. *Sociolinguistic Theory: Linguistic Variation and its Social Significance*, 3rd ed, Oxford UK and Cambridge US: Blackwell. First ed. 1995, second ed. 2003.

Embleton, Sheila, Dorin Uritescu and Eric S. Wheeler. 2012. Effective Comparisons of Geographic and Linguistic Distances. In Gabriel Altman, Peter Grzybek, Sven Naumann, and Relja Vulanović (eds.). *Synergetic Linguistics. Text and Language as Dynamic Systems*. Vienna: Praesens Verlag. pp. 225-232

Embleton, Sheila, Dorin Uritescu and Eric S. Wheeler. 2017. An Expanded Quantitative Study of Linguistic vss Geographic Distance Using Romanian Dialect Data. Proceedings of Qualico 2016. In Lu Wang, Reinhard Köhler, Arjuna Tuzzi (eds.), *Structure, Function and Process in Texts*, Lüdenscheid, Germany: RAM-Verlag, pp. 25-33.

Nerbonne, John, Peter Kleiweg, Franz Manni and Wilbert Heeringa. 2007. Projecting Dialect Distances to Geography: Bootstrap Clustering vs. Noisy Clustering. http://www.let.rug.nl/~nerbonne/papers/Nerbonne-Kleiweg-Manni.pdf

Wheeler, Eric S. 2007. Language Change in a Communication Network. pp 689-698 in Peter Grzybek, Reinhard Köhler (editors). *Exact Methods in the Study of Language and Text*. Berlin and New York: Mouton de Gruyter.

Data Presentation in the REDE SprachGIS: Possibilities and Limits (*Regionalsprache.de*)

Robert Engsterhold, Jeffrey Pheiff, Maria Luisa Krapp

The purpose of this talk is to introduce the online platform *Regionalsprache.de* (cf. REDE) and its browser application the REDE SprachGIS, a geographic information system for linguists. We will discuss the possibilities users have to consult existing data that the REDE SprachGIS makes digitally available and to present their own linguistic data cartographically with point-text, point-symbol, or area maps. In addition to these rather qualitative approaches to mapping, users can also opt for a quantitative approach in visualizing their data in the form of pie-chart maps, bar-chart maps, or choropleth maps.

In addition, we would also like to discuss new possibilities that go hand in hand with digital data presentation. Users are no longer dependent on the representation of data in printed dialect atlases, but they can now "interact" with the underlying data. For example, users can recode the data according to their wishes; they also have more options available to more comfortably analyze the data with frequency charts for example, or by highlighting particularly important variants. Furthermore, they have direct access to additional materials such as recordings, other atlases, survey materials, bibliographic information to benefit their analyses.

In our presentation, we will discuss these issues from a technical perspective, illustrate them with several application examples of new features of the REDE SprachGIS and show how to quickly and easily access central results of the project. In this context, we will demonstrate how users can compile "speaking" linguistic maps like the Digitaler hessischer Sprachatlas (cf. DHSA) and access interactive maps of the vertical spectra between the dialects and the standard language, including sound samples of german speakers in various contexts. Additionally, we will introduce approaches to more complex data visualizations and show how to import, export, and map data in GeoJSON format in the application.

References

DHSA = Digitaler Hessischer Sprachatlas (DHSA). (2019ff.) Von Jürgen Erich Schmidt. Bearbeitet von Dennis Beitel. Herausgegeben von Regionalsprache.de (REDE III), Marburg: Forschungszentrum Deutscher Sprachatlas. URL: <https://www.regionalsprache.de/dhsa/, Stand: 13.05.2022.

REDE = Regionalsprache.de (REDE III). Forschungsplattform zu den modernen Regionalsprachen des Deutschen. (2020ff.) Von Jürgen Erich Schmidt / Joachim Herrgen / Roland Kehrein / Alfred Lameli / Hanna Fischer. Bearbeitet von Robert Engsterhold, Heiko Girnth, Simon Kasper, Juliane Limper, Georg Oberdorfer, Tillmann Pistor, Anna Wolańska. Unter Mitarbeit von Dennis Beitel, Milena Gropp, Maria Luisa Krapp, Vanessa Lang, Salome Lipfert, Jeffrey Pheiff, Bernd Vielsmeier. Studentische Hilfskräfte. Marburg: Forschungszentrum Deutscher Sprachatlas. URL: <https://www.regionalsprache.de>, Stand: 13.05.2022.

”Loud, crass, and punctuated” or “Rich, fast, and educated”?: Heterogeneity in Language Regard

Benjamin Jones, University of Washington
Nicole Chartier, Amazon
Betsy E. Evans, University of Washington

The assumption of structured heterogeneity is a fundamental aspect of the variationist approach to understanding language production (Weinreich, Labov & Herzog 1968); however, this notion is not often extended to research on evaluations of language. That is to say, although attitudes are used as an explanatory factor in linguistic change, variation with regards to attitudes is often masked by single-point measures of a whole community. In recent decades sociolinguistic research paradigms have been reasonably critiquing the lack of studies examining variation in the evaluations of linguistic variables within a given speech community (e.g. Levon, 2018). The aim of this paper is to unmask the inter-group variation of attitudes to further our theoretical constructs of language variation and change.

Non-linguists’ evaluations of language variation, or ‘language regard’ represent beliefs about language but also reveal implicit beliefs that listeners have about speakers. Depending on the method of inquiry into speakers’ language regard, the relationship between an individual’s experience with a geographic location and their perception of linguistic variation in that place can also be explored. Language regard research combined with geographic space is often referred to as Perceptual Dialectology (hereafter PD). PD is well suited for examining the heterogeneity in attitudes to linguistic variation (Preston, 2015). However, socially relevant information, such as socioeconomic status (SES) or regional identity, is infrequently incorporated in as an explanatory variable. We have found only a few studies where such variables are considered in the analysis of PD data (e.g. Demirci & Kleiner (1999) for gender, Montgomery (2012) for location, Evans, Dunbar & Chartier (forthcoming) for travel experience, Fernandes & Ravindranath (2013) for age, Cukor-Avila (2018) for gender and age). In order to demonstrate that respondents' socio-demographic background can play an important role in understanding attitudes expressed in PD research, 398 PD maps were collected online using an integrated GIS application (FLOM, n.d.) to explore New Englanders’ perceptions of linguistic variation in that region (Chartier & Jones, 2018). Detailed demographic information, including age, socioeconomic status (SES), and a regionality index (Chambers 2000:10-13; Chambers and Heisler 1999: 40-46) was solicited along with respondents’ computer-assisted drawn maps of linguistic variation in New England. Results from Chartier & Jones (2018) showed that respondents from New England marked similar perceived dialect regions on their maps, yet evaluated those same regions differently. This follow-up study examines how age, SES, and regionality demonstrate the existence of structured heterogeneity in PD evaluations of speech in the same way that it exists in speech production. These results suggest that the inclusion of sociodemographic factors in PD analysis is necessary for understanding the role of evaluation in linguistic variation and production. This has important implications for variationist theory with regards to how we understand language regard and change in any geographic region.

References

Chambers, J. K. 2000. Region and language variation. *English World-Wide* 21: 1-31.

Chambers, J. K. and Troy Heisler. 1999. Dialect topography of Quebec City English. *Canadian Journal of Linguistics* 44: 23-48.

Chartier, N. & Jones. B. 2018. "Six Views of New England: Mapping Perceptions of New England Speech." Poster presented at New Ways of Analyzing Variation 47 (NWAV-47), New York University, October 19.

Cukor-Avila, P. 2018. A Variationist Approach to Studies of Language Regard. In B. Evans, E. Benson, & J. Stanford (Eds.), *Language Regard: Methods, Variation and Change* (pp. 31-61). Cambridge: Cambridge University Press.

Evans, B., Dunbar, M. & Chartier, N. (forthcoming). Cardiffians' perceptions of English in the UK. *Journal of Linguistic Geography.*

FLOM (nd.) Folk Linguistic Online Mapping Application. Retrieved from http://depts.washington.edu/flom/

Levon, E. (2018). The Systematicity of Emergent Meaning. Plenary talk: New Ways of Analyzing Variation (NWAV) 47, New York, NY.

Montgomery, C. (2012). The effect of proximity in perceptual dialectology. *Journal of Sociolinguistics* 16(5): 638–668. doi: http://dx.doi.org/10.1111/josl.12003

Preston, D. R. (2015). Does language regard vary. In *Responses to language varieties. Variability, processes and outcomes*, 3-36.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical Foundations for a Theory of Language Change. In W., Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics: A symposium* (pp.97-195). Austin: University of Texas Press.

Panel: German in Austria

Speech repertoires and varietal spectra - Eliciting phonological variation on the dialect-standard-axis

Johanna Fanta-Jende (University of Vienna)

Researchers in variationist linguistics are able to draw on a large number of established survey methods for the analysis of variation in intra-speaker and inter-speaker language use. Different survey methods yield very different types of data with respect to naturalness, comparability and their practical application in research. The exploration of the social-vertical dimension in contrast to the classical areal-horizontal dimension presents certain challenges with respect to established methodologies.

There are several thorough studies focusing on the 'architecture' of the dialect-standard-axis for specific regions in Germany (cf. Lenz 2003, Lameli 2004 and Kehrein 2012), comparable endeavours for the Austrian context still represent a major research desideratum. Austria, however, may be regarded as the "ideal sociolinguistic research laboratory" (Lenz 2018, 269) due to its high dynamics on the dialect-standard axis. This presentation would like to contribute to the description of the structure and dynamics of the areal-horizontal and in particular the 'social-vertical' language dimension in Austria.

The empirical input consists of language data from 'autochthonous' speakers of various sociodemographic backgrounds in rural areas of Bavarian Austria, representing different dialect regions. The data is collected in various survey settings: an interview led by a foreign academic, an unguided conversation among friends, two translation tasks and reading-aloud tasks.

Based on various selected phonological phenomena, the author likes to demonstrate how a multi-method approach enables accessing different parts of individuals' language repertoires. Important research questions revolve around the following aspects: How do active members of the Austrian speech community "move" along their individual "spectrum of linguistic possibilities (Macha 1991). How can different parts of the vertical language spectrum be captured by a selection of various 'natural' and 'standardized survey settings? In a second step, the talk will discuss the relationship of intra- and inter-individual variation: How can the individual repertoires be used to derive assumptions about the overall vertical spectra of the selected dialect regions? And how can the applied methodological insights be taken for further general research into the structure and dynamics of the complex language spectrum?

References

Kehrein, Roland. 2012. *Regionalsprachliche Spektren im Raum. Zur linguistischen Struktur der Vertikale* (Zeitschrift für Dialektologie und Linguistik, Beihefte 152). Stuttgart: Steiner.

Lameli, Alfred (2004): Standard und Substandard. Regionalismen im diachronen Längsschnitt. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik: Beihefte, 128).

Lenz, Alexandra N. 2003. *Struktur und Dynamik des Substandards. Eine Studie zum Westmitteldeutschen (Wittlich/Eifel)* (Zeitschrift für Dialektologie und Linguistik, Beihefte 125). Stuttgart: Steiner.

Lenz, Alexandra N. 2018. The Special Research Programme ‘German in Austria. Variation – Contact – Perception’. In Ammon, Ulrich & Marcella Costa (eds.). *Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language Choice in Tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l’Europe* (Sociolinguistica 32). Berlin & Boston: de Gruyter, 269–277.

Macha, Jürgen. 1991. *Der flexible Sprecher. Untersuchungen zu Sprache und Sprachbewußtsein rheinischer Handwerksmeister*. Köln: Böhlau.

Doing Sociophonetics with Linguistic Atlas Project Data

Josef Fruehwald, University of Kentucky

Recent trends sociophonetic research have seen us expanding the size of our data sets through the use of semiautomatic techniques like forced alignment and automatic vowel formant analysis. In some cases, canny use of archival recordings has allowed for dueling time depths to analyses: “100” or “130” years of sound change (Labov et al, 2013; Hay et al, 2015). The use of Linguistic Atlas data has the potential to unlock similar time depths for sociophonetic research across North America, but LAP data, like any other historical data, does pose a certain technological hurdle for researchers wishing to unlock this potential. The currently most common tools to use for these semiautomated techniques (the Monteal Forced Aligner (McAullife et al 2018) and the FAVE suite (Rosenfelder et al 2015) currently require full time-aligned orthographic transcripts as input. For this presentation, we carry out a feasibility study of utilizing fully automated speech-to-text systems on LANCS data. We will explore how currently available systems such as DARLA (Reddy & Stanford, 2014), CLOx (Wassink et al 2018) and wav2vec-U (Baevski et al 2021) perform on archival data. The word error rate will be calculated to evaluate the accuracy of these automated transcriptions, and the transcripts will be corrected for forced alignment with MFA and vowel formant analysis with FAVE-extract. Any identifiable errors or shortcomings of these automated systems, especially as they relate to the age or original media of recording will be tracked and documented, again to provide best practice recommendations.

The linguistic expression of persuasion across varieties of English

Robert Fuchs, University of Hamburg

There is a substantial amount of evidence regarding register variation in a number of languages and their varieties that is based on Biber's Multidimensional Model (1988; 1995). This line of research has also been extended to postcolonial varieties of English, with some studies comparing a large number of varieties (Kruger & Van Rooy, 2016; Xiao, 2009) and others focussing on particular varieties, such as East African English and Australian English (Kruger & Smith, 2018; Van Rooy et al., 2010).

Within this framework, one dimension is the overt expression of persuasion, and previous research has demonstrated, for example, that Indian English employs relatively few markers of overt persuasion in formal registers, while Hong Kong English employs comparatively many (factor 4 in Xiao, 2009).

However, previous research has focussed exclusively on a quantitative comparison of register dimensions (e.g. more/less overt expression of persuasion), not on qualitative differences in how these register dimensions are expressed in different varieties. Thus, the present study asks how persuasion is linguistically expressed across varieties of English and how such differences can be explained.

In order to answer this question, data on four L1 and six L2 varieties of English drawn from the International Corpus of English (Greenbaum, 1991) will be investigated with the Multidimensional Analysis Tagger (Nini, 2015). Unlike in previous research, separate analyses will be conducted for all ten varieties in order to reveal potential differences in how persuasion is overtly expressed in these varieties. Preliminary results indicate (1) greater differences between varieties in spoken than in written language and (2) that speakers and writers of L1 varieties use more similar means of overt persuasion than speakers and writers of L2 varieties, suggesting a certain degree of indigenisation (Schneider, 2007) of linguistic markers of overt persuasion in these varieties.

References

Biber, Douglas (1988) Variation across Speech and Writing. Cambridge: Cambridge University Press.
Biber, Douglas (1995) Dimensions of Register Variation: A Cross-linguistic Comparison. Cambridge: Cambridge University Press.
Greenbaum, S. (1991). ICE: The international corpus of English. *English Today*, *7*(4), 3-7.
Kruger, H., & Smith, A. (2018). Colloquialization versus densification in Australian English: a multidimensional analysis of the Australian Diachronic Hansard Corpus (ADHC). *Australian Journal of Linguistics*, *38*(3), 293-328.
Kruger, H., & Van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, *37*(1), 26-57.
Kruger, H., & Van Rooy, B. (2018). Register variation in written contact varieties of English: A multidimensional analysis. *English World-Wide*, *39*(2), 214-242.
Nini, A. (2015). Multidimensional analysis tagger (version 1.3). *URL http://sites. google. com/site/multidimensionaltagger*.
Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.
Van Rooy, B., Terblanche, L., Haase, C., & Schmied, J. (2010). Register differentiation in East African English: A multidimensional study. *English World-Wide*, *31*(3), 311-349.

Xiao, R. (2009). Multidimensional analysis and the study of world Englishes. *World Englishes*, *28*(4), 421-450.

Um, uh, and variation in American English

Matt Hunt Gardner
University of Oxford
matt.gardner@ling-phil.ox.ac.uk
&
Benedikt Szmrecsanyi
KU Leuven
benedikt.szmrecsanyi@kuleuven.be

Spontaneous speech takes place in real time, which means that speakers have to continually plan utterances both prior to and while speaking. Unsurprisingly, this is a demanding task. As a consequence, disfluencies are non-trivial in spontaneous speech; across studies about 6% of uttered words are disfluent [1][2]. Goldman-Eisler [3] shows that almost half of people's speaking time is made up of pausing and overt disfluencies like *um* and *uh*. There is overwhelming evidence from the psycholinguistic litterature that increased cognitive load coincides with increased rates of overt disfluencies (including filled pauses) and extended speech planning time (unfilled pauses). These two measures are in turn used as metrics for cognitive load. Exerting explicit executive control during production increases cognitive load, therefore choice-making is linked to disfluency and speech planning time. Grammatical planning involves both automatic processes and explicit control, and both must include probabilistic constraint-based variant selection. Variant substitution triggered by attention paid to speech or audience accommodation during the self-monitoring process is due to explicit executive control, adding to cognitive load, and potentially precipitating overt disfluencies or extended planning time. This leads to the hypothesis that grammatical alternations that are subject to style-shifting or audience design (i.e., sociolinguistic markers [4]) will coincide with a higher frequency of overt disfluencies and require more planning time compared to grammatical alternations that are not subject to style-shifting or audience design (i.e., sociolinguistic indicators [4]). Further, the greater the number of variable contexts subject to executive control within an utterance, the greater the number of overt disfluencies or the longer the amount of required planning time is expected.

No research has been conducted thus far to investigate whether specifically *grammatical* choice, i.e., the presence of grammatical variation, may cause increased disfluency in spontaneous speech. Against this backdrop we ask the question: does sociolinguistic variation make planning speech harder?

To answer this question we draw on the well-studied Switchboard Corpus of American English [5], which contains about 240 hours of recorded conversations between 542 Americans from all across the country. For each conversation we isolate variable contexts for 20 different morphosyntactic variables (N=46,867) and investigate correlations with both filled pauses (N=42,695) and speech planning time.

Surprisingly, we find that a greater number of variable contexts per 100 words coincides with fewer disfluent phenomena, suggesting variation instead facilitates speech production. This relationship is consistent across multiple varieties, and although factors like sex and age do have an effect on how many *um's* and *uh's* occur, the relationship between them and variable contexts does not. Our findings bolster arguments that variation is deeply embedded within the speech production process and is an integral component of the language faculty.

[1] Jean E. Fox Tree. "The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech". In: *Journal of Memory and Language* 34.6 (1995), pp. 709–738.

[2] Heather Bortfeld et al. “Disfluency Rates in Conversation. Effects of Age, Relationship, Topic, Role, and Gender”. In: *Language and Speech* 44.2 (2001), pp. 123–147.

[3] Frieda Goldman-Eisler. *Psycholinguistics. Experiments in Spontaneous Speech*. New York: Academic Press, 1968.

[4] William Labov. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press, 1972.

[5] J.J. Godfrey, E.C. Holliman, and J. McDaniel. “SWITCHBOARD: telephone speech corpus for research and development”. In: *[Proceedings] ICASSP-93. 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing* (1992).

Exploring language use and linguistic attitudes over a decade in later life: a generational “lect”? (Annette Gerstenberg)

While the "pressures of the standard language market" are identified as a driving force of linguistic change in adulthood (Wagner, 2012:379), it is assumed that "the pressures to conform to societal norms may weaken once more" at the end of a professional career (Cheshire, 2005:1556). The hypothesized relaxation after retirement has been observed in French by Gadet (2003:55).

At the very heart of research across the lifespan lies the question of the standard itself: is the the codified standard of the grammars the “standard”, or is it the current linguistic usage norm? This question will be explored on the basis of a continental French panel analysis, a language in which a diglossic linguistic situation has been postulated, such that a "congealed" high variety artificially perpetuating the historical linguistic norm contrasts with the common "demotic" low variety (Massot & Rowlett 2013).

Against this complex background, the question of what characterizes older speakers and their "generational lect" is highly illuminating when it comes to French. This paper explores the use of a core variable in French sociolinguistics: *ne* deletion (Armstrong & Smith 2002), specifically amongst older speakers, in comparison with communal *ne* deletion change. Negation reflects the normative orientation to 20th century hexagonal French as it was consistently taught at school, especially until the 1960s, and which continued to be vital in French society as a marker of fr. *bon usage*, ‘good usage’.

The 20 speakers included in the analysis consist of 10 heterosexual couples recorded in 34 interviews, at two different points in time (2005/2015; aged 58/68 to 86/96). Interviews touch on topics in the individuals’ biographies, their current situations, activities, and thoughts on language.

The effect of subject type (noun vs. pronoun), embedding and lemma are explored with special attention to high-frequency, potentially pragmaticized, constructions (Hansen & Malderez 2003; Bybee 2002). In addition, metrics of lexical statistics are used to control for possible change in linguistic register between the first and second interview. Finally, metalinguistic statements are included in the analysis.

The results allow us to situate speakers as representing a generational group within the speech community of contemporary French, and to trace individual trajectories in the post-retirement phase. For this purpose, speaker biographies and life trajectories between the first and second interviews provide further qualitative information, allowing for a nuanced understanding of differentiated use of linguistic means in old age.

The results show that language use and internal grammars continue to develop even with speakers in their 60s, 70s and 80s, as most speakers continue to progressively align themselves with the current standard of spoken language (Buchstaller & Wagner 2017). However, the learned norm of the 20th century remains formative in the 21st century. Furthermore, complementary qualitative analysis shows that this norm is an important part of the speakers' perception of a generational lect.

References

Armstrong, Nigel & Alan Smith. 2002. The Influence of Linguistic and Social Factors on the Recent Decline of French 'ne'. *Journal of French Language Studies* 12. 23–41.

Buchstaller, Isabelle & Suzanne E. Wagner. 2017. Introduction: Using Panel Data in the Sociolinguistic Study of Variation and Change. In Suzanne Wagner & Isabelle Buchstaller (eds.), *Panel Studies of Variation and Change* (Routledge studies in language and intercultural communication), 1–18. New York: Routledge.

Bybee, Joan L. 2002. Main Clauses are Innovative, Subordinate Clauses are Conservative: Consequences for the Nature of Constructions. In Joan L. Bybee & Michael Noonan (eds.), *Complex Sentences in Grammar and Discourse. Essays in Honor of Sandra A. Thompson*, 1–17. Amsterdam: Benjamins.

Cheshire, Jenny. 2005. Age- and Generation-Specific Use of Language. In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier & Peter Trudgill (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*. HSK 3.2, 2nd edn. (Handbücher zur Sprach- und Kommunikationswissenschaft 3.2), 1552–1563. Berlin, New York: de Gruyter.

Gadet, Françoise. 2003. *La variation sociale en français.* Paris: Ophrys.

Hansen, Anita B. & Isabelle Malderez. 2003. Le 'ne' de négation en région parisienne: une étude en temps réel. *Langage et société* 107(4). 5–30.

Massot Benjamin & Rowlett Paul. 2013. Le débat sur la diglossie en France: aspects scientifiques et politiques. *Journal of French Language Studies* 23(1). 1–16.

Wagner, Suzanne E. 2012. Age grading in Sociolinguistic Theory. *Language and Linguistics Compass* 6.

Studying enregisterment in the speech community: the 'London talks' interviews

Johanna Gerwin (Queen Mary University London)

The 'London talks' interviews are part of an ongoing research project studying the enregisterment of London dialects by members of the local speech community. 'Enregisterment' describes how a linguistic repertoire becomes noticeable and differentiable from the rest of language to speakers (Agha 2003; Silverstein 2003). A 'register' in this sense is 'a way of speaking linked with a social situation, a set of linguistic forms linked with and constitutive of a context' (Johnstone 2017:17). These links are created, maintained and (re-)negotiated by each member of a speech community through a variety of observable metadiscursive activities, such as the invention and use of register names, phonetically motivated dialect spellings, or the creation of stereotypical dialect speakers ('characterological figures' representing the social values of the register) (Agha 2007:151).

The fieldwork collects narrative data on aspects of meta-discourse to establish how London speakers experience and conceptualize their own and fellow Londoners' ways of speaking, labelled, e.g., *Cockney*, *Estuary English* or *Multicultural London English/MLE*. All are scientifically described London-based varieties (Cheshire, Kerswill, Fox, & Torgersen 2011; Altendorf 2012; Fox 2012) as well as socially and culturally enregistered ways of speaking (Kerswill 2014). In the research interviews, informants are encouraged to elaborate on the linguistic shibboleths they consider typical of their own and other speakers' dialects, the social and place indexes they associate with them, and who they consider as prototypical dialect speakers.

The interviews are currently being conducted all over London amongst a broad sample of the local population and are publicised as the project 'London talks' (for more information and a blog featuring some of the informants visit www.londontalksresearch.co.uk). The conversations consist of a part involving elicited commentary on 5 audio samples of different London voices, and another consisting of questions on the informants' own dialect and individual dialect biography.

Based on a sample of 20 participants, it can be shown that London varieties are enregistered differently for different members of the speech community, but that a community consensus can be established (cf. also Johnstone, Andrus, & Danielson 2006). For example, of the above labels, only 'Cockney' is widely known and linked to a clear place index, whereas the others are generally referred to by lay coinages, such as 'slang', 'patois', or 'South London' for MLE or 'south-eastern bland' for 'Estuary English' and localised all over the city. Furthermore, the social indexes and evaluations of these registers are highly variable and strongly dependent on the informants' own dialect socialisation.

The 'London talks' fieldwork forms part of a wider study on the historical enregisterment of London English, which diachronically analyses meta-discourse about London varieties from the beginning of the 19th century onwards and in a variety of published and broadcast media and cultural products. Whereas (mass) media meta-discourse provides insight into the culturally and socially shaped registers from a 'sender' perspective, the interviews focus on dialect conceptualisation by individuals in the speech community and thus approach the questions of enregisterment from the 'receiver' perspective.

References

Agha, Asif (2007). *Language and Social Relations*. Cambridge: Cambridge University Press.

——— (2003). The social life of cultural value. *Language and Communication* 23:231–73.

Altendorf, Ulrike (2012). Varieties of English: Estuary English. In Alexander Bergs & Laurel Brinton (eds.), *Historical Linguistics of English (HSK 34.1)*, 1913–27. Berlin/New York: De Gruyter.

Cheshire, Jenny; Kerswill, Paul; Fox, Sue; & Torgersen, Eivind (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics* 15(2):151–96.

Fox, Sue (2012). Varieties of English: Cockney. In Alexander Bergs & Laurel Brinton (eds.), *Historical Linguistics of English (HSK 34.1)*, 2013–31. Berlin/New York: Mouton de Gruyter.

Johnstone, Barbara (2017). Enregistering Dialect. In Lieselotte Anderwald & Jarich Hoekstra (eds.), *Enregisterment - Zur sozialen Bedeutung sprachlicher Variation*, 15–28. Frankfurt et al.: Peter Lang.

Johnstone, Barbara; Andrus, Jennifer; & Danielson, Andrew E. (2006). Mobility, Indexicality, and the Enregisterment of 'Pittsburghese'. *Journal of English Linguistics* 34(2):77–104.

Kerswill, Paul (2014). The objectification of 'Jafaican'. The discoursal embedding of Multicultural London English in the British media. In Jannis Androutsopoulos (ed.), *Mediatization and Sociolinguistic Change*, 427–55. Berlin/Boston: DeGruyter.

Silverstein, Michael (2003). Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23:193–229.

Socio-demographic trajectory, vowel normalization, and the marriage of auditory and acoustic approaches in assessing lifespan change

James Grama, Lea Bauernfeind, Carina Ahrens, Mirjam Eiswirth and Isabelle Buchstaller
Sociolinguistics Lab, University of Duisburg-Essen

Tyneside English has a long history of sociolinguistic investigation (Milroy *et al.* 1994; Watt 1998; Beal *et al.* 2012; Warburton 2020). These works demonstrate that the vowel system in the North East is undergoing a widespread process of levelling towards supralocal forms (e.g., Haddican *et al.* 2013). At the same time, some speakers' maintenance of localised forms has been interpreted as a strategic choice in the performance of locally relevant identities that tend to be conditioned by factors such as gender and age (Watt 2002; Buchstaller *et al.* 2017). Ongoing research by the TUULS project (Llamas *et al.* 2017) has been investigating the extent to which geographical mobility across the lifespan in older speakers impacts their use of localized forms. However, with the exception of Buchstaller *et al.* (2017), previous findings have largely been based on apparent-time analyses or induced from comparisons with legacy data. This paper is the first to explore the extent to which individual speakers follow the trajectories in the Northern vowel system observed in the community at large across their own lifespan.

We report on a novel panel corpus which covers twelve speakers from Tyneside who were re-recorded twice or three times, starting at early adulthood, adulthood immediately pre-retirement, and post-retirement. Our analysis reports on approximately 4,000 tokens of two locally salient vowels in the North East—FACE and GOAT—across these time slices (1971, 2013, 2019). Our focus is twofold: first, we attack methodological questions of comparability when dealing with stark differences in recording quality (1970s v. 2010s), as well as the acoustic ramifications of (not) undertaking vowel normalization of the same speakers measured at different points across their lifespans. Second, we demonstrate that socio-demographic trajectories play a significant role in the articulation of lifespan changes. While working-class speakers show an increase in the proportion in localized vowel realizations over their lifespans, middle-class speakers show evidence of retrenchment towards the standard, followed by a tail back towards localized forms post-retirement (Downes 1984, Buchstaller 2006). Further evidence suggests that these changes in proportion are paralleled in F1/F2 space, with variants more in line with the standard emerging at T2, then abating in T3. Overall, results across the lifespan for middle-class speakers suggest that phonetic changes pattern alongside changes in proportion for at least one speaker. Furthermore, our analysis provides empirical evidence for the hypothesized U-shaped pattern, a bedrock of sociolinguistic theorizing (Downes 1984; Buchstaller et al. 2017).

References

Beal, Joan, Lourdes Burbano-Elizondo and Carmen Llamas. 2012. *Urban North-Eastern English: Tyneside to Teesside*. Edinburgh, U.K.: Edinburgh University Press.

Buchstaller, Isabelle, Anne Krause, Anja Auer, and Stefanie Otte. 2017. Levelling across the life-span?: Tracing the FACE vowel in panel data from the North East of England. *Journal of Sociolinguistics*, 21(1): 3–33.

Downes, William (1998). Language and Society. 2nd edn. Cambridge: Cambridge University Press.

Fromont, Robert and Jennifer Hay. 2012. LaBB-CAT: An annotation store. *Proceedings of the Australasian Language Technology Workshop*, 113–117.

Haddican, Bill, Paul Foulkes, Vincent Hughes and Hazel Richards. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change* 25: 371–403.

Llamas, Carmen, Dominic Watt, Peter French, Almut Braun, & Duncan Robertson. 2017. Routinised mobility and vowel change at the individual and community level in the North East of England. Paper presented at the *4th International Workshop on Sound Change (WSC4)*, University of Edinburgh.

Milroy, James, Lesley Milroy, Sue Hartley and David Walshaw. 1994. Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change* 6: 327–357.

Watt, Dominic. 1998. Variation and change in the vowel system of Tyneside English. Unpublished PhD dissertation. Newcastle, U.K.: Newcastle University.

Watt, Dominic. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics* 6: 44–63.

Warburton, Jasmine. 2020. Phonetic variation in the North East of England: On intra-regional differences in /u/, /ə/ and l/ realizations. Newcastle, U.K.: Newcastle University.

Cross-Linguistic Homogeneity and Heterogeneity in Pronominal Expression

Gregory R. Guy, Rafael Orozco & Danny Erker

Variationist sociolinguistics has long recognized that both features of linguistic contexts and speech communities' social characteristics condition variant choice. Recent work (Labov 2010; Tamminga et al. 2016) adds *cognitive constraints* as a third category in this typology. Much work in the field (Carvalho et al. 2015) assumes —implicitly or explicitly— that certain constraints have consistent effects across speech communities. The present paper probes that assumption using a comparative approach to constraints on the alternation between null and overt pronominal subjects in Spanish and Portuguese. We analyzed 44,605 tokens from sociolinguistic interviews with 208 speakers in 7 locales: Barranquilla, Boston, Lisbon, Medellín, New York, São Paulo, and Xalapa. We hypothesize that *cognitive* constraints, based on universal properties of the human mind, should have the most consistent effects across languages, while *linguistic* constraints will be affected by each language's structural idiosyncrasies, and *social* constraints will reflect particular speech communities' properties, practices and ideologies.

Our findings support this hypothesis. *Priming* —reflecting that property of cognition/neurobiology underlying human sensitivity to repetition— has a consistent cross-linguistic and cross-community effect: the realization as null or overt in one clause favors the same realization in successive clauses. *Reference chains* are similarly systematic: new referents favor overt pronominal subjects in all our datasets, a consequence, we argue, of the common cognitive task of tracking referents across discourse. Linguistic constraints on pronominal expression are more diverse: Spanish-speaking communities have consistent effects of *reflexivity* (reflexive verbs have lower pronoun rates), *person/number* (plural subjects have lower pronoun rates), and *tense/mood* (imperfectives favor overt pronouns while preterites disfavor them). These constraints diverge in the substantially different Portuguese pronominal system, where a new pronoun *a gente,* is replacing first-person plural *nós*, reflexives are rare, and 2nd sg *tu* is non-existent in São Paulo. Portugal and Brazil also differ in the treatment of second-person singular polite *você*: it favors overt forms in Brazil but disfavors them in Portugal. Notwithstanding, a functional analysis —a presumably cognitive requirement that more overt pronouns are needed when verbal inflections are less distinctive — is partially supported by the results from all speech communities.

Social distribution also varies considerably across speech communities. One prominent effect is *dialect*: Caribbean speakers use the most overt pronouns among the Spanish dialects, and Brazil has over twice the pronoun rate of Portugal (68% vs 32%). *Gender* differences although prominent in Lisbon, São Paulo, Barranquilla and New York —with women favoring overt subjects— are absent in the other communities. *Age grading* is found in Portuguese, and the Spanish monolingual corpora, where younger speakers favor lower pronoun rates. Our findings suggest that similar usage patterns will arise across languages and communities for a given linguistic variable insofar as they are linked to general aspects of human cognition. Crosslinguistic and cross-community heterogeneity, in contrast, is expected to emerge in relation to specific structural properties of individual language varieties. Similarly, the social dimensions routinely implicated in synchronic linguistic variation and in historical change are likely to be vectors of cross-community heterogeneity, reflecting local interactional roles, norms, and dynamics. (500 words)

References

Carvalho, Ana M., Rafael Orozco, & Naomi Lapidus Shin. 2015. *Subject Pronoun Expression in Spanish: A Cross-Dialectal Perspective*. Washington: Georgetown University Press.

Labov, William. 2010. *Principles of linguistic change, volume 3: Cognitive and cultural factors* (Vol. 36). John Wiley & Sons.

Tamminga, M., MacKenzie, Laurel, & Embick, David. 2016. The dynamics of variation in individuals. *Linguistic variation*, *16*(2), 300–336. https://doi.org/10.1075/lv.16.2.06tam

Structural and social variable use of the negators *la?*, *laa*, and *wa-laa* in Syrian Arabic

Rania Habib
Syracuse University

Abstract

This study investigates the variable structural and social distribution of the negators *la?*, *laa*, and *walaa* in Syrian Arabic (SA), answering the following research questions:

1. What are the functions and/or structural contexts that trigger the use of the investigated negators?
2. How frequent each negator with each one of these functions and/or contexts?
3. Is there correlation between each negator and certain functions and/or structural contexts?
4. Are there differences between children and adults in their implementation of these functions and/or structural contexts regarding each negator?
5. Are there sex and/or age differences in the use of these negators?

The study analyzes 1972 tokens gleaned from the naturally occurring speech of 50 children and 22 adults with equal gender distribution in each generation and equal numbers of children and males and females in each of four age groups 6-8, 9-11, 12-14, and 15-18. The three negators are treated separately due to differences in functions and/or contexts and lack of inter-dependency. Negators are coded according to their function and/or contexts for each speaker. Statistical analyses are performed to determine the effects of sex and age and any correlations between frequency of negators and functions and/or contexts.

la? and *laa* share seven functions: answering yes/no/tag questions, negating propositions, contrastive negation, agreement, emphatic negation, repair, and interjection. However, absolute prohibition is performed only with *la?*. *laa* can express prohibition; negate verbs, nouns, PPs, adverbs, demonstratives, adjectives, pseudo-verbs; form the (impersonal) negative copula *laanaa* and *laaħee* (usually formed with *maa*); and in classical expressions. *walaa*, known as negative coordinator, can function as a stand-alone negator, performing prohibition, negating nouns, verbs, PPs, pseudo-verbs, passive participles, demonstratives, adjectives, pronouns, quantifiers, active participles, and forming (impersonal) negative copulas. *la?* negates most frequently propositions and yes/no/tag questions. It occurs as *laa?* almost categorically among children and as *la?* almost categorically among adults. *laa* performs most frequently prohibition, followed by negating yes/no/tag questions and propositions. *walaa* negates most frequently nouns, verbs, and PPs.

Age emerged as statistically significant among children regarding *laa?*; use decreases as age decreases. Gender emerged as statistically significant among adults; men use less *la?* than women. In the combined data, age and gender merged as statistically significant; males use less *la?* than females, and children use more than adults. Age emerged as statistically significant regarding *laa* and *walaa*; children use less *laa* and *walaa* than adults.

These findings partially reflect the complexity and great variability of the negative system in SA both linguistically and socially. They show that certain functions/contexts favor certain negators more frequently than other functions/contexts. The differences in functions/contexts provide evidence that *laa* and *la?* are different negators in SA, whereas both would be considered *laa* in Modern Standard Arabic. *walaa* can function as a separate negator, not only as negative coordinator. The findings also show generational differences regarding *la?*, *laa* and *walaa* and gender difference regarding *la?*. Although this study does not deal with phonological variation, a shift towards the use of the elongated *laa?* among children is observed.

Mapping regional variation of speech rate using automatic measurements of amplitude envelope peaks

We investigate the region-specific interweaving of speech rate and phonetic reduction on the basis of standard-intended reading pronunciation in the German-speaking countries of central Europe (Hahn & Siebenhaar, 2019). The data base for this research are 1652 recordings of men and women in two age groups (17-20 and 50-60 years respectively) available (Kleiner, 2015), all of which have already been automatically aligned with WebMAUS (Kisler et al., 2017). For more reliable evaluations of sound durations, the segmentations at sound level were also corrected manually for approximately 600 of these recordings (the group of young men).
On the basis of these segmentations and annotations specific reduction phenomena are investigated, correlated to speech rate and displayed in geolinguistic space. For the speech rate various measurements are used such as articulation and speaking rates in phones or syllables/second.
Due to the amount of workload involved in manual correction, not all data can be included in the study, which is why it is not yet possible to compare speech rates between the sexes and age groups in terms of geolinguistic space.
We therefore examine in a first step whether automatic procedures such as the calculation of the peaks of the amplitude envelopes (He & Dellwo, 2016, 2017) can provide sufficiently reliable data to be able to include the missing recordings in the analysis and to rise the local density of the data points. Envelopes were calculated by low-pass filtering (cut-off: 10Hz) a full-wave rectified speech signal. A peak point was defined as the sample that is preceded and followed by lower amplitude samples and that has a minimum amplitude of one tenth of the maximum amplitude sample in the signal. Inter-peak intervals were defined as the intervals between two peak points starting and ending with the first and last peak in an utterance respectively (c.f. fig 1). First tests are promising and achieve high correlations between the manually segmented data and the automatically measured data.
In a second step, the measured speech rate values are transferred to geolinguistic maps using arcGIS (ESRI, 2015) and REDE SprachGIS (Schmidt et al., 2008ff.) to test and compare regional patterns of the groups in question.
The aim of these comparisons is to show that the regional patterns already worked out for young men do not appear to be exclusive to this group of speakers, but also show similar patterns for women and older speakers. In this way it can be substantiated that regionality is an important factor for the variation of speaking speed. Moreover, it can be justifiably assumed that the patterns of reduction phenomena and the temporal microstructures observed for the group of the young speakers can be transferred to the other speaker groups.

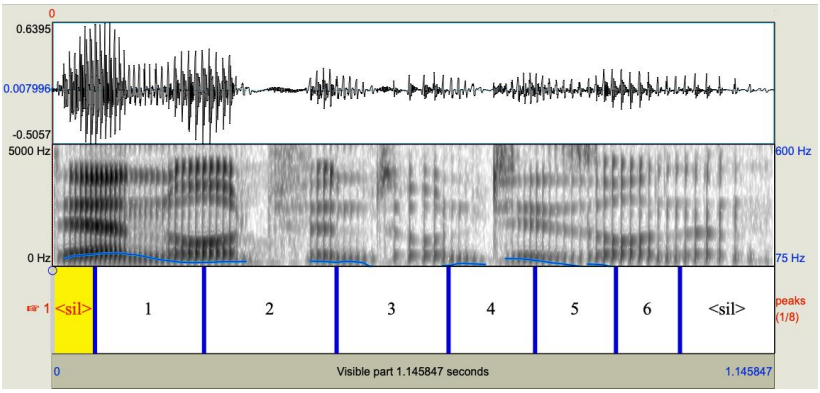

Figure 1: Waveform and spectrogram of the utterance “Der Nordwind und die Sonne” produced by a male speaker with peak-to-peak text TextGrid containing 7 peaks (boundaries), hence 6 peak-to-peak intervals.

References

ESRI. (2015). ArcGIS Desktop (Version 10.3). ESRI (Environmental Systems Research). Redlands, CA.

Hahn, M., & Siebenhaar, B. (2019). Spatial Variation of Articulation Rate and Phonetic Reduction in Standard-Intended German. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019 (pp. 2695–2699). Australasian Speech Science and Technology Association Inc.

He, L., & Dellwo, V. (2016). A Praat-Based Algorithm to Extract the Amplitude Envelope and Temporal Fine Structure Using the Hilbert Transform. In Interspeech 2016 (pp. 530–534). ISCA.

He, L., & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. The Journal of the Acoustical Society of America, 141(5), 488–494.

Kisler, T., Reichel, U., & Schiel, F. (2017). Multilingual processing of speech via web services. Computer Speech & Language, 45, 326–347.

Kleiner, S. (2015). 20. „Deutsch heute“ und der Atlas zur Aussprache des deutschen Gebrauchsstandards. In R. Kehrein, A. Lameli, & S. Rabanus (Eds.), Regionale Variation des Deutschen: Projekte und Perspektiven (pp. 489–518). De Gruyter.

Schmidt, J. E., Herrgen, J. & Kehrein, R. (2008ff.). Regionalsprache.de (REDE). Forschungsplattform zu den modernen Regionalsprachen des Deutschen. Teil 6: REDE SprachGIS – Das forschungszentrierte sprachgeographische Informationssystem von Regionalsprache.de. Forschungszentrum Deutscher Sprachatlas.

Classification of Kansai dialects using phonetic distance

Some Japanese dialectologists have attempted to classify Japanese dialects using vocabulary, suffix, vowel shift, and proportion of standard Japanese use (e.g., Inoue, 2001; Kindaichi, 1964; Tojo, 1953). They descriptively revealed some variations in the dialects and classified them. In order to further understand how the dialects vary, we implemented a phonetics-based metric and examined how Kansai dialects spoken by old generations are classified using a density-based clustering method called HDBSCAN (Campello et al., 2013)

We implemented *aline* distance (Downey et al., 2008) in order to measure phonetic distances between words in standard Japanese and Kansai dialects. The advantage of aline algorithm is that it considers weight of phonetic features (e.g., dental, palatal, nasal, front, and central) when it compares pairs of words, and generates distance scores. Downey et al. (2008) demonstrated to what extent words in languages spoken in eastern Indonesia (i.e., Rindi and Sumba) phonetically diverged from their cognates (proto-Austronesian words). Here, we examined whether aline distance becomes a metric for classifying dialectal variation.

We designed a questionnaire containing 76 questions in order to elicit written forms of words, phrases, and sentences. We distributed it to 791 Japanese speakers in Kansai areas (i.e., Hyogo, Kyoto, Mie, Nara, Osaka, Shiga, and Wakayama) by post and received answers from them. We converted the written data to International Phonetic Alphabets, and calculated aline distance using the alineR package (Downey et al., 2017) in R language (R Core Team, 2021). We removed all participants who missed answering any question in the questionnaire, leaving 491 participants.

We initially ran a principle component analysis (PCA) and identified that ten words (*watashi*, *ore*, *shindoi*, *gokiburi*, *hikigaeru*, *benjo*, *higanbana*, *katazukeru*, *konai*, and *hisashiburi*) repetitively appeared in the retained PCA components. Hence, we selected these words and ran PCA again. The results of Horn's parallel analysis demonstrated that four PCA components to be retained in further analyses. The loadings of principle components (contribution > 10%) in each components did not overlap among the four principle components except PC4.

In order to identify the optimal HDBSCAN model (tuning the hyper parameter, *minPts*), we assessed internal cluster metrics (e.g., Calinski Harabasz, CDbw, Dunn, and Silhouettte) and external cluster metrics (e.g., Czekanowski Dice, Folkes Mallows, and Jaccard) with 500 bootstrap samples. The results demonstrated that the optimal minPts was 31, and that there were three clusters among the Kansai-dialect speakers. One group (n = 164) resided in the Japan-Sea side of Hyogo, Kyoto, and Shiga prefectures as well as Mie prefectures. Another group (n = 100) resided in

the southern part of Osaka and Nara, and Mie prefecture. The other group (n = 36) generally resided western part of Kansai area.

This study identified lexical items, which play a role in phonetically investigating dialectal variation, and demonstrated that Kansai dialects can be classified into three groups. In addition, this study suggests that phonetic distance (aline) can be another metric to understand how dialects vary in a language.

(486 words)

References

Campello, R., J. G. B., Moulavi, D., & Sander, J. (2013). Densiity-based clustering based on hierarchical density estimates. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 160–72. Springer. doi:10.1007/978-3-642-37456-2_14

Downey, S. S., Sun, G., & Norquest, P. (2017). alineR: an R package for optimizing feature-weighted alighments and linguistic distances. *The R Journal*, **9** (1), 138-152.

Downey, S. S., Hallmark, B., Cox, M. P., Norquest, P., & Lansing, J. S. (2008). Computational feature-sensitive reconstruction of language relationships: developing the ALINE distance for comparative historical linguistic reconstruction. Journal of Quantitative Linguistics, 15(4), 340-369.

Inoue, F. (2001). *Keiryōgakuteki hōgengaku*. Tokyo, Meijishoin.

Kindaichi, H. (1964). *Watashi no hōgenkukaku*. In M. Tojo (Ed.), Nihongo Hōgenkukaku, Tokyo, Tokyodo shuppan.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Tōjo, M. (1953). Nihon hōgengaku. Tokyo, Hirokawakōbunkan.

LED-A: a web app for measuring distances in the sound components among local dialects

Wilbert Heeringa, Vincent van Heuven, Hans Van de Velde

In order to explore a dialect landscape and to reveal spatial patterns dialectologists measure linguistic distances among local dialects. Kessler (1995) found the Levenshtein distance very suitable for measuring linguistic distances among Irish Gaelic dialects using phonetic transcriptions. Other scholars followed him by applying the method to dialects from other language families. The Levenshtein distance is a numerical value of the cost of the least expensive set of insertions, deletions or substitutions that would be needed to transform one string into another (Kruskal 1999). This distance measure is available in the online web app Gabmap (Nerbonne et al. 2011, Leinonen et al. 2015).

In this poster we present a new web app – LED-A - that shares the features of Gabmap that are often used by dialectologists such as cluster analysis, multidimensional scaling, beam maps, area maps and RGB maps (maps that visualize the dialect landscape as a continuum). Our web app, however, differs in the design of the user interface, and the app includes features that are not found in Gabmap.

As for the user interface, we aimed to maximize flexibility, user-friendliness and intuitiveness. As for the features, four different variants of the Levenshtein distance are readily available without the need of coding a configuration file. Among others, PMI Levenshtein is included, which learns segment distances based on the alignments that are generated by the algorithm (see Wieling et al. 2009, Wieling 2012). Both aggregated and individual word distances can be obtained on the basis of whole words or only on the basis of vowel or consonant substitutions or indels.

When creating maps, it is sufficient to upload the coordinates of the places, i.e. no coordinates that constitute the outline are required. Multiple map backgrounds can be chosen from.

References

Kessler, B. (1995). Computational dialectology in Irish Gaelic. In *Proceedings of the 7th Conference of the European Chapter of the Association for Computational Linguistics*, pages 60–67, Dublin. EACL.

Kruskal, J. B. 1999. “An overview of sequence comparison.” In *Time Warps, String edits, and Macromolecules. The Theory and Practice of Sequence Comparison* edited by D. Sankoff, and J. Kruskal, 2nd ed., 1–44. Stanford: Center for the Study of Language and Information. 1st edition appeared in 1983.

Leinonen, T., Çöltekin, Ç., & Nerbonne, J. (2016). Using Gabmap. *Lingua*, 178, 71-83.

Nerbonne, J., Colen, R., Gooskens, C., Kleiweg, P., & Leinonen, T. (2011). Gabmap-a web application for dialectology. *Dialectologia: revista electrònica*, 65-89.

Wieling, M., Prokić, J., and Nerbonne, J. (2009). Evaluating the pairwise alignment of pronunciations. In Borin, L. and Lendvai, P., editors, Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education, (LaTeCH - SHELT&R 2009). Workshop at the 12th Meeting of the European Chapter of the Association for Computational Linguistics. Athens, 30 March 2009, pages 26—34.

Wieling, M. B. (2012). A quantitative approach to social and geographical dialect variation. PhD thesis University of Groningen.

Dialect perception versus dialect production: how are they related?

Frans Hinskens, Wilbert Heeringa

Abstract

Due to the influence of the standard language on the one hand and increased mobility on the other, Dutch local dialects have become less differentiated and fused to larger wholes, (supra)regional koinai or 'regiolects' (Hoppenbrouwers 1990, Hinskens 1993, Auer & Hinskens 1996, Hinskens, Auer & Kerswill 2005, Heeringa & Hinskens 2014, 2015). In this paper we study whether this change is perceived by the speakers themselves and, if so, to which extent.

Janson (1983) writes that 'for an individual in a situation of change, perception seems to lag behind production'. Therefore, we also study how the perception is related to the speaker's production.

We conducted a perception experiment where speakers of local Dutch dialects listened to recordings of their own dialect, of other local dialects in the same area, and of local dialects from another dialect group. There were 151 participants, 90 male listeners with an average age of 55 years and 61 female listeners with an average age of 44 years, mainly speakers of southern local Dutch dialects. The participants listened to eight recordings of older male speakers and to eight recordings of younger female speakers. They rated the difference of the local dialects the recordings of which they heard in comparison to their own dialect on a five-point Likert scale where 0='equal' and 4='very strongly different'. Below, we refer to the ratings as scores.

Using the recorded material we calculated dialectometric distances at the lexical level, the morphological level and the level of the sound components separately for the older male speakers and the younger female speakers.

Using this material, we address the following questions:

1. *Is the difference in dialect production between younger female speakers and older male speakers perceived by the listeners?* We test whether the differences between the scores of local dialects within the listeners' own dialect area and the scores of local dialects outside the listeners' own dialect area are on average smaller when they listen to the recordings of the younger female speakers than when they listen to the recordings of the older male speakers.

2. *Are dialect differences less well perceived by younger female listeners than by older male speakers?* We focus on the recordings of the older male speakers and test whether the differences between the scores of local dialects within the listeners' own dialect area and the scores of local dialects outside the listeners' own dialect area are smaller for the younger female listeners than for the older male listeners.

3. *Does perception lag behind production?* We focus on the younger female listeners and test whether their scores correlate better with the corresponding dialectometric measurements obtained on the basis of the older male speakers than with the corresponding measurements obtained on the basis of the younger female speakers (See Table 1).

perception		production	correlation
younger female listeners	vs.	production older male speakers	r_{OM}
younger female listeners	vs.	production younger female speakers	r_{YF}

Table1. Perception lags behind production if $r_{OM} > r_{YF}$.

References

Auer, P. & F. Hinskens (1996). 'The convergence and divergence of dialects in Europe. New and not so new developments in an old area.' In: U. Ammon, K. J. Mattheier & P. H. Nelde (eds.), *Sociolinguistica, International Yearbook of European Sociolinguistics*, volume 10, *Convergence and divergence of dialects in Europe*. Tübingen: Max Niemeyer Verlag, 1-30.

Heeringa, W. & F. Hinskens (2014). 'Convergence between dialect varieties and dialect groups in the Dutch language area.' In: B. Szmrecsanyi and B. Wälchli (eds.), *Aggregating dialectology, typology, and register analysis; linguistic variation in text and speech, Linguae et Litterae* 28, Berlin and Boston: De Gruyter, 26-52 and 452-453.

Heeringa, W. & F. Hinskens (2015). Dialect change and its consequences for the Dutch dialect landscape. How much is due to the standard variety and how much is not? *Journal of Linguistic Geography* 3(1), 20-33.

Hinskens, F. (1993). 'Dialectnivellering en regiolectvorming. Bevindingen en beschouwingen.' In: F. Hinskens, C. Hoppenbrouwers & J. Taeldeman (eds.), *Dialectverlies en Regiolectvorming*, special issue of *Taal en Tongval*, number 6, 40-61.

Hinskens, F., P. Auer & P. Kerswill (2005). 'The study of dialect convergence and divergence: conceptual and methodological considerations.' In: P. Auer, F. Hinskens & P. Kerswill (eds.), *Dialect change. The convergence and divergence of dialects in contemporary societies*. Cambridge: Cambridge University Press, 1-48.

Hoppenbrouwers, C. (1990). *Het regiolect; van dialect tot Algemeen Nederlands*. Muiderberg: Coutinho.

Janson, T. (1983). Sound change in perception and production. *Language* 59(1), 18-34.

A directional shift in a linguistic change:
A longitudinal study on English-speaking expatriates in Japan

Keiko Hirano
University of Kitakyushu, Japan

This paper attempts to demonstrate that the direction of linguistic change in a dialect-contact environment can shift over time. This provisional analysis reports on linguistic change occurring in an English-speaking expatriate community in Japan in which dialect contact (Britain, 2018; Trudgill, 1986, 2004) among English varieties occurs by comparing corpus data from 2000 and 2001 (Hirano, 2013; Hirano & Britain, 2020) with more recent data. Speakers' choice of possessive verbs (*have got*, *have* and *got*) (Tagliamonte, 2003, 2013; Tagliamonte et al., 2010) and obligatory verbs (*must, have got to, have to* and *got to*) (Tagliamonte, 2013; Tagliamonte & D'Arcy, 2007) are examined. This longitudinal study is based on three sets of linguistic data: (1) a corpus of English-language conversations collected in 2000 from young British and American English speakers who had recently arrived in Japan [Data1], (2) a corpus collected in 2001 from the same speakers after they had lived in Japan for a year [Data2] and (3) a corpus collected in 2019 from British and Americans who had worked and lived in Japan for over ten years [Data3].

The analysis of possessive verbs extracted from the three sets of data of the British English speakers shows that they increased their use of *have got* from Data1 to Data2 but decreased their use in Data3 (55%-62%-49%). Their total use of *have* and *got* decreased in Data2 but increased in Data3 (45%-38%-51%). The American English speakers maintained their total use of *have* and *got* in Data2 but slightly decreased their use in Data3 (88%-88%-81%), while their use of *have got* remained unchanged in Data2 but increased in Data3 (12%-12%-19%). These changes suggest that the British English speakers were using more typically "British" grammatical constructions in Data2 (one year after their arrival in Japan), while the American English speakers were maintaining more typically "American" constructions a year later. The analysis of Data3 (after 10 years or longer living in Japan), however, suggests an alteration of the direction of this linguistic change. Both the British and Americans were adopting the use of verbs that have strong associations with the other nationality's style of English. A similar tendency was observed among the British English speakers in terms of the choice of verbs of obligation; Data2 from the British English speakers showed an increase in the use of more typically "British" grammatical constructions (*have got to* and *must*), but Data3 revealed an increase in the adoption of *have to* and *got to*, which have strong associations

with American English. The analysis found that the American English speakers began using *have got to*, which is a more typically "British" construction, in Data2 and even more so in Data3. These changes in the use of possessive and obligatory verbs among the three sets of data indicate that the direction of a linguistic change in a dialect-contact environment is not always unidirectional but may shift over the long term.

REFERENCES

Britain, David. (2018). "Dialect contact and new dialect formation". In C. Boberg, J. Nerbonne and D. Watt (eds), *Handbook of Dialectology*. Oxford: Wiley Blackwell. 143–158.

Hirano, Keiko. (2013). Dialect contact and social networks: Language change in an Anglophone community in Japan. Frankfurt: Peter Lang.

Hirano, Keiko, and Britain, David. (2020). Accommodation and social networks: Grammatical variation among expatriate English speakers in Japan. In Y. Asahi (ed.), *Proceedings of Methods XVI: Papers from the sixteenth international conference on methods in dialectology, 2017*. Berlin: Peter Lang. pp.91–104.

Tagliamonte, Sali A. (2003). Every place has a different toll': Determinants of grammatical variation in cross-variety perspective. In G. Rohdenburg and B. Mondorf (eds.), *Determinants of grammatical variation in English*. Berlin: Mouton de Gruyter. 531–554.

Tagliamonte, Sali A. (2013). *Roots of English: Exploring the History of Dialects*. Cambridge: Cambridge University Press.

Tagliamonte, Sali. and D'Arcy, Alexandra. (2007). The modals of obligation/necessity in Canadian perspective. *English World-Wide* 28(1): 47–87.

Tagliamonte, Sali, D'Arcy, Alexandra, and Jankowski, Bridget. (2010). "Social work and linguistic systems: Marking possession in Canadian English". *Language Variation and Change* 22: 149–173.

Trudgill, Peter. (1986). *Dialects in Contact.* Oxford: Basil Blackwell.

Trudgill, Peter. (2004). *New dialect formation: The inevitability of colonial Englishes*. Oxford: Oxford University Press.

Sounds in the City: Perceptions of Ethnicity in Toronto English

Changing patterns of immigration have increased the ethnolinguistic diversity of Canada's largest cities, leading to long-term maintenance of heritage languages (HL) and the development of ethnically marked ways of speaking the majority language. Members of ethnolinguistic minority groups who grow up in 'ethnic enclaves' (neighborhoods where HLs are spoken regularly) tend to have higher degrees of orientation to their background and may be more likely to use speech features associated with it. Research on the sociolinguistic consequences of urban ethnolinguistic diversity has been conducted (e.g., Hoffman and Walker 2010) but, while anecdotes and public discourse point to awareness of ethnically marked speech, there is little systematic research on perceptions of speech of individual ethnolinguistic groups. Perceptual studies of ethnolinguistic groups outside of Canada have largely focused on "multiethnolects" (e.g., Kircher and Fox 2019), or on well-established varieties (e.g., Purnell, Idsardi and Baugh 1999).

We report results of a perceptual study of ethnically marked speech in Toronto, Canada's largest and most ethnically diverse city. Drawing on methods from previous work (e.g., Campbell-Kibler 2009, Levon 2014), we tested listener ability to identify the ethnic background of 18 native speakers of Toronto English from five of the largest ethnic groups (British/Irish, Chinese, Italian, Portuguese and Punjabi), stratified by sex and degree of ethnic orientation (EO). Listeners from Toronto heard short voice clips (~15 sec.) and were asked to identify the speaker's ethnic background from a list and judge several characteristics of the speaker: likelihood of befriending the speaker, occupation, whether the speaker was from Toronto and how well they spoke English.

Results from almost 500 participants confirm listener awareness of ethnically marked ways of speaking and greater ability to identify speakers who identify more strongly with their ethnic background. Some speakers and ethnic backgrounds are more salient than others and listeners from the same background as the speaker were *not* better judges of ethnicity (with one exception: Chinese participants better recognized Chinese speakers). The majority of participants correctly identified the High EO Chinese, Italian and Punjabi speakers but not the others.

Quality of speaking is correlated with being heard as from Toronto, and Punjabi, Italians and Chinese (of both Low and High EO) are identified as "from Toronto" more than those of British/Irish descent. People are more likely to choose as a possible friend those speakers who were identified as *not* belonging to one of the 5 specified ethnic groups, as well as speakers identified as Chinese. However, this does not match the actual identify of the speakers selected as most befriendable: the Low EO Punjabi, Chinese and Italian, and the High EO Portuguese. Our examination of perceptions of likely occupation, plot-tested to correspond to certain personality traits (e.g., social worker for 'sympathetic', entrepreneur for 'ambitious) did not reveal clear patterns.

These survey results will be augmented with reports from focus groups in which community members will discuss how they identify speakers of different communities. These findings deepen our understanding of the sociolinguistic consequences of ethnolinguistic diversity on social identity and group interactions.

References

Campbell-Kibler, K. (2009) The nature of sociolinguistic perception. *Language Variation and Change* 21: 135–56.

Hoffman, M. F. (2017) "In the front and in the back": The role of ethnicity in back vowel fronting in Toronto English. presented at the Annual Meeting of the American Dialect Society, Austin Texas.

Hoffman, M.F. & Walker, J.A. (2010). Ethnolects and the City: Ethnolinguistic variation in Toronto English. *Language Variation and Change* 22: 37-67.

Kircher, R., & Fox, S. (2019). Attitudes towards multicultural London English: Implications for attitude theory and language planning. *Journal of Multilingual and Multicultural Development, 40*(10), 847-864. doi:http://dx.doi.org.ezproxy.library.yorku.ca/10.1080/01434632.2019.1577869

Levon, E. (2014). Categories, stereotypes, and the linguistic perception of sexuality. *Language in Society* 43: 539–566.

Purnell, T., Idsardi, W. & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. Journal of Language and Social Psychology, 18(1), 10-30.

Methods Mainz 2022.08. Abstract 9 chkd

Diffusion and obsolescence of dialect vocabulary in 250 years

Fumio INOUE
Yasushi HANZAWA

In this paper, we analyze the diffusion process of standard and dialectal forms in terms of geography and age distribution based on computational lexicology. Our purpose is to elucidate the main patterns of lexical changes since before the modernization. This topic can be placed in the fields of linguistic geography and historical sociolinguistics. The basic data are drawn from two large-scale questionnaire surveys. These surveys inquired about the words recorded in a dialect glossary *Hamaogi* which was compiled 250 years ago in Tsuruoka city in northern Japan. Lexicological information was added to the data, and MCA (Multiple Correspondence Analysis) and Cluster Analysis were applied to the whole set of data.

This paper considers, based on this data, generational differences in usage of the words and changes in the usage rate of standard Japanese forms which have replaced words recorded in the dialect glossary. Of the 420 words recorded in *Hamaogi*, many have been replaced by standard Japanese forms, and some by other dialectal forms (new dialect forms). The standard forms were originally taken up to examine the reverse process of dialect decline, but analysis of the data required us to deal with the phenomenon of obsolescence instead. Contrary to expectations, some standard forms were used more often by the older generation and less by the younger generation. These are obsolete words which are related to the old lifestyle. By comparing these tendencies of *Hamaogi* words with the nationwide dialect distribution, we determined that generally those words used only in a small area have declined while words used in a wide area have survived. Also, *Hamaogi* words with a high survival rate showed a seemingly contradictory tendency to be replaced with standard Japanese forms. We discovered this tendency to be the result of a competing process distinct from survival

and replacement: the obsoletion of words. This is a process in 250 years by which words reflecting the pre-modern lifestyle are forgotten, and applies to standard forms as much as dialect forms. The process of obsoletion or dialect attrition seems to govern the linguistic change, and even the linguistic change in progress now may be disturbed by obsoletion.

“Das ist dann schon total cool zu sagen, so *Machanot*”
Revealing speakers’ justifications for linguistic choices

Esther Jahns
Universität Potsdam/Carl von Ossietzky Universität Oldenburg

In this talk I show how in a multilingual community speakers’ justifications for linguistic choices concerning loans from different donor languages can be revealed through meta-discussion and a task based on single lexical items from these languages.

Like other contemporary Jewish communities (Bunin Benor and Hary 2017; Kahn and Rubin 2016), German Jews make use of what a what Benor (2008: 1068) has defined as a “distinctively Jewish linguistic repertoire”. This repertoire consists mainly of lexical items from Yiddish and Hebrew that are integrated into German. In addition to the expression and construction of Jewish identity, the repertoire offers possibilities for inter- and intraspeaker variation as, due to the different donor languages, there are often two and sometimes more variants possible for a respective concept. This is a prerequisite for taking over a social meaning as it allows for the speakers to position themselves towards other speakers (Eckert 2012, 2008; Johnstone, Andrus and Danielson 2006) and or to express language ideologies towards the donor languages through the choice of a certain variant. In order to reveal language ideologies that Silverstein (1979: 193) defines as “sets of beliefs about language articulated by users as a rationalization or justification of perceived language structure and use” it is therefore important to grasp speakers’ own perception and interpretation of variation.

Therefore, I developed a method that was influenced by methods of Perceptual Dialectology (Cramer 2016). In a first step, I collected lexical items from the repertoire mainly through expert interviews. A selection of these items were presented to the 12 Jewish speakers in Berlin that I conducted semi-structured qualitative interviews with. The participants were asked to evaluate each item and categorize it according to their individual use. The aim of these interviews was twofold; to get first insights to speakers’ individual use and to grasp their explicit and implicit explanations for the use and avoidance of distinct items. The format of the semi-structured interview made it possible to dig deeper whenever it seemed necessary and to enhance meta-linguistic comments from the respective interviewee.

Through an analysis of the transcribed interviews influenced by methods of Grounded Theory (Charmaz 2010) speakers’ shared language ideologies towards Hebrew and Yiddish, but also other categories that are of relevance when it comes to their linguistic choices, could be revealed.

Thus, I can show through the applied method that speakers' perception of variation concerning the use of elements from their distinctively Jewish linguistic repertoire and in consequence their own linguistic choices are heavily influenced by shared clusters of language ideologies towards the donor languages. The shared multilingual resources as well as the discourse about them allows speakers to exploit variation in a meaningful way.

References

Benor, Sarah B. 2008. Towards a New Understanding of Jewish Language in the Twenty-First Century. *Religion Compass* 2: 1062–1080.

Bunin Benor, Sarah and Benjamin Hary. 2017. A Research Agenda for Comparative Jewish Linguistic Studies. In Benjamin H. Hary and Sarah Bunin Benor (eds.) *Languages in Jewish Communities, Past and Present* (Contributions to the Sociology of Language /CSL] Band 112). Berlin: de Gruyter Mouton. 672–694.

Charmaz, Kathy. 2010. *Constructing grounded theory*: *A practical guide through qualitative analysis*. Los Angeles, Calif.: SAGE.

Cramer, Jennifer. 2016. Perceptual Dialectology. In *Oxford Handbook Online*.

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12: 453–476.

Eckert, Penelope. 2012. Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation. *Annual Review of Anthropology* 41: 87–100.

Johnstone, Barbara, Jennifer Andrus and Andrew E. Danielson. 2006. Mobility, Indexicality and the Enregisterment of "Pittsburghese". *Journal of English Linguistics* 34: 77–104.

Kahn, Lily and Aaron D. Rubin (eds.) (2016). *Handbook of Jewish languages* (Brill's handbooks in linguistics 2). Leiden, Boston: Brill.

Silverstein, Michael. 1979. Language structure and linguistic ideology. In Paul R. Clyne, William Hanks and Carol Hofbauer (eds.) *The Elements: A Parasession on Linguistic Units and Levels*. Chicago: Chicago Linguistic Society. 193–247.

Individual dialect attrition as an outcome of mobility: A methodological case study

E Jamieson, Laura Domínguez, Glyn Hicks (University of Southampton) & Monika Schmid (University of York)

When people migrate to an area where a different dialect is spoken, they may accommodate their speech towards their new interlocutors (Giles 1979). Over time, these individual acts of short-term accommodation can become *long-term accommodation* (Auer & Hinskens 2005). Such adaptation of a speaker's native dialect—even when they are not conversing with the group they have been accommodating towards—can be considered *L1 attrition* (Schmid & Köpke 2017): changes to an adult L1 grammar obtaining when a speaker receives extensive exposure to a variety distinct from the one they acquired as a child. Attrition may involve suppression of forms from the first dialect, or acquisition of forms from the second dialect (Auer & Hinskens 2005); both have been attested in phonology (e.g. Barden & Grosskopf 1998), but more rarely in morphosyntax (e.g. Otheguy & Zentella 2012, Domínguez & Hicks 2016). How can we best investigate the questions of what syntactic phenomena are more amenable to attrition, and why?

In this talk we introduce a multimethod study investigating potential grammatical attrition in 30 Standard Southern British English (SSBE) speakers who settled in Belfast 15+ years ago. Specifically, we investigate possible acquisition of the Northern Subject Rule (NSR), a feature of Belfast English (Henry 1995) where *-s* agreement is permitted with 3rd person plural DP subjects, but not 3rd person plural pronominal subjects.

1. The boys go/goes out.
2. They go/*goes out.

Following Hicks & Domínguez's (2020) model of grammatical attrition, we predict attrition—in this context, acquisition of NSR—to be possible, due to existing features shared by the grammars of both varieties, assembled differently onto lexical items (Adger & Smith 2010). We present results from two tasks, with all stimuli recorded by an SSBE speaker.

1) A self-paced listening task (Ferreira et al. 1996): an aural version of a self-paced reading task, designed for cases where written stimuli are inappropriate. The time it takes a participant to aurally process each individual word in a sentence is measured, with anomalous information (e.g. agreement mismatches) expected to increase processing time. If participants have undergone attrition, we predict no increase in processing time for verbs exhibiting NSR agreement, as compared to a group of SSBE speakers in England.

2) An acceptability judgment task: participants hear sentences and judge their acceptability on a 1-5 scale (Schütze 1996). If participants have undergone attrition, we predict higher rates of acceptability for NSR examples as compared to a group of SSBE speakers in England.

We also incorporate a network score, combining % of participants' self-reported close Belfast contacts with a network density measure (Evans 2004), and a measure of participants' attitude, developed from Keijzer's (2007) sociolinguistic questionnaire for language attrition. In taking these methods together, we present a test case of a methodology for exploring potential long-term changes to individuals' dialect grammars following migration. In doing so, we tap into an understudied dimension of (bi)dialectal variation, adding to understanding of the potential linguistic outcomes (Britain 2009) of increased social and geographic mobility in the 21st century.

References

Adger, D. & Smith, J. 2010. Variation and the minimalist program. In Cornips, L. & Corrigan, K. (eds.) *Syntax and variation: Reconciling the biological and the social*. Amsterdam: John Benjamins. 149-178.

Auer, P. & Hinskens, F. 2005. The role of interpersonal accommodation in a theory of language change. In Auer, P., Hinskens, F. & Kerswill, P. (eds.) *Dialect change: Convergence and divergence in European languages*. Cambridge: CUP. 335-357.

Barden, B. and Großkopf, B. 1998. *Sprachliche Akkommodation und soziale Integration: Sächsische Übersiedler und Übersiedlerinnen im rhein-/moselfränkischen und alemannischen Sprachraum*. Tübingen: Niemeyer.

Britain, D. 2009. One foot in the grave? Dialect death, dialect contact and dialect birth in England. *International Journal of the Sociology of Language*. 196/197. 121-155.

Domínguez, L. & Hicks, G. 2016. Synchronic change in a multidialectal community: Evidence from Spanish null and postverbal subjects. In Cuza, A., Czerwionka, L. & Olson, D. (eds.) *Inquiries in Hispanic linguistics: From theory to empirical evidence*. Amsterdam: John Benjamins.

Evans, B. 2004. The role of social network in the acquisition of local dialect norms by Appalachian migrants in Ypsilanti, Michigan. *Language Variation and Change.* 16. 153-167.

Ferreira, F., Henderson, J., Anes, M., Weeks, P. & McFarlane, D. 1996. Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving-window technique. *Journal of Experimental Psychology*. 22(2). 324-335.

Giles, H. 1979. *Language: Contexts and consequences*. Keynes: Open University Press.

Henry, A. 1995. *Belfast English and Standard English: Dialect variation and parameter setting*. Oxford: OUP.

Hicks, G. & Domínguez, L. 2020. A model for L1 grammatical attrition. *Second Language Research*. 36(2). 143-165.

Keijzer, M. 2007. *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. PhD thesis: Vrije Universiteit Amsterdam.

Otheguy, R. & Zentella, A. 2012. *Spanish in New York: Language contact, dialect levelling and structural continuity*. Oxford: OUP.

Schmid, M. & Köpke, B. 2017. The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism.* 7(6). 637-667.

Schütze, C. 1996. *The empirical basis of linguistics: Grammaticality judgments and linguistics methodology*. Chicago: University of Chicago Press.

The vowel space as a sociolinguistic variable for comparing an ethnolect across dialect regions

Lisa Jeon (*University of North Texas*)
Andrew Cheng (*Simon Fraser University*)
Dot-Eum Kim (*University of Georgia*)

The analysis of vowel space area (VSA) calculated using corner vowels is a useful sociolinguistic variable for studying vocalic variation and sound change in American English. Prior research has shown that VSA has a relation with social factors including place-related identity (Labov, 1963), regional dialect (Fox & Jacewicz, 2017), and gender and sexuality (Heffernan, 2010; Pierrehumbert et al., 2004). The present study explores the utility of examining VSA for a comparative analysis of Korean Americans in three different regions. While recent work in sociophonetics highlights the effect of race and ethnicity on vocalic variation, there is no research that directly compares an ethnolect of American English across dialect regions.

We address this gap by investigating the vowel spaces of 51 Korean Americans based in three urban U.S. cities: Los Angeles (n=27), Houston (n=12), and Atlanta (n=12). The phonetic data come from conversational speech gathered in interviews with 1.5 and 2nd generation Korean Americans, aged 18-55. We extract vowel formant measurements using forced alignment and automated formant tracking (McAuliffe et al., 2017; Shue et al., 2011; Reddy & Stanford, 2015). We use Lobanov normalization (Lobanov, 1971) and calculate the median F1/F2 measurement for each of five corner vowels (FLEECE, TRAP, LOT, GOAT, and GOOSE). Following D'Onofrio et al. (2019), we calculate each speaker's complete vowel space area using Heron's method and the Euclidean distance between corner vowels. The VSA values are then compared across region, gender, and vowel to shed light on the ways in which an ethnic group might demonstrate internal variation based on geography.

Linear models fit to the data show that Korean Americans from each region pattern differently in terms of both vowel space size and spread, with between-group variation influenced by both region and gender. Mean vowel duration as a measure of speech rate also appears to be an influencing factor. Holistic vowel space compression occurs for Korean Americans in California and in Texas, but not for those in Georgia. We connect these broader VSA patterns with speakers' differing orientations to metropolitan, Southern, and supra-regional Korean American identity. Variability in Korean Americans' vowel spaces may index region, but what it means to be and sound Korean American clearly differs among groups.

Our study underscores the value of analyzing vocalic variation in terms of overall vowel space, particularly for comparative studies looking at ethnolectal variation across different regions. We argue that using methods that treat the size and dispersion of VSA itself as a sociolinguistic variable can identify aspects of vocalic variation and sound changes in progress that may not be generalizable through the analysis of individual vowels (D'Onofrio, Pratt, & Van Hofwegen, 2019).

Bibliography

D'Onofrio, Annette, Pratt, Teresa, & Van Hofwegen, Janneke. 2019. Compression in the California Vowel Shift: Tracking generational sound change in California's Central Valley. *Language Variation and Change*, *31*.2: 193-217.

Fox, Robert Allen, & Jacewicz, Ewa. 2017. Reconceptualizing the vowel space in analyzing regional dialect variation and sound change in American English. *The Journal of the Acoustical Society of America* 142: 444–459.

Jacewicz, Ewa, Fox, Robert A., & Salmons, Joseph. 2007. Vowel space areas across dialects and gender. In J. Trouvain & W. J. Barry (Eds.) *Proceedings of the 16th International Congress of Phonetic Sciences:* 1465–1468. Saarbrueken: University of Saarland.

Kim, Ji Young, & Nicole Wong. 2020. "(Divergent) Participation in the California Vowel Shift by Korean Americans in Southern California." *Languages* 5.53: 1-27.

Labov, William. 1963. The Social Motivation of a Sound Change. *Word* 18:1–42.

Lobanov, Boris M. 1971. Classification of Russian vowels spoken by different speakers, *Journal of the Acoustical Society of America* 49: 606–608.

McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner, & Morgan Sonderegger 2017. Montreal Forced Aligner: Trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association.*

Pierrehumbert, Janet B., Bent, Tessa, Munson, Benjamin, Bradlow, Ann R., & Bailey, J. Michael. 2004. The influence of sexual orientation on vowel production. *The Journal of the Acoustical Society of America* 116: 1905–1908.

Reddy, Sravana & Stanford, James. 2015. A Web Application for Automated Dialect Analysis. In *2015 Proceedings of NAACL-HLT*.

Shue, Yen-Liang, Keating, Patricia, Vicenik, Chad, & Yu, Kristine. 2011. VoiceSauce: A program for voice analysis, *Proceedings of the ICPhS XVII*: 1846-1849.

Mapping techniques used to explore dialect variation and change in the project 'Swiss-German Dialects Across Time and Space'

Péter Jeszenszky, Carina Steiner, Adrian Leemann

Center for the Study of Language and Society, University of Bern, Bern, Switzerland

Swiss German dialects are known for their regional variation within a relatively small area. They enjoy high prestige as opposed to Standard German, but there is dialect change and levelling going on (cf. Christen, 1998). In addition, one of the main driving factors of dialect change in Swiss German is hypothesised to be the spatial mobility of the population. Due to this variation and change of spatial character it is crucial to assess linguistic and sociodemographic records in space as well.

In this contribution we present different mapping techniques we employ in the SDATS project. 'Swiss-German Dialects Across Time and Space' is a 5-year project supported by the Swiss National Science Foundation, aiming at the investigation of variation and change in Swiss German since the last comprehensive survey of Swiss German dialects conducted in the 1950s, the 'Sprachatlas der deutschen Schweiz' (SDS). The SDATS corpus (Leemann et al. 2020) includes structured interviews with 1000 participants (belonging to two age cohorts, 20-35 and 60+ years old) in 125 reference localities (i.e. 8 per locality; Jeszenszky, Steiner & Leemann, 2021) in German-speaking Switzerland. The SDATS survey elicited ~300 linguistic items, and participants filled out a 300+ item metadata questionnaire.

For different kinds of publications, we analyse various kinds of data for which diverse mapping techniques are needed, some of which will be showcased in our conference contribution. Please follow the link[1] below for the figures mentioned further on. Owing to the sociolinguistic aims of the project, we have eight speakers at each reference locality, rather than only one. This makes it more difficult to use point-based techniques for mapping such as symbolisation (Figure 1). Thus, one thousand data points have to be visualised in an aggregated manner in most cases. The area over which we can aggregate may be the reference localities, cantons or even larger regions (Figure 2). Data loss can be prevented using diagrams which are, however, often suboptimal for comparison. Voronoi- (or Thiessen-) polygons around reference localities are often used for extending reference points' visibility in space and thereby also signalling the potential influence of the point in space. In these polygons the most dominant variant is represented by colour (Figure 3). Beyond this interpolation technique, we also use others, such as the *k* nearest neighbours (*knn* – Figure 4) and the kriging interpolation (Figure 5). Age often means large differences in variant usage. Mapping the two age cohorts separately shows us apparent-time change and reduces the number of data points with the same coordinates. For the purpose of real-time comparison, we often accompany such maps with ones based on SDS data. The two largest-scale visual products of the project will be an interactive online atlas and a physical printed atlas. A prototype of the former and preliminary maps (Figure 6) from the latter will also be presented at the conference, along with offline and online platforms at which to implement them the different mapping techniques discussed.

[1] https://doi.org/10.6084/m9.figshare.19874089.v1

REFERENCES

Christen, H. (1998). Convergence and divergence in the Swiss German dialects. Folia Linguistica, 32(1–2), 53–68. https://doi.org/10.1515/flin.1998.32.1-2.53

Jeszenszky, P., Steiner, C., & Leemann, A. (2021). Reduction of Survey Sites in Dialectology : A New Methodology Based on Clustering. Frontiers in Artificial Intelligence, 4(642505), 1–23. https://doi.org/10.3389/frai.2021.642505

SDATS = Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., & Messerli, J. (2020). SDATS Corpus – Swiss German dialects across time and space. Retrieved from https://osf.io/s9z4q

SDS = Sprachatlas der deutschen Schweiz. (1962–2003). Hotzenköcherle, R. Schläpfer, R., Trüb, R., and Zinsli, P. (Eds.). Bern: Francke (Vols. 1–6), Basel: Francke (Vols. 7, 8)

Effects of mobility on dialect change: Introducing the Linguistic Mobility Index

Péter Jeszenszky, Carina Steiner, Adrian Leemann

Center for the Study of Language and Society, University of Bern, Bern, Switzerland

Connection patterns across space – therefore potential contact among people with different dialect backgrounds – are crucial drivers of dialect change (Chambers, 2002). Through the ever-increasing mobility of society in the last century, the effects of diverse linguistic connections on dialects have been intensifying (Milroy, 2002). The quantitative effects of mobility itself have been, however, under-researched in sociolinguistics (Beaman, 2019), partly due to the lack of quantitative mobility data about survey participants.

Our contribution introduces the *Linguistic Mobility Index* (LMI). Based on various items of metadata elicited about survey participants' linguistic biography, LMI aims to quantify the kind of mobility that allows a participant to be exposed to linguistic influences from outside their reference locality, potentially impacting their dialect.

LMI's concept is implemented based on the SDATS corpus (Leemann et al. 2020). The corpus includes structured interviews with 1000 participants (belonging to two age cohorts, 20-35 y.o. and 60+) in 125 reference localities (Jeszenszky, Steiner & Leemann, 2021) in German-speaking Switzerland. The SDATS survey elicits ~300 linguistic items, and participants fill out a 300+ item metadata questionnaire. Part of this metadata reveals linguistic influences related to the participant's mobility.

LMI models the effects of localities associated with the following factors, constructed from metadata items: the origins of parents (Cheshire et al., 1999), the origin of the long-term life partner, places of residence in other dialect areas along with their duration, and the places of current education or workplace (e.g., Britain, 2013), age and education level. The weights of these factors are determined using the volume of the exposure, the type of relationship the participant has (had) with the local dialect the factors are associated with, and its linguistic distance to the participant's reference location, based on the Language Atlas of German-speaking Switzerland (SDS, 1962-2003).

In practice, a participant having low LMI means that the person is little exposed to direct linguistic influences outside their reference locality, never lived, studied or worked outside the reference locality, and both of their parents grew up in the reference locality. Linguistically it would correspond to an idealistic sedentary scenario, which would ease the preservation of the *local default dialect* spoken at their reference locality. In turn, high LMI means a parent or partner from, residence in or long exposure to other dialect areas, including reasons of work or (higher) education.

The meaningfulness of LMI is validated by testing its predictive power on dialect change between this local default dialect (items recorded in SDS) and items recorded in SDATS. Using mixed-effects models, we analysed the effects of LMI on dialect change considering the SDATS survey's design variables (age, sex, educational background) and show that LMI performs well at predicting lexical change: higher mobility as measured by the LMI predicts higher odds for lexical change. This success allows us to suggest the adoption of LMI in other studies of language variation and change. Among others, LMI could potentially test if linguistically mobile people become leaders of language change.

REFERENCES

Beaman, K. V. (2019). Identity and mobility in linguistic change across the lifespan: the case of Swabian German. In A. Ziegler, S. Edler, & G. Oberdorfer (Eds.), Urban Matters. Current Approaches of International Sociolinguistic Research (Studies in Language Variation) (pp. 1–42).

Britain, D. (2013). The role of mundane mobility and contact in dialect death and dialect birth. In M. Hundt & D. Schreier (Eds.), English as a Contact Language (pp. 165–181). https://doi.org/10.1017/CBO9780511740060.010

Chambers, J. K. (2002). Dynamics of dialect convergence. Journal of Sociolinguistics, 6(1), 117–130. https://doi.org/10.1111/1467-9481.00180

Cheshire, J., Gillett, A., Kerswill, P., & Williams, A. (1999). The role of adolescents in dialect levelling. In Social Research.

Cheshire, J. (2005). Age and generation-specific use of language. In U. Ammon, N. Dittmar, K. Mattheier, & P. Trudgill (Eds.), Sociolinguistics: an introductory handbook of the science of language and society. (2nd ed., pp. 1552–1563). Berlin: Mouton de Gruyter.

Jeszenszky, P., Steiner, C., & Leemann, A. (2021). Reduction of Survey Sites in Dialectology : A New Methodology Based on Clustering. Frontiers in Artificial Intelligence, 4(642505), 1–23. https://doi.org/10.3389/frai.2021.642505

Milroy, L. (2002). Introduction: Mobility, contact and language change - Working with contemporary speech communities. Journal of Sociolinguistics, 6(1), 3–15. https://doi.org/10.1111/1467-9481.00174

SDATS = Leemann, A., Jeszenszky, P., Steiner, C., Studerus, M., & Messerli, J. (2020). SDATS Corpus – Swiss German dialects across time and space. Retrieved from https://osf.io/s9z4q

SDS = Sprachatlas der deutschen Schweiz. (1962–2003). Hotzenköcherle, R. Schläpfer, R., Trüb, R., and Zinsli, P. (Eds.). Bern: Francke (Vols. 1–6), Basel: Francke (Vols. 7, 8)

The Angola Language Mapping Project: A Countrywide Sociolinguistic Survey

Linda Jordan
SIL International

Isata Manuel
Instituto de Línguas Nacionais

In 2012, the Angolan Ministry of Culture initiated an ongoing language mapping project with the purpose of providing information that will help promote access to education and the use of indigenous Angolan languages in all domains. The main objective is to identify the distinct languages in Angola and enumerate their dialects as an aid to decision making regarding potential language development programs.

Little linguistic research was conducted in Angolan territory during the colonial era and the subsequent civil war. Redinha's (1962) work on the distribution of ethnolinguistic groups drew on prior sources as well as his own research and brought together the best information available at the time. Fernandes and Ntondo (2002) created an overview of Angolan languages, which is the most detailed resource produced since independence.

In 2011, the Institute of National Languages distributed a preliminary fact-finding survey to be completed by the provincial culture directorates, requesting a list of languages, dialects and ethnic groups present in each municipality of a province. The results of this survey were then used to target specific areas for fieldwork. The researchers collected wordlists for lexicostatistical and comparative analysis, conducted comprehension testing, and facilitated participatory group interviews for a perceptual dialectology study.

Weighing the currently available data, the countrywide survey has identified 50 distinct languages spoken within Angola's borders. Within 14 of these languages, it has been possible thus far to identify a total of 65 dialects and plot them on a series of provincial maps.

The main product of the language mapping project is an up-to-date language atlas of Angola (forthcoming), showing the area where speakers of each language are in the majority and the distribution of known dialects. It is hoped that this will not only provide useful information but also encourage further research and documentation of Angola's ethnolinguistic diversity.

References

Fernandes, João and Zavoni Ntondo. 2002. *Angola: Povos e línguas*. Luanda: Editorial Nzila.

Redinha, José. 1962. *Distribuição étnica de Angola*. Luanda: Imprensa Nacional de Angola.

Fifty Years of PRICE Variation in Kentuckiana

Brian José & Anabel Büscher

The Center for Language Education and
Research at INDIANA STATE UNIVERSITY

Introduction

Monophthongization of the PRICE (AY) diphthong is a well-known stereotype of English in the Southern United States. It has been argued to be the initiating event of the *Southern Shift* (Labov, Ash, & Boberg 2006). Labov et al define three stages of the Shift: Stage 1 consists of AY glide deletion before voiced obstruents and word-finally; Stage 2 additionally involves a reversal of the mid front vowels FACE (EY) and DRESS (EH) on the F1 and F2 dimensions; Stage 3 further requires F1 and F2 reversal of the high front vowels FLEECE and KIT. This paper examines AY glide reduction in a speech community just outside of America's cultural South, in southern Indiana, where previous research (José 2020, to appear) revealed that Stage 2 of the Southern Shift has almost been attained –EY and EH are inverted on F1 but not on F2– but didn't consider the PRICE vowel (Stage 1).

Research Questions; Methods, Data

Therefore, this paper seeks to round out the description of the Southern Shift in this community by addressing the following research questions:

What is the extent of AY glide reduction there?
Is there any evidence of change through real and/or apparent time?

Our approach, then, combines synchronic and diachronic perspectives. Interviews from one of three time periods –the 1960s (n=4), the early 2000s (n=14), and the mid 2010s (n=20)– were transcribed orthographically. Random samples of the AY diphthong were preliminarily coded as either diphthongal or glide-weakened (McGivern & José 2019). The data consists of nearly 1500 tokens: an average of 39 and a range of 18 to 58 tokens per speaker. Speakers are divided into three age groups: elders, adults, and teenagers. We also hope to complement the auditory analysis with Euclidean distance measurements of glide lengths, but that work has not yet been done.

Partial, Early Results

The initial main results, excluding teenagers who are represented only in the 2010s, reveal the following patterns. Rates of AY ungliding in pre-voiceless contexts (e.g., *price*) are consistently lower than in pre-other contexts (e.g., *pry*, *prize*, *prime*); the lone exception is the 1960s elders. Pre-voiceless AY ungliding has decreased substantially into the 2010s; it's only here that AY ungliding ever falls below 20%. The lone exception is the 2010s elders, at just above 20%. Rates of AY ungliding are noticeably elevated in the 2000s; it's only in pre-voiceless contexts, particularly for elders and women, that they don't tower above the 1960s and 2010s. The results support the conclusion that this speech community has passed Stage 1 of the Southern Shift, even if it hasn't fully achieved Stage 2. However, this is coupled with a decline in glide weakening in pre-voiceless contexts over time.

References

José, Brian. 2020. A real-time perspective on the Southern vowel shift in Kentuckiana. In Y. Asahi (ed), *Proceedings of Methods XVI*. Berlin: Peter Lang. 179-193.

José, Brian. To appear. A real-time trend study of the Southern Vowel Shift in Kentuckiana. *American Speech*.

Labov, William; Ash, Sharon; Boberg, Charles. 2006. *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton de Gruyter.

McGivern, Samantha; José, Brian. 2019. Look! 50 Years of 'AY' glide reduction in Southern Indiana: A first listen. Poster presented at the 2019 Sycamore Undergraduate Research Fellowship [SURF] Exposium, Indiana State University, 04 December.

Exploring developmental norms of vowel production in the Gheg dialect of Albanian

Enkeleida Kapia[1,2], Josiane Riverin-Coutlée[1], Conceição Cunha[1] and Jonathan Harrington[1]
[1]Institute of Phonetics and Speech Processing, Ludwig-Maximilians University of Munich (Germany)
[2]Academy of Albanological Sciences (Albania)

Establishing developmental norms in speech production is essential in assessing impaired speech in children [cf. 8], but best practice guidelines also suggest that these should be language and/or dialect specific [2-9]. This is because norms are not necessarily transferable from one dialect to another, e.g. if they differ in vowel inventory size [1]. This study is a first step towards describing vowel production of typically-developing Gheg-speaking children, for whom no developmental norms have been established yet. Gheg, one of the two main dialects of Albanian, is spoken by 3.5 million people in central and northern Albania, including the capital Tirana, and Kosovo; developmental norms could thus benefit thousands of Gheg-speaking children. This study focuses on two indices often correlated with speech intelligibility and (a)typical development [e.g. 3-5-6-7-11]: a) vowel space area (VSA), and b) formant centralization ratio (FCR), both derived from acoustic characteristics of the corner vowels /i, u, a/.

The subjects were 48 Gheg-speaking children (24F) aged 6 to 8 years old who attended primary schools in the greater Tirana area, 26 of whom were recorded twice (1st and 2nd grades). This age group was prioritized because impaired speech tends to be diagnosed only when Albanian-speaking children enter primary school [10]. Additionally, 28 adults (24F) served as control. All speakers completed a picture-naming task that included words with stressed /i, u, a/, from which the first two formants were measured at the temporal midpoint.

Descriptive results for the VSA and FCR scores (Table 1) will be discussed in relation to norms in other languages. Sex differences in formant values of children's vowels despite minimal anatomical differences [4], the presence of hyper-articulation in children compared to adults (lower FCR and higher VSA), and measurable progression towards adult norms within a year (reduction of VSA) will also be addressed.

References

[1] Al-Tamimi, J., & Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. *Proceedings of Interspeech 2005*, 2465–2468.
[2] Chung, H., Kong, E. J., Edwards, J., Weismer, G., Fourakis, M., & Hwang, Y. (2012). Cross-linguistic studies of children's and adults' vowel spaces. *Journal of the Acoustical Society of America*, *131*(1), 442–454.
[3] Kent, R. D., & Rountrey, C. (2020). What acoustic studies tell us about vowels in developing and disordered speech. *American Journal of Speech-Language Pathology*, *29*(3), 1749–1778.
[4] Fitch, W. T., & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *106*(3), 1511–1522.
[5] Liu, H.-M., Tsao, F.-M., & Kuhl, P. K. (2005). The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy. *Journal of the Acoustical Society of America*, *117*(6), 3879–3889.

[6] Pettinato, M., Tuomainen, O., Granlund, S., & Hazan, V. (2016). Vowel space area in later childhood and adolescence: Effects of age, sex and ease of communication. *Journal of Phonetics*, *54*, 1–14.
[7] Sapir, S., Ramig, L. O., Spielman, J. L., & Fox, C. (2010). Formant Centralization Ratio: A Proposal for a new acoustic measure of dysarthric speech. *Journal of Speech, Language, and Hearing Research*, *53*(1), 114–125.
[8] Schölderle, T., Haas, E., & Ziegler, W. (2020). Age norms for auditory-perceptual neurophonetic parameters: A prerequisite for the assessment of childhood dysarthria. *Journal of Speech, Language, and Hearing Research*, *63*(4), 1071–1082.
[9] Storkel, H. L. (2019). Using developmental norms for speech sounds as a means of determining treatment eligibility in schools. *Perspectives of the ASHA Special Interest Groups*, *4*(1), 67–75.
[10] UNICEF. (2018). *We All Matter! Situation Analysis of Children with Disabilities in Albania*. https://www.unicef.org/albania/reports/we-all-matter
[11] Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, *50*(6), 1510–1545.

Table 1. Mean and (standard deviation) for vowel space area (VSA) and formant centralization ratio (FCR) scores. Higher VCA indicates expanded vowel space. Higher FCR indicates more centralization.

	cross-generational				longitudinal			
	men	women	boys	girls	boys Yr1	boys Yr2	girls Yr1	girls Yr2
VSA (Hz^2)	203 931 (52 081)	310 443 (74 188)	442 470 (118 245)	580 919 (121 509)	497 414 (76 874)	401 018 (100 376)	609 334 (113 523)	558 978 (141 000)
FCR	1.022 (0.071)	1.000 (0.060)	0.918 (0.059)	0.905 (0.046)	0.900 (0.047)	0.929 (0.060)	0.918 (0.044)	0.916 (0.048)

„Es werden im wesentlichen [sic!] nur Wörter aufgenommen, welche deutlich unterschiedlich zum Hochdeutschen sind.“
On the verticality of laypersons' dialect collections and the possibilities of its measurement

Within the last few years we have noticed an increasing amount of dialect collections, a lot of them appearing on the web, that have been made by laypersons. These word lists with „translation“, usually compiled within years of work and often by several dialect-speaking and -interested persons, mostly have the purpose to save dialectal words and phrases from oblivion. But also the – from the laypeople‘s point of view – special features of the respective dialect are documented in this way. So far, such collections have hardly received scientific attention (cf. Zilz 2010), not least because their compilation cannot stand up to scientific criteria. Indirectly, however, they contain information with scientific evaluation potential: By writing down those words that, in the eyes of the laypeople, belong to their dialect and that make it up, they give insight into their ideas of what they consider to be part of their particular variety.

Although the question of how laypeople conceptualize dialect(s) is of central interest in perceptual dialectology, it has so far only been approached by analyzing laypersons' conceptions of the spatial dimension of (their) dialect(s) as well as its naming and attributions (mental maps, pile-sort-method; cf. e. g. Niedzielski/Preston 2000, Lameli/Purschke/Kehrein 2008, Tamasi 2003). The presentation however focuses on the conceptualization of different Tyrolean dialects by analyzing the collections laypeople made of their own dialect. A first pilot study in this regard, presented at the International Symposium LaienWissenSprache in Kiel (cf. Schmidtke 2019), has already shown that laypersons have quite a wide-ranging idea of their diatopic variety: On the one hand, the collections do not only contain basic dialectal words, but also numerous elements of what is in use in a larger geographical radius and can thus be classified as regiolect, substandard, or colloquial (e. g. *ausrichtn* (süddt., österr., coll.) [ausrichten] ‚to speak ill of sb.‘). On the other hand, elements of the standard language, with more or less phonetic distance to it, are also classified as belonging to one‘s own diatopic variety (e. g. *loata* [Leiter] ‚ladder‘, *ausramen* [ausräumen] ‚to clear out sth.‘).

Thus, the vertical extent on the dialect-standard-axis is of particular interest: Where can individual collections be located on this scale? The question is approached by measuring the Levenshtein distance of the respective words. Initially, this has been done by using the tool Peter Kleiweg offers on the web (http://www.let.rug.nl/kleiweg/L04/). Based on a study carried out as part of a diploma thesis (Kapferer 2020) that focuses on an automation-supported analysis the presentation in a second step also evaluates possibilities and limitations of such a computational analysis.

Literature:

Tamasi, Susan L. (2003): Cognitive patterns of linguistic perceptions. [dissertation]. Athens: University of Georgia. https://esploro.libs.uga.edu/esploro/outputs/9949334114802959

Kapferer, Stefanie (2020): Entwurf eines computativen Messverfahrens zur Berechnung des Dialektalitätswertes von Laiendialektsammlungen. Samt Konzeption und

Implementierung des Verfahrens in Form des *dialect measurement tools*. Unpubl. Diplomarbeit. Innsbruck.
Lameli, Alfred/Purschke, Christoph/Kehrein, Roland (2008): Stimulus und Kognition. Zur Aktivierung mentaler Raumbilder. Linguistik Online 35, 55–86.
Niedzielski, Nancy A./Preston, Dennis R. (2000): Folk linguistics (= Trends in Linguistics. Studies and Monographs 122). Berlin, New York: De Gruyter.
Schmidtke, Henrik (2019): Internationale Fachtagung „LaienWissenSprache“, Kiel, 05.–07. September 2019. In: ZDL 86/3, 331-338.
Zilz, Wilfried; Thielemann, Otto (2010): Das Ortswörterbuch von Eltze in der Region Hannover. Praktische und theoretische Studien zur syntopischen Lexikographie. (= Veröffentlichungen des Instituts für Historische Landesforschung der Universität Göttingen, 55). Teilw. zugl.: Göttingen, Univ., Diss., 2007. Bielefeld: Verl. für Regionalgeschichte.

Linguistic diversity across languages and registers: A corpus-linguistic basis for investigating emerging grammars in language-contact situations

Martin Klotz
Humboldt-Universität zu Berlin

Annika Labrenz
Humboldt-Universität zu Berlin

Anke Lüdeling
Humboldt-Universität zu Berlin

Heike Wiese
Humboldt-Universität zu Berlin

Findings from language variation in language-contact situations can be difficult to interpret, since it is challenging to obtain suitable comparisons. For noncanonical patterns found in the language use of bilingual speakers particularly, it is important to take into account bilingual speakers' full repertoires, including informal and formal registers, and to use matching repertoire data from monolingual speakers for comparisons. We present methods of data elicitation and corpus-linguistic processing that allow for this, for the example of adolescent and adult heritage speakers of four languages (Greek, Russian, Turkish, German) in the context of two majority languages (English in the US, and German in Germany) and their monolingual counterparts in five countries (Greece, Russia, Turkey, US, Germany), and discuss methodological implications.

Our data has been obtained with the method described in Wiese (2017) ("Language Situations"), which yields naturalistic, yet controlled and comparable productions for informal and formal, written and spoken registers, covering key domains of speakers' repertoires (cf. also Biber and Conrad, 2009).

To allow for qualitative and quantitative linguistic analyses and systematic, broad-scale comparisons of potential new options and noncanonical patterns, all data is integrated into a single, unified, multi-layer corpus using the ANNIS corpus search engine (Krause, Leser, and Lüdeling, 2016). The corpus features multiple annotation layers of different kinds, multiple segmentations, and aligned multi-modal data. We present the corpus architecture and discuss challenges for the corpus-linguistic infrastructure posed by the integration of data from different languages, scripts, and registers including computer-mediated written language and informal spoken language. We illustrate potential conflicts between standardized representations and the explorative approach towards grammatical patterns. We show how the corpus supports the exploration and analysis of new grammatical options in cross-linguistic and within-language investigations across registers and speaker groups. By accessing rich metadata not only potential new dialects can be identified by grouping grammatical patterns, but also crucial extra-linguistical factors can be argued for as properties of the speaker communities.

References

Biber, Douglas and Susan Conrad (2009). Register, Genre, and Style. Cambridge: CUP.

Krause, Thomas, Ulf Leser, and Anke Lüdeling (2016). "graphANNIS: A Fast Query Engine for Deeply Annotated Linguistic Corpora". JLCL 31.1, pp. 1–25.

Wiese, Heike (2017). Language Situations: A method for capturing variation within speakers' repertoires. To appear in: Yoshiyuki Asahi (ed.), Methods in Dialectology XVI. Frankfurt a. M.: Peter Lang.

Linguistic diversity across languages and registers: A corpus-linguistic basis for investigating emerging grammars in language-contact situations

Martin Klotz
Humboldt-Universität zu Berlin

Annika Labrenz
Humboldt-Universität zu Berlin

Anke Lüdeling
Humboldt-Universität zu Berlin

Heike Wiese
Humboldt-Universität zu Berlin

Findings from language variation in language-contact situations can be difficult to interpret, since it is challenging to obtain suitable comparisons. For noncanonical patterns found in the language use of bilingual speakers particularly, it is important to take into account bilingual speakers' full repertoires, including informal and formal registers, and to use matching repertoire data from monolingual speakers for comparisons. We present methods of data elicitation and corpus-linguistic processing that allow for this, for the example of adolescent and adult heritage speakers of four languages (Greek, Russian, Turkish, German) in the context of two majority languages (English in the US, and German in Germany) and their monolingual counterparts in five countries (Greece, Russia, Turkey, US, Germany), and discuss methodological implications.

Our data has been obtained with the method described in Wiese (2017) ("Language Situations"), which yields naturalistic, yet controlled and comparable productions for informal and formal, written and spoken registers, covering key domains of speakers' repertoires (cf. also Biber and Conrad, 2009).

To allow for qualitative and quantitative linguistic analyses and systematic, broad-scale comparisons of potential new options and noncanonical patterns, all data is integrated into a single, unified, multi-layer corpus using the ANNIS corpus search engine (Krause, Leser, and Lüdeling, 2016). The corpus features multiple annotation layers of different kinds, multiple segmentations, and aligned multi-modal data. We present the corpus architecture and discuss challenges for the corpus-linguistic infrastructure posed by the integration of data from different languages, scripts, and registers including computer-mediated written language and informal spoken language. We illustrate potential conflicts between standardized representations and the explorative approach towards grammatical patterns. We show how the corpus supports the exploration and analysis of new grammatical options in cross-linguistic and within-language investigations across registers and speaker groups. By accessing rich metadata not only potential new dialects can be identified by grouping grammatical patterns, but also crucial extra-linguistical factors can be argued for as properties of the speaker communities.

References

Biber, Douglas and Susan Conrad (2009). Register, Genre, and Style. Cambridge: CUP.

Krause, Thomas, Ulf Leser, and Anke Lüdeling (2016). "graphANNIS: A Fast Query Engine for Deeply Annotated Linguistic Corpora". JLCL 31.1, pp. 1–25.

Wiese, Heike (2017). Language Situations: A method for capturing variation within speakers' repertoires. To appear in: Yoshiyuki Asahi (ed.), Methods in Dialectology XVI. Frankfurt a. M.: Peter Lang.

Physiological change or identity formation? A sociophonetic study of bilingual transmasculine voices

Remco Knooihuizen & Max Reuvers, Rijksuniversiteit Groningen

For transgender men, the changes that occur to their voice as the result of hormone replacement therapy (HRT) can be an important part of their identity: the administration of extraneous testosterone causes physiological changes that result in a lowering of pitch, which allows these speakers to be more easily perceived as men. Recent research on transmasculine speakers in both speech therapy and sociolinguistics has focused on charting change in gendered linguistic variables that is thought to be linked to physiological processes (e.g., pitch) or to social processes (e.g., the production of /s/) (Azul, 2015; Azul et al., 2017, 2018; Damrose, 2009; Van Borsel et al., 2000; Weirich & Simpson, 2018; Zimman, 2016, 2017ab, 2018).

Studies have so far focused on the analysis of monolingual speech. However, bilingual production may provide an additional window on transmasculine voice change. Although people may be variably aware of sociolinguistic patterns in their L2, the sociolinguistic production of gender (in particular masculinity) differs between languages (Boyd, 2018), while physiological processes necessarily remain the same, independent of language. Comparing the same speakers across languages, then, gives us additional evidence that can help us explore which changes are part of agentive identity formation, and which are physiological.

In this paper, we present data from a longitudinal study on voice change in five transgender men. They were interviewed monthly over a period of two years from the onset of HRT, resulting in about 50 hours of speech in L1 Dutch and L2 English. The data is analysed for changes in pitch (Henton, 1989; Hillenbrand & Clark, 2009; Traunmüller & Eriksson, 1995; Zimman, 2017a), vowel formants (Hillenbrand & Clark, 2009; Pierrehumbert et al., 2004; Puts et al., 2012), and production of /s/ in both languages (Fuchs & Toda, 2010; Levon et al., 2017; Zimman, 2017ab).

Differential developments in Dutch and English show not only a distinction between physiological and sociolinguistic change — the former showing parallel change in both languages, and the latter showing differences — but also sheds light on how speakers respond to changes in how they are perceived in society.

References

Azul, David. 2015. Transmasculine people's vocal situations: A critical review of gender-related discourses and empirical data. *International Journal of Language & Communication Disorders* 50(1). 31–47. 10.1111/1460-6984.12121

Azul, David, Aron Arnold & Christina Neuschaefer-Rube. 2018. Do transmasculine speakers present with gender-related voice problems? Insights from a participant-centered mixed-methods study. *Journal of Speech, Language, and Hearing Research* 61. 25–39. 10.1044/2017_JSLHR-S-16-0410

Azul, David, Ulrika Nygren, Maria Södersten & Christiane Neuschaefer-Rube. 2017. Transmasculine people's voice function: A review of the currently available evidence. *Journal of Voice* 31(2). 261.e9–261.e23. 10.1016/j.jvoice.2016.05.005

Boyd, Zac. 2018. *Cross-linguistic variation of /s/ as an index of non-normative sexual orientation and masculinity in French and German men.* Ph.D. dissertation, The University of Edinburgh.

Damrose, Edward J. 2009. Quantifying the impact of androgen therapy on the female larynx. *Auris Nasis Larynx* 36(1). 110–112. 10.1016/j.anl.2008.03.002

Fuchs, Susanne & Martine Toda. 2010. Do differences in male versus female /s/ reflect biological or sociophonetic factors? In Susanne Fuchs, Martine Toda & Marzena Żygis (eds.), *Turbulent sounds: An interdisciplinary guide*, 281–302. Berlin: De Gruyter.

Henton, Caroline G. 1989. Fact and fiction in the description of female and male pitch. *Language & Communication* 9(4). 299–311. 10.1016/0271-5309(89)90026-8

Hillenbrand, James M. & Michael J. Clark. 2009. The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics* 71(5). 1150–1166. 10.3758/APP.71.5.1150

Levon, Erez, Marie Maegaard & Nicolai Pharao. 2017. Introduction: Tracing the origin of /s/ variation. *Linguistics* 55(5). 979–992. 10.1515/ling-2017-006

Pierrehumbert, Janet B., Tessa Bent, Benjamin Munson, Ann R. Bradlow & J. Michael Bailey. 2004. The influence of sexual orientation on vowel production. *Journal of the Acoustical Society of America* 145(6). EL574–EL580. 10.1121/1.5113566

Puts, David A., Coren L. Apicella & Rodrigo A. Cárdenas. 2012. Masculine voices signal men's threat potential in forager and industrial societies. *Proceedings of the Royal Society B* 279. 601–609. 10.1098/rspb.2011.0829

Traunmüller, Hartmut & Anders Eriksson. 1995. The frequency range of the voice fundamental in the speech of male and female adults. http://www2.ling.su.se/staff/hartmut/f0_m&f.pdf

Van Borsel, J., G. De Cuypere, R. Rubens & B. Destaerke. 2000. Voice problems in female-to-male transsexuals. *International Journal of Language & Communication Disorders* 35(3). 427–442. 10.1080/136828200410672

Weirich, Melanie & Adrian P. Simpson. 2018. Gender identity is indexed and perceived in speech. *PLoS ONE* 13(12). e0209226. 10.1371/journal.pone.0209226

Zimman, Lal. 2016. Sociolinguistic agency and the gendered voice: Metalinguistic negotiations of vocal masculinization among female-to-male transgender speakers. In Anna M. Babel (ed.), *Awareness and control in sociolinguistic research*, 253–277. Cambridge: Cambridge University Press. 10.1017/CBO9781139680448.013

Zimman, Lal. 2017a. Gender as stylistic bricolage: Transmasculine voices and the relationship between fundamental frequency and /s/. *Language in Society* 46(3). 339–370. 10.1017/S0047404517000070

Zimman, Lal. 2017b. Variability in /s/ among transgender speakers: Evidence for a socially grounded account of gender and sibilants. *Linguistics* 55(5). 993–1019. 10.1515/ling-2017-0018

Zimman, Lal. 2018. Transgender voices: Insights on identity, embodiment, and the gender of the voice. *Language and Linguistics Compass* 12(8). e12284. 10.1111/lnc3.12284

Cross-Border Language Contacts in the Polish-Belarusian Border Region in the 21st Century

Katarzyna Konczewska (Institute of Polish Language Polish Academy of Sciences)

This paper focuses on border effects involving varieties of two closely related Slavic languages in contact, Polish and Belarusian, in the little-studied northern part of the contemporary Polish-Belarusian border region. The preliminary results of research allow us to posit the prevalence of productive bilingualism with diglossia in this area and to define the sociolinguistic situation as exoglossic, unbalanced, and four-component. The material for the study was collected by the author in 2015-2019 during dialectological interviews in communities on both sides of the Polish-Belarusian border, which it the lateral dialectology area of the Balto-Slavic contact zone. Microarea studies are key regarding linguistic contact in the transitory zones since they allow, e.g. for the tracing of directions, the depth of local dialect infiltration.

The microarea under investigation is characterized by its relative inaccessibility, as well as heterogeneity of local residents in terms of national identity and religious affiliation. At present, the autochthonous local population is comprised of both Eastern Orthodox and Roman Catholics, mainly Poles and Belarusians in terms of national self-identification, who are the descendants of peasants and the petty gentry. A unique feature of the area under investigation is that for more than five hundred years it was an integral unit within various state formations; it was divided by a political border only in 1948. Uneven settlement processes due to landscape features, as well as historical and political factors influenced the formation of specific, multicomponent sociolinguistic situations on each side of the border.

In this presentation I will examine the pluricentric languages common in the area, as well as the linguistic codes used by their native speakers. While the theory and methodology of research on language and dialect contact in border regions have been addressed in the scholarly literature (e.g. Woolhiser 2005), Konczewska (2021) has shown that the peculiarities of the formation and development of the area under investigation would benefit from a more individualized approach.

Hypotheses concerning the course of linguistic contacts in peripheral areas are the key elements in the research of linguistic contact in the greater Baltic area. The verification of such hypotheses will optimize research quality and make new knowledge available. In these studies I strive to go pass beyond the research models of traditional linguistics, taking the work of ethnographers and ethnohistorians into account as well.

Konczewska, Katarzyna. 2021. Polacy i język polski na Grodzieńszczyźnie. Kraków: Instytut Języka Polskiego PAN.
Woolhiser, Curt. 2005, *Political Borders and Dialect Convergence/Divergence in Europe*. In: P. Auer, F. Hinskens, P. Kerswill (eds.), *Dialect Change: Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 236-262.

The Endangered State of the Japanese Narada Dialect and its Background

KONISHI Izumi (The University of Tokyo)
MIKI Yosuke (Komazawa University)
YOSHIDA Noriko (Jissen Women's University)

Background, Aim, and Methods

The Narada dialect is spoken in Narada village in Yamanashi Prefecture, Japan. Its most prominent feature is its word-accent system, which distinguishes it from the Yamanashi dialect. According to UNESCO's Language Vitality and Endangerment Framework, it can be classified as a "severely endangered" dialect as it is spoken only by elderly people.

We have been conducting descriptive and sociolinguistic research interviews with Narada residents since the 1990s. In this presentation, we report the past and present sociolinguistic state of the Narada dialect. We also discuss the background factors that are endangering the dialect.

Results and Conclusion

The Narada dialect features a word-accent system with "raising kernel," i.e., pitch rising, whereas the Yamanashi dialect features a word-accent system with "lowering kernel," i.e., pitch falling; this "lowering kernel" system is also present in the Tokyo dialect (Uwano 2012). Interestingly, the position of the accent (rising kernel) in the Narada dialect corresponds to the position of the accent (lowering kernel) in the Yamanashi dialect (Konishi 1999; Table 1). It is believed that the accent system of the Narada dialect emerged as an outgrowth of the Yamanashi dialect system.

Table 1. Word-accents of the Narada, Yamanashi, and Tokyo dialects.

	Narada	Yamanashi	Tokyo
"nose"	hana=	hana=	hana=
"rain"	a[me	a]me	a]me
"older sister"	a[ne	a]ne	ane=
"cloud"	kumo[	kumo]	ku]mo

[: rinsig kernel,]: lowering kernel, =: kernelless

We found that the residents of Narada were bidialectal, speaking both Narada and Yamanashi dialects. They spoke the Narada dialect when conversing with other Narada residents and the Yamanashi dialect when conversing with people from other regions. They acquired the Yamanashi dialect while interacting with elementary school children from other villages. None of them required systematic learning or a great deal of effort to acquire it. The acquisition of the Yamanashi dialect

was made easier by the correspondence of accent positions, which appear to have been internalized by speakers of the Narada dialect, thereby facilitating dialect switching.

Currently, there are approximately 10 or fewer speakers of the dialect. In addition, a few years ago, several families with no ties to this area started to reside in the village. As a result, the village has been revitalized; however, the dialect is seldom used within the village, even among the residents of Narada. Some of our consultants stated that speaking with us for our surveys was the only time that they used the Narada dialect. In other words, the Narada dialect is no longer a language of daily life.

In conclusion, the word-accent correspondence formerly aided speakers of the Narada dialect by making it easier for them to acquire the Yamanashi dialect; however, it is currently accelerating the disappearance of the dialect.

References

Konishi, Izumi (1999) *Narada-hōgen akusento no genzai* (The current state of word-accent of the Narada dialect), *Jinbun-Gakuhō*, 320, 25-40.

Uwano, Zendo (2012) Three types of accent kernels in Japanese, *Lingua*, 122, 1415-1440.

UNESCO Atlas of the World's Languages in Danger, http://www.unesco.org/languages-atlas/

The interviewer's role in socio-linguistic interviews regarding non-linguists' conceptualizations of speech repertoires

Koppensteiner, Wolfgang & Rita Stiglbauer

It is a commonplace in linguistics that speakers adapt their utterances to their interaction partners both on the linguistic as well as on the content level. However, the (possible) effects of the interviewer on attitudinal expressions by the interviewee and what is actually (not) said in socio-linguistic interviews have not been subject to thorough investigations yet. This presentation addresses this research gap, focusing on the impact the interviewer's (country of) origin as an influencing factor for (meta-communicative) expressions of the interviewee's attitudinal utterances in socio-linguistic interviews (cf. Briggs 2005, Tagliamonte 2006, Deppermann 2013) in a pluricentric constellation. (That means in a context, in which various centres (regions, countries) share a common standard variety with linguistic peculiarities for each centre (cf. Clyne 1989, Ammon 1995). The interview data used for this presentation stems from the special research programme 'German in Austria. Variation – Contact – Perception', that conducted approx. 150 interviews with speakers in rural areas all over Austria. These interviews elicited the speakers' attitudes towards languages and German in Austria in general and towards individual conceptualizations of language variation and speech repertoires in particular (cf. Koppensteiner & Lenz 2017). All interviewers conducting these interviews were born and socialized in Austria. In a follow-up study using a sample of 10 interviewees, these interviews were 'repeated' varying only in one decisive parameter: This time, an interviewer from Germany replicated the elicitation approximately one year after the first interview. Major other parameters (e.g. interview guideline, setting, technical equipment and location) of these second interviews remained unchanged. The interviews were thoroughly processed, annotated and content-analytically investigated making use of CAQDAS tools. In addition to the comprehensive comparison of conceptualizations of individual speech repertoires, (incl. 'perceived' variational patterns) within both interview rounds, aspects of the interviewees' meta-linguistic and implicit reflections regarding the interviewer and the interview situation were thoroughly examined. Results provide evidence for a set of contextual layers co-determining attitudinal and other utterances within socio-linguistic interviews, driven by the following main research questions:

- Which impact does the interviewer have on the verbalization of language attitudinal utterances in general? In particular, which impact does the interviewer's (country of) origin have on such utterances with regard to aspects of pluricentricity?
- Do informants discuss the interviewers' (linguistic and socio-linguistic) background and if yes, how? Do they discuss the methodological approach, i.e. the (socio-linguistic) interview situation, per se and if yes, how?

References:

Ammon, Ulrich (1995): *Die deutsche Sprache in Deutschland, Österreich und der Schweiz*. Berlin/New York: de Gruyter.

Briggs, Charles L. (2005): Sociolinguistic Interviews. In: Ammon, Ulrich / Dittmar, Norbert / Mattheier, Klaus J. / Trudgill, Peter (eds.): *Soziolinguistik. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft. Bd. 2. 2*. Berlin/New York: de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 3.2), 1052–1062.

Clyne, Michael (1989): Pluricentricity: National Variety. In: Ammon, Ulrich (ed.): *Status and Function of Languages and Language Varieties*. Berlin: de Gruyter, 357–371

Deppermann, Arnulf (2013): Interview als Text vs. Interview als Interaktion. *Forum: Qualitative Sozialforschung* 14/3.

Koppensteiner, Wolfgang / Lenz, Alexandra N. (2017): Theoretische und methodische Herausforderungen einer perzeptiv-attitudinalen Standardsprachforschung. Perspektiven aus und auf Österreich. In: Sieburg, Heinz / Solms, Hans-Werner (eds.): *Das Deutsche als plurizentrische Sprache. Ansprüche - Ergebnisse – Perspektiven.* Erich Schmidt Verlag (Zeitschrift für deutsche Philologie Sonderheft), 43–68.

Tagliamonte, Sali A. (2006): *Analysing Sociolinguistic Variation.* Cambridge: Cambridge University Press.

“And now say *that* in Swedish”: Using conversation groups to elicit multilingual speech from Swedish Estonian teens

Mari-Liis Korkus

This presentation explores conversation groups, a mixed approach for gathering multilingual interactive data, and reports research findings on the bilingual language use of Swedish Estonian teenagers (Korkus 2021). The study’s central aim was to investigate code-switching functions, their patterns, occurrences, and meanings, in an interactive situation. The recorded conversations followed both a gamified and a discussion-like setting. This talk aims to assess the effectiveness of conversation groups and to suggest improvements to develop an optimal data collection strategy for researching interactions between multilingual speakers.

Data collection took place in 2019. Stockholm-based informants (n = 5) with Estonian roots, aged 15–17, acted as informants for this study and spoke in two conversation groups; each group talked for approximately two hours. First, speakers played a trilingual word-guessing game where one informant explained a word in Swedish, Estonian, or English (e.g., *kogukond* ‘community’, *skolstrejk* ‘school strike’) to the other informant(s). After this, speakers discussed at least one given topic selected by the author (e.g., climate change, youth employment, differences between Sweden and Estonia). Participants were free to use both Swedish and Estonian throughout the interaction.

The conversation recordings (four hours and eight minutes in total) included 1786 utterances, of which 439 (24.6%) contained code-switching. Code-switching was used most during the gamified setting (62.4%), where speakers explained words in different languages. The analysis showed six code-switching functions: semantic specifications (n = 176), quasi-translations (n = 95), vocabulary limitations (n = 63), expressiveness (n = 29), cross-utterance language harmony (n = 26), and wordplay (n = 6).

The data demonstrate that conversation groups can effectively elicit multilingual speech; however, some adjustments are necessary. The central issue with conversation groups was that this approach produced somewhat biased results. Speakers were in a position where they had to interact with others while speaking multiple languages, i.e., the environment was manipulated and did not emulate an everyday setting. While the discussions allowed the speakers to interact more freely, the word-guessing game proved to be more restrained. The analysis showed that, in most cases, code-switching had semantic associations (40.1%): language switches occurred when the speaker was referring to a particular topic, person, or action. The fact that the speakers primarily discussed specific topics in conversation groups can explain this result.

This talk proposes two ideas for further development. A future study should replicate the conversation group setting with different speakers to see if the results remain consistent. Four informants in the presented study were born and raised in Sweden, and one moved there in 2015. Contrary to expectations, the analysis revealed that the latter speaker produced most of the utterances containing code-switching (n = 131). Alternatively, the author suggests other methods for eliciting more natural multilingual interactions, for example, via semi-structured group interviews (see Labov 1984; Johnstone 2000).

References

Korkus, M.-L. (2021). A sociopragmatic approach to code-switching: A case study of Swedish-Estonian teenagers [Sotsiopragmaatiline sissevaade koodivahetusele: näiteid rootsieesti teismeliste keelest]. − *Eesti Rakenduslingvistika Ühingu aastaraamat = Estonian Papers in Applied Linguistics 17*, 141−157. http://dx.doi.org/10.5128/ERYa17.08.

Labov, W. (1984). Field Methods of the Project on Linguistic Change and Variation. In J. Baugh & J. Sherzer (Eds.), *Language in use: Readings in sociolinguistics* (pp. 28–53). Englewood Cliffs: Prentice Hall.

Johnstone, B. (2000). *Qualitative Methods in Sociolinguistics*. Oxford, New York: Oxford University Press.

The Fractal Structure of Language: How Many Dialects?
William A. Kretzschmar, Jr., University of Georgia

In previous study of dialect survey data (e.g. Kretzschmar 2009, 2015), the frequency profiles of variant lexical responses to the same cue are all patterned in nonlinear A-curves. Phonetic transcriptions from the Atlas also have the same A-curve frequency profiles, as have the distributions of measurements of vowels in F1/F2 space. Moreover, these frequency profiles are scale-free, or fractal, in that the same A-curve patterns occur at every level of scale. A-curve patterns describe the distribution of all linguistic features we have observed—lexical, phonetic, and grammatical—for a survey overall, for different groups of speakers, for individual speakers, and even for separate environments in which vowels occur. These findings challenge the boundaries that linguists have traditionally drawn for dialects, whether geographic, social, or phonological, and demand that we use a new model for understanding language variation. Instead of using statistics to try to match our generally shared impressions of dialects, we should realize that there is an unlimited number of dialects, and our choice to focus on one or another of them should follow, not from popular perception, but clear definitions of the population of speakers we wish to observe.

References

Kretzschmar, William A., Jr. 2009. *The Linguistics of speech*. Cambridge: Cambridge UP.
Kretzschmar, William A., Jr. 2015. *Language and complex systems*. Cambridge: Cambridge UP.

Dialectometry with topic models

Olli Kuparinen & Yves Scherrer

Dialectometrical analyses are most often based on dialect atlases compiled systematically over the last 150 years. Although such data have many benefits, such as good geographical coverage, the atlases tend to be old and only exist for some languages. Dialect corpora collected in semi-directed interviews have thus become important data sources more recently. In this study, a topic model approach to discover differences between dialects from interview data is presented as an alternative to analyses based on dialect atlases.

Topic models are often used to find latent semantic structure in a collection of text documents. Co-occurring words (e.g., *dog, bone, fetch*) in multiple documents are assumed to constitute a topic (dogs). Because the interest has been in semantic similarity, the model has been used on normalized and lemmatized language data to prevent modeling the same word in its different forms.

For dialects, the interest lies in these differing forms (i.e., structural differences). This means the model can be used on phonetically transcribed data directly, and it will find components that correspond to different dialectal features or combinations of them. It is then easy to do traditional dialectometrical analysis, and see which components are used where. Topic models have been used on dialect data before, but the analysis has focused on lexical variation in social media (Eisenstein et al. 2010) or the linguistic features have been searched for before the modeling (Kuparinen et al. 2021). The approach presented here uses transcribed data directly, without pre-processing steps.

The approach is tested on corpora from three languages: Finnish, Norwegian and Swiss German. The datasets include Samples of Spoken Finnish (Institute for the Languages in Finland 2014), Norwegian Dialect Corpus (Johannessen et al. 2009) and Archimob Corpus (Scherrer et al. 2019), all of which include interview transcriptions from several locations in their respective areas. The modeling is tested on different levels of the transcriptions: complete words, character n-grams (sequences of characters) and automatically segmented words. The model used in the study is non-negative matrix factorization (NMF; Lee & Seung 1999). It returns two distributions: one presenting the components over features (which features are used in which components) and one presenting the documents over components (which components are used in which documents).

The results are very promising and show that such a model can find important dialectal differences directly from interview transcriptions. When modeling based on complete words, the model finds differences in frequent words, such as personal pronouns (*minä, mää, mie* 'I' in Finnish) or negation (*itte, ikkje, ittje* 'not' in Norwegian). When using the character level as input, the model discovers phonological differences, such as diphthong opening or reduction (*mi**ä**s* 'man', ***ae**na* 'always') in Finnish and l-vocalization in Swiss German (a**u**so 'so'). The automatic segmentation of data is used to combine both levels and find important words as well as important character sequences. The languages differ somewhat, with Finnish and Norwegian producing clearer divisions based on complete words, and Swiss German based on character-level differences.

References

Eisenstein, J., B. O'Connor, N. A. Smith & E. P. Xing. (2010). A latent variable model for geographic lexical variation. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*: 1277–1287.

Institute for the Languages of Finland (2014). *Samples of Spoken Finnish* [speech corpus]. Kielipankki. Retrieved from http://urn.fi/urn:nbn:fi:lb-2020112937

Johannessen, J. B., J. Priestley, K. Hagen, T. A. Åfarli & Ø. A. Vangsnes. (2009). The Nordic dialect Corpus - an Advanced Research Tool. In Jokinen, K. and E. Bick (eds.): *Proceedings of the 17th Nordic Conference of Computational Linguistics NODALIDA 2009. NEALT Proceedings Series* Volume 4.

Kuparinen, O., J. Peltonen, L. Mustanoja, U. Leino & J. Santaharju. (2021). Lects in Helsinki Finnish. A probabilistic component modeling approach. *Language Variation and Change*, 33(1): 1-26

Lee, D. D. & H. S. Seung (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755): 788-791.

Scherrer, Y., T. Samardžić & E. Glaser (2019). Digitising Swiss German - How to process and study a polycentric spoken language. In: *Language Resources and Evaluation* 53(4).

Hanna Lappalainen
University of Eastern Finland

Lotta Aarikka
University of Turku

How to study the relationship between language use and class distinctions?

Finland, like other Nordic societies, has traditionally been considered very equal, and talk of social classes and social inequality has long been almost a taboo. However, rapid changes in the economic structure and working life and migration both within Finland and from elsewhere have contributed to an increase in income disparities. Social segregation is not only manifested in differences in livelihoods but is also heard and reflected in language use and social relations. However, up-to-date information on the relationship between class distinctions and language use is lacking and the acquisition of the information is challenging.

Our paper discusses methods for identifying linguistic phenomena associated with social classes and class divisions in Finland. The aim is to approach social class from a holistic and intersectional point of view, so that not only traditional extralinguistic variables (e. g. occupation and education) but also aspects of ethnicity and immigration are included. Our study has emerged among a multidisciplinary research group which gathers and sociolinguists and sociologists from Finland. Our aim is to answer to the following questions:

1. How do class distinctions manifest in the use of Finnish and Finnish-Swedish?

2. Which linguistic and social resources are recognized and interpreted significant from the perspective of class distinctions?
3. How are classes constructed linguistically, socially and culturally?

In our paper, we will concentrate on recognition of the (linguistic) resources of Finnish associated with class distinctions (see question 2) from the methodological point of view. We will discuss *how* to find relevant linguistic phenomena. We have collected pilot data by questionnaires and listening tests and analyzed interview data in order to aim information on both conscious and unconscious concepts of Finns with different social and linguistic backgrounds.

In our paper, we will focus on the comparison of different methods and results achieved by them. Our preliminary results show that the following linguistic phenomena are frequently associated with class distinctions: grammatical correctness vs. grammatical errors, the extent of vocabulary, the use of standard vs. dialects as well as the use of swear words or slang. In our paper, we will evaluate different methods and consider possibilities to develop them.

Literature

Bourdieu, P. 1987: What Makes a Social Class? On the Theoretical and Practical Existence of Groups. *Journal of Sociology* 32, 1–17.

Casasanto, L. Staum – Grondelaers, S. – van Hout, R. 2015. Got class? Community shared conceptualizations of social class in evaluative reactions to sociolinguistic variables. – A. Prikhodkine & D. R. Preston (eds.), *Responses to language varieties. Variability, processes and outcomes*, 159–173. Amsterdam: John Benjamins.
Chun, C. W. 2019: Language, discourse, and class. What's next for sociolinguistics? *Journal of Sociolinguistics* 23, 332–345.
Lo, Adrienne 2020: Systems, features, figures. Approaches to language and class vs. language and race. *Journal of Sociolinguistics* 24, 293–307.

Use all the LAP data!

Moving toward inductive discovery of patterns and connections in the data of the Linguistic Atlas Project

Mark Richard Lauersdorf, University of Kentucky

A call for data-driven, inductive discovery of patterns and connections in large datasets would not strike anyone as novel in the world of "big data" (cf. Kitchin 2014), and it really isn't even new in the study of language variation (see, for example, 35 years ago in Horvath 1985, Horvath and Sankoff 1987). It can be argued, however, that data-driven approaches have not achieved the status of "standard tool" in the toolkit of many linguists, even while such approaches present enticing new opportunities for investigating language variation in its social context (e.g. Lauersdorf 2018). In making an argument for data-driven approaches to the study of language variation, it is also important to emphasize the need to "use all the data", both the obvious and the "hidden" data, in order to maximize the effectiveness of the inductive process (Lauersdorf 2018; 2021).

The data of the Linguistic Atlas Project (LAP) is truly "big data", unstructured, semi-structured, and structured, and with all of the "volume, velocity, and variety" that one would expect (Laney 2001). The rich data in the LAP includes: language data at all structural levels (phonetic/phonological, morphological, syntactic, semantic, lexical); socio-cultural data concerning not only the interviewees, but also the interviewers; image data representing objects and physical contexts; and time-space (geospatial and temporal location) data; stored in a full range of audio-visual media and objects from raw audio recordings, to textual fieldnotes with image illustrations, to formally published atlas volumes and cartographic representations. And this rich data environment exists iteratively across the multiple regional atlas projects that constitute the LAP. With this level of data complexity and the potential for related pieces of information to be scattered across the various data types and their media manifestations, the LAP presents an ideal testbed for deploying a data-driven, inductive approach that uses all the data – the entirety of the linguistic and socio-cultural materials assembled and produced in the atlas process – to facilitate discovery of complex patterns of language variation.

This presentation will provide a structured overview of the LAP as "big data" and will sketch a "big-picture" conceptualization of a data-driven approach to mining and analyzing that data for (historical) dialectology and (historical) sociolinguistics.

Horvath, Barbara. 1985. *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.

Horvath, Barbara and David Sankoff. 1987. Delimiting the Sydney speech community. *Language in society* 16.2: 179–204.

Kitchin, Rob. 2014. Big Data, new epistemologies and paradigm shifts. *Big data & society* 1.1, article 1. https://doi.org/10.1177/2053951714528481 (accessed 25 May 2022).

Laney, Doug. 2001. 3D data management: Controlling data volume, velocity and variety. Meta Group Research Note. https://community.aiim.org/blogs/doug-laney/2012/08/25/deja-vvvu-gartners-original-volume-velocity-variety-definition-of-big-data (accessed 25 May 2022).

Lauersdorf, Mark Richard. 2018. Historical (standard) language development and the writing of historical identities: A plaidoyer for a data-driven approach to the investigation of the sociolinguistic history of (not only) Slovak. In: Stephen M. Dickey and Mark Richard Lauersdorf, eds. *V zeleni drželi zeleni breg: Studies in honor of Marc L. Greenberg*. Bloomington, Indiana: Slavica Publishers (appeared 2019), 199–218.

Lauersdorf, Mark Richard. 2021. Historical sociolinguistics and the necessity of interdisciplinary collaboration. In: Allison Burkette and Tamara Warhol, eds. *Crossing Borders, Making Connections: Interdisciplinarity in Linguistics*. Berlin: De Gruyter Mouton, 207-230.

Introduction: German in Austria – Variation across registers and structural levels

Alexandra N. Lenz (University of Vienna)

This talk introduces the special section 'German in Austria – Variation across registers and structural levels' and provides an insight into the special language situation regarding German in Austria. It also formulates research questions, which are discussed from different perspectives in the subsequent talks.

German in Austria provides a perfect research laboratory to investigate language variation across the entire 'vertical' variation on the dialect/standard axis. First, there is regional variation between five different dialect areas (Central Bavarian, South Bavarian, South-Central Bavarian transition zone, Alemannic, and Alemannic-Bavarian transition zone) concerning local dialects. Second, there are inter-regional differences with regard to the 'vertical' dialect/standard repertoires. While Bavarian dialect speakers' typically have diaglossic repertoires, resulting from the Bavarian dialect/standard continuum, the Alemannic dialect region is characterized by diglossic spectra (Auer 2018). However, recent data suggest a restructuring of these spectra, indicating 'concentration zones' in Bavarian continua as well as intermediate variants in Alemannic diglossia (Lenz 2019, Fanta-Jende 2020). In addition and third, as a 'non-dominant center' of German (Clyne 1995), Austrians are exposed to nationally diverging standard varieties, namely the Austrian German and the German German standard variety (Muhr 2007, Krech et al. 2009, Ammon et al. 2016).

The central question discussed by the panel relates to suitable methodological approaches taken to investigate the (socio-)linguistic variability in Austria across various registers on the one hand and across various structural levels on the other hand. The different talks present multi-method approaches taken to examine variation in Austrian speakers' repertoires with a focus on phonetics/phonology, morphology, syntax and lexis. These include, amongst others, different questionnaire-based methods, translation tasks, language production experiments, (formal) interviews and (informal) conversations with friends. Thus, in addition to the results of the individual studies, benefits and limitations of the methodological approaches are discussed.

On the one hand, the introductory talk aims at a synopsis of previous research results with regard to their assertions on individual phenomena, selected registers and individual system levels. On the other hand, it intends to identify research desiderata that still need to be addressed in future research. They form the background against which the contributions of the panel will be discussed. The discussion of the introductory talk as well as of the entire panel will be guided by the following research questions: Which (parallel and different) structures do the linguistic level of phonetics/phonology, morphology, syntax and lexis exhibit regarding patterns of areal/horizontal and vertical/social variation? Which features (from which levels) (do not) correlate with each other? Which phenomena and which structural levels show what degree of stability or dynamisms? Which methods do we need to answer these questions?

References

Ammon, Ulrich, Hans Bickel & Alexandra N. Lenz. 2016. *Variantenwörterbuch des Deutschen: Die Standardsprache in Österreich, der Schweiz, Deutschland, Liechtenstein, Luxemburg, Ostbelgien und Südtirol sowie Rumänien, Namibia und Mennonitensiedlungen*. Berlin & Boston: de Gruyter.

Auer, Peter. 2018. Dialect change in Europe – Leveling and Convergence. In Boberg, Charles, John Nerbonne & Dominic Watt (eds.). *The Handbook of Dialectology*. Hoboken, NJ: Wiley Blackwell, 159–176.

Clyne, Michael. 1995. *The German language in a changing Europe*. Cambridge: Cambridge University Press.

Fanta-Jende, Johanna. 2020. Varieties in Contact. Horizontal and vertical dimensions of phonological variation in Austria. In Lenz, Alexandra N. & Mateusz Maselko (eds.). *Variationist Linguistics meets Contact Linguistics*. Vienna: Vienna University Press, 203–240.

Krech, Eva-Maria, Eberhard Stock, Ursula Hirschfeld & Lutz C. Anders (eds.). 2009. *Deutsches Aussprachewörterbuch: Mit Beiträgen von Walter Haas, Ingrid Hove und Peter Wiesinger*. Berlin & New York: de Gruyter.

Lenz, Alexandra N. 2019. Bairisch und Alemannisch in Österreich. In Herrgen, Joachim & Jürgen Erich Schmidt (eds.). *Sprache und Raum. Ein internationales Handbuch der Sprachvariation. Vol. 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt* (Handbücher zur Sprach- und Kommunikationswissenschaft. 30.4). Berlin & Boston: de Gruyter, 318–363.

Lenz, Alexandra N., Ludwig Maximilian Breuer, Matthias Fingerhuth, Anja Wittibschlager & Melanie Seltmann. 2019. Exploring syntactic variation by means of "Language Production Experiments"- Methods from and analyses on German in Austria. *Journal of Linguistic Geography* 7/2, 63–81.

Muhr, Rudolf. 2007. *Österreichisches Aussprachewörterbuch. Österreichische Aussprachedatenbank*. Frankfurt am Main et al.: Peter Lang.

Panel: Advancing Studies in Sociolinguistic Variation and Change in German

From diglossia to diaglossia – Western speech repertoires in Austria revised

Alexandra N. Lenz, Johanna Fanta-Jende and Florian Tavernier (University of Vienna)

This presentation focuses on the German language area in Austria, more precisely on the hitherto less researched western part of the country. The key element of this talk is a hypothesized transition of the 'vertical' dialect/standard axis from a former diglossic towards a diaglossic structure. While in a diglossic spectrum dialect and standard language are structurally and functionally clearly separated from each other, a diaglossic spectrum is characterized by gradual transitions and 'intermediate' registers between the two extreme poles dialect and standard (cf. Auer 2011; Lenz 2010).

On the one hand, diaglossic spectra are assumed in Austria for the entire Bavarian language area, relying mostly on data from the Central Bavarian east of the country and the area around Austria's capital Vienna (cf. Lenz 2019). On the other hand, a diglossic constellation is generally assumed for the western part of Austria following the variation patterns of the adjacent Alemannic areas in German-speaking Switzerland (cf. Auer 2005: 15; Christen 2019: 273-275). For the neighboring High Alemannic Waldshut-Tiengen in southwestern Germany, however, Kehrein's (2012) results point to a synchronously observable de-diglossization (see analogously West Flemish, cf. Ghyselen 2007). This de-diglossization becomes apparent by an intergenerational comparison as diglossic language patterns are confirmed for the older generation while intersituational shifts on the dialect/standard axis characterize the language repertoires of young speakers.

The empirical basis of this talk, which is dedicated to vertical-social relations on the dialect/standard axis in Austria, is provided by the corpus of the Special Research Program "German in Austria. Variation – Contact – Perception" (cf. Lenz 2018). A total of 24 'autochthonous' speakers (age- and gender-balanced) from three rural locations in western Austria were recorded in six differing survey settings each. The aim of these settings was to evoke various registers of the individual spectra of linguistic variation, which should also enable cross-speaker comparisons. In this mix of methods, rather "free" conversational settings (conversations among friends and interviews) are supplemented by more standardized and controlled translation, reading and sentence completion tasks.

Our quantitative as well as qualitative analyses focusing on the phonetic-phonological level provide evidence for the hypothesis of a restructuring of the vertical spectrum "in vivo" from an older diglossic spectrum to a younger diaglossic model of the dialect/standard repertoires.

References

Auer, Peter. 2005. Europe's sociolinguistic unity, or. A typology of European dialect/standard constellations. In Delbecque, Nicole, Johan van der Auwera & Dirk Geeraerts (eds.). *Perspectives on Variation. Sociolinguistic, Historical, Comparative.* Berlin & New York: de Gruyter, 7–42.

Auer, Peter. 2011. Dialect vs. standard: A typology of scenarios in Europe. In Kortmann, Bernd & Johan van der Auwera (eds.). *The languages and linguistics of Europe: A comprehensive guide.* Berlin: De Gruyter Mouton, 485–500.

Christen, Helen. 2019: Alemannisch in der Schweiz. In Herrgen, Joachim & Jürgen Erich Schmidt (eds.). *Sprache und Raum. Ein internationales Handbuch der Sprachvariation.* Vol. 4: *Deutsch*. Berlin & Boston: de Gruyter Mouton, 246–279.

Ghyselen, Anne-Sophie. 2016. From Diglossia to Diaglossia: A West Flemish Case-Study. In Marie-Hélène Côté et al. (eds.). *The Future of Dialects*. Language Science Press, 35–62.

Kehrein, Roland. 2012. *Regionalsprachliche Spektren im Raum. Zur linguistischen Struktur der Vertikale*. Stuttgart: Franz Steiner.

Lenz, Alexandra N. 2010. Emergence of Varieties through Restructuring and Reevaluation. In Auer, Peter & Jürgen Erich Schmidt (eds.). *Language and Space. An International Handbook of Linguistic Variation*. Vol. 1: *Theories and Methods*. Berlin & New York: de Gruyter Mouton, 295–315.

Lenz, Alexandra N. 2018. The Special Research Programme „German in Austria. Variation – Contact – Perception“. In Ammon, Ulrich & Marcella Costa (eds.). *Yearbook Sociolinguistica* 32. Berlin & Boston: de Gruyter, 269–277.

Lenz, Alexandra N. 2019. Bairisch und Alemannisch in Österreich. In Herrgen, Joachim & Jürgen Erich Schmidt (eds.). *Sprache und Raum. Ein internationales Handbuch der Sprachvariation.* Vol. 4: Deutsch. (Handbücher zur Sprach- und Kommunikationswissenschaft. 30.4). Berlin & Boston: de Gruyter Mouton, 318–363.

From pronoun to zero marking:
variation in the expression of the 1st person subject in South Estonian Seto

Liina Lindström, Helen Plado, Maarja-Liisa Pilvik
University of Tartu

The expression of subject pronouns is a phenomenon which is known to vary significantly from one language to another with regard to both the available grammatical means for expressing the pronominal subjects (e.g., independent pronouns, affixes, clitics) and the obligatory *vs*. optional use of these means (Siewierska 2004, Dryer 2013).

In our paper, we set our focus on the Seto language – a South Estonian variety, which belongs to the Finno-Ugric language family –, and investigate the variation in the expression of 1st person singular subject pronouns in both affirmative and negative clauses. Estonian is similar to other Finno-Ugric languages in that it enables the expression of person both as independent pronouns as well as verbal suffixes, which are developed from pronouns (Janhunen 1982). However, Seto as a South Estonian variety makes almost no use of the verbal suffixes, and can simultaneously also drop the pronoun, which makes solving the person reference fully context-dependent.

We compare older Seto data in the Corpus of Estonian Dialects (CED), recorded in the 1960s-1970s, to newer Seto data from the past decade in order to 1) detect potential changes in the usage frequency of the overt 1SG subject pronoun, and 2) assess the effect and relative importance of different morphosyntactic, semantic, and cognitive/discourse-related factors characterizing the situational and contextual settings of the corpus usage events where reference to the 1st person is made and therefore, potentially affecting the choice between overt and covert pronoun use.

In contemporary spoken Estonian, pronouns occur with 82% of 1st person verb forms (Duvallon & Chalvin 2004), while there is remarkable variation in that regard in Estonian dialects (Lindström et al. 2009). Our earlier studies based on the CED data have shown that the use of explicit pronoun with the 1st person verb form can range from 24% to 71% with Seto having the lowest percentage. We hypothesize that the use of the 1st person pronoun has increased over time due to the rapid population decline in the Seto-speaking community and the growing influence of Estonian and Russian.

There is a multitude of factors whose individual and complex collective effects have been observed on the choice of subject pronoun expression in studies about Estonian, Finnish, and Karelian, for example (Lindström et al. 2009, Helasvuo 2014, Helasvuo & Kyröläinen 2016, Väänänen 2016, Uusitupa 2021). In this paper, we employ mixed-effects logistic regression analysis for assessing the potential effect of important factors from previous research on pronominal subject expression, such as tense, polarity, verb semantics, form, and distance of the previous reference to the 1SG, etc. We expect that the factors linked to cognitive processing of the sequential structure of the conversation emerge from the analysis as the strongest co-determinants of subject pronoun use in Seto. Distance from the previous reference to the 1st person, in particular, is expected to influence the outcome because contextual clues become more uncertain the more time has passed from the last explicit reference.

References:

Dryer, Matthew S. 2013. Expression of Pronominal Subjects. In: Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck

Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/101, Accessed on 2021-12-15.)
Duvallon, Outi & Antoine Chalvin. 2004. La réalisation zéro du pronom sujet de première et de deuxième personne du singulier en finnois et en estonien parlés. *Linguistica Uralica* 40(4). 270–286.
Helasvuo, Marja-Liisa 2014. Agreement or crystallization: Patterns of 1st and 2nd person subjects and verbs of cognition in Finnish conversational interaction. *Journal of Pragmatics* 63. 63–78.
Helasvuo, Marja-Liisa & Aki Kyröläinen 2016. Choosing between zero and pronominal subject: modeling subject expression in the 1st person singular in Finnish conversation. *Corpus Linguistics and Linguistic Theory* 12(2). 263–299.
Janhunen, Juha 1982. On the structure of Proto-Uralic. *Finnisch-ugrische Forschungen* 44. 23–42.
Lindström, Liina, Mervi Kalmus, Anneliis Klaus, Liisi Bakhoff & Karl Pajusalu 2009. Ainsuse 1. isikule viitamine eesti murretes. *Emakeele Seltsi aastaraamat* 54(2008). 159–185.
Siewierska, Anna 2004. *Person*. New York: Cambridge University Press.
Uusitupa, Milla 2021. When Finnish and Karelian person systems come together: Comparing open person constructions in Border Karelian dialects. Eesti ja soome-ugri keeleteaduse ajakiri ESUKA / Journal of Estonian and Finno-Ugric Linguistics 12(1). 397–431.
Väänänen, Milja 2016. *Subjektin ilmaiseminen yksikön ensimmäisessä persoonassa. Tutkimus suomen vanhoista murteista*. (Turun Ylipoiston Julkaisuja - Annales Universitatis Turkuensis - Ser. C 430). Turku: Turun yliopisto.

3. Comparing human and computer classification of phonetic features in the Scottish/English border region

Carmen Llamas, Dominic Watt (University of York) and Georgina Brown (Lancaster University)

The *Accent and Identity on the Scottish/English Border* (AISEB) project (UK ESRC RES-062-23-0525) examined the links between phonological variation and local, regional and national identities at the extreme ends of the political border between England and Scotland (Eyemouth and Berwick in the east; Gretna and Carlisle in the west). The border is said to coincide with the most tightly-concentrated bundle of dialect isoglosses in the English-speaking world, turning Scotland into a 'dialect island' (Aitken 1992). The border therefore represents a prime context for the investigation of language and identity. In this paper we will discuss one of the tests used as part of AISEB's speech perception strand. Under the researcher's supervision, participants were asked to classify short audio samples according to the perceived origin of the speaker as a way of gaining insights into the socio-geographical associations and relative salience of local pronunciation variants, to complement those obtained using the Social Category Association Test (SCAT) described in Llamas, Watt & MacFarlane (2016). As an additional, objective means of probing the notion of salience, we compare the findings for our human participants to those yielded by a prototype automated accent classification system, Y-ACCDIST (Brown & Wormald 2017).

Aitken, A.J. (1992). Scots. In McArthur, T. (ed.). *The Oxford Companion to the English Language*. Oxford: Oxford University Press, pp. 893-899.

Brown, G. & Wormald, J. (2017). Automatic sociophonetics: Exploring corpora with a forensic accent recognition system. *Journal of the Acoustical Society of America* 142 (1): 422-433.

Llamas, C., Watt, D. & MacFarlane, A. (2016). Estimating the relative sociolinguistic salience of segmental variables in a dialect boundary zone. *Frontiers in Psychology*, August 2016, 7: 1163. DOI: 10.3389/fpsyg.2016.01163.

A moribund Japanese colonial koiné in the Pacific:
a panel study of language obsolescence

Kazuko Matsumoto
University of Tokyo

David Britain
University of Bern

This paper reports early results from our research on structural obsolescence in the postcolonial Japanese variety spoken on Palau in the Western Pacific. During Japanese rule (1914-1945), radically different dialects of Japanese were brought by settlers, who eventually accounted for the vast majority of the population in the capital Koror. Due to intensive contact with the children of these settlers, many Palauan children acquired a Japanese colonial koiné as part of their linguistic repertoire. In 1945, all Japanese settlers were expatriated, with English becoming the official colonial language. Today just a few very elderly Palauans survive to remind us of the once vibrant Japanese speech community.

Two comparable data sets from 28 elderly Palauans (17 fluent- and 11 semi-speakers) collected at a 10-year interval (Dataset I collected in 2000 and Dataset II collected in 2010) are analysed in order to examine to what extent their vernacular Japanese acquired during childhood (in the case of semi-speakers) and consolidated in their late teens and early twenties (in the case of fluent-speakers) is stable or obsolescing, given increasingly rare opportunities to use Japanese. These two speaker groups enable us to assess the extent to which obsolescence proceeds differently when the original command of the language differed. The linguistic variable we investigate here is (g), with variants [g], [ŋ] and [k]; and consider a range of linguistic and social constraints on variation.

The analysis of Dataset I demonstrates that despite the absence of [g] in the Palauan phonetic inventory, Palauan Japanese speakers had (a) largely acquired the irregular Japanese Eastern Dialect system, which at the time of settlement had wide socio-spatial currency in Japan ([g] word-initially, but [ŋ] word-internally), (b) added the Palauan variant [k] for /g/ as a substrate feature. These results confirm the expected outcomes both of the founder principle (Mufwene 1996) and of intensive dialect contact (Trudgill 2004) as well as substrate effects.

In this paper, we present a comparative analysis of Datasets I and II to examine, as obsolescence progresses, in which direction Palauan Japanese is travelling: paying attention to the effects of substrate Palauan phonology, the demography and dialectology of Japanese settlers as well as the speakers' original command of the language, we show whether speakers orient more and more to Palauan phonology (i.e. a 'nativisation' route, Dorian 1978) or, instead, orient increasingly towards the majority ingredient dialects at the time of koiné formation (the 'concentration model' route, Wolfram 2002).

References:

- Dorian, N. (1978) The fate of morphological complexity in language death: Evidence from East Sutherland Gaelic, *Language* 54: 590-609.
- Mufwene, S. (1996). The founder principle in creole genesis. *Diachronica* 13(1): 83–134.
- Trudgill, P. (2004). *New-dialect formation: The inevitability of colonial Englishes*. Oxford: Oxford University Press.
- Wolfram, W. (2002). Language death and dying. In J. Chambers et al. (eds.) *Handbook of language variation and change*. Oxford: Blackwell. 764-788.

The limits of perception, Kevin B. McGowan, University of Kentucky

We have learned, from the last roughly 70 years of speech perception research, that variation is an often-useful aspect of a richly structured speech signal. Much of what used to be considered overwhelming to listeners or a troubling source of noise is now understood to be informative: whether variation is attributable to phonetic context (e.g. Beddor, McGowan, Coetzee, Boland, and Brasher 2013), talker differences (e.g. Nygaard and Pisoni 1998), or social information (e.g. Mack & Munson 2011; Sumner, Kim, King, and McGowan 2014). Listeners' knowledge of the sources of variability gives structure to the variation in speech and allows for the once bafflingly consistent percepts we observe across listeners. In this talk I will address an apparent paradox in the perception literature: that expert listeners are somehow simultaneously expert ignorers of and exquisite attenders to fine phonetic detail.

Every speech sound has multiple phonetic cues that listeners can use during perception. Many of these cues are sufficient to allow a particular percept, but none is indispensable. Lisker (1986), for example, offers 16 distinct acoustic features which can differentiate word-medial voiced and voiceless stops (but none of which, including the presence of voicing, is required for a voiced percept). Beddor has suggested that groups of listeners in a population will covertly attend to cues to the presence of a nasal consonant differently (Beddor 2009). And, indeed, the same speech sound in the same language will be cued differently for different listeners (Beddor et al 2013), and these differences are linked to how those same listeners then produce these phonetic cues when they take their turn as talkers (Beddor et al 2018).

Becoming an expert listener, either via L1 or L2 acquisition, has been construed as a systematic process of becoming an expert ignorer of cues and variation that are irrelevant to the speech sounds of one's target language. We see this in children (Narayan et al. 2010; Werker et al. 2012), bilingual adults (Schulman 1983), and we see it reflected in the difficulties people have acquiring new distinctions in an L2 (e.g. Baese Berk et al. 2020).

At the same time, listeners will use perceived (or implied) gender (Strand & Johnson 1996, Hay et al 2019), sexuality (Mack & Munson 2011), nationality (Niedzielski 1999, McGowan 2015; Gnevsheva 2017), emotion (Kim & Sumner 2017), age (Drager 2011, Hay et al. 2019), persona (D'Onofrio, 2015), or identity (McGowan & Babel 2020) conveyed by 'irrelevant' or sub-categorical phonetic information to inform perceptions of speech sounds and will use perceptions of speech sounds, interactively, to inform perceptions of social category (e.g. Lambert et al. 1960; Sumner et al. 2014).

Given the multiplicity of cues that exist for each category, the ways in which these cues overlap, and the ways in which these cues interact, we don't yet have a theory of language that allows us to refer to some speech cues as "linguistic" and to others as "non-linguistic" with any kind of confidence or rigor. Just as traditional speech perception had to let go of the idea of particular, invariant cues to each speech sound category, sociophonetic speech perception needs to let go of the idea that there are particular cues to social categories.

Baese‐Berk, M. M., McLaughlin, D. J., & McGowan, K. B. (2020). Perception of non‐native speech. Language and Linguistics Compass, 14(7), e12375.

Beddor, P. S. (2009). A coarticulatory path to sound change. Language, 785-821.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. The Journal of the Acoustical Society of America, 133(4), 2350-2366.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. Language, 94(4), 931-968.

Best, C. T. (1995). Learning to perceive the sound pattern of English. Advances in infancy research, 9, 217-217.

D'Onofrio, A. (2015). Perceiving personae: Effects of social information on perceptions of TRAP-backing. University of Pennsylvania Working Papers in Linguistics, 21(2), 5.

Drager, K. (2011). Speaker age and vowel perception. Language and Speech, 54(1), 99-121.

Gnevsheva, K. (2017). Within-speaker variation in passing for a native speaker. International Journal of Bilingualism, 21(2), 213-227.

Hay, J., Walker, A., Sanchez, K., & Thompson, K. (2019). Abstract social categories facilitate access to socially skewed words. PloS one, 14(2), e0210793.

Kim, S. K., & Sumner, M. (2017). Beyond lexical meaning: The effect of emotional prosody on spoken word recognition. The Journal of the Acoustical Society of America, 142(1), EL49-EL55.

Lambert, W. E., Hodgson, R. C., Gardner, R. C., & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*(1), 44.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling/b/versus/p/in trochees. Language and speech, 29(1), 3-11.

Mack, S., & Munson, B. (2012). The influence of/s/quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp'stereotype. Journal of Phonetics, 40(1), 198-212.

McGowan, K. B. (2015). Social expectation improves speech perception in noise. Language and Speech, 58(4), 502-521.

McGowan, K. B., & Babel, A. M. (2020). Perceiving isn't believing: Divergence in levels of sociolinguistic awareness. Language in Society, 49(2), 231-256.

Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. Developmental science, 13(3), 407-420.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. Journal of language and social psychology, 18(1), 62-85.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. Perception & psychophysics, 60(3), 355-376.

Schulman, R. (1983). Vowel categorization by the bilingual listener. PERILUS Working Papers III: 81–99.

Strand, E. A., & Johnson, K. (1996, October). Gradient and visual speaker normalization in the perception of fricatives. In KONVENS (pp. 14-26).

Werker, J. F., Yeung, H. H., & Yoshida, K. A. (2012). How do infants become experts at native-speech perception?. Current Directions in Psychological Science, 21(4), 221-226.

Exploring the effects of age and gender in sociolinguistic perception: Evidence from Tyneside English

Johanna Mechler, Universität Duisburg-Essen

Sociolinguistic research has shown that both intralinguistic factors (e.g., differences in the perception of morphosyntactic and phonetic variables, see Levon & Buchstaller 2015) and extralinguistic factors, such as age, socioeconomic status, and gender (e.g., Campbell-Kibler 2007), play a vital role in the perception process. The aim of the current study is to expand this paradigm by exploring how the perception of (ing), a well-documented stable linguistic feature, is mediated both age and gender, using sound samples from Tyneside English speakers.

Tyneside English is well described regarding the processes of variation and change in production (Watt 2002, Buchstaller & Corrigan 2015, inter alia). To date, however, the variety has only been subject to relatively few perception experiments, none of which examines the relevance of speaker and listener age for sociolinguistic perception of change in progress (Buchstaller & Levon 2015, Levon, Buchstaller, Mearns 2020). This study addresses this gap by exploring the relevance of maturation in the perception of stimuli-rich speech samples from the North East of England. It draws on naturally occurring sound samples from eight speakers who differ in age and gender. Careful sampling of the large LaVaLi corpus, that contains approximately 100 hours of sociolinguistic interviews, ensures that the guises were stringently controlled for sound quality, fundamental frequency as well as voice quality. Each speech stimulus is about ten seconds in length.

Using an online survey surface (Psychopy), the samples were played to two listener groups in an inter-speaker design: half of the informants listening to the stimuli were asked to rate the speakers on a “professionalism” scale in the context of a job application. The other half of informants were told that the speakers were introducing themselves to a local volunteer group via voice message. This listener group was then asked to respond according to perceived friendliness. Finally, participants answered questions regarding their attitudes towards the North East varieties (e.g. North, North East, Newcastle) and – to explore the impact of cognitive factors – they filled out the diagnostic questions of the BAPQ (Hurley et al. 2007). In a mixed effects regression model, the effects of age, gender, attitudes, and results of the BAPQ were tested. Results suggest that the perception of vernacular linguistic features is highly contingent on both speaker and listener age.

While previous work has focused on the effect of (perceived) speaker gender, the present paper reveals the intersectional nature of gender and age in sociolinguistic perception. These findings allow us to develop a more holistic understanding of the cognitive underpinnings of age-related variability in language perception. They also produce important perceptual information to inform our models of language change across the lifespan.

References:

Buchstaller, Isabelle & Karen P. Corrigan. 2015. Morphosyntactic features of Northern English. In Raymond Hickey (ed.), Researching Northern English, 71-98. Amsterdam; Philadelphia: John Benjamins.

Campbell-Kibler, Kathryn. 2007. Accent, (ING), and the social logic of listener perceptions. American Speech 82(1): 32-64.

Hurley, Robert, Molly Losh, Morgan Parlier, J. Steven Reznick & Joseph Piven. 2007. The Broad Autism Phenotype Questionnaire. Journal of Autism and Developmental Disorders 37. 1679-1690.

Levon, Erez and Isabelle Buchstaller. 2015. Perception, cognition, and linguistic structure: The effect of linguistic modularity and cognitive style on sociolinguistic processing. Language Variation and Change 27(3). 319-348.

Levon, Erez, Isabelle Buchstaller & Adam Mearns. 2020. Towards an integrated model of perception: Linguistic architecture and the dynamics of sociolinguistic cognition. In Karen Beaman, Isabelle Buchstaller, Susan Fox & James Walker (eds.), Advancing socio-grammatical variation and change: In honour of Jenny Cheshire, 32-54. London: Routledge.

Watt, Dominic. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. Journal of Sociolinguistics 6(1). 44-63.

Daniela Mereu & Silvia Dal Negro

(Free University of Bozen-Bolzano)

Different methods for spontaneous speech elicitation: conversational and linguistic issues

This contribution aims to discuss some methodological issues related to the influence of the different techniques of spontaneous data elicitation on the conversational and linguistic structure.

The notion of spontaneity refers to the mode of speech production in relation to the conditions of data elicitation (Warner 2012). Spontaneous speech covers a wide range of speech production modes, because it includes anything that is not read or prepared, such as Map Tasks, interviews and conversations. However, different data elicitation strategies will produce changes in the speech event; consequently, we expect that different types of spontaneous speech show linguistic differences due to the strategy used.

To investigate this hypothesis, we focused on three recordings of Italian spontaneous speech, obtained in different ways: a semi-structured interview (duration: 36'57''; effective speech: 32'1''; tokens: 5635), a dialogue (duration: 33'02''; effective speech: 27'; tokens: 4713), and a Map Task (duration: 11'54''; effective speech: 9'4''; tokens: 1636). To limit the sources of variation, the same pair of speakers has been recorded for the three events: two 50-year-old friends, native speakers of Italian, born and living in Bolzano.

First, we analyzed some indicators considered as cues of spontaneity, such as the amount of silence, speech overlaps, disfluencies (interruptions and repetitions), laughter and speech laugh, speech rate (Kouwenhoven et al. 2018). The values obtained were normalized for the duration of each recording or the amount of effective speech, depending on the parameter examined. We found significative differences among the three conversations, related to their structure. We will give here only an example.

The highest amount of interruptions (1.83%) and overlaps (5.14%) for the Map Task, especially compared to dialogue (0.89% and 1.49%, respectively), can be explained with the different tasks requested: while in the first the speakers have a specific goal to reach and they need to be precise with the indications to their interlocutor, in the dialogue, speakers had plenty of time to manage the conversation and there was no a specific task, except that of talking about different topics.

Secondly, we will provide an analysis on the distribution of discourse markers (DMs) in the three conversations, distinguishing among different functions, i.e. interactional, metatextual and cognitive (Bazzanella 1995). An exploratory investigation showed that among the interactional DMs, in all conversations, turn-taking devices and back-channel expressions are used (e.g. *allora* 'so', *capito* 'understood'), but in the Map Task we reported the absence of connectives with a phatic function underling a shared knowledge (e.g. *sai* 'you know'), used both in dialogue and interview. Furthermore, the metatextual connectives, related to the organization of the ongoing discourse (Fraser 2009), tend to be used much more in the dialogue and interview rather than in Map Task. As the results seem promising, we will extend the conversational and DMs analysis to a larger dataset including three recordings for every speech event.

Although preliminary, these results show that the data elicitation method can be a relevant factor in the analysis of the distribution and the frequency of certain linguistic phenomena related to the conversational structure.

References

Bazzanella, C. (1995), “I segnali discorsivi”. In Renzi, L., Salvi, G. & Cardinaletti, A. (eds.), Grande grammatica italiana di consultazione. Bologna: Il Mulino, 225–257.

Fraser, B. (2009), “An Account of Discourse Markers”. International Review of Pragmatics 1: 293–320.

Kouwenhoven, H., Ernestus, M. & van Mulken, M. (2018), “Register variation by Spanish users of English. The Nijmegen Corpus of Spanish English”. Corpus Linguistics and Linguistic Theory 14 (1), 35–63.

Warner, N. (2012), “Methods for studying spontaneous speech”. In Cohn, A., Fougeron, C. & Huffman, M. (eds.), Handbook of Laboratory Phonology. Oxford: Oxford University Press, 612–633.

The relationship between ratings data and feature recognition in a dialect of English
Chris Montgomery, University of Sheffield, UK

This paper presents research relating to the noticeability of regional features in two samples of speech from the Isles of Scilly. The Isles of Scilly are a group of islands off the South West coast of England, which have a population of around 2,200 people. The islands have an interesting settlement history, with compulsory education introduced there before the rest of England. The variety of English spoken there has for centuries attracted comments that relate to its 'refinement' and difference to the English spoken in its near neighbour, Cornwall. The following quotes provide examples of such commentary:

(1) …the Language of Scilly refines upon what is spoken in many Parts of Cornwall; probably from the more frequent Intercourse of the Inhabitants, some more than others, with those who speak the Standard English best… (Heath 1750: 436)

(2) The Islanders are remarkable for speaking good English—far preferable, at least, to what is generally heard amongst the humbler classes of any county, at some distance from the metropolis… (Woodley 1822: 105)

(3) The English spoken today (1979) by natives of the Isles of Scilly … is scarcely removed from Standard (southern) English, using a slightly modified 'received pronunciation' (R.P.) as of educated persons. (Thomas 1979: 109)

Research conducted by Moore and Carter (2015; 2017; 2018) has demonstrated that, far from being unmarked and dissimilar to the English spoken in Cornwall, Scillonian English is actually quite similar to that variety, especially amongst residents of the islands who were educated on the islands. The persistence of the notions of 'Standardness' and unmarkedness in contemporary metalinguistic commentary about Scillonian English (e.g. Taylor 2016) sets up an interesting question: are listeners able to notice the regionality that exists in Scillonian speech?

This paper presents new analysis of data collected using a method that permitted 103 respondents to listen to and rate two guises constructed from an interview with an older male Scillonian speaker before identifying in real time features that they considered to be regional. Respondents were then able to review these reactions in order to clarify what they had reacted to. This produced a dataset comprising ratings data and reactions data for each guise. The two guises differed primarily by topic, with one guise dealing with Scillonian island life, and the other focussing on (non location-specific) farming activities.

I focus on the ratings data, and the main ratings components of 'Status' and 'Solidarity', as well as the topic and location cues in each guise, in order to explain the different levels of attention paid to different features in the guises. To do this, I address notions of awareness and 'salience' (Drager & Kirtley 2016), noticeability, priming, and 'surprisal' (Rácz 2013). In doing so, I explore the relationship between ratings data and topic, context, and stereotypes, and how all of these factors interact with the ways in which listeners notice regional features in speech.

References

Drager, Katie & M. Joelle Kirtley. 2016. Awareness, Salience, and Stereotypes in Exemplar-Based Models of Speech Production and Perception. In Anna M. Babel (ed.), *Awareness and Control in Sociolinguistic Research*, 1–24. Cambridge: Cambridge University Press.

Heath, Robert. 1750. *A Natural and Historical Account of the Islands of Scilly; Describing their Situation, Number, Extent, Soil, Culture, Produce, Rareties, Towns, Fortifications, Trade,*

Manufacture, Inhabitants. Their Government, Laws, Customs, Grants, Records, and Antiquities. London: R. Manby & H. S. Cox.
Moore, Emma & Paul Carter. 2015. Dialect contact and distinctiveness: The social meaning of language variation in an island community. *Journal of Sociolinguistics* 19(1). 3–36.
Moore, Emma & Paul Carter. 2017. 'The Land Steward Wouldn't Have a Woman Farmer.' In Chris Montgomery & Emma Moore (eds.), *Language and a Sense of Place: Studies in Language and Region*, 258–280. Cambridge: Cambridge University Press.
Moore, Emma & Paul Carter. 2018. Natural phonetic tendencies and social meaning: Exploring the allophonic raising split of price and mouth on the Isles of Scilly. *Language Variation and Change* 30(3). 337–360.
Rácz, Péter. 2013. *Salience in sociolinguistics: A quantitative approach*. Berlin: Mouton de Gruyter..
Taylor, Colin. 2016. *The Life of a Scilly Sergeant*. Century.
Thomas, Charles. 1979. A glossary of spoken English in the Isles of Scilly. *Journal of the Royal Institute of Cornwall* 8. 109–47.
Woodley, George. 1822. *View of the Present State of the Scilly Islands: Exhibiting their Vast Importance to the British Empire*. London: London: Longman and Co.

New visualization techniques in dialect geography: The merging of analytic and synthetic mapping in the virtual and interactive cartography of VerbaAlpina

Christina Mutter | Project VerbaAlpina, Munich University (LMU)

Cartography is the most plausible form of representation of dialect variation and - indirectly - of dialectal change. In dialectology, two traditions have developed since the 19th century that use 'analytic' or 'synthetic' maps both of which have advantages and disadvantages. Analytic maps are typical of the Romance tradition; they reflect the single linguistic utterances as a whole, as for example is the case of the linguistic atlas of Italy and Southern Switzerland (AIS). Here the focus is on the documentation of the single utterances and the reader has to understand the geographical distribution of more abstract types from the single utterances by himself. This form of visualization has the advantage that it provides very detailed information that is reliable to the source, traceable and verifiable; analytic atlases are dialectal corpora avant la lettre. On the other hand, analytic maps are rather confusing due to the large number of utterances. The so-called synthetic maps, which mostly use point symbols for visualization, primarily represent the Germanic tradition, as for example the linguistic atlas of Vorarlberg (VALTS); in this case, single utterances are documented only occasionally if they are strongly marked. However, the geographical distribution of selected features of the single utterances is directly apparent through the allocation of symbols. Quantitative relationships, especially those of dialectometric nature, can only be represented in this way. A disadvantage due to the feature selection is the rather limited transparency of synthetic maps which are also very suggestive.

In order to address the described disadvantages of traditional cartography, mapping methods are needed that make it possible to merge the two mentioned visualization traditions. Such a mapping tool, for example, is the virtual interactive map of the DFG long-term project VerbaAlpina which investigates the Alpine region in its cultural and linguistic unity. The online mapping of VerbaAlpina (https://www.verba-alpina.gwi.uni-muenchen.de/) is based on georeferenced linguistic utterances and combines the two methods. While at first glance 'synthetic' maps are displayed, at second glance, after clicking on the individual symbols on the map, the user also has access to the single utterances, which guarantees empirical transparency; in addition, the utterances are enriched with metadata and links to reference dictionaries. The visualization of language data via the interactive map also allows to combine different data sources and to display not only data from linguistic atlases but also data from dictionaries and crowdsourcing. Thus, via the interactive map, vocabulary can be viewed not only from an onomasiological perspective (as in traditional geolinguistics) but also from a semasiological one. The visualization of the linguistic data can always be done qualitatively as well as quantitatively. In addition, the interactive map offers the function of creating so-called synoptic maps. This function enables the user to save a certain selection of maps on a synoptic combination map in order to visualize the range of any linguistic and non-linguistic features in context.

This talk will present the described visualization techniques offered by the virtual interactive map of VerbaAlpina in more detail.

References

Gabriel, Eugen (1985-2004): Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus, vol. 1-5, Vorarlberger Landesbibliothek, Bregenz

Jaberg, Karl / Jud, Jakob (1928-1940): Sprach- und Sachatlas Italiens und der Südschweiz, Zofingen, vol. 1-7

Krefeld, T.: s.v. "Kartographie", in: VerbaAlpina-de 19/1, Methodologie, https://doi.org/10.5282/verbaalpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DK%2334

Krefeld, T.: s.v. "Synoptische Karte", in: VerbaAlpina-de 19/1, Methodologie, https://doi.org/10.5282/verbaalpina?urlappend=%3Fpage_id%3D493%26db%3D191%26letter%3DS%2356

Lameli, Alfred/ Kehrein, Roland/ Rabanus, Stefan (2010): Language and Space. Vol. 2: Language Mapping. An international Handbook of Linguistic Variation, Berlin/New York, De Gruyter

Stephan Lücke (2019): Einführung in die Geolinguistik (ITG/slu). Lehre in den Digital Humanities. Version 1 (13.01.2019, 09:44). URL: https://www.dh-lehre.gwi.uni-muenchen.de/?p=78314&v=1.

Comparative study of 2nd person singular address in Spanish: A translation task study

Due to challenges with analysis of second person singular (2PS) variation in extemporaneous corpora, quantitative studies commonly rely on elicitation. Self-report questionnaires (e.g., Brown & Gilman 1960) are the most frequent method, but discourse completion tasks (Newall 2016), role plays (Lamanna 2012), and subjective reaction tests (Moyna & Loureiro-Rodríguez 2018) are also used. Studies usually explore single local sites, so cross-linguistic and cross-dialectal comparisons (Carricaburo 2015) are necessarily speculative. Artificiality, in particular participants' focus on form, are also concerns.

We present early results of a cross-dialectal study of Spanish 2PS in Colombia, Mexico, Spain, Chile, and New York City using an innovative translation task designed to mitigate those challenges. Participants (n=472) identified as having sufficient bilingual proficiency received 50 items consisting of a picture of one person talking to another, accompanied by an introduction and English quote, which they were instructed to translate, containing *you*. They were not told of the purpose of the task.

Raw data are exponents of the Spanish 2PS variants—pronouns or verbal agreement morphemes—they produced translating *you* classified as:

- *ustedeo* (formal),
- *tuteo* (informal)
- voseo (informal)
- sumercedeo (varies)

Analysis constructed one Lmer model for participant effects, and one model per region of scene effects.

Participant Model: Fixed factors included Region, Gender, Social Class, Age, Speech Act Type, and Setting. Participant was a random factor.

Scene Models: Fixed factors included Gender, Relative Ages, Affect, Relationship Distance, Status, Setting (e.g., home) and Relationship Types (e.g., granddaughter to grandmother); Scene was a random factor.

Overall, *tuteo* dominated responses, and *voseo* and *sumercedeo* were rare even in Chile and Colombia where they are used. Participant analysis identified Region and Age as significant. Regional percentages of *ustedeo* follow:

Colombia	Chile	Mexico	NYC	Spain
38.1%	25.4%	18.7%	14.0%	9.8%

Age showed slow monotonic decline in *ustedeo*. Both tendencies conform to prior research.

Scene factors showed expected co-linearity of Relative Ages, Social Distance, and Relative Status. The question was which would be picked for the best-fit model. These varied as shown below; either speaking lower to higher in Status or with greater Social Distance predicted more *ustedeo*.

Colombia Status	Chile Status	Mexico Distance	NYC Distance	Spain Distance

We conclude that the instrument successfully compares trends across these Spanish varieties in 2PS variation regarding *ustedeo* versus informal variants. It confirms prior speculations regarding regional tendencies in *ustedeo* rates and its apparent time decline. However, *voseo*, which is stigmatized particularly in Chile, is clearly underrepresented. Apparently, the mental models used by participants in their translations do not reflect actual usage despite our efforts to capture informality in scenes. We were not entirely able to overcome the artificiality problem.

Nevertheless, the tasks successfully tease out factors—Relative Age, Status, and Social Distance—that have often been lumped together under notions of politeness or power and solidarity. These results show differences in priority for those factors cross-regionally. As such, our research invites further explorations for the determinants of those factors as we expand coverage to other Spanish varieties and different languages.

Brown, Roger and Albert Gilman. 1960. The pronouns of power and solidarity. In Thomas A. Sebeok (ed.), *Style in Language,* 253-276. Cambridge: MIT Press.

Carricaburo, Norma. 1997. *Las fórmulas de tratamiento en el español actual* (2nd edn.). Madrid: Arco Libros.

Lamanna, Scott. 2012. Second person singular pronoun use during closed role plays by Colombians in the North Carolina Piedmont triad. In Richard J. File-Muriel and Rafael Orozco (eds.), *Colombian Varieties of Spanish,* 169-203. Madrid/Frankfurt am Main: Iberoamericana/Vervuert.

Moyna, María Irene & Loureiro-Rodríguez, Verónica. 2018. La técnica de máscaras emparejadas para evaluar actitudes hacia formas de tratamiento en el español de Montevideo. *Revista Internacional de Lingüística Iberoamericana*, *15*(2 (30)), 47–82.

Newall, Gregory. 2016. Second person singular forms in Cali Colombian Spanish: Enhancing the envelope of variation. In María Irene Moyna and Susana Rivera Mills (eds.), *Forms of Address in the Spanish of the Americas,* 149-170. Amsterdam: John Benjamins.

When your brain tricks your ear

Nancy Niedzielski
Rice University

This paper reports on several experimental techniques that allow researchers to determine how and to what degree language attitudes underlie speech perception. These techniques have revealed complementary processes: first, expectations regarding features that listeners believe should be present, based on phenomena such a visual cues or knowledge of speaker social categories, can lead to the perception of those features, regardless of whether or not they actually occur in the stimuli. Experiments based on the well-known McGurk Effect, whereby visual cues regarding consonant production influence and can even override auditory cues, reveal that speech perception is more than a matter of acoustic processing. We are not surprised, then, when other types of information are integrated with auditory information. Numerous types of experiments reveal that information regarding speaker age, gender, ethnicity, and region can affect how speech stimuli is perceived. For instance, different photographs presented in conjunction with the same auditory stimuli can tell us how preconceived beliefs regarding who is producing that stimuli affects its perception. Thus, beliefs and attitudes affect how listeners interpret acoustic signals, which can help listeners to retrieve acoustic information that is, for instance, masked by noise, but can also lead to perceptions that do not match the actual acoustics present in the signal.

These perceptual experiments can reveal that the reverse is also possible: various techniques can reveal cases where, on the one hand, language attitudes research suggests that listeners are not aware of acoustic features, but experimental tests show that these same listeners in fact are aware of them. Respondents may suggest, for instance, that they are unaware of vowel differences between speakers of different social groups, or that they do not notice features such as mergers, while in perceptual tests, knowledge of such phenomena is uncovered. That is, listeners often have an implicit knowledge of features in speakers' varieties that they are unable to report explicitly. We discuss the findings and implications of previous work on how attitudes towards and beliefs about language variation affect perceptual processes and also suggest new ways that this might be explored, including how older techniques might be modified to gain an even greater understanding of the integration of the social with other types of information involved in human speech perception.

The 'atlas of everyday German in Salzburg':
An update on (areal) variation in a central region of Austria

Konstantin Niehaus (Salzburg), Julian Blaßnigg (Salzburg), Peter Mauser (Salzburg)

'Everyday' language encompasses all 'lects' which are spoken in informal and routine settings and can be produced spontaneously (cf. Elspaß 2010: 419). 'Everyday' German differs between the German-speaking countries and regions: While speakers in northern Germany tend to use (near-)standard German, speakers in Switzerland chiefly speak local dialects; today's situation in southern Germany and Austria is diaglossic, i.e. speakers shift between standard, dialect, and a broad spectrum of recent regiolects (cf. Kehrein 2019). The macro-areal variation found in 'everyday' German was first documented in the WDU (*Wortatlas der deutschen Umgangssprachen*) and, more recently, is being documented since the 2000s by the AdA (*Atlas zur deutschen Alltagssprache*), in particular on the lexical level (cf. also König, Elspaß & Möller 2020: 232–245).

Yet, both atlases cover the variation of more restricted areas only at large but not in greater detail. This holds also true for the central Austrian federal state of Salzburg where everyday German ranges from several different dialects like West Central, South Central, and South Bavarian to more widespread regiolects to areal standard German, frequently spoken especially in Salzburg city (cf. also Lenz 2019).

We will present a project (funded by the government of Salzburg, 2020–2023) which will establish a new atlas for the Austrian region of Salzburg, the *Atlas zur Salzburger Alltagssprache* (ASA, 'atlas of everyday German in Salzburg'). It draws on the method of the AdA and utilizes an online questionnaire on 76 mostly lexical but also phonetic and grammatical features. Data collection was conducted in co-operation with the biggest regional newspaper (*Salzburger Nachrichten*) in four rounds in late 2019 and includes social information on the participants' residence, age, gender, level of education, profession, mobility, and family heritage. In contrast to earlier dialectological studies on the region (cf. Scheutz 2007, 2016), our questionnaire reached a huge response of 1,500 to 5,000 participants each round.

Thus, the data allows for quantitative sociolinguistic and geolinguistic analysis (based on the approaches in Pickl et al. 2019 and Pröll et al. 2021). We will present first results in that regard and concentrate on lexical regional variation across Salzburg, particularly on the level of infrastructural 'planning regions' (*Planungsregionen*) since the Alpine landscape should be a decisive factor. In addition, we will discuss social factors where applicable and address the

distinction of 'lects' in Salzburg, i.e. we will give empirical insights on the diaglossic situation based on apt items of the questionnaire. As the data also allows for a more detailed cartography than previous studies, we will briefly touch on issues of visualization that arise with the tools ArcGIS and Geoling (https://www.geoling.net).

Ultimately, our goal will be a popular book publication in the fashion of Leeman et al. (2018) which does not, unlike earlier atlases, impose the traditional opposition (local dialects spoken by older people *versus* regiolect or standard spoken by younger people) but instead takes into account the actual diaglossic situation.

References:

AdA = Elspaß, Stephan & Robert Möller. 2003–2021. *Atlas zur deutschen Alltagssprache*. URL: http://www.atlas-alltagssprache.de/ [10.12.2021].

Elspaß, Stephan. 2010. Alltagsdeutsch. In Hans-Jürgen Krumm, Christian Fandrych, Britta Hufeisen & Claudia Riemer (eds.), *Deutsch als Fremd- und Zweitsprache. Ein internationales Handbuch*. 2nd, revised edition. Vol. 1. Berlin, New York: de Gruyter. 418–424.

Kehrein, Roland. Areal Variation im Deutschen „vertikal". In Joachim Herrgen & Jürgen Erich Schmidt (eds.), *Sprache und Raum. Ein internationales Handbuch der Sprachvariation.* Band 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt. Berlin, Boston: de Gruyter. 121–158.

König, Werner, Stephan Elspaß & Robert Möller. [19]2020. *dtv-Atlas deutsche Sprache*. München: dtv.

Leemann, Adrian, Stephan Elspaß, Robert Möller & Timo Grossenbacher. 2018. *Grüezi, Moin, Servus. Wie wir wo sprechen*. Reinbek bei Hamburg: Rowohlt.

Lenz, Alexandra N. 2019. Alemannisch und Bairisch in Österreich. In Joachim Herrgen & Jürgen Erich Schmidt (eds.), *Sprache und Raum. Ein internationales Handbuch der Sprachvariation.* Band 4: Deutsch. Unter Mitarbeit von Hanna Fischer und Brigitte Ganswindt. Berlin, Boston: de Gruyter. 318–363.

Pickl, Simon, Simon Pröll, Stephan Elspaß & Robert Möller. 2019. Räumliche Strukturen alltagssprachlicher Variation in Österreich anhand von Daten des „Atlas zur deutschen Alltagssprache (AdA)". In Lars Bülow, Ann Kathrin Fischer & Kristina Herbert (eds.), *Dimensionen des sprachlichen Raums: Variation – Mehrsprachigkeit – Konzeptionalisierung*. Frankfurt am Main, Oxford u.a.: Peter Lang. 39–59.

Pröll, Simon, Stephan Elspaß & Simon Pickl. 2021. Zum variationslinguistischen Verhältnis von Stadt und Land. Ein Beitrag zu einer sprachlichen Raumtypologie am Beispiel Österreichs. *Linguistik Online* 110 (05/2021), 293–329.

Scheutz, Hannes (ed.). 2007. *Drent und herent. Dialekte im salzburgisch-bayerischen Grenzgebiet. Mit einem sprechenden Dialektatlas auf CD-ROM*. Unter Mitarbeit von Sandra Aitzetmüller und Peter Mauser. Salzburg: Salzburg – Berchtesgadener Land – Traunstein EU Regio.

Scheutz, Hannes (ed.). 2016. Salzburger Sprachatlas. URL: https://www.sprachatlas.at/salzburg/index.html [10.12.2021].
WDU = Eichhoff, Jürgen: *Wortatlas der deutschen Umgangssprachen*. Vol. I (1977), vol. II (1978). Bern: Francke; vol. III (1993). München: Saur; vol. IV (2000). Bern, München: Saur.

Title: Minimal minimal pairs: Phonetic contrast in Unterland Yiddish vowels

Presenters: Chaya R. Nove, Ben Sadock

Abstract: Among Eastern European Yiddish dialects, the varieties spoken in the sector of the Transcarpathian region known in Yiddish as the Unterland, namely, the border area of present-day Slovakia, Hungary, Romania and Ukraine, have been the least studied (Sadock & Masor, 2018; Weinreich, 1964). Scholars that did examine Unterland Yiddish (UY) classified it, along with the Yiddish of Poland, as Central Yiddish based on a number of distinctive features, including a length contrast in the peripheral vowels. These studies also found considerable variation and dialect mixing overall, a likely consequence of the area's unique geographical, political, cultural and linguistic circumstances (see e.g., Krogh, 2012; Weinreich, 1964).

The present study maps the UY vowel system in phonetic space and analyzes the acoustic correlates of the contrast in the long-short vowel pairs {/iː/, /i/} e.g., as in /ziːn/ 'son' and /zɪn/ 'sun', {/uː/, /u/} e.g., /ʃtruːf/ 'punish' and /ʃlʊf/ 'sleep,' and {/aː/, /a/} e.g., /haːnt/ 'today' and /hant/ 'hand'. The data consist of audio segments extracted from recordings of Holocaust testimonies of twelve survivors from the Transcarpathian region of Eastern Europe, conducted on behalf of the USC Shoah Foundation Visual History Archives between 1996 and 1998. The audio was transcribed and segmented using an acoustic model trained on a larger Yiddish dataset. The duration and first and second formant frequencies of the vowels were then extracted using Fast Track, a new Praat plug-in, and analyzed statistically. The results show slight systematic differences in spectrum and a surprisingly weak contrast in duration, especially for the /i/ and /a/ pairs. Furthermore, regression models show a significant effect of gender, with female speakers exhibiting smaller durational differences between long-short /i/ and /a/ than males. Based on these patterns, we hypothesize that the length contrast in the pre-war Yiddish of the Transcarpathian region was undergoing change and possibly on the verge of collapse.

While vowels have long been at the center of scholarship on Yiddish phonology and vowel length is one of the primary features distinguishing between the northern and southern dialects of Eastern European Yiddish, this project is the first to analyze these vowels acoustically. In fact, Yiddish may be the sole Germanic language whose sound system has not yet been subjected to a thorough acoustic analysis (see, however, Bleaman 2018 for an analysis of stop consonants and Kleine 1998 on Standard Yiddish [g] and [ɔ]). This project thus fills an important gap in Yiddish linguistics, while also illustrating how archival recordings can be utilized to increase our understanding of understudied dialects. Moreover, the results of this study have implications for dialect classification, exposing systematic differences among the subdialects of the Central Yiddish dialect region that may have resulted from language contact and geopolitical change. Finally, this phonetic description of Unterland Yiddish can function as a baseline for examining contemporary dialects of Yiddish that derive from that region (e.g., Hasidic Yiddish of New York).

References

Beranek, F. (1936). [Yiddish in Czechoslovakia]. *YIVO Bleter*, *IV*, 63–75.
Bleaman, I. L. (2018). *Outcomes of minority language maintenance: Variation and change in*

New York Yiddish [Doctoral Disseration]. New York University.
Kleine, A. (1998). Toward a "Standard Yiddish Pronunciation." An Instrumentally Aided Phonetic Analysis. In M. S. Schmid, J. R. Austin, & D. Stein (Eds.), *Historical Linguistics 1997: Selected papers from the 13th International Conference on Historical Linguistics (ICHL), Düsseldorf, 10-17 August* (pp. 201–211). Amsterdam; Philadelphia: John Benjamins Publishing.
Krogh, S. (2012). How Satmarish is Haredi Satmar Yiddish? In M. Aptroot, E. Gal-Ed, R. Gruschka, & S. Neuberg (Eds.), *Leket: Yidishe shtudyes haynt/Jiddistik heute/Yiddish Studies Today* (p. 664). Düsseldorf University Press.
Sadock, B., & Masor, A. (2018). Bobover Yiddish: "Polish" or "Hungarian?" *Journal of Jewish Languages*, *6*, 89–110.
Weinreich, U. (1964). Western traits in Transcarpathian Yiddish. In *For Max Weinreich on his Seventieth Birthday, Studies in Jewish Languages, Literature and Society* (pp. 245–264). The Hague: Mouton.

Toward an In-Depth Analysis of Subject Pronoun Expression: A View from Xalapa, Mexico

Rafael Orozco, Paul Kidhardt & Robert Bayley

The alternation between null and overt pronominal subjects, i.e., subject pronoun expression (SPE) has been investigated by variationist sociolinguists since the 1970s. The effects of the verb on SPE have been explored since the 1980s, mainly in terms of lexical categories (Bentivoglio 1980; Enríquez 1984), establishing that psychological or mental activity verbs favor overt subjects while external activity verbs favor null subjects (Carvalho et al. 2015). Meanwhile, social constraints on SPE have not been sufficiently explored (Orozco & Hurtado 2021). This study seeks to answer questions that emerged from increased SPE research during the 21st century. We probe the effects of eight predictors (four internal and four external) on 3,760 tokens from a socially-stratified corpus of Spanish spoken in Xalapa, Mexico. The constraints explored include grammatical person and number, TMA, age, sex, education, and socioeconomic status.

Our results reveal an overall pronominal rate of 25% and internal conditioning congruent with findings throughout the Hispanic World (cf. Carvalho et al. 2015; Lastra & Martín-Butragueño 2015; Otheguy & Zentella 2012). Grammatical number and person exerts the strongest pressure among linguistic constraints while speaker age does so among social predictors. The robust effect of age sets Xalapa apart from most other Hispanic speech communities, as SPE is not known to be strongly conditioned by social predictors (Carvalho et al. 2015:). When age does condition SPE, its effect is not as strong as it is in this speech community. Concurrently, the pronominal rate found among teenagers (10%) is below the lowest overall pronominal rate in the Hispanic world. The lower pronominal rate among younger speakers is consonant with findings in other varieties of Spanish such as Dominican (Alfaraz 2015), Mexico City (Lastra & Martín-Butragueño 2015), Colombian (Orozco & Hurtado 2021), and Peninsular Spanish (Prada Perez 2015). Moreover, circumstantiated analyses of the intersection between age, education, socioeconomic status and grammatical person and number uncovers that speakers currently in college favor the first person plural while those with a postgraduate education favor the first person singular. This finding suggests that university students seem to have an increased sense of collectivity whereas those with further education appear more individualistic. The effects of pronominal subject + verb collocations reveal, among other things, opposing tendencies for forms corresponding to a single verb. For instance, *era* 'I was' promotes overt subjects whereas *son* 'they are' has the opposite effect. Thus, our analysis debunks the premise that the effect of the verb on SPE can be accurately analyzed in terms of lexical categories or verb infinitives.

Our analytical approach is both flexible and fine-grained, thus improving our explanatory power of pronominal expression. We use binary logistic multilevel models and logistic regression followed by average marginal effects for SPE to provide a more intuitive, high resolution comparative panorama than prior studies. Marginal results improve interpretability, reliability between studies, and causal relevance, especially for interactions. Further, the study of the effects of age shall help determine, among other things, whether we are in the presence of an acquisitional feature also possibly present in other pro-drop languages. (500 words)

References

Alfaraz, Gabriela. 2015. Variation of Overt and Null Subject Pronouns in the Spanish of Santo Domingo. In Ana M. Carvalho, Rafael Orozco y Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*, 3-17. Washington DC: Georgetown.

Bentivoglio, Paola. 1980. *Why* canto *and not* yo canto*? The problem of first-person subject pronoun in spoken Venezuelan Spanish*. Los Angeles: University of California Los Angeles M.A. Thesis.

Carvalho, Ana M., Rafael Orozco & Naomi Lapidus Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington: Georgetown University Press.

Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.

Lastra, Yolanda & Pedro Martín Butragueño. 2015. Subject Pronoun Expression in Oral Mexican Spanish. In Ana M. Carvalho, Rafael Orozco y Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*, 41-59. Washington DC: Georgetown University Press.

Orozco, Rafael & Luz Marcela Hurtado. Variable subject pronoun expression revisited: This is what the Paisas do. *Proceedings of the Linguistic Society of America* 6(1) (20 March 2021). 713-727. https://doi.org/10.3765/plsa.v6i1.5006

Otheguy, Ricardo. & Ana. C. Zentella 2012. *Spanish in New York: language contact, dialectal leveling, and structural continuity*. Oxford: Oxford UP.

Tracking language change in real time: Challenges for community-based research in the 21st century

Katharina Pabst, Sam Brunet, Alison Chasteen, Sali A. Tagliamonte
University of Toronto

Tracking language change in real time is challenging (but see Cukor-Avila & Bailey 2017; Nahkola & Saanilahti 2004; Sankoff 2017; Sundgren 2002; *inter alia*) because projects that attempt it are incredibly "difficult, time-consuming, and expensive" (Cieri & Yaeger-Dror 2017:53). Researchers who have successfully built longitudinal corpora (e.g. G. Sankoff and colleagues) report numerous problems, including difficulties tracing participants, attrition due to mortality, and a lack of resources that prohibits long-term planning (Sankoff 2017). However, we discovered that there are additional challenges in the 21st century. In this presentation, we discuss the many hurdles we faced in building a longitudinal corpus in a large North American city. Data collection took place before the Covid-19 pandemic, but the lessons we learned still apply.

Our project is based on a 1.2 million-word corpus of sociolinguistic interviews first collected in 2003-4. Of the 99 speakers we attempted to re-interview, we were able to find only 65. Traditional strategies such as contacting people via mail were not successful. The most effective strategy for tracking down participants was using their original addresses in order to find their current phone numbers. Another valuable method was to contact former interviewers who had recruited participants from their own social networks. Social media such as LinkedIn and Facebook yielded mixed results, as did strategic Google searches. Of the 65 original participants we found, only 14 agreed to participate again. This highlights the need to find ways to keep participants engaged in the research process. In order to facilitate interpretation of our panel data, we also created a trend study. This led to another challenge, finding people that matched the original speakers as they were back then and as they are now.

To date, we have learned a great deal about recent societal change, community-based corpus construction and sociolinguistic methods, leading us to make a number of key recommendations for future studies. First, make judicious use of the original interviews and interview reports to identify details (such as hobbies or workplaces) that will make it easier to find participants again and facilitate rapport when re-interviewing. Second, with regard to tracking down participants, we recommend building ways to track individuals, e.g., by asking participants to indicate their willingness to be contacted again on consent forms and what method of communication they prefer (Wagner & Tagliamonte 2017). We also suggest finding ways to stay in touch with interviewers and interviewees, such as returning to the community on a regular basis (e.g. Cukor-Avila and Bailey 2017), organizing popular interest talks that update participants on the results of the study so that they are motivated to continue to participate more fully in the future. Generally, it is important to have an outward facing component to the research program that focusses on the societal impacts of language variation and change so that the general public can appreciate the importance of the research, e.g. a dedicated website, writing press releases, etc. (see also Pichler et al. 2018).

References

Cieri, C., & Yaeger-Dror, M. (2017). Alternative sources of panel study data: Opportunities, caveats and suggestions. In S. E. Wagner & I. Buchstaller (eds.), *Panel studies of variation and change* (pp. 53-72). New York, NY & London: Routledge.

Cukor-Avila, P. & Bailey, G. (2017). The effect of small Ns and gaps in contact on panel survey data. In S. E. Wagner & I. Buchstaller (eds.), *Panel studies of variation and change* (pp. 210-45). Routledge studies in language change. New York, NY & London: Routledge.

Nahkola, K. & Saanilahti, M. (2004). Mapping language changes in real time: A panel study on Finnish. *Language Variation and Change*, *16*(2), 75-92.

Pichler, H., Evans Wagner, S., & Hesson, A. (2018). Old age language variation and change: Confronting variations ageism. *Language & Linguistics Compass, 12*(6), 1-21.

Sankoff, G. (2017). Before there were corpora: The evolution of the Montreal French Project as a longitudinal study. In S. E. Wagner & I. Buchstaller (eds.), *Panel studies of variation and change* (pp. 21-52). Routledge studies in language change. New York, NY & London: Routledge.

Sundgren, E. (2002). Aterbesok i eskilstuna: en undersokning av morfologisk variation och forandring i nutida talsprak (Unpublished doctoral dissertation). Uppsala University, Uppsala.

Wagner, S. E. & Tagliamonte, S. A. (2017). What makes a panel study work? Researcher and participant in real time. In S. E. Wagner & I. Buchstaller (eds.), *Panel studies of variation and change* (pp. 213-232). New York, NY & London: Routledge.

Dialectology in work. Research Using the Visual Interface of the DMW-Project

Nicole Palliwoda (Kiel), Helmut Spiekermann (Münster)

The *Dialektatlas Mittleres Westdeutschland (DMW)* (Spiekermann et al. 2020) is a long-term (2016-2032) project funded by the North Rhine-Westphalian Academy of Culture and Sciences. Its main goal is to document the Low und Middle German dialects in North Rhine-Westphalia and the neighboring regions in Lower Saxony and Rhineland-Palatinate, mainly in form of a digital, speaking atlas based on approx. 3000 explorations in 1000 places. The data is gained by using a question book (Fragebuch), consisting of about 800 questions on phonological, morphological, syntactic and lexical features. The project is divided into four modules of which the second (Exploration, 2018-2015) and the third (Analysis, 2026-2030) form the central part of the empirical work. Even though a linguistic analysis actually requires a more or less completed Exploration of the data, the DMW-Project aims to present preliminary dialectological insights and findings from the collected data while the exploration is still in progress. This is realized by preview maps, which show dynamically the present state of exploration and editing (transcription) of the data. The advantages – and disadvantages – of this approach are the topic of the given talk.

The preview maps fulfil two main purposes: (1) They are directed to people, who do not have a linguistic training but are interested in the development of the local dialects. The popular scientific notation (POP-Notation) used in the preview maps is based on the Latin alphabet and enables people to read the transcription easily. (2) The preview maps are useful to linguistics as well. Even though the data is not yet fully analyzed and the transcription is simplified, it nevertheless allows the examination of linguistic problems, such as the current course of the isoglosses of the *Rheinischer Fächer*. This will be shown by some examples.

Bibliography

Palliwoda, Nicole (2021): Dialektatlas Mittleres Westdeutschland (DMW): Atlaskarten sind online zugänglich. In: Zeitschrift für Germanistische Linguistik (ZGL). 49/2. S. 433-434.

Spiekermann, Helmut/Tophinke, Doris/Vogel, Petra M./Wich-Reif, Claudia (2020): Dialektatlas Mittleres Westdeutschland. In: Bundesraat för Nedderdüütsch & Niederdeutschsekretariat (Hrsg.): Niederdeutsch in der Wissenschaft - aktuelle Projekte und Lehre. Hamburg. S. 50-51.

Internet

Dialektatlas Mittleres Westdeutschland (DMW): www.dmw-projekt.de.

Xanthi Katsanta Dimitris Papazachariou

High Vowel Loss in the South-Greek dialect of Agrinio: Acoustic analysis within the framework of Dialect Contact

With this paper, we present the study of High Vowel Loss in the dialect Of Agrinio (South Greece), as well as our effort to interpret the realisation of the phenomenon within the Dialect Contactframework.

High Vowel Loss –i.e. the deletion of unstressed [i] & [u]- has been defined as one of the few stereotypical characteristics of Northern Greek dialects, the isogloss that splits Greece into two main areas, i.e. Northern Greek dialects, that delete unstressed [i]s & [u]s, even within the prosodic word (as in: /ðikos$_{\#}$mu/ "mine"→ [θkozm]), and Southern Greek dialects, which do not show High Vowel Loss, especially in the middle of the prosodic word (as: /ðikos$_{\#}$mu/ "mine"→ [ðikozmu]), (Newton 1972& Trudgill 2003 among others).

Nevertheless, acoustic analysis on casual speech of 60 informants (10' from each informant, or 10 hours of casual speech altogether) verified the existence ofHigh Vowel Loss in a Southern Greek dialect, i.e. the dialect of Agrinio.

Linguistic parameters (as the variant of the deletion –four different variants of the phenomenon-, thelocation of the variant within the prosodic word, its location in relation to the lexical stress, the surrounding consonants –and consequently the possible consonantal clusters- the type of word's grammatical category, as well as the word itself) were taken into consideration, as well as social parameters (like age, gender, education/social group, homogenous Vs mixed origin, as well as neighbourhood status).

Our analysis shows:

i. the existence of four different variants of the phenomenon of High Vowel Loss, from non-deletion to full deletion. The other two intermediate variants are vowel reduction and deletion with trace.
ii. The correlation of deletion with trace of [i] with particular morphological units, (like the articles [ti] [tis] [tu]), as well as between a consonant$_{[+stop]}$ and its next consonant, where the trace of the deletion seems to block the creation of an homosyllabusC$_{+stop}$C cluster.
iii. The gradience of the phenomenon, as it has also been found by other similar acoustic analyses on Northern Greek dialects (Topintzi&Baltazani 2012, Kainada&Baltazani 2015, Lengeris et. al. 2016,Παπαζαχαρίου 2019). Nevertheless, our study furtherrevealsa sociolinguistic correlation, i.e. highly educated young people from new and mixed neighbourhoods produce significantly smaller percentages of deletion with trace and full deletion in contrast to middle-aged/old speakers from old neighbourhoods.

The statistically different behavior between well educated, young speakers from new and mixed neighborhoods and old people with low education from the old neighborhoods can be interpreted as part of the process of dialect levelling (Trudgill 1986), especially after 60s, when many new incomers moved in the city, mainly in the new neighborhoods that were created at that time.

References

Newton, B. 1972. *The generative interpretation of dialect: a study of Modern Greek Phonology*. Cambridge: Cambridge University Press.

Kainada, E. & M. Baltazani. 2015. The vocalic system of the dialect of Ipiros. Στο G. Kotzoglou, K. Nikolou, E. Karantzola, K. Frantzi, I. Galantomos, M. Georgalidou, V.

Kourti-Kazoullis, C. Papadopoulou& E. Vlachou (επιμ.), *Proceedings of the 11th international conference on Greek linguistics,* (pp. 101-123). Rhodes: Laboratory of Linguistics of the Southeastern Mediterranean.

Lengeris, A., Kainada, E., Baltazani, M. & P. Iverson. 2016. Vowel raising, deletion and diphthongization in Kozani Greek. In A. Ralli, N. Koutsoukos, & S. Bompolas (eds.). *Proceedings of the 6th Modern Greek Dialects and Linguistic Theory,* (pp.93-101).Patras: Univeristy of Patras Press.

Παπαζαχαρίου, Δ. 2019. Αποκοπή των άτονων υψηλών φωνηέντων στην ποικιλία της Αγιάσου της Λέσβου: Μελέτη της κώφωσης από τη σκοπιά της ακουστικής φωνητικής [HighVowelLossinthedialectofAgiasos (Lesbos): Anacousticstudyofvoweldeletion]. In Α. Αρχάκης, Ν. Κουτσούκος, Γ. Ξυδόπουλος& Δ. Παπαζαχαρίου (eds.) *Γλωσσική Ποικιλία: Μελέτες αφιερωμένες στην Αγγελική Ράλλη [LinguisticVariation: StudiesinhonorofAngelaRalli]*, (σσ. 497-514).Αθήνα: Κάπα Εκδοτική.

Topintzi, N. &M. Baltazani. 2012. The acoustics of high-vowel loss in a Northern Greek dialect and typological implications. In: P. Hoole, L. Bombien, M. Pouplier, Ch. Mooshammer& B. Kühnert. (eds.), *Consonant clusters and structural complexity,* (pp 373-402). Berlin: Mouton de Gruyter.

Trudgill, P. 1986. *Dialects in contact*. Oxford: Basil Blackwell.

Trudgill, P. 2003. Modern Greek dialects: a preliminary classification. *Journal of Greek Linguistics* 4: 45-64.

Comparing language attitudes and folklinguistic methods, and their insights into variation within Australian English

Since 1960, the primary methodology for understanding the social evaluation of language has been subjective reaction tests (SRTs) or the matched/verbal guise technique. Folklinguistic approaches allow more open reporting in contrast to the scale-based approach of SRTs (Garrett, 2010), but little work has been completed contrasting these indirect/direct methodologies, despite suggestions they may lead researchers to fundamentally different understandings (e.g., see conclusions in Garrett, Coupland, & Williams, 2003). Research usually employs samples of very different accents (e.g., McKenzie, 2015) or audio manipulated to vary in just one variant (e.g., Campbell-Kibler, 2008). This study investigates the effect of method on language evaluations and further stretches the usual project design by using similar, unmodified voices (four young women from Melbourne).

The study was designed to examine the differences in SRT and folklinguistic data and the response to similar, local, voices. Data were collected from visiting secondary school students from a large number of schools. The participants included in this study were all 15–20 years of age and assessed as (likely) Australian English speakers based on detailed background information they provided. The programme for the visitors was repeated on two days. Those present the first day completed the SRT (participants, $n = 159$). On the second day, student-participants ($n = 213$) commented on the speakers in an open response item, after they were played the same four audio files. Participants were currently studying Australian English and had been exposed to the ideas of linguistics, making this open task easier. Previous analyses have shown that such cohorts still evaluate language in similar ways to those with no training in linguistics (Penry Williams, 2019). The day-one data, nine semantic differential scales and an assessment of sociolect/broadness, were analysed for their means and standard deviations as individual measures and in meaningful collections of the nine items. Folklinguistic data were analysed via the keywords methodology (Garrett, 2010), quantifying comment content and therefore allowing for comparable analyses.

Results show that the two tasks produce quite different evaluations. These are presented for their insights into variation in Australian English. Discussion then focusses on comparing the results from the two tasks. The personal traits of the SRT were not a focus in the written data, in which the students pointed out specific linguistic features (especially use of *like* and high rising tunes), demographic factors (age group, ethnicity, nationality, social class) and sociolect labels, alongside some negative evaluations. The influence of the similar voices is examined, especially in terms of if this resulted in participants creating diversity when it did not appear or opting out in later items. Whether the SRT results relate to evaluations of

linguistic features made salient in the open data is also explored. The contemplation of limitations focusses on the role of the content of the speakers' narratives.

The study adds to ongoing discussions regarding dimensions of attitudes (Pharao & Kristiansen, 2019; Rosseel & Grondelaers, 2019), considering the boundaries of direct/indirect methods.

References

Campbell-Kibler, K. (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society, 37*(5), 637–659.

Garrett, P. (2010). *Attitudes to language*. New York: Cambridge University Press.

Garrett, P., Coupland, N., & Williams, A. (2003). *Investigating language attitudes: Social meanings of dialect, ethnicity, and performance*. Cardiff: University of Wales Press.

McKenzie, R. M. (2015). UK university students' folk perceptions of spoken variation in English: The role of explicit and implicit attitudes. *International Journal of the Sociology of Language, 2015*(236), 31–53.

Penry Williams, C. (2019). *Folklinguistics and social meaning in Australian English*. Abingdon; New York: Routledge.

Pharao, N., & Kristiansen, T. (2019). Reflections on the relation between direct/indirect methods and explicit/implicit attitudes. *Linguistics Vanguard, 5*(s1)

Rosseel, L., & Grondelaers, S. (2019). Implicitness and experimental methods in language variation research. *Linguistics Vanguard, 5*(s1)

Regional phonetic variation in the speech of young urban Russians: An exploratory study of quantitative and qualitative vowel reduction in Moscow and Perm

Margje Post
Universitetet i Bergen, Norway
margje.post@uib.no

Most young urban Russians speak with little or no local characteristics in their speech, but small regional differences are likely to be present even in this language with a strong standard language ideology, especially in prosody. Contemporary Central Standard Russian is known for its unusual word prosody. Words have a heavy nucleus: the first pretonic syllable is unusually prominent, forming a salient contrast, together with the stressed syllable, with unstressed syllables in other, weak positions, which are heavily reduced, both in quality and quantity (Zlatoustova 1981, Kodzasov 1999). This means that effectively, there are two degrees of vowel reduction: a moderate degree for the first pretonic syllable (and for onsetless and final open syllables, contexts promoting long vowel duration) and a radical degree of reduction for any other unstressed syllable (Crosswhite 2000). This two-degree reduction is strong in many traditional rural dialects in Central Russia, but less so in other parts of Russia (Potebnja 1866; Vysotskij 1973; Al'muxamedova and Kul'šaripova 1980). Two recent studies suggest that this regional variation in prosodic word shape between Central Russian and non-central varieties is still present in modern urban Russian (Grammatčikova et al. 2013; Erofeeva 2005). However, in these two earlier studies, the number of speakers and vowels measured was very low.

We compared vowel quality and duration in the speech of a larger set of speakers to confirm that regional differences persist in modern urban speech, and to explore different prosodic factors that may play a role (position in the word, position in the sentence, accentual status). We recorded 32 adolescents, born in 1998 or 1999, in Moscow (central variety) and Perm (non-central variety). They read 10 sentences containing words with a CV_2-CV_1-'CV_0C structure with pretonic vowels /o/ and /a/ after non-palatalized consonants (which merge in most modern varieties of Russian), in several prosodic conditions.

The main finding is that, as expected, the Muscovites make a much larger difference between the first pretonic (V1) and second pretonic (V2) vowels than the pupils from Perm, even in this formal speaking style. In the Moscow data, the first pretonic is almost twice as long as the second pretonic, whereas in the Perm words, the two vowels have almost equal duration. This difference between Moscow and Perm speech is stable across all speakers and prosodic conditions, i.e. irrespective of the word's position in the utterance and its accentual status. The results were corroborated by a small auditory study.

Our vowel quality data confirm the parallel between quantity (duration) and quality (F1 and F2), both inside the second pretonic vowel (cf. Barnes 2006 on phonetic undershoot in Moscow speech) and between the second and first pretonics, with short second pretonic

vowels being less open than their longer first pretonic neighbours, but in Moscow, the latter can have the same open quality as the – even longer – stressed vowels.

These findings will be discussed in the context of theoretical models of language variation and change (cf. Auer 2005; Krause 2010).

References:

Al'muxamedova, Z. M. and R. E. Kul'šaripova (1980). *Redukcija glasnyx i prosodija slova v okajuščix russkix govorax* [*Vowel reduction and word prosody in Russian* okan'e *dialects*]. Kazan: Kazan University Press.

Auer, Peter (2005). Europe's sociolinguistic unity, or: A typology of European dialect/standard constellations. In *Perspectives on Variation. Sociolinguistic, Historical, Comparative*, eds. N. Delbecque, J. van der Auwera and D. Geeraerts. Berlin/New York: De Gruyter, 7–42.

Barnes, J. (2006). *Strength and weakness at the interface: Positional neutralization in phonetics and phonology*. Berlin: Mouton de Gruyter.

Crosswhite, K. M. (2000). Vowel reduction in Russian: A unified account of standard, dialectal, and 'dissimilative' patterns. In: *University of Rochester Working Papers in the Language Sciences (Spring 2000)*. Rochester, NY, University of Rochester, 107–171.

Erofeeva, E. V. (2005). *Idiomy kak verojatnostnye struktury: sociolingvističeskij aspekt, na materiale fonetičeskogo urovnja.* [*Idioms as probabilistic structures: Their sociolinguistic aspect, based on phonetic data.*] Doctoral dissertation, Saint Petersburg State University.

Grammatčikova, E. V., S. V. Knjazev, L. V. Luk"anova, S. K. Požarickaja (2013). Ritmičeskaja struktura slova i mesto realizacii tonal'nogo akcenta v regional'nyx variantax sovremennogo russkogo literaturnogo jazyka [Rhythmic word structure and pitch accent alignment in regional variants of modern Standard Russian.] In: *Aktual´nye voprosy teoretičeskoj i prikladnoj fonetiki*, eds. A. V. Archipova and I. M. Kobozeva. Moscow: Buki-Vedi, 69–90.

Kodzasov, S. (1999). Russian. In: *Word Prosodic Systems in the Languages of Europe*, ed. H. v.d. Hulst. Berlin, De Gruyter, 852–870.

Krause, M. (2010). Zur Typologie von Sprachsituationen: Binnensprachliche Variation zwischen Standard und Dialekt im heutigen Russland. *Wiener Slawistischer Almanach* 65: 53-81.

Potebnja, A. A. (1866). O zvukovyx osobennostjax russkix narečij [On the phonetic characteristics of the Russian dialects]. In: *Dva issledovanija o zvukax russkogo jazyka. I. O polnoglasii. II. O zvukovyx osobennostjax russkix narečij*. Voronež, Tip. V. Gol'dštejna: 53–140.

Vysotskij, S.S. (1973). O zvukovoj strukture slova v russkix govorax [On the sound structure of the word in Russian dialects]. In: *Issledovanija po russkoj dialektologii*, edited by S. F. Bromlej. Moscow: Nauka, 17–41.

Zlatoustova, L. V. (1981). *Fonetičeskie edinicy russkoj reči [The phonetic units of Russian speech]*. Moscow: Moscow University Press.

What lies underneath: Silversteinian nths, indexical fields, sociolinguistic monitors, and attitudinal cognitoria, Dennis R Preston, University of Kentucky

Only a few proposals have been advanced for cognitive or procedural models of the relationship between the perception of linguistic facts and nonlinguistic ones. One reason for this is surely the fact that most studies that integrate features from demography, identity, style, and stance in variationist studies have done so with a production bias, although there are increasing studies of perception within the variation and change paradigm.

This paper outlines the theoretical underpinnings of some of those few proposals, taking into consideration the preceding papers in this session that have touched on basic aspects of speech perception, idexicalities, and misdirection ("priming"). The ones examined, including discussion of their considerable overlap, are the following:

1) Silverstein's (linguistic anthropological) notion of indexicality (2003), including the Irvine-Gal model of iconization (Irvine 2001) and Eckert's notion of an "indexical field" (2008), the latter exemplified experimentally and in discourse in Campbell-Kibler's 2008 -ING study

2) Social psychological and folk linguistic accounts of attitudes and language regard, particularly Niedzielski & Preston's "folk linguistic triangle" (2000) and Preston's elaboration of it as an "attitudinal cognitorium," exemplified from experimental and discoursal evidence in, e.g., Preston 2010,

3) Labov's "sociolinguistic monitor," outlined in detail and updated in Levon and Fox 2014.

This outline concludes by proposing some "best-practices" (and "new opportunities") considerations of what is at stake:

1) Careful distinctions must be made between what is to be processed (the "attitude object") and how the processing is triggered ("noticing") and takes place, including consideration of more and less conscious operations as well as looking more closely at the multiple levels of consciousness and number of underlying beliefs that may be involved in only one response or characterization.

2) Careful consideration must be given to the granularity of the linguistic object considered, including attention to the possibilities of "inadvertent priming," both in accompanying linguistic information as well as in the task itself in an experimental setting

3) Greater reliance should be made on what is revealed about the relationship of social to linguistic meanings as revealed in discourse, not just in overt metalinguistic talk but in the subtler pragmatic characteristics of such talk (e.g., implication, presupposition, entailment) that may reveal underlying attitudinal or regard features involved in such processing (e.g., Rodgers 2016).

Dr. Heinrich Ramisch
Englische Sprachwissenschaft
Universität Bamberg
An der Universität 9
96045 Bamberg

Proposal for a paper to be read at Methods XVII (University of Mainz, 1-5 August 2022)

In search of good data:
Past tense and past participle forms from a variationist/phonetic point of view

Various studies in corpus linguistics (e.g. Levin 2009, De Clerck & Vanopstal 2015) examine verb regularisation processes in different varieties of English. In fact, forms such as *learned/learnt* or *burned/burnt* are one of the most frequently cited grammatical differences between (Standard) American English and British English, with *–ed* being regarded as more American and *–t* as typically British. In his detailed study, Levin (2009) shows that certain syntactic and semantic factors have contributed to the maintenance of the variation in British English, for example the transitive or intransitive use of the verbs or a difference in aspect (durative vs. punctual). As the variation between *learned/learnt* or *burned/burnt* is based on a phonological difference in the spoken language (voiced vs. voiceless ending) it is rather important to ascertain whether the forms to be found in the different corpora reliably represent the pronunciation of the verb-forms. Rather surprisingly, studies of this type generally do not include a phonetic analysis, but merely rely on written forms. Therefore, it will be a major aim of this paper to explore the relationship between the written verb-forms and their actual pronunciation in the spoken language.

To test the influence of the phonetic environment, I carried out a study with 20 American speakers (undergraduate students). The results clearly reveal such an influence. In a frequent syntactic structure such as *learned* plus *to*-infinitive (e.g. *Joe only learned to swim when he was a teenager*), the most usual realization by the speakers was an amalgamation of the final sound of *learned* with the initial /t/ of the infinitive marker *to*, with the result that the /t/ constitutes the onset of the unstressed syllable /tə/. Apart from presenting the findings of this study, the paper will generally consider the complex relationship between spelling and pronunciation that may cause problems in the analysis of verb-forms. By examining different varieties of English (both standard and nonstandard, traditional and modern), it can equally be demonstrated that a phonetic analysis is mandatory.

Finally, I will also examine an interesting and innovative technique to collect spoken corpus data. UCLA's *NewsScape Library of Digital TV News* provides a large dataset of TV news that allows linguists to carry out full text searches on the basis of the subtitles broadcast together with the news programmes (cf. Uhrig 2018).

References:

De Clerck, Bernard and Klaar Vanopstal. 2015. "Patterns of regularisation in British, American and Indian English: A closer look at irregular verbs with *t/ed* variation." In: Peter Collins (ed.). *Grammatical Change in English World-Wide*. Amsterdam: John Benjamins, 335-372.

Levin, Magnus. 2009. "The formation of the preterite and the past participle." In: Günter Rohdenburg and Julia Schlüter (eds). *One Language, Two Grammars? Differences between British and American English.* Cambridge: CUP, 60-85.

Ramisch, Heinrich. 2014. "At the crossroads of variation studies and corpus linguistics: the analysis of past tense and past participle forms." In: Sarah Buschfeld et al. (eds). *The Evolution of Englishes. The Dynamic Model and beyond*. Amsterdam: John Benjamins, 301-311.

Uhrig, Peter. 2018. "NewsScape and the Distributed Little Red Hen Lab: A Digital Infrastructure for the Large-Scale Analysis of TV Broadcasts." In: Anne-Julia Zwierlein et al. (eds). *Anglistentag 2017 Regensburg*. Trier: WVT, 99-114.

Subjective dialect division in Indian varieties of English

Giuliana Regnoli (University of Regensburg)

Recent research about the Indian diaspora has addressed questions of how the internal heterogeneity of speech communities and intra-linguistic differences inform the understanding of community boundaries, attitudinal orientations, and identity development (Hundt & Sharma, 2014). However, little attention has been given to the mapping of perceptual landscapes in Indian varieties of English - a method which is almost always neglected in perceptual studies of New Englishes (NEs).

The present paper aims to shed light on the relationship between belief and use in language from a folk linguistic perspective. It sets out to investigate a transient multilingual community of Indian university students located in Heidelberg, Germany, and their direct and indirect attitudes towards Indian English (IndE) linguistic variation in their home country. The study used a mixed-method research design by combining data collection techniques from the fields of language attitudes and perceptual dialectology. Direct and indirect attitudes were investigated first by means of map-drawing tasks and then with a Verbal Guise test (VGT) for 60 respondents. In order to present a detailed description of Indian English (IndE) perceived variation, the data set of the VGT was based on the speech of educated speakers from two typologically distinct language groups, the Indo-Aryan one, which comprises languages spoken primarily in Northern India, and the Dravidian one, i.e., languages spoken in Southern India. The data collected with the aforementioned methods were analysed following the principles of quantitative approaches. While the analysis of the data gathered from the 'draw-a-map' task corresponded to the perception of accent variation, i.e., the elicitation of the speakers' overt stigmatisations and stereotypes of IndE accents, that of the VGT corresponded to its production, i.e., the correlation between indirect attitudes and accent variation.

The study provided the first perceptual map of India showing aggregated English dialect areas. Reported results showed that the community members operationalise perceptual differences on a North vs. South basis and that Northern IndE accents are perceived as more correct than Southern IndE ones for most attitude traits. Moreover, the study shed new light on the social stereotypes of the participants, emphasising their use of a similar cognitive template for area identifications. The triangulation of the 'draw-a-map' task and the VGT allowed to determine accurate folk perceptions and provided a full picture of overt and covert stigmatisations towards accent variation, which has not been accounted for in attitude studies of NEs.

Keywords: attitudinal cognitorium, folk awareness, folk linguistics, Indian Englishes; perceptual dialectology

References

Hundt, M. & Sharma, D. (2014). Introduction. In M. Hundt & D. Sharma (Eds.), *English in the Indian Diaspora* (pp. 1–8). Amsterdam: Benjamins.

In between Quebec and Hexagonal French: A longitudinal study of front vowels

(in special session Panel research: Methodological challenges, practices and ways forward)

Josiane Riverin-Coutlée and Jonathan Harrington
Institute of Phonetics and Speech Processing, Ludwig-Maximilians University of Munich

The 21st century is that of mobility and digitization. For dialectologists, this means having to deal with speakers exposed to increased linguistic diversity, but also unprecedented availability of speech data. In this study, we combine these two realities of our time to show that longitudinal data collected from platforms such as YouTube may offer a rich perspective on lifespan phonetic changes resulting from social mobility (e.g. Kwon, 2018; MacKenzie, 2017).

The general hypothesis is that the type of profession and the audience that an adult speaker addresses in a professional capacity may influence lifespan phonetic changes (Buchstaller & Wagner, 2018; Gerstenberg & Voeste, 2015). We expand on a previous acoustic study of the speech of Michaëlle Jean, in which we found changes in the realization of a dialect feature of Quebec French (QF) as a function of the stages of her career (Riverin-Coutlée & Harrington, in press). In the current study, our aim is to find out whether other features changed over a similar timeline.

We focus on acoustic characteristics of the front unrounded vowels, specifically: 1) the location of /e/ in the acoustic space, which tends to be closer to tense [i] in QF than in Hexagonal French (HF) (e.g. Riverin-Coutlée & Roy, 2020 vs. Storme, 2017); 2) duration and formant dynamics of the vowel in FÊTE, produced as a short and monophthongal /ɛ/ identical to that in FAITE in HF, but with a greater duration and a diphthongal quality, /a͡ɛ/, in QF (Côté, 2012; Martin, 1995; Riverin-Coutlée & Roy, 2020).

The speech data consists of 62 recordings (4h 39m) spanning three decades (1988-2021) and divided into five career stages. The recordings were orthographically and phonemically transcribed, then forced-aligned using a series of BAS tools (Kisler et al., 2017), and structured into a database using EMU-SDMS (Winkelmann et al., 2017). Segment boundaries and formant detection were manually corrected for 2690 tokens of the front unrounded vowels /i, e, ɛ, a͡ɛ, a/ in word-final syllable. F1 and F2 were estimated at 11 time points from vowel onset to offset, then transformed into three DCT coefficients representing the mean, linear slope and curvature of the trajectories (Watson & Harrington, 1999). These coefficients and the log-transformed duration were set as response variables in linear mixed-effect regression models with *Vowels* and *Career Stages* as fixed effects, and *Words* as random effect.

The results showed that /e/ evolved from a QF-like small distance from [i] in the first career stage, to a larger distance, and back to a smaller distance. No substantial change was observed for /a͡ɛ/, which was mainly distinguished from /ɛ/ through length, not quality. This first suggests high sensitivity to fine properties of the vowel space, as the proximity of /e/ to [i] is not known as a salient feature of QF and seems to evolve in coordination with the tense-lax split studied in Riverin-Coutlée & Harrington (in press). Second, the stability of the length contrast between /ɛ, a͡ɛ/ shows that features of QF were not suppressed *en bloc* as Jean's career became international, but their otherwise similar spectral properties suggest a spoken accent that is neither entirely QF nor HF. Overall, these results emphasize the relevance of social and linguistic factors in explaining lifespan phonetic stability and change.

References

Buchstaller, I. & Wagner, S. E. (2018). Introduction: Using panel data in the sociolinguistic study of variation and change. In S. E. Wagner & I. Buchstaller (Eds.), *Panel Studies of Variation and Change* (pp. 1–18). Routledge.

Côté, M.-H. (2012). Laurentian French (Quebec): Extra vowels, missing schwas and surprising liaison consonants. In R. Gess, C. Lyche, & T. Meisenburg (Eds.), *Phonological Variation in French: Illustrations from Three Continents* (pp. 235–274). John Benjamins.

Gerstenberg, A. & Voeste, A. (2015). Investigating the lifespan perspective. In A. Gerstenberg & A. Voeste (Eds.), *Language Development: The Lifespan Perspective* (pp. 1–8). John Benjamins.

Kisler, T., Reichel, U. D. & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347.

Kwon, S. (2018). Phonetic and phonological changes of Noam Chomsky: A case study of dialect shift. *American Speech*, *93*(2), 270–297.

MacKenzie, L. (2017). Frequency effects over the lifespan: A case study of Attenborough's r's. *Linguistics Vanguard*, *3*(1).

Martin, P. (1995). L'opposition entre /ɛ/ (bref) et /ɛː/ (long) en français actuel du Québec. *La linguistique*, *31*(2), 33–45.

Riverin-Coutlée, J. & Harrington, J. (in press). Phonetic change over the career: A case study. *Linguistics Vanguard*.

Riverin-Coutlée, J. & Roy, J.-P. (2020). A descriptive account of the Quebec French diphthong FÊTE. *Journal of the International Phonetic Association*.

Storme, B. (2017). The *loi de position* and the acoustics of French mid vowels. *Glossa: A Journal of General Linguistics*, *2*(1), 64–89.

Watson, C. I. & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, *106*(1), 458–468.

Winkelmann, R., Harrington, J. & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, *45*, 392–410.

A Unified Account of the Low Back Merger Shift

Rebecca Roeder and Matt Hunt Gardner

The low back merger (LBM) is attested as a widespread phenomenon across North American varieties of English (e.g., Labov et al., 2005; Becker, 2019). However, little previous work has examined the LBM as a change in progress. Among the studies that have investigated the genesis of the merger (e.g., Bigham, 2010; Durian, 2012; Boberg, 2017), focus has been on the LOT/THOUGHT merger, with little emphasis on PALM as a vowel implicated in the process, likely because PALM is now merged with LOT in most North American varieties of English, despite historically merging with BATH and START in Standard British English (Boberg, 2010, p. 128). Boberg (2010) observes that the independence of PALM from LOT-THOUGHT prevents the LBM shift in Eastern New England (p. 155), highlighting the importance of PALM to this vowel shift. The current study examines the LBM as a change in progress in Victoria, British Columbia, where some older speakers realize PALM differently than LOT or THOUGHT. Findings are based on automated measurements of wordlist data archived in the Synchronic Corpus of Victoria English (D'Arcy, 2015) from 29 speakers (13 men, 16 women), born between 1913 and 1941.

Comparison of vowel mean charts across individuals reveals the following three patterns: fully merged low back vowels; LOT/THOUGHT merged with PALM distinct; and all three low back vowels distinct. TRAP (non pre-voiced-velar, non pre-nasal) consistently appears in the lower front quadrant of the vowel space instead of in a retracted low central position, supporting the hypothesis that TRAP retraction does not begin until the LBM is complete. Previous research found both retracted TRAP and the LBM to be ubiquitous in Victoria speakers born after 1941 (Roeder et al., 2018). Complementary evidence comes from quantitative analysis, which finds no group-level correlations between F1/F2 of TRAP and F1/F2 of LOT, THOUGHT, or PALM among these older speakers.

These results fit a phonological explanation for the LBM shift based on principles of the Modified Contrastive Specification Theory (Dresher et al., 1994), the Contrastive Hierarchy Theory (Dresher, 2009), and the Successive Division Algorithm (Dresher, 2009), according to which phonemes are specified only for contrastive features, and it is only these contrastive features that are active within the phonology (Gardner, 2011; Roeder & Gardner, 2013). Following this theory, TRAP—as the only low vowel in the system once the low back vowels are merged—is unspecified for the feature [± Back], allowing for phonetic dispersion and focalization (Becker-Kristal, 2010) to draw TRAP toward the low central region of the vowel space. Via analogy, the phonetic reinforcement of [± Peripheral] is triggered. [± Peripheral] is the highest ranked constraint and the only constraint differentiating TRAP and the LBM vowels. In other words, as TRAP becomes more central, so too do all lax vowels. This research contributes to ongoing work on the mechanisms involved in the formation of the Third Dialect of English (Labov, 1991), as well as to interdisciplinary work at the interface between sociophonetics and phonological theory.

References

Becker, K. (ed.). (2019). *The low-back-merger shift: Uniting the Canadian Vowel Shift, the California Vowel Shift, and Short Front Vowel Shifts across North America* (Publication of the

American Dialect Society 104). Durham, NC: Duke University Press.

Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion theory: Insights from a large cross-linguistic corpus*. University of California, Los Angeles Ph.D. dissertation.

Boberg, C. (2010). *The English language in Canada: Status, history and comparative analysis*. New York: Cambridge University Press.

Boberg, C. (2017). Archival data on earlier Canadian English. In Hickey, R. (Ed.), *Listening to the past: Audio records of accents of English*, 375–394. Cambridge University Press.

Bigham, D. (2010). Correlation of the low-back vowel merger and trap-retraction. *University of Pennsylvania Working Papers in Linguistics*: Vol. 15 : Iss. 2 , Article 4. Available at: https://repository.upenn.edu/pwpl/vol15/iss2/4

D'Arcy, A. (2015). At the crossroads of change: Possession, periphrasis and prescriptivism in Victoria English. In Collins, P. (ed.), *Grammatical change in English world-wide*, 43-64. Amsterdam: John Benjamins.

Dresher, E. B. (2009). *The Contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Dresher, E. B., Piggott, G. & Rice, K. (1994). Contrast in phonology: Overview. *Toronto Working Papers in Linguistics* (13). iii–xvii.

Durian, D. (2012). A new perspective on vowel variation across the 19th and 20th centuries in Columbus, OH. Columbus, OH: The Ohio State University PhD dissertation.

Gardner, M. H. (2011). Beyond the Phonological Void: Contrast and the Canadian Shift. Generals paper, University of Toronto.

Labov, W. (1991). The three dialects of English. In Eckert, P. (Ed.), *New ways of analyzing sound change*, 1-4. San Diego: Academic.

Labov, W., Ash, S., & Boberg, C. (2005). *The Atlas of North American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.

Roeder, R., & Gardner, M. H. (2013). The phonology of the Canadian Shift revisited: Thunder Bay and Cape Breton. *University of Pennsylvania Working Papers in Linguistics*: Vol. 19: Iss. 2, Article 18. Available at: https://repository.upenn.edu/pwpl/vol19/iss2/18

Roeder, R., Onosson, S., & D'Arcy, A. (2018). Joining the Western Region: Sociophonetic Shift in Victoria. *Journal of English Linguistics* Vol 46: Iss. 2, pp. 87-112.

Measuring language attitudes towards ethnolectal features in Swiss-German-speaking children: A mixed-methods approach

Melanie Röthlisberger (University of Zürich, Switzerland), Eline Zenner (KU Leuven, Belgium), Laura Rosseel (Vrije Universiteit Brussel, Belgium)

Background: This paper reports on a mixed-methods project that investigates how children aged 6 to 12 use and evaluate ethnolectal features in Swiss-German. Features of ethnolectal Swiss-German (e.g. example 1) are typically discussed in connection with adolescents with a migrant background (e.g., Tissot et al. 2011) and in the way these features are appropriated by non-migrant adolescents for stylistic purposes (Auer 2002; Schmid 2017). Little is known about the way adolescents acquire the usage and the social meaning attached to these forms.

(1) Omission of prepositions, articles, pronouns or auxiliaries, as in *Chani bleistift?* <Can I pen?> for *Chani en bleistift ha?* <Can I have a pen?>

Aim: This project investigates Swiss-German preadolescents' production, perception and evaluation of ethnolectal features through a mixed method approach, (a) taking stock of the linguistic repertoire available to children (production study) and, (b) measuring the social meanings attached to these features (evaluation study).

Production study: To identify the linguistic features of interest, we investigate anecdotal claims about the usage of ethnolectal features as defined by Auer (2002, see also Tissot et al. 2011) using spoken data collected through the diapix task (Baker & Hazan 2011) and free storytelling. For the production-oriented part, two studies were conducted in and around the city of Winterthur in one urban and one rural primary school with children aged 6-12. Preliminary analysis of the sampled speech indicates that the use of ethnolectal features is already common among children under 12. This is the case in both areas; this usage, however, seems to be restricted to migrant children in the rural area while also being used among non-migrant children in the urban areas.

Evaluation study: The social evaluation children attach to these features is assessed in a second, experimental step that includes a visually enriched and child-friendly version of the matched guise technique and a language awareness test (N=86). Results show that the youngest children have no preference for either the Swiss German or the ethnolectal guise and seem to lack awareness of the ethnolectal features. However, that awareness develops with age and so does a more positive overall evaluation of the Swiss German guise compared to the ethnolectal guise. The latter is furthermore increasingly associated with social meanings of 'non-Swissness'. Interestingly, social meanings of urbanity, which have been reported in adult populations, do not (yet?) appear in the evaluations held by pre-adolescents.

Implications: Obtaining a better understanding of the acquisition trajectory of Swiss-German ethnolectal features, more generally informs us about innovation and language change (cf. Cheshire et al. 2011) and about ongoing restructuring processes in Swiss-German (see, e.g. Leemann et al. 2014).

References

Auer, Peter. 2002. "Türkenslang": Ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen. In Annelies Häcki Buhofer (ed.), *Spracherwerb und Lebensalter*, 255–264.

Tübingen/Basel: Francke.
Baker, Rachel & Valerie Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43(3). 761–770. doi:10.3758/s13428-011-0075-y.
Cheshire, Jenny, Paul Kerswil, Sue Fox & Eivind Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics* 15(2). 151–196.
Leemann, Adrian, Marie-José Kolly, Iwar Werlen, David Britain & Dieter Studer-Joho. 2014. The diffusion of /l/-vocalization in Swiss German. *Language Variation and Change* 26(2). 191–218.
Schmid, Stephan. 2017. Differenzierungsprozesse im Sprachgebrauch von Jugendlichen in der Deutschschweiz: zur sozialen Interpretation von ethnolektalen Sprechweisen in Schweizer Medien. *Bulletin Suisse de Linguistique Appliquée (Vals-Asla)* Nspécial(t. 1). 105–116.
Tissot, Fabienne, Stephan Schmid & Esther Galliker. 2011. Ethnolektales Schweizerdeutsch: Soziophonetische und morphosyntaktische Merkmale sowie ihre dynamische Verwendung in ethnolektalen Sprechweisen. In Elvira Glaser & Natascha Frey (eds.), *Dynamic des Wandels - Wandel und Variation: Akten des 3. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD)*, 319–344. Stuttgart: Steiner.
Zenner, Eline, Rosseel, Laura, & Dirk Speelman. 2020. Starman or Sterrenman: an acquisitional perspective on the social meaning of English in Flanders. *International Journal of Bilingualism* 25(3): 568-591.

Mixed-method analyses of reduced article forms in the South- and Central-Bavarian transition area

Elisabeth Scherr & Arne Ziegler (Graz)

Some variants of spoken German, including the Bavarian dialects, are attested to have an inventory of reduced forms of the definite article. Of particular interest, besides the weakening or drop of the vowels, is the deletion of the dental plosive *d-* in some regions (cf. amongst others Weiß 1998, Scheutz 1988 or Zehetner 1985):

(1) 4023:	*in dem wold*	*aiso*	***as** WOLDfest;*	South-/Central Bavarian transition area
	in dem Wald	also	**das** WALDfest	Standard German
	‚in the woods	well	the «party in the woods»‘	

Instead of assuming a revocation or reduction of the German article system in these cases (cf. e.g. Nübling 2005 or Leiss 2010), we will

(i) conduct detailed quantitative analyses (frequency analysis, Chi-Square-test, contingency analysis) to trace the occurrences of reduced article forms alongside the vertical dialect-standard-axis in the South-/Central-Bavarian transition area;
(ii) empirically grasp the language-internal reasons that suggest a functionally driven use of definite articles without *d-* by quantitative-qualitative mixed-method analyses.

This latter assumption is based on previous dialectological studies suggesting that the two paradigms – full article forms on the one hand and reduced articles on the other – are systematically used, driven by underlying functional distribution principles (cf. Hartmann 1980, Ebert 1971 on the Fering dialect or Dirani (2020) for South Hessian): Full forms of the definite article would be used in contexts of concrete deictic reference (pragmatic definiteness) whereas reduced forms would have a tendency of being used with unique nouns, associative-anaphoric or generic reference (semantic definiteness).

Our study is based on two corpora on German spoken language. The corpus „Urban Languages in Austria“ is compiled in the course of the long-term project „Vienna and Graz. Cities and their influential force“ (Austrian Science Fund F6004). The data of the second corpus „Youth Languages in Austria“ (Austrian Science Fund P 25683) integrates the speech production of autochthonous dialect speakers. The combinatory analysis of both corpora not only allows a detailed investigation of the parameters *age*, *gender* and *degree of formality*, but also sheds light on the difference between rural, rural-urban and urban communication in the South-/Central-Bavarian transition area.

The results show that definite articles without the dental *d-* are highly frequent in dialectal speech (with a share of around 80%). More detailed analyses, however, show no significant functional difference in the distribution of the aforementioned paradigms: Reduced articles are equally used in contexts of semantic as well as pragmatic definiteness. This situation is reversed when focussing on communication setting with higher degrees of formality: As expected, the share of articles without *d-* is reduced here (55% in the agglomeration areas and around 36% in urban regions). Mixed methods analyses show, however, that they tend to be used in contexts of semantic definiteness. Reduced definite articles are therefore neither a purely dialectal feature nor are they a symptom of a revocation process; rather, their „reallocation“ (Britain 2002) could indicate a possible language change in process being rooted in dialectal speech and successively spreading to vertically „higher“ settings.

BRITAIN, DAVID (2002): Diffusion, levelling, simplification and reallocation in past tense BE in the English Fens. In: *Journal of Sociolinguistics* 6 (1), 16–43.

DIRANI, SEYNA MARIA (2020): Zum unterschiedlichen Artikelgebrauch im Südhessischen. Empirie trifft auf Theorie. In: SPEYER, AUGUSTIN und JULIA HERTEL (eds.): *Syntax aus Saarbrücker Sicht*. Wiesbaden/Stuttgart: Steiner, 41–60.

EBERT, KAREN (1971): *Referenz, Sprechsituation und die bestimmten Artikel in einem Nordfriesischen Dialekt (Fering)*. Dissertation, Universität Kiel.

HARTMANN, DIETRICH (1980): Über Verschmelzungen von Präpositionen und bestimmtem Artikel. Untersuchungen zu ihrer Form und Funktion in gesprochenen und geschriebenen Varietäten des heutigen Deutsch. In: *Zeitschrift für Dialektologie und Linguistik* 47 (2), 160–183.

LEISS, ELISABETH (2010): Koverter Abbau des Artikels im Gegenwartsdeutschen. In: BITTNER, DAGMAR / LIVIO GAETA (eds.): *Kodierungstechniken im Wandel. Das Zusammenspiel von Analytik und Synthese im Gegenwartsdeutschen*. Berlin/New York: de Gruyter, 137–158.

NÜBLING, DAMARIS (2005): Von in die über in'n und ins bis im. Die Klitisierung von Präposition und Artikel als ‚Grammatikalisierungsbaustelle'. In: LEUSCHNER, TORSTEN / TANJA MORTELMANS / SARAH DE GROODT (eds.): *Grammatikalisierung im Deutschen*. Berlin/New York: de Gruyter, 105–132.

SCHEUTZ, HANNES (1988): Determinantien und Definitheitsarten im Bairischen und Standarddeutschen. In: STEIN, PETER K. / ANDREA WEISS / GEROLD HAYER (eds.): *Festschrift für Ingo Reiffenstein zum 60. Geburtstag*. Göppingen: Kümmerle, 231–258.

WEIß, HELMUT (1998): *Syntax des Bairischen: Studien zur Grammatik einer natürlichen Sprache*. Tübingen: Niemeyer.

ZEHETNER, LUDWIG (1985): *Das bairische Dialektbuch*. München: Beck.

Corpus-based computational dialectology with normalization

Yves Scherrer, Department of Digital Humanities, University of Helsinki
yves.scherrer@helsinki.fi

Dialectological research increasingly focuses on corpus-based approaches. Dialect corpora typically consist of transcribed interviews and aim to represent realistic, everyday speech grounded in linguistic context (Szmrecsanyi & Anderwald 2018). However, dialect corpora do not lend themselves well to quantitative studies because the different interviews are not directly comparable: if informant A does not use word *x*, it may just be that A chose to talk about topics that did not require the use of word *x*, and not that *x* does not exist in A's dialect.

In his seminal work on quantitative corpus-based dialectology, Szmercsanyi (2013) relies on syntactic annotation to make dialect corpora comparable and to infer geographical distributions of syntactic patterns. In this paper, we focus on the quantitative analysis of more traditionally studied linguistic levels, namely phonology and morphology. Consequently, we propose to use **orthographic normalization** rather than syntactic annotation to provide comparability. Several dialect corpora, e.g., the Swiss German *ArchiMob* corpus (Scherrer et al. 2019) or the *Samples of Spoken Finnish* collection (Institute for the Languages of Finland, 2014) include either manual or semi-automatic normalization annotations on the word level.

Normalization is the annotation of every dialectal word with a canonical word form, for example the standardized spelling of the word, as illustrated in the following example from ArchiMob:[1]

Transcription:	jaa	de	het	me	no	gluegt	tänkt	dasch	ez	de	genneraal
Normalization:	ja	dann	hat	man	noch	gelugt	gedacht	das ist	jetzt	der	general
Gloss:	yes	then	has	one	again	looked	thought	this is	now	the	general

For our analysis, we align the transcribed words and their normalized counterparts on the character level, yielding correspondences between transcribed and normalized characters and character n-grams. The frequency distributions of these correspondences vary across dialects and thus can serve as a basis for comparisons between dialects. For example, in some Swiss German dialects, /l/ becomes /u/ in certain phonological contexts. In order to define the geographical area in which this /l/-vocalization occurs, it is not sufficient to compute the relative frequency of /u/ in each text, because /u/ also occurs in other, irrelevant phonological contexts. Normalization allows us to define phonological contexts easily and hence to restrict our search to those occurrences of /u/ that are aligned with normalized /l/. This gives us a clearer and more accurate picture of the geographical extent of /l/-vocalization.

The success of this analysis depends essentially on two factors: the character alignment method and the automatic discovery of dialectologically relevant alignments. Ideally, the character alignment method can identify many-to-many correspondences, such as those occurring between a diphthong and a long vowel. We will apply alignment methods initially developed for phrase-based statistical machine translation (Koehn et al. 2003; Tiedemann 2009) and grapheme-to-phoneme conversion (Jiampojamarn et al. 2007) to this task. We will also test different weighting schemes to discover and extract character (n-gram) correspondences that show regional variation. These findings can then be compared with

[1] The normalization language used in ArchiMob is similar, but not identical to Standard German.

traditional atlas-based dialect classifications.

References

Institute for the Languages of Finland (2014): *Samples of Spoken Finnish (speech corpus)*. Kielipankki. http://urn.fi/urn:nbn:fi:lb-2020112937

S. Jiampojamarn, G. Kondrak, T. Sherif (2007): *Applying many-to-many alignments and hidden Markov models to letter-to-phoneme conversion.* In: Proceedings of HLT-NAACL 2007, 372-379.

P. Koehn, F. J. Och, D. Marcu (2003): *Statistical phrase-based translation*. In: Proceedings of HLT-NAACL 2003, 127-133.

Y. Scherrer, T. Samardžić & E. Glaser (2019). *Digitising Swiss German - How to process and study a polycentric spoken language*. In: Language Resources and Evaluation. 53(4).

B. Szmrecsanyi (2013). *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge University Press.

B. Szmrecsanyi & L. Anderwald (2018). "Corpus-based approaches to dialect study". In: C. Boberg, J. Nerbonne & D. Watt (eds.), *The Handbook of Dialectology*. Wiley-Blackwell.

J. Tiedemann (2009). *Character-based PSMT for closely related languages*. Proceedings of EAMT 2009, 12-19.

Identifying keywords and phrases in British COVID-19 newspaper discourse

Julia Schilling & Robert Fuchs
University of Hamburg
julia.schilling@uni-hamburg.de, robert.fuchs@uni-hamburg.de

The COVID-19 pandemic has upended life around the globe, leading to intense public debate and a flurry of lexical innovation across many languages. (Socio-)Linguists quickly started to document and analyze COVID-19 discourse (Baines et al. 2021; Saraff et al. 2021), but there is as yet no systematic analysis of the lexical items and discourse patterns that characterize British COVID-19 discourse. We address this research gap through a systematic comparative analysis of public discourse during the COVID-19 pandemic. Through a big data approach, we identify not just distinct keywords and phrases linked to the pandemic but also track their development over time and across regions.

As news can offer an insight into and simultaneously influence the public's perception of the COVID-19 pandemic, our analysis focuses on discourse in regional and national English newspapers. The starting point of the analysis is a contrastive keyword analysis of the discourse of every month of 2019 with its equivalent in 2020 and 2021, comparing pandemic with pre-pandemic discourse, while filtering out seasonal effects (e.g. discussion of *snow* in January). Our data comprises 10% of all articles from 51 national and regional English newspapers published between January 2019 and October 2021, producing a corpus of approximately 386,118 articles and 229,347,771 tokens. Rather than collecting newspaper articles based on a pre-existing list of keywords, we use a data-driven approach to identify COVID-19 related n-grams ($1 \leq n \leq 4$) for each month of the pandemic based on log likelihood and log ratio. We then assign these keywords to semantic fields such as COVID-19 NAMES (e.g. *Covid-19*, *SARS-CoV-2*), PUBLIC HEALTH INSTRUCTIONS (e.g. *self-isolation*, *quarantine*), and VACCINATION and examine their development over time using statistical measures such as median, IQR, standard deviation, and skewness of the distribution.

This analysis yielded over 300 1-grams, 350 2-grams, 200 3-grams, and 100 4-grams related to the COVID-19 pandemic. Results indicate that the lexis of COVID-19 discourse in British newspapers significantly varies not only over time, but also within semantic fields of discourse and across regions.

Baines, Annalise; Ittefaq, Muhammad & Mauryne Abwao (2021) “#Scamdemic, #Plandemic, or #Scaredemic: What Parler Social Media Platform Tells Us About COVID-19 Vaccine“. *Vaccines* 9 (421), pp. 1-16.

Saraff, Sweta; Singh, Tushar & Ramakrishna Biswal (2021) “Coronavirus Disease 2019: Exploring Media Portrayals of Public Sentiment on Funerals Using Linguistic Dimensions”. *Frontiers in Psychology* 12:626638.

A measure for heterogeneity in spatial language variation

An entropy-like measurement method for spatial distribution of language items

Schönberg Andreas[1] & Alfred Lameli [2]

[1]PhD student at the department Deutscher Sprachatlas, Philipps University of Marburg, Germany
[2]Professor at and Director of the department Deutscher Sprachatlas, Philipps University of Marburg, Germany

Abstract

Dialectometric studies usually ask about the internally consistent groups of dialects within a language area (see Goebl 1984). However, when dealing with larger sets of geographically specified language data, the problem arises of identifying those regions that are particularly prone to variation or particularly sensitive to language change. The question then is not so much about stability in an area (typically indicated by the definition of clusters), but about instability. More recent dialectometric studies have introduced a number of solutions to this problem, for example, based on resampling techniques (see, e.g., Wieling & Nerbonne 2015). In our project, we follow an approach based on the concept of entropy (e.g., Prokić & Nerbonne 2008) that, in contrast to other studies (Prokić et al. 2009), is not applied to strings of tokens, but geographic distributions.

Our study deals with data from a historical language survey of German dialects at 2500 sites in the regions of Baden (Germany) and Elsass (France). These data are interesting from the perspective that they contain information on different age groups and thus enable analyses on language change (= apparent time; in contrast, analyses in real time become possible by comparison with both more traditional and more recent surveys in the same region).

In order to identify areas which are more sensitive to language change than others we use an entropy-like measure for the identification of heterogeneity/uniformity in spatial language distributions. More concrete, we use a nearest neighbor approach resulting, first, for every linguistic variable of our corpus (e.g., morphemes, lexemes) in a normalized global index with higher values indicating a more homogeneous spatial distribution and lesser values indicating a variative state. We use this global measure for the automated detection of linguistic items with higher/lesser language variation.

Furthermore, a transformation into a local measure of spatial variation makes it possible, second, to automatically identify individual regions with particularly high language variation (typically the transition zones between areas of linguistic variants). This is used, for example, to predict language change or to test the correlation of spatial variation that occurs for different linguistic phenomena. Applying this measure to a collection of multiple linguistic phenomena leads to a new perspective on the structuring of linguistic space highlighting not so much the clusters of linguistic similarity, but the zones of particular linguistic dynamics. The paper will introduce the measure and discuss some examples.

References

Goebl, Hans (1984): Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF. Tübingen: Niemeyer.

Prokić, Jelena, John Nerbonne, Vladimir Zhobov, Petya Osenova, Kiril Simov, Thomas Zastrow & Erhard Hinrichs (2009): The computational analysis of Bulgarian dialect pronunciation. In: Serdica. Journal of Computing(3), 269-298.

Prokić, Jelena and John Nerbonne (2008): Recognizing Groups among Dialects. In: International Journal of Humanities and Arts Computing(2), 153-172.

Wieling, Martijn & John Nerbonne (2015): Advances in Dialectometry In: Annual Review in Linguistics 1(1), 243-264.

A cognitive geographic approach to dialectology: Cognitive distance as a predictor for perceptual dialect distance

Hedwig Sekeres & Martijn Wieling

In dialectology, the central relationship under investigation is often that between dialect distance and geographic distance (Heeringa & Nerbonne, 2013). Nevertheless, other approaches may be better suited to represent contact situations, such as (historical) travel distance (Gooskens, 2004) or 'rice paddy distance' (Stanford, 2012) and have been successfully used to explain dialect variation.

In this study, we consider a different type of distance that is commonly used in the field of cognitive geography to explain perceptual dialect differences. Cognitive geography is based on the assumption that an individual's mental representation of their environment has a greater effect on their behaviour than the actual environment (Montello, 2018). A commonly used metric in cognitive geography is the cognitive distance: the geographic distance between two places as estimated by an individual (Montello, 1991). The present study is the first to our knowledge to incorporate the theoretical framework and this distance metric from cognitive geography in dialect research. Although the individual and social aspects of language are an important component of research in dialectology, and recently even quantitative dialectology (e.g., Wieling, 2012; Wieling et al., 2011), the individual aspects of geography have not been widely considered. In this study, the existing parallel between real and perceptual dialect distances that exists in dialectology is extended to the spatial component of dialect research by including both geographic and cognitive distances in the analysis. This study aims to assess whether the use of cognitive distance adds to our understanding of dialect variation.

A total of 850 participants from two provinces in the north of the Netherlands were willing to estimate the geographic distance to seven locations in the same region. They were subsequently asked to rate the similarity of dialect fragments from these locations to the dialect of the location in which they grew up. Participants were not aware that the speakers they rated came from the same locations for which they had to provide distance estimates. Additionally, for each participant, the geographic distance between the location in which they grew up and the seven speaker locations was calculated. A linear mixed-effects regression model was built to predict perceptual dialect distance from both cognitive distance and geographic distance.

The resulting model indicates that geographic distance is more predictive of perceptual dialect distance than cognitive distance, but that there is also a significant interaction between cognitive and geographic distance. Cognitive distance is more predictive of perceptual dialect distance when geographic distance is low than when geographic distance is high. Furthermore, an exploratory analysis revealed that gender and proficiency in the participants' local dialect were predictive of perceptual dialect distance as well. Our findings indicate that cognitive distance can be used to explain dialect variation when the area under investigation is small, and consequently that the framework of cognitive geography can be usefully employed in dialectological research.

Works cited

Gooskens, C. (2004). Norwegian dialect distances geographically explained. In B.-L. Gunnarson, L. Bergström, G. Eklund, S. Fridell, L. H. Hansen, A. Karstadt, B. Nordberg, E. Sundgren, & M. Thelander (Eds.), *Language Variation in Europe. Papers from the Second International Conference on Language Variation in Europe, ICLAVE 2* (pp. 195–206).

Heeringa, W., & Nerbonne, J. (2013). Dialectometry. In F. Hinskens & J. Taeldeman (Eds.), *Language and Space: Dutch*, pp. 624–646). De Gruyter Mouton.

Montello, D. R. (1991). The measurement of cognitive distance: Methods and construct validity. *Journal of Environmental Psychology*, *11*(2), 101–122. https://doi.org/10.1016/S0272-4944(05)80071-4.

Montello, D. R. (2018). Behavioral and cognitive geography: introduction and overview. In D. R. Montello (Ed.), *Handbook of behavioral and cognitive geography* (pp. 3–15). Edward Elgar Publishing Limited.

Stanford, J. N. (2012). One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change*, *24*(2), 247–278. https://doi.org/10.1017/S0954394512000087.

Wieling, M. (2012). *A Quantitative Approach to Social and Geographical Dialect Variation.* PhD dissertation. Groningen: University of Groningen.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613. https://doi.org/10.1371/journal.pone.0023613.

Low Saxon corpus-based dialectometry

Janine Siewert
University of Helsinki
janine.siewert@helsinki.fi

In connection with our research project on diachronic and synchronic variation in Low Saxon, we investigate dialect similarity in 19th and 21st century Low Saxon based on data from Germany and the Netherlands. Traditionally, Low Saxon dialect classification has mostly been based on phonological and morphological traits, like the ones presented by Schröder (2004). In this study, however, we are focusing on the orthographic and the (morpho-)syntactic side and compare how these relate to each other and to the more traditional classifications.

The majority of our dataset is taken from the the LSDC dataset (Siewert et al., 2020), from relevant prose texts from Leopold and Leopold (1882)[1] and the Twentse Taalbank (van der Vliet, 2021). Our overall dataset covers eight dialect regions from the 19th, 20th and 21st century, but in this study, we use the 19th and 21st century data from the five major West Low Saxon dialect groups: Dutch North Saxon, German North Saxon, Dutch Westphalian, German Westphalian and Eastphalian. Overall, these consist of 34,460 sentences and 345,131 tokens from the 19th century, and 44,740 sentences and 740,849 tokens from the 21st century which we have converted to CoNLL-U format and automatically PoS tagged.

One interesting area to pay attention to with respect to dialect distance is the Dutch-German border. Like Goossens (2019) observed, the Low Saxon dialects along the border have started to diverge under the influence of the majority languages. According to him, this divergence is most pronounced at the lexical level, but convergence towards the majority language has also been attested in phonology, morphology and syntax. While studies on the divergence of dialects along the border often focus on the occurrence and frequency of particular traits based on interviews, cf. Smits (2011), this study addresses the overall (dis)similarity in prose texts.

Dialect similarity at the orthographical level based on character n-grams will be compared to dialect distance based on PoS tag sequences to investigate if these lead to different dialect groupings. Malmasi and Zampieri (2017) observed in their experiments for identifying Swiss German dialects that approaches based on character n-grams outperform word-based ones and, in their study on British dialects, Wolk and Szmrecsanyi (2016) have employed part-of-speech n-grams for corpus-based dialectometry concluding that this approach can achieve results comparable to manually selected features. We will combine these with clustering approaches on the one hand and principal component analysis (PCA) on the other hand.

As in the 19th century, school education and majority language media played a smaller role in

[1] Digitised by dbnl: https://dbnl.nl/tekst/leop008sche00_01/

everyday life compared with today, we hypothesize that the effect of language contact with Dutch and German is less visible in the morphology and syntax of 19th century Low Saxon, even though the border is probably already clearly discernable at the orthographic level. Therefore, we will investigate how these results compare to more modern data from the 21st century.

References

Ingrid Schröder. Niederdeutsch in der Gegenwart: Sprachgebiet – Grammatisches – Binnendifferenzierung. In Dieter Stellmacher, editor, *Niederdeutsche Sprache und Literatur der Gegenwart*, pages 35–97. Georg Olms Verlag, Hildesheim, Zürich and New York, 2004.

Janine Siewert, Yves Scherrer, Martijn Wieling, and Jörg Tiedemann. LSDC - a comprehensive dataset for low Saxon dialect classification. In *Proceedings of the 7th Workshop on NLP for Similar Languages, Varieties and Dialects*, page 25–35, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics (ICCL). URL https://www.aclweb.org/anthology/2020.vardial-1.3.

Joh. A. Leopold and L. Leopold. *Van de Schelde tot de Weichsel*. J.B. Wolters, Groningen, 1882.

Goaitsen van der Vliet. Twentse taalbank. http://www.twentsetaalbank.nl/, 2021. Accessed: 2021-12-15.

Jan Goossens. „Dialektverfall" und „Mundartrenaissance" in Westniederdeutschland und im Osten der Niederlande. In Gerhard Stickel, editor, *Varietäten des Deutschen: Regional- und Umgangssprachen*, pages 399–404. De Gruyter, 2019. doi: doi:10.1515/9783110622560-023. URL https://doi.org/10.1515/9783110622560-023.

Tom Smits. Dialectverlies en dialectnivellering in nederlands-duitse grensdialecten. *Taal en Tongval*, 63(1):175–196, 2011.

Shervin Malmasi and Marcos Zampieri. German dialect identification in interview transcriptions. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 164–169, 2017.

Christoph Wolk and Benedikt Szmrecsanyi. Top-down and bottom-up advances in corpus-based dialectometry. *The future of dialects: Selected papers from Methods in Dialectology XV*, 1:225, 2016.

A Variationist Linguistic Landscape Study of English language use in Viennese public space

Barbara Soukup (University of Vienna)
Andreas Baumann (University of Vienna)

This contribution reports on a large-scale project that applies state-of-the-art variationist sociolinguistic methodology to quantitative linguistic landscape (LL) study on an unprecedented level of consistency and rigor (cf. Soukup 2016, 2020). Study context is the investigation of written English vs. German language use in the public space of Vienna, Austria. Data collection adhered to two central variationist tenets: (I) the 'count-all' Principle of Accountability (Labov 1969); applied via (II) hypothesis-driven stratified sampling (Sankoff 2005). Regarding (I), because the dependent variable under investigation is written language choice (English vs. German), all stationary items (signs, objects, displays) bearing any identifiable written text (=all possible places of occurrence of language choice) were recorded. Regarding (II), data collection covered a survey area operationalizing three main hypotheses about LL items catering to local audiences in Vienna: (1) AGE – predicting a higher presence of English in areas where many young adults, vs. seniors, live; (2) MULTILINGUALISM (more English where more people of different linguistic backgrounds come together), regarding (2a) RESIDENTS, as well as (2b) TOURISTS; and (3) COMMERCIALISM (more English in shopping vs. residential areas, due to its strong association with global consumerism and commerce). To implement these hypotheses, 3x2 Viennese administrative districts ('Gemeindebezirke') were selected (matching districts that load very high or very low on the parameters from hypotheses 1, 2a, and 2b), and within each district, one shopping street and one residential street (hypothesis 3). On each street, a 200m-stretch centering on the lengthwise midpoint was surveyed under the count-all principle, yielding 17,091 datapoints.

For data analysis, a generalized linear model was constructed featuring English language use as dependent variable, and AGE, MULTILINGUALISM, TOURISM, and COMMERCIALISM as primary predictor variables. A multimodel-inferencing approach (Burnham & Anderson 2002) based on a set of candidate models was chosen to compute (a) the relative variable importance of the predictors in all candidate models and (b) statistical significance and effect sizes of all predictors in the averaged model. An additional generalized linear model together with hierarchical clustering was employed to assess the effects of and the relationships amongst a set of secondary variables regarding the recorded LL items, including their MATERIALITY (e.g.

enamel or wood); text APPLICATION method (e.g. printed or handwritten); SIZE; physical LOCATION (e.g. on a wall or on street furniture); PLACEMENT type (legal or illegal); and ACTIVITY context (e.g. shop sign vs. traffic sign).

Findings show that among the primary predictors, TOURISM and MULTILINGUALISM have the strongest impact on language use, such that TOURISM promotes English usage, and MULTILINGUALISM actually inhibits it (suggesting the local 'lingua franca' to be German). Overall, however, the item-related ('secondary') variables are more relevant than the areal-level ('primary') variables, such that e.g. the activity context of public administration disfavors English use, while the arts as well as food/accommodation industries enhance it; and small items are more likely to contain English than large ones.

This contribution discusses in detail the study's methodological setup, innovations, and statistical analyses, as well as general implications for variationist LL research.

References:

Burnham, K.P. & Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. New York: Springer.

Labov, W. 1969. 'Contraction, Deletion, and Inherent Variability of the English Copula', *Language* 45 (3), 715–762.

Sankoff, D. 2005. Problems of Representativeness/Probleme der Repräsentativität. In U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (eds.), *Sociolinguistics/Soziolinguistik, Vol. 2*, 2nd edn., 998-1002. Berlin: Walter de Gruyter.

Soukup, B. 2016. English in the Linguistic Landscape of Vienna, Austria (ELLViA): Outline, rationale, and methodology of a large-scale empirical project on language choice on public signs from the perspective of sign-readers . Views 25, 1-24.

Soukup, B. 2020. Survey area selection in Variationist Linguistic Landscape Study (VaLLS): A report from Vienna, Austria . Linguistic Landscape 6 (1), 52-79.

Lexical Variation in the Dialects of Malta and Gozo

This paper is concerned with regional lexical variation in Maltese. Previous literature on the topic consists of a number of undergraduate theses that cover specific semantic fields, such as agricultural tools, construction and culinary terminology (see, among others, Chetcuti 2003, Borg 2011, Attard & Spagnol 2014). Such studies do find some regional patterns, but the data are not explicitly compared across a wide range of localities that cover most of the Maltese and Gozitan dialects. Also, whilst specific semantic fields confirm the existence of lexical variation in determined niches, they are limited in establishing the extent of lexical variants of more generic concepts widely used in everyday life and affirming regional linguistic variation.

In this paper I present the results drawn from a lexical survey conducted among 1,000 speakers, aged 60 or over, from 25 localities in Malta and Gozo in an attempt to provide a general idea of Maltese geolinguistics by mapping them out using Gabmap. I introduce a general method for the comparison of geosynonyms to outline common patterns of regional lexical variation. In particular, I focus on the results of a quantitative analysis of lexical data derived from everyday concepts spread across a number of localities in order to give an overview of the degree of lexical variation present in different regions of the Maltese Islands.

References

Aquilina, Joseph & Benedikt Isserlin (1981) Survey of Contemporary Dialectal Maltese. Volume 1: Gozo. Leeds: Leeds University Press.

Attard, Keith & Spagnol, Michael (2014) Differenzi lessikali bejn il-Malti u l-Għawdxi. Leħen il-Malti. L-Għaqda tal-Malti – Università.

Borg, Albert (2011) Lectal Variation in Maltese. In Sandro Caruana, Ray Fabri & Thomas Stolz (eds.), Variation and Change. The Dynamics of Maltese in space, time and society. Akademie Verlag, 11-31.

Chetcuti, Flavia (2003). Lexical Variation in the Semantic Field of Agricultural Tools. B.A. (Hons), University of Malta.

I could just be sat here all day: A corpus-based study on recent changes in the use of pseudo-passives with SIT and STAND in British English dialects

ULRIKE STANGE
JOHANNES GUTENBERG-UNIVERSITÄT, MAINZ

The present study focuses on the use of pseudo-passives with SIT and STAND in British English. These are equivalent in meaning to the standard form with the present progressive:

(1) a. so then we **were** just **sat** here watching *I'm a Celebrity* [BNC2014 ST8H: 917]
b. and this dog **was** just **sitting** in this little champagne bucket [BNC2014 SAQD: 387]

(2) a. this man **was stood** there with this chainsaw [BNC2014 SAQD: 659]
b. and he **was standing** next to me and he kept making conversation [BNC2014 SKPP: 332]

Originally a feature of Northern English dialects (Klemola 1999), earlier studies based on the spoken section of the BNC1994 have shown that it has spread to other areas in Britain (Klemola 1999), notably the Southwest. This spread is more advanced for pseudo-passives with SIT, which are also found in the Midlands and in London (Stange 2014; see also Kortmann/ Szmrescanyi 2004 and Rohdenburg/Schlüter 2009 for the distribution of pseudo-passives in British English dialects).

Drawing on spoken data from the BNC1994 (demographic subset, 5m words) and the BNC2014 (11m words), this paper traces the diachronic development of pseudo-passives with SIT and STAND and their standard counterpart in British English in the last twenty years. Potential changes considered in the analysis include the geographical distribution of BE *sat* and BE *stood* and their frequency of use. In addition, the data are coded for speaker information (age, gender, socio-economic status, dialect) to see whether their use can be associated with a particular group of speakers. The analyses are based on c. 2,250 occurrences, c. 600 of which constitute pseudo-passives with SIT and STAND. First results show that, for both male and female speakers, BE *stood* is increasing in frequency at the expense of BE *standing* in real time and in apparent time. BE *sat*, in contrast, has considerably decreased in frequency in real time and in apparent time, while BE *sitting* remains stable. Despite this trend, BE *sat* is still more frequent in the 2014 data than BE *stood* (13 vs. 8 occ. pmw). Furthermore, pseudo-passives with SIT and STAND are newly attested in Eastern England dialects in the BNC2014 data, which indicates that this structure has now spread across the whole island. Pseudo-passives are thus apparently a feature that is maintained rather than effaced in the process of dialect levelling.

References

Klemola, Juhani (1999) "Still sat in your car? Pseudo-passives with *sat* and *stood* and the history of non-standard varieties of English English." *Sociolinguistica* 13: 129-140.

Kortmann, Bernd & Benedikt Szmrecsanyi (2004) "Global synopsis – morphological and syntactic variation in English". In: Kortmann, Bernd, Kate Burridge, Rajend Mesthrie & Edgar Schneider (eds.), *A Handbook of Varieties of English*, Vol. 2: Morphology and Syntax. Berlin/New York: Mouton de Gruyter, 1142-1202.

Rohdenburg, Günter & Julia Schlüter (2009) "New Departures." IN: Rohdenburg, Günter & Julia Schlüter (eds.) *One Language, Two Grammars*. Cambridge: Cambridge UP, 364-423.

Stange, Ulrike (2014) "*I was sat there talking all night*: A Corpus-Based Study on Factors Governing Intra-Dialectal Variation in British English." *English Language and Linguistics* 20 (3): 511-531.

Preposition-drop and ditransitives in North West British English

Jon Stevenson

The present paper reports on a two pronged investigation into syntactic variation in the North-West of England between Manchester and Liverpool. First, a substantial Twitter dataset reveals a distinct boundary between these two centres, marked by the relative frequency of competing syntactic variants. Second, a mass-participation grammaticality judgement survey, taken by sixth form students (aged 16-18) in 19 schools across the region, probes the underlying representation of these structures for individual speakers. Whilst previous studies have tended to focus on place, the present work additionally looks for correlations in acceptance pattens across speakers, differentiated at the level of postcode area, age and social class.

A number of structures are investigated: ditransitives with variant surface object orders (1), their passivised counterparts, and preposition-(determiner)-drop phenomena (2). Previous research has shown that Manchester TGDs are akin to GTDs (Haddican, 2010), while Liverpool TGDs by the same diagnostics are more likely PDAT with a NULL preposition, a property found elsewhere Liverpool English (2a) (Myler, 2013; Biggs, 2016).

(1) Ditransitives
   a. Someone gave it me but I've not tested it. (TGD)
   b. I'm listening to the album, John lent me it (GTD)
   c. Someone sent it to me as a joke (PDAT)
   d. John sent the letter the bank (TGD-full-DP)
   e. He shouted the results the crowd (TGD-shout)

(2) Preposition/determiner dropping
   a. Imagine going the pub and asking for a pint of carling [Liverpool]
   b. going home to change then going library [Manchester/London]

Twitter data confirm high rates of preposition-dropping in Liverpool and environs. However, the picture is complicated by the prevalence of preposition-determiner dropping (2b) in Manchester and a notable dispreference for pronominal TGDs (1a) in Liverpool. Frequency distributions for both phenomena reveal marked boundaries where the proportion of variants change dramatically over just a few miles. Interestingly, the boundary for ditransitives (1) is markedly different to that for preposition/determiner dropping (2); this may be explained by (2) being more sociolinguistically salient than (1).

Survey results both corroborate Twitter findings and demonstrate considerable inter-speaker variation in the underlying analysis of each variant which supports an analysis akin to 'covert representational variability' (MacKenzie, 2019). Meanwhile, there is an expected dispreference for the TGD in Liverpool whilst acceptance of full-DP TGDs (1d) and TGDs with shout/donate type verbs (1e) is divided between speakers, but tends to be preferred in the region between Manchester and Liverpool. Support for theme-passives ("it was given him") is low in most places, especially with full-DPs, while pronominal TGDs (1a) do seem to behave like GTDs for most speakers whether or not they permit dropped prepositions elsewhere.

Preposition-determiner dropping (2b) is available in Manchester and appears to mirror that found in London, as described in Hall (2019), while preposition-dropping (2a) in Liverpool and environs appears closer to the more restricted sort described in Myler (2013) than in Biggs (2016).

References

Biggs, A. (2016). Locating variation in the dative alternation. *Linguistic Variation*, 16(2):151–182.

Haddican, B. (2010). Theme–goal ditransitives and theme passivisation in British English dialects. *Lingua*, 120:2424–2443.

Hall, D. (2019). The impersonal gets personal. *Natural Language & Linguistic Theory*.

MacKenzie, L. (2019). Perturbing the community grammar: Individual differences and community-level constraints on sociolinguistic variation. *Glossa: a journal of general linguistics*, 4(1):28.

Myler, N. (2013). On coming the pub in the North West of England: Accusative unaccusatives, dependent case, and preposition incorporation. *The Journal of Comparative Germanic Linguistics*, 16:189–207.

Standard-dialect variation and ideas about linguistic norms in lexicographic data

Philipp Stöckle

This paper deals with grammatical variation found in a huge data set which was originally created – and is still being used – for lexicographic purposes: the database of the "Wörterbuch der bairischen Mundarten in Österreich (WBÖ)" (Dictionary of Bavarian dialects in Austria). Large parts of the data were collected by volunteers in a questionnaire-based survey which lasted several decades (between 1913 and 1937). In the course of the survey, the so-called "Sammler" (collectors) noted the answers of the informants on little paper slips (cf. Stöckle 2021), which were later on digitized (i.e., manually transferred into an xml-format database; cf. Bowers/Stöckle 2018) and can be used for search queries. Besides lexicographic information (such as lemma, pronunciation, etc.), the paper slips contain many sample sentences in the respective dialects which contain different grammatical phenomena that can be analyzed independently (cf. Stöckle 2020). Moreover, in many cases the collectors added translations into standard German which sometimes deviate from the dialectal original versions, as the following examples illustrate:

Subjunctive II

1) a. *Dǫssn ischt an Ockr, den wos i go nĭt* ***heargab*** (original)
   b. *Jenes ist ein Acker, den ich gar nicht* ***hergeben würde*** (translation)
   'That's the field I would not give away at all.'

***geben*/*tun* as PUT verbs**

2) a. *si muəss əs kχǫarɒ riχtɒ und in t mīl* ***tĩə*** (original)
   b. *Sie muß das Korn richten und in die Mühle* ***geben*** (translation)
   'She must prepare the grain and put it in the mill.'

The first example displays a variant of the subjunctive II, the second shows different realizations of so-called PUT verbs, in this case *tun* ('to do') and *geben* ('to give'). Each of the examples includes the original dialectal form (1a and 2a) as well as the translations made by the collectors (1b and 2b). Although in both cases equivalent forms would exist in the standard variety, the translations deviate from the dialectal variant.

So why did the collectors use different constructions in their translations, especially in cases where similar constructions would be appropriate in the standard variety? It is assumed that at least one of the reasons may have been ideas about certain linguistic norms the collectors followed in order to make the "best" translation. This is especially remarkable, since recent findings show that some of the deviating variants (as the subjunctive II with *würde* or the use of *geben* as a PUT verb) are frequently used in present spoken German (cf. Breuer/Wittibschlager 2020, Lenz in print).

In a second step, the findings from the WBÖ data will be contrasted with data from recent surveys (cf. Lenz in print; Stöckle/Wittibschlager accepted), taking under consideration the different methods of data collection. In my paper I will show that there can be found a continuation between the two levels of comparison (dialect–standard within the WBÖ data vs. WBÖ data–recent data), and that – despite the different methodological standards and different purposes – the historical data can serve as a basis of comparison which can help to understand recent developments in language variation and change.

References:

- BOWERS, JACK / PHILIPP STÖCKLE (2018): TEI and Bavarian dialect resources in Austria: updates from the DBÖ and WBÖ. In: FRANK, ANDREW U. / CHRISTINE IVANOVIC / FRANCESCO MAMBRINI / MARCO PASSAROTTI / CAROLINE SPORLEDER (eds.): Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2), 25–26 January 2018 Vienna, Austsria. Wien: Gerastree Proceedings, 45–54.
- BREUER, LUDWIG M. / ANJA WITTIBSCHLAGER (2020): The variation of the subjunctive II in Austria – Evidence from urban and rural analyses. In: Linguistic Variation 20(1), 136–171.
- LENZ, ALEXANDRA N. (in print): Ein "Austriazismus" auf Erfolgskurs - 'geben' als Basisverb der zielorientierten Objektsbewegung. In: SCHERR, ELISABETH / ARNE ZIEGLER (eds.): In Stadt und Land. Perspektiven variations- und soziolinguistischer Forschung. (Special Issue Linguistik Online).
- STÖCKLE, PHILIPP (2021): Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In: LENZ, ALEXANDRA N. / PHILIPP STÖCKLE (eds.): Germanistische Dialektlexikographie im 21. Jahrhundert. Stuttgart: Steiner.
- STÖCKLE, PHILIPP (2020): Dialektvariation an der Schnittstelle von Syntax, Morphologie und Lexik – der Konjunktiv II in den bairischen Dialekten Österreichs und Südtirols. In: CHRISTEN, HELEN / BRIGITTE GANSWINDT / JOACHIM HERRGEN / JÜRGEN ERICH SCHMIDT (eds.): Regiolekt – Der neue Dialekt? Akten des 6. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD). Stuttgart: Steiner, 149–173.
- STÖCKLE, PHILIPP / ANJA WITTIBSCHLAGER (accepted): Zur Sprachdynamik des Konjunktivs im Bairischen in Österreich. In: BÜLOW, LARS / STEPHAN ELSPAß / PHILIP VERGEINER (eds.): Formen und Funktionen des Konjunktivs in Österreich aus variationslinguistischer Perspektive. (Special Issue Linguistik Online).

Fei schee: The social meaning of intensifier use in Swabian

James M. Stratton & Karen Beaman
Purdue University & Universität Tübingen

Intensifiers are devices which provide speakers with the opportunity to impress, persuade, praise, and generally influence the interlocutor's reception of a message. It therefore comes as no surprise that intensifiers index social meaning, with factors such as gender and age influencing their use (e.g., Fuchs, 2017; Stratton, 2020). While intensifier variation has been explored widely in English (e.g., Tagliamonte, 2008; Fuchs, 2017), to date there has been only one variationist sociolinguistic analysis of German intensifiers (Stratton, 2020), with little to no research on their use in regional German dialects. To this end, the present study uses variationist sociolinguistic methods to examine the use of intensifiers in Swabian German, a variety spoken by about 800,000 speakers in southwestern Germany.

To examine language use across the lifespan, following the traditional sociolinguistic interview, 20 speakers of Swabian German were first recorded in 1982 and then re-recorded in 2017. Speakers originated from two speech communities, Stuttgart, an urban center with approximately one million inhabitants, and Schwäbisch Gmünd, a semi-rural center with a population of around 60,000. Each intensifiable adjective was coded binomially for intensification, as well as linguistic (e.g., syntactic position, semantic classification), social (e.g., gender, age, geographic mobility, education, community origin), and demographic factors (e.g., place of birth, residences lived, years in each location).

Preliminary results from the distributional analysis indicate that *ganz* 'quite', *so* 'so', and *sehr* 'very' were the top three intensifiers, a finding which is consistent with the general distribution of intensifiers in standard German (Stratton, 2020). However, clear changes can be observed over time, with *ganz* 'quite' and *sehr* 'very' decreasing in popularity, and *so* 'so' and *echt* 'really' increasing in frequency over time. Meanwhile, geographic mobility (i.e., the number and length of residential moves over the speaker's lifespan) appears to play a role in lexical decisions, with *gut* (e.g., *die sind gut froh* 'they are very/well happy') and *fei* (*fei schee* 'very nice') retained in the repertoires of speakers with the lowest mobility indices.

For the multivariate analysis, a binary mixed effects logistic regression was run in *Rbrul* (Johnson, 2009), with intensification as the dependent variable. Gender was found to have a significant effect, with women using amplifiers (e.g., *so* 'so', *echt* 'really') more frequently than men, and men using downtoners (e.g., *e bissle* 'a bit') more frequently than women. This finding corroborates Stratton (2020) which found women to have a tendency to scale up the meaning of an adjective more frequently than men, while men tend to scale down the meaning of an adjective more frequently than women. Speaker community also demonstrated a significant effect, with speakers from Stuttgart showing the highest intensification rate, suggesting that speakers in urban environments feel the pressure to intensify more frequently than speakers in semi-rural centers. All in all, this study shows that speakers use intensifiers to index different social meanings (i.e., gender identity, sense of place and belonging) as they construct their own identities.

Selected References

Fuchs, R. (2017). Do women (still) use more intensifiers than men? Recent change in the sociolinguistics of intensifiers in British English. *International Journal of Corpus Linguistics* 22, 345–374.

Johnson, D. (2009). Getting off the Goldvarb standard: Introducing Rbrul for mixed effects variable rule analysis. *Language and Linguistics Compass 3*(1). 359-383.

Stratton, J. (2020). Adjective Intensifiers in German. *Journal of Germanic Linguistics* 32(2), 183-215.

Tagliamonte, S. (2008). So different and pretty cool! Recycling intensifiers in Canadian English. *English Language and Linguistics* 12, 361–394. (Special issue of English Language and Linguistics, Intensifiers, Ed. by Belén Méndez-Naya).

Phonological Change in the Upper Rhine Area in the 20th Century from a Sociolinguistic Point of View - The Influence of the Individual on Variation in Written Questionnaires

Maj-Brit Strobel, Albert Ludwig University of Freiburg

In written (indirect) data collection, homogeneity of the sample group is usually (methodologically) assumed and its heterogeneity is not made an object of scientific interest. This is also due to the typically low control over selection of respondents and lack of information on them. However, the rediscovered "Maurer questionnaires", which were sent to schools in Alsace (France) and Baden (Germany) in 1941, provide an indirectly collected data set that contains considerable metadata on the respondents. The “Maurer data” can thus be used to test sociolinguistic hypotheses on language change and variation in the first half of the 20th century.

In contrast to most sociolinguistic surveys, the “indirect method” does not provide a spatially uniform distribution of clearly definable groups of people; in most cases, there is only one questionnaire per location. Therefore, the influence of the sociolinguistic parameters must be examined separately from the spatial variation, even though the data is not distributed equally in space. For this purpose, a quadrat count method is applied. This way, statistically comparable maps for the different groups, e.g. older vs. younger respondents, are created (cf. e.g. methods with a grid over the study area in Auer, Baumann and Schwarz 2011, Lameli, Glaser and Stoeckle 2020, Pheiff and Kasper 2020).

The paper focuses on the influence of the social parameters age, occupation, and 'biographical mobility' (place of birth) of the individual on phonological variation. The following questions will be addressed:

- Can – in combination with comparisons in real time – phonological change be dated more precisely?
- Did the dialectal decline in favor of near-standard variants in Baden (cf. Streck 2012; Schwarz 2015) and the divergence of Baden and Alsatian dialects (cf. Auer, Pfeiffer, Breuninger 2017) begin already before 1945?
- What are the potentials and limitations of such a dataset for sociolinguistic studies and what conclusions can be drawn about the indirect method?

First results indicate, on the one hand, horizontal change in apparent time, e.g. the spread of (Rhine-)Franconian variants in Lower Alsace. On the other hand, broadly scattered variation of standard (phoneme) variants is evident in some phenomena, especially in Baden. The latter rather points to a transcription problem than to phonological change. In this context, increasing standard competence seems to play a role, which is reflected in 'misspellings' decreasing in time (cf. among others Ganswindt 2017 and Strobel 2021 on the discussed data).

Literature

Auer, Peter, Peter Baumann and Christian Schwarz (2011): Vertical vs. horizontal change in the traditional dialects of southwest Germany. A quantitative approach. In: Taal en Tongeval 63(1), 13–41.

Auer, Peter, Martin Pfeiffer and Julia Breuninger (2017): Neuere Entwicklungen des Alemannischen an der französisch-deutschen Grenze im Oberrheingebiet. In: Christen, Gille & Purschke (ed.): Räume, Grenzen, Übergänge, 27–44.

Ganswindt, Brigitte (2017): Landschaftliches Hochdeutsch. Rekonstruktion der oralen Prestigevarietät im ausgehenden 19. Jahrhundert. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik Beihefte. 168).

Lameli, Alfred, Elvira Glaser and Philipp Stoeckle (2020): Draging areal information from a corpus of noisy dialect data. In: Journal of Linguistic Geography 8, 31–48.

Pheiff, Jeffrey and Simon Kasper (2020): Die Erhebung der regionalsprachlichen Syntax des Deutschen: horizontal, indirekt, vertikal und online. In: Niederdeutsches Wort 60, 35–87.

Schwarz, Christian (2015): Phonologischer Dialektwandel in den alemannischen Basisdialekten Südwestdeutschlands im 20. Jahrhundert. Eine empirische Untersuchung zum Vokalismus. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik Beihefte. 159).

Streck, Tobias (2012): Phonologischer Wandel im Konsonantismus der alemannischen Dialekte Baden-Württembergs. Stuttgart: Steiner (Zeitschrift für Dialektologie und Linguistik Beihefte. 148).

Strobel, Maj-Brit (2021): Die Verschriftungen in der Dialekterhebung Friedrich Maurers in Baden und im Elsass als Evidenz für die Verbreitung der Standardlautung. In: Zeitschrift für Germanistische Linguistik 49(1).

Applying the state-of-the-art tonal distance metrics to large dialectal data

Matthew Sung, Jelena Prokic & Yiya Chen
Leiden University

From Seguy's (1971, 1973) early dialectometric studies to the application of Levenshtein distance in dialectometry (e.g. Heeringa 2004) nowadays, the calculation of phonetic distances between dialects has largely been focused on segments. Despite the fact that tonal languages make up to 70% of the languages in the world (Yip 2002: 1), tones are still largely neglected or simplified in comparative dialectological studies. For instance, Stanford (2012) treats any two tones as either the same or different, there is no in between distances.

In the current literature, there are two studies that propose new tone distance measures, namely Yang & Castro (2008) and Tang (2009). Yang & Castro (2008) has found that their Onset-Contour(-Offset)/OC(O) representation correlates best with mutual intelligibility based on applying Levenshtein distance on a range of tonal representations, including Chao's (1930) tone letters, tones as autosegments (e.g. Duanmu 1994), tones as approximated pitch targets (Xu & Wang 2001). Tang (2009) on the other hand tested several approaches on 15 Chinese dialects, including an inventory-based comparison, Levenshtein distance on Chao's (1930) tone letters and Yang & Castro's (2008) OCO tone representation. In addition, she compared Cheng's (1991) published distances based on 17 dialects (see Cheng 1991, Tang 2009 for more). By counting the misclassifications in the split between Mandarin and non-Mandarin dialects, she found that Cheng's distances work the best, with only 1 misclassification. However, she also noted that this method "fails to reflect any of the internal taxonomy" after distinguishing the Mandarin and non-Mandarin dialects (Tang 2009: 137). Thus far, no existing study measures tone distances for the purpose of dialect classification.

In this presentation we examine the existing tone distance measures and apply them on a large, newly compiled, dialect dataset. The data that we will be using comes from Zhan & Cheung (1987), Zhan & Cheung (1994), Zhan & Cheung (1998), Shao (2016), Chen & Lin (2009a), Chen & Lin (2009b) and Xie (2007), and it consists of 123 Yue and Pinghua dialects represented with over 120 words each. Our results show that the current state-of-the-art method proposed by Yang & Castro can distinguish less than 50% of the tones in our data, which makes it unsuitable for classifying dialects at a lower level. In addition, we will also show how other representations (binary comparison and Levenshtein distance on Chao's (1930) tone letters) perform. Lastly,

we will compare for the first time segmental dialect classification and tonal dialect classification and see how much similarity the two linguistic levels share.

References:

Chao, Y. R. (1930). "A system of tone letters". *Le maître phonétique*, 8(45), 24-27.

Chen, H. & Lin, Y. (2009a). *Yue yu ping hua tu hua fang yin zi hui di 1 pian: Guangxi yue yu, Guinan ping hua bu fen 粵語平話土話方音字彙第 1 編: 廣西粵語、桂南平話部分 [The Lexicon of Yue, Pinghua and Tuhua Volume 1: Guangxi Yue and Guinan Pinghua]*. Shanghai: Shanghai Education Publishing.

Chen, H. & Lin, Y. (2009b). *Yue yu ping hua tu hua fang yin zi hui di 2 pian: Guangxi yue yu, Guinan ping hua bu fen 粵語平話土話方音字彙第 2 編: 桂北、桂東及周邊平話、土話部分 [The Lexicon of Yue, Pinghua and Tuhua Volume 1: Guangxi Yue and Guinan Pinghua]*. Shanghai: Shanghai Education Publishing.

Cheng, C. C. (1991). "Quantifying affinity among Chinese dialects". *Journal of Chinese Linguistics Monograph Series*, (3), 76-110.

Duanmu, S. (1994). "Against contour tone units". *Linguistic Inquiry*, 25(4), 555–608.

Heeringa, W. (2004). *Measuring dialect pronunciation using Levenshtein distance*. Groningen: University Library Groningen. [PhD dissertation, University of Groningen.]

Séguy J. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de linguistique romane* 35, 335–57.

Séguy J. 1973. La dialectométrie dans l'atlas linguistique de la Gascogne. *Revue de linguistique romane* 37, 1–24.

Shao, H. (2016). *Yue xi zhan mao di qu yue yu yu yin yan jiu 粵西湛茂地區粵語語音研究 [The Phonological Study of the Yue Dialects spoken in the Zhan-Mao area in Western Guangdong]*. Guangzhou: Sun Yat-Sen University Press.

Stanford, J. N. (2012). "One size fits all? Dialectometry in a small clan-based indigenous society". *Language Variation and Change*, (2), 247-278.

Tang, C. (2009). *Mutual intelligibility of Chinese dialects: an experimental approach*. Netherlands Graduate School of Linguistics. [PhD dissertation, Leiden University.]

Xie, J. (2007). *Guangxi han yu fang yan yan jiu 廣西漢語方言研究 [Studies on the Chinese dialects in Guangxi]*. Guangxi People's Publishing.

Xu, Y. and Wang, Q. E. (2001). "Pitch targets and their realization: Evidence from Chinese". *Speech Communication*, 33, 319–337.

Yang, C., & Castro, A. (2008). "Representing tone in Levenshtein distance". *International Journal of Humanities and Arts Computing*, 2(1-2), 205-219.

Yip, M. (2002). Tone. Cambridge University Press.

Zhan, B. & Cheung, Y. (1987). *A Survey of Dialects in the Pearl River Delta, Vol. 1, Comparative Morpheme-Syllabary*. People's Publishing House of Guangdong.

Zhan, B. & Cheung, Y. (1994). *A Survey of Yue Dialects in North Guangdong*. Jinan University Press.

Zhan, B. & Cheung, Y. (1998). *A Survey of Yue Dialects in West Guangdong*. Jinan University Press.

A real world method for historical sociolinguistics

Sali A. Tagliamonte, University of Toronto
Laura Rupp, Vrije Universiteit Amsterdam

Many linguists have argued for the value of synchronic dialects and living language materials to provide "an important picture of varying stages in the (recent) development" of a language (Tagliamonte, 2006). This presentation focuses on dialect features that are obsolescing. There is a modest body of research on fading dialect features (e.g. Schilling-Estes & Wolfram, 1999; Jankowski & Tagliamonte, 2017); however, few works have detailed the precise methodology for conducting this type of research nor its benefits.

First, is the inherent value of giving 'voice' to rare and dying features because they typically embody cultural identity. Second, documenting the linguistic features in understudied dialects enriches the knowledge base of human languages (Henry, 1995). Third, is that fading dialect features provide insights into the underlying processes of language change. While previous research has reported relatively negative trajectories of loss (e.g. Schilling-Estes & Wolfram, 1999; Jankowski & Tagliamonte, 2017), there is also evidence of novel pathways of innovation (Rupp & Tagliamonte, to appear). These findings put a positive 'spin' on the possible directions of change in a global situation of severe language loss and highlights how timely this work is for current scholarship.

How can analysts discover and study rare features before they are lost to future generations? Building on a foundation of linguistic theory, variationist sociolinguistics and dialectology, we outline a collaborative research program in which we have developed a step-by-step methodology for studying language features on the verge of demise: from identifying to documenting to analyzing them. To begin with, what does it take to recognize a dying feature and where does one look? A first step involves anthropological observation, community-based fieldwork or searching on-line data bases, dictionaries and other sources. In a contemporary community, seek out older male speakers in blue collar jobs and compare them to the younger speakers. At the same time, legacy recordings may exist that can significantly augment the synchronic perspective. Once an obsolescing feature has been observed, data extraction using manual methods is critical since every token matters. In analyzing the feature, a key consideration is how far does the feature go back in time with the same meaning and what is its geographical distribution? Consulting the now readily available compendia of historical data such as the Oxford English Dictionary or the Corpus of Historical American English (Davies, 2010) is essential. Just as important is the old-fashioned method of contacting fellow researchers for evidence in their materials. The advantages of studying small numbers of tokens is being able to conduct detailed discourse-pragmatic analysis; focusing on distributional patterns (instead of complex statistics), and delving deeply into co-existing features.

With these steps of our methodology in mind, we will use for illustration the obsolescing feature of double demonstratives in rural Ontario, Canada (e.g. *... and all-of-a-sudden* ***this here fox*** *was in the window*; Tagliamonte, 2013-2018). In the process we will demonstrate the discovery of patterned variation and underlying systemic developments that offer new explanations for language change.

References

Davies, M. (2010). The Corpus of Historical American English. https://www.english-corpora.org.

Henry, A. (1995). *Belfast English and Standard English: Dialect variation and parameter setting.* New York & Oxford: Oxford University Press.

Jankowski, B. & Tagliamonte, S. A. (2017). A lost Canadian dialect: The Ottawa Valley 1975-2013. In Säily, T., Nurmi, A., Palander-Collin, M. & Auer, A. (Eds.), *Exploring future paths for historical sociolinguistics.* Amsterdam: John Benjamins. 239-274.

Oxford English Dictionary. www.oed.com

Rupp, L. & Tagliamonte, S. A. (to appear). Loss and emergence: (Double) demonstratives and indefinite *this* in Ontario dialects. *Journal of Pragmatics*.

Schilling-Estes, N. & Wolfram, W. (1999). Alternative models of dialect death: Dissipation vs. contraction. *Language* 75(3): 486-521.

Tagliamonte, S. A. (2006). Historical change in synchronic perspective: The legacy of British dialects. In van Kemenade, A. & Lou, B. (Eds.), *Handbook on the history of English.* Malden and New York: Blackwell. 477-506.

Tagliamonte, S. A. (2013-2018). Social Determinants of Linguistic Systems. Insight Grant: Social Sciences and Humanities Research Council of Canada #435 130 065.

Palestinian Arabic in the diaspora: Evidence from dialect contact in Lebanon

Yasmine Abou Taha and Stephen Levey

Population displacement and migration triggered by protracted conflict have led to extensive contact between speakers of different varieties of Levantine Arabic (Horesh & Cotter 2016). Observing that these social processes open up avenues for exploring contact-induced change in the Arabic-speaking world, we investigate the possibility that Palestinian Arabic (PA), as spoken by Palestinian refugees in Beirut, is converging with Lebanese Arabic (LA), the majority and socially dominant variety.

Our synchronic data come from 45 hours of spontaneous speech recorded from 39 speakers of PA and 27 speakers of LA, stratified by age, sex, and level of education. To extend the time-depth of the analysis, we additionally examine two diachronic sources representing each comparison variety: (i) a sub-set of the *Palestinian Oral History Archive* (POHA), comprising recordings elicited in Lebanon from first-generation Palestinian refugees born as early as 1897; and (ii) a series of plays belonging to the *Lebanese Popular Theatre Corpus* (LPTC), performed in colloquial LA and televised between the early 1960s and 1983.

Drawing on the framework of comparative variationist sociolinguistics (e.g., Poplack & Tagliamonte 2001), our linguistic focus is on a socially salient phonological variable involving the raising of /a:/ to [e:] in word-medial position, a stereotypical feature of LA (Naïm 2006), but traditionally absent from the PA spoken in Beirut (Hennessey 2011). We also examine two morpho-syntactic variables, the expression of verbal negation and future temporal reference, whose structural congruence in the two contact varieties should ostensibly enhance the possibility of convergent change (Thomason 2001:76). To the extent that convergence has taken place, we hypothesize that it should be detectible in socially salient aspects of segmental phonology (Trudgill 1986:20) and that it should have affected multiple linguistic components (Thomason 2001: 92-93), as gauged from detailed quantitative examination of the *internal structure* of the targeted variable systems in each variety (Poplack 2020:47).

Results reveal evidence of contact-induced change in PA in the raising of /a:/ to [e:] in word-medial position, and dialect levelling affecting the system of verbal negation, where there is a reduction in socially marked variants used by educated third-generation Palestinians. By contrast, the expression of future temporal reference in PA diverges from LA, notably in relation to the dramatic increase in the use of the proclitic future marker, *ħa-,* also observed in varieties of PA spoken outside Lebanon (AbuAmsha 2016).

Among the major factors shaping the outcomes of dialect contact in Beirut, we implicate the social characteristics (generation, level of education, mobility) of speakers, as well as the communal insularity of Palestinian refugee camps. Although our general findings do not impeach the influential view that extra-linguistic factors are primordial determinants of contact-induced change (Thomason & Kaufman 1988; Thomason 2001), our results relating to the variable expression of futurity caution that: (i) internal, structural constraints act with external ones in determining the outcomes of contact (Sankoff 2013:502); and that (ii) contact-induced change cannot be adduced from shared surface correspondences alone, contra Leddy-Cecere (2018).

REFERENCES

AbuAmsha, D. 2016. The future marker in Palestinian Arabic: An internal or contact-induced change? In L. Hracs (ed.) *Proceedings of the 2016 Annual Conference of the Canadian Linguistic Association.* < http://cla-acl.ca/actes-2016-proceedings/>

Hennessey, A. 2011. The linguistic integration of the Palestinian refugees in Beirut: A model for analysis. Unpublished M.A. dissertation, American University of Beirut.
Horesh, U. & Cotter, W. 2016. Current research on linguistic variation in the Arabic-speaking world. *Language and Linguistics Compass* 10(8): 370-381.

Leddy-Cecere, T.A. 2018. Contact-induced grammaticalization as an impetus for Arabic dialect development. Unpublished PhD dissertation, University of Texas at Austin.

Naïm, S. 2006. Beirut Arabic. In K. Versteegh, M. Eid, A. Elgibali, M. Woidich & A. Zaborski (eds.), *Encyclopedia of Arabic Language and Linguistics*, *Vol I*. Brill: Leiden, pp. 174-286.

Poplack, S. 2020. A variationist perspective on language contact. In E. Adamou & Y. Matras (eds.), *Routledge Handbook of Language Contact*. London: Routledge, pp.46-62.

Poplack, S. & Tagliamonte, S. 2001. *African American English in the Diaspora*. Oxford: Wiley-Blackwell.

Sankoff, G. 2013. Linguistic outcomes of bilingualism. In J.K. Chambers & N. Schilling (eds.) *The Handbook of Language Variation and Change.* Oxford: Wiley-Blackwell, pp. 501-518.

Thomason, S.G. 2001. *Language Contact: An Introduction.* Washington, DC: Georgetown University Press.

Thomason, S.G. & Kaufman, T. 1988. *Language Contact, Creolization, and Genetic Linguistics.* Berkeley/Los Angeles: University of California Press.

Trudgill, P. 1986. *Dialects in Contact.* Oxford: Blackwell.

[500 words, excluding references]

Abstract

Individual sensitivity to change in the lingua franca use of English

*Irene Taipale[1], Mikko Laitinen[1,2]

[1]University of Eastern Finland, [2]Center for Data Intensive Sciences and Applications, Linnaeus University

Correspondence:

Irene Taipale
irene.taipale@uef.fi

Despite the persistent dichotomy between native and non-native English being questioned, the study of variation and change largely focuses on native varieties of the language (Li 2020). Most studies aiming to detect patterns of linguistic regularity also utilize large data sets that attempt to minimize the presence of the individual. Recent corpus studies show, however, that advanced non-native users of English may display heightened sensitivity to features undergoing frequency shifts similarly to native speakers (Laitinen 2016). Our examination of variation and change in English as a lingua franca (ELF) is operationalized by inspecting Americanization, which is a fruitful source of enquiry since American English (AmE) predominates with several patterns of linguistic change in codified varieties of English (Leech et al. 2009; Baker 2017; Gonçalves et al. 2018).

We utilize geo-tagged tweets retrieved from the Nordic Tweet Stream, a real-time monitor corpus freely available for research and re-use at https://cs.uef.fi/nts/. The material consists of the idiolects of 150 individual Finnish users who actively tweet in English from three geographically varying areas: the countryside, mid-size towns and large cities. By making use of one of the first non-native English data sources that enable the inspection of intranational regional variation, we offer new and unique perspectives for the study of regional variation in English. While Americanization closely overlaps with other processes, such as colloquialization, the degree of Americanization can be quantified by using a simplified dichotomy between AmE and BrE. Our focus is on spelling and lexico-grammatical and morphological variables, such as V *+ -ing* | V + infinitive (e.g. *start doing* | *start to do*) and expanded predicates (e.g. *take a look* | *have a look*). The quantitative observations show that, particularly in the case of grammatical features, ELF speakers appear to have generally adhered to ongoing linguistic change.

References:

Baker, P. (2017). *American and British English: Divided by a common language?* Cambridge: Cambridge University Press.

Gonçalves, B., Loureiro-Porto, L., Ramasco, J., Sánchez, D. (2018). 'Mapping the Americanization of English in space and time', *PLoS ONE* 13(5). https://doi.org/10.1371/journal.pone.0197741.

Laitinen, M. (2016). "Ongoing changes in English modals: On the developments in ELF", in *New Approaches in English Linguistics: Building Bridges*, eds O. Timofeeva, S. Chevalier, A.-C. Gardner, and A. Honkapohja (Amsterdam: John Benjamins)*, 175–196.*

Leech, G., Hundt, M., Mair, C., and Smith, N. (2009). *Change in Contemporary English: A Grammatical Study*. Cambridge: Cambridge University Press.
Li, W. 2020. Multilingual English users' linguistic innovation. *World Englishes* 39(2), 236–248.

New criteria for selecting a local dialect speaker: how much important the grandparents' information?

Akiko TAKEMURA

This study aims at setting new criteria for selecting a local dialect speaker by inquiring about the origin of their parents, which accommodates to the reality in big cities. When selecting a dialect speaker in Japan, it is often said that the person who represents a local dialect must have locally-born parents and grandparents. However, how many people can meet the criteria in big cities? In view of the difficulty to find such dialect speaker, it is high time that this criteria must be simplified.

In the past, Chambers and Heisler (1990) proposed the Regional Index to determine whether or not the person can be a representative of the dialect. The influence of parental origin is considered, but not a priority. However, the parental origin influences children's acquisition of the local dialect (Payne 1976, Stanford 2008).

This study shows the results of two surveys. The first survey is based on the data collected in 1965 in Hokkaido, Japan, by NINJAL. The objective was to see if the immigrant family members acquire the standard Japanese through three generations. Hokkaido, situated in the northern part of Japan, was once a land of Ainu, indigenous people. In around 1870, people from the mainland Japan started to settle in for development. Therefore, there were many immigrants from all over Japan. In this survey, three immigrant families participated and in each family three family members (first, second and third generation) read out loud the word list. They documented their accent. Using this data, the author carried out a quantitative study to see how much the dialect of the first generation is transmitted to second and third generation. The results show that the correspondence rate for the dialect of the first generation is different from family to family, but overall the correspondence rate decreases toward third generation. The correspondence rate between first and third generation was between 13% and 30%. In other words, the dialect of the first generation was partially

transmitted.

Then, how much the parental origin influences the children's acquisition of local dialect? The author carried out a survey in Kagoshima, Japan, to see how much children whose parents are not locally born acquire the local dialect accent. The result indicates that the children whose parents are locally born acquire the local accent much better than the children whose parents are not locally born. The children whose mother was locally born showed better performance in acquisition of local accent than the children whose father was locally born.

In conclusion, these results suggest that we do not have to refer to the origin of the grandparents when choosing a dialect speaker because the grandparent's dialect was not much transmitted to third generation. However, the parental origin influences children's dialect acquisition. Therefore, when selecting a dialect speaker, we should ask the person's parental origins. Ideally, the person's parents come from the same dialect region, but it is also acceptable the case where the person's mother comes from the same dialect region.

(495 words)

References:

Chambers, J. K. and Heisler, Troy (1990) Dialect Topography of Quebec City English. *Canadian Journal of Linguistics* 44(1). 23-48.

National Institute for Japanese Language and Linguistics (NINJAL) (1965) Kyōtsūgoka no katē: Hokkaidō ni okeru oyako sandai no kotoba [The process of language standardization: language through three generations in Hokkaidō]. *Kokuritsu kokugo kenkyūsho hōkoku* [NINJAL Reports] 27. Kokuritsu kokugo kenkyūsho [National Institute for Japanese Language and Linguistics].

Payne, Arvilla (1976) *The acquisition of the phonological system of a second dialect*. Ph.D dissertation. University of Pennsylvania.

Stanford, James. N. (2008) Child dialect acquisition: New perspectives on parent/peer influence. *Journal of Sociolinguistics* 12. 567-506.

Between dialectometry and sociolinguistics: the analysis of internal border effects

Esteve Valls (Universitat Internacional de Catalunya)

Following the realisation that the north-western dialect continuum of Catalan is splitting along the political border between Catalonia and Aragon, in Spain (as shown, for example, by Valls et al. 2013), this research upholds the view that internal borders should be incorporated into border studies, since they often trigger processes of linguistic convergence and divergence which alter the most common patterns of linguistic diffusion. The language change is analysed in apparent-time using a combination of dialectometric techniques that constitutes an innovation within the field of border effects, and which, in the specific case of Catalan in Aragon, illustrates the usefulness of dialectometry in detecting processes of structural hybridisation in areas where the vitality of the language is most seriously undermined. Lastly, this investigation evinces the need to further develop a form of social dialectometry that not only answers sociolinguistic questions, but also makes it possible to objectively evaluate the social motivations fuelling the ongoing changes —an attempt to bring dialectometry and sociolinguistics closer together that we explore by using generalised additive mixed-effects regression modelling, in line with Wieling et al. (2011).

Valls, Esteve; Wieling, Martijn & Nerbonne, John. (2013). Linguistic advergence and divergence in north-western Catalan: A dialectometric investigation of dialect leveling and border effects. *Literary and Linguistic Computing* 28(1).

Wieling, Martijn; Nerbonne, John & Baayen, Harald. (2011). Quantitative Social Dialectology: Explaining Linguistic Variation Socially and Geographically. *PLoS ONE*, 6(9).

Disguised variation: the case of /e/ in Hungarian dialectology

Fruzsina S. Vargha
Hungarian Research Centre for Linguistics (Budapest)

Standard Hungarian has a high and a low short front vowel, /i/ and /ɛ/, but most Hungarian dialects also have a mid /e/. Standard Hungarian and those dialects where there is no distinction between /ɛ/ and /e/ have a „one-e" system, and the dialects with /e/ vs. /ɛ/ distinction have a „two-e" system (Kontra–Vargha 2014). In the main resource of Hungarian dialect pronunciation, The Atlas of Hungarian Dialects (Deme–Imre: 1968–1977), the pronunciation of /ɛ/ is transcribed with great spatial variation (Vargha 2020: 454), but, to the contrary, the transcription of /e/ is phonetically homogenous. In his monograph about Hungarian dialects, Imre already states that the apparent lack of variation in the pronunciation of /e/ might be the consequence of a consistently erroneous transcription method of these vowels (Imre 1971: 272). While Atlas data are registered with a detailed phonetic transcription, /e/ is almost always transcribed as [e], as if the transcribers' only aim were to merely indicate the phonemic distinction.

Data collection for the Atlas was made with paper and pencil (respondents' answers were transcribed on the spot by the fieldworkers). However, one or two hours of interviews (mainly spontaneous speech data) were also recorded at 352 locations during 1960 and 1964. The usefulness of older recordings in dialectology and in the study of dialect change was demonstrated in several studies (e.g. Brissos 2014, Purnell 2013, Thomas 2017), and the acoustic analysis of the field recordings might lead to the verification of the transcribed atlas data. The quality of the Hungarian recordings is not always good enough, but in most cases, F1 and F2 are reliably measurable manually.

Formant measurements of /e/ and /ɛ/ are made at 30 locations and the results are compared to the vowel qualities deduced quantitatively from the atlas data. While the transcriptions of /ɛ/ are largely confirmed, normalized F1 and F2 values of /e/ and /ɛ/ also demonstrate a regional variation for the phonetic realization of both vowels. In western dialects, where /ɛ/ is realized more open (with higher F1 and lower F2), /e/ is typically more open as well. Thus /ɛ/ sounds of north-eastern dialects are acoustically nearly identical to western realizations of /e/.

The determination of exact vowel qualities and the comparison of F1 and F2 values of different vowels can contribute to a better understanding of historical vowel changes (merger of /ɛ/ and /e/ in today „one-e" systems), and might lead to predictions about the future of /e/ and /ɛ/ distinction in dialect areas where it is still maintained.

Diffusion of Viennese Monophthongization in Austria's traditional dialects

Philip C. Vergeiner (University of Salzburg), Lars Bülow (University of Vienna), Jan Luttenberger (Austrian Academy of Science), David Britain (University of Bern)

In this paper, we investigate the geographical and structural diffusion of Viennese Monophthongization (Moosmüller & Vollmann 2001; Moosmüller & Scheutz 2013). By means of a new numerical measure to assess and compare formant movement in 18 lexical items, we provide evidence that Viennese Monophthongization is an ongoing, regular sound change transforming [aɛ̯] and [aɔ̯] gradually into [æː] and [ɒː] in the dialects of (Eastern) Austria.

The data are based on direct dialect recordings of 76 speakers in two age-groups in 19 rural locations of eastern and central Austria. Results indicate that [æː] and [ɒː] are diffusing in a wave-like fashion from Vienna (where Viennese Monophthongization originated, cf. Gartner 1900). Even though Viennese Monophthongization is reported to have been established in other bigger cities for more than 30 years (Moosmüller and Vollmann 2001, Moosmüller and Scheutz 2013), the data show no evidence for diffusion from these cities (cf. for different models of spatial diffusion Britain 2012). There are also other factors affecting the degree of formant movement: The phonetic-phonological environment (stress and the following consonant) explains most of the variance in the data, whereas no frequency effects (Phillips 2006) could be found. Furthermore, we identified social identity, cultural space (Horvath and Horvath 2001), and gender-related network structures as language external factors.

References:

Britain, David. 2012. Diffusion. In Alexander Bergs & Laurel J. Brinton (eds.), English historical linguistics. An international handbook (Handbücher zur Sprach- und Kommunikationswissenschaft 1–2), 2031–2043. Berlin: De Gruyter Mouton.

Gartner, T. 1900. Lautbestand der Wiener Mundart. Zeitschrift für hochdeutsche Mundarten 1. 141–147.

Horvath, Barbara M. & Ronald J. Horvath. 2001. A multilocality study of a sound change in progress. The case of /l/ vocalization in New Zealand and Australian English. Language Variation and Change 13(1). 37–57. https://doi.org/10.1017/S0954394501131029 (accessed 29 July 2021).

Moosmüller, Sylvia & Ralf Vollmann. 2001. "Natürliches Driften" im Lautwandel: die Monophthongierung im österreichischen Deutsch. Zeitschrift für Sprachwissenschaft 20(1). 42–65.

Moosmüller, Sylvia & Hannes Scheutz. 2013. Chain shifts revisited. The case of monophthongisation and E-merger in the city dialects of Salzburg and Vienna. In Peter Auer, Javier Caro Reina & Göz Kaufmann (eds.), Language variation – European perspectives IV: Selected papers from the Sixth International Conference on Language Variation in Europe (ICLaVE 6), Freiburg, June 2011 (Studies in Language Variation 14), 173–186. Amsterdam: Benjamins.

Phillips, Betty S. 2006. Word frequency and lexical diffusion. Hampshire: Palgrave Macmillan.

Conference: Methods XVII, Johannes Gutenberg-University Mainz

Teenagers' language in Estonia: the use of English in a newly compiled corpus

Virve Vihman, Kadri Koreinik, Liina Lindström, Aive Mandel, Maarja-Liisa Pilvik & Kristiina Praakli
University of Tartu

Teenagers are avid linguistic innovators, and have been shown to be the drivers of some linguistic changes (Eckert 2003; Androutsopoulos 2008; Tagliamonte 2016). Accessing samples of teen language poses a methodological challenge, yet it is crucial for a better understanding of youth language and its role in variation and change. Moreover, most of the research in this area has focussed on the language of L1 English-speaking youth, but their linguistic practices in English are influencing young people's language use much more broadly. English is a *lingua franca* used in entertainment and social media channels, and social media has grown to play a major role in young people's interactions. Hence, examining teenage language outside of the English-speaking world is important for understanding new language contact phenomena that come not from direct spoken communication between interlocutors, but rather from computer-mediated communication (CMC). In this talk, we ask how English affects the language use of young L1 Estonian-speaking youth in communication amongst themselves. We use case studies of lexical usage to identify the extent to which English has affected young people's usage: the rate and lexical diversity of code-switching in the corpora and the functions of the use of *nagu* 'like'.

The paper draws on data from a project investigating teenage language use in Estonia, in which two subcorpora, spoken language and digital messaging (DM), have been compiled through a participatory approach. The dual corpora can offer insights into contact with English as well as the language varieties used in spoken and DM registers. The project aims to identify variation by age, gender and geographic area. Participants (n = 131) range in age from 9 to 18 (mean=13.68, SD=2.40), are 71% female (94 female, 37 male) and represent four regions of Estonia. The language samples include 97 hours of spoken conversations and 204,210 words of DM texts.

We analyse the rate and lexical diversity of English words and phrases and the distribution of items across the dataset to investigate differences across ages and genders. We found that all major parts of speech are represented in the English-language lexical items, as well as categories such as taboo words, interjections and discourse markers. Although we do not find differences in amount of code-switching between genders, based on analysis of a sub-sample of the corpus, preliminary results suggest differences in the lexicon and the functions of code-switching.

We also report on a study of the functions of *nagu*, which is used in similar functions as English 'like', including the quotative 'be like'. The word is used overall much more frequently in the teen corpus than in the Balanced Corpus of Standard Written Estonian and the Phonetic Corpus of Spontaneous Speech. We trace its use across the age groups in our study and functions, to determine whether the use of *nagu* shows ongoing change. Preliminary results indicate an increase in the use of the word across our age groups.

References:

Androutsopoulos, J. 2008. Research on Youth-Language. In U. Ammon, N. Dittmar, K.J. Mattheier & P. Trudgill (eds.), *An International Handbook of the Science of Language and Society*, 496-1505. De Gruyter Mouton.
Eckert, Penelope 2003. Language and Adolescent Peer groups. *Journal of Language and Social Psychology*, 22(1): 112-118.
Tagliamonte, Sali A. 2016. *Teen talk: The language of adolescents.* Cambridge: Cambridge University Press.

Allophones, not exemplars, are the main carriers of social meaning in accent processing

Hielke Vriesendorp

Which linguistic units are the main carriers of social meaning in accent processing? Sociolinguistic research on the cognitive processing of language variation has provided evidence in support of exemplar theory. This phonological theory posits that speech is processed by matching linguistic input to detailed memories of words, or 'exemplars', which include contextual and social information. Walker and Hay (2011), for example, find that words that are typically used by older speakers are recognised quicker when they are pronounced by old sounding voices. This implies social and contextual detail influences language processing, and therefore that sociolinguistic processing is integral to linguistic processing more broadly. However, much less is known about how exemplar models work in the processing of social meaning itself (i.e. what social information – region, social class, persona, stance, etc. – is signalled or evoked by linguistic variants). This is particularly true in the context of so-called 'hybrid exemplar models'. In these models, listeners are posited to not just use highly specific exemplars in language processing, but to also abstract over patterns in these exemplars and use these for speech processing as well (as found for example by Ernestus 2014). Which of these are used when we process social meaning?

The current paper investigates this through a large-scale accent recognition task which compared accent recognition accuracy for high-frequency and low-frequency lexical stimuli, as well one accent recognition task with non-word stimuli. If there are differences between performance in high and low lexical frequency stimuli, that would suggest that lexical exemplars or other lexical representations are central to accent processing. Recognition on the basis of non-word stimuli would suggest segmental representations such as allophones (can) carry accent information.

In the experiments, British listeners were asked to recognise three different groups of English accents: Yorkshire, General American, and Standard English. They heard isolated words, pronounced by 42 different speakers. In the first experiment the critical items fell into two conditions: high-frequency

lexical items and low-frequency lexical items. And in the second experiment all stimuli were non-words. It was possible to control for speaker voice, word length, intonation, and distinctive accent features by playing respondents two closely matched stimuli at separate points in the experiment, and a closely matched non-word in the second experiment. For example, they would hear the high-frequency word *ask* and low-frequency *flask* pronounced by the same speaker of SSBE, with the same intonation and voice quality, each at a different point in the experiment. In the non-word experiment the stimulus *nask* was used.

The two experiments found that recognition was the same between high-frequency and low frequency stimuli, weakening the idea that lexical exemplars are the driving force behind accent recognition. Accent recognition was distinctly above chance in the non-word stimulus task, suggesting segmental representations such as allophones are able to carry accent information. Still, recognition was lower in the non-word stimuli, suggesting that having an abstract phonological word form as an 'anchor point' for segmental information is helpful to the recognition and processing of social meaning.

References

- Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 142, 27–41.
- Walker, A., and Hay, J. (2011). Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology*, 2(1), 219–237.

Longitudinal sociolinguistic projects as brands

Suzanne Evans Wagner, Betsy Sneller, and Jack Rechsteiner

Michigan State University, USA

Since March 2020, sociolinguists have creatively adopted remote data collection techniques to replace face-to-face methods (e.g. Hall-Lew et al 2021, Leeman et al 2021). Our research group has turned from gathering regional speech recordings via sociolinguistic interviews to collecting self-recorded audio 'diaries' via a mobile app. However, the opportunities afforded by this change—such as greater geographical reach—have been counterbalanced by challenges (Sneller 2022). For our longitudinal project, MI Diaries, primary challenges included making potential participants aware of our project, gaining their trust, and encouraging them to remain with the project. Fortunately, these challenges are not unique to sociolinguists. Companies and non-profit organizations must also recruit and retain customers/donors, and they have established many successful strategies in this regard. We therefore took a cross-disciplinary approach and envisioned our project as a *brand.*

MI Diaries needed to build a brand of safety and honesty, so participants would feel comfortable sharing stories from their lives with a non-visible researcher. To address this, we constructed a social media strategy, defined our mission, vision, and values, created a visual identity, managed our press and publicity, and utilized promotional incentive programs. As a result, we are successfully recruiting, retaining and engaging participants. Over the last 18 months, we have received more than 2,000 submissions from over 250 participants, adding up to 375+ hours of audio.

In this presentation, we provide an overview of branding theory (Sammut-Bonnici 2015) and its application to public sector brands, including research projects (Boenigk & Becker 2016). We propose that longitudinal sociolinguistic projects can benefit from brand management practices to meet their research goals. Further, as we discuss with reference to our own community outreach activities, a strong brand can support the broader mission of sociolinguistics to educate the public about social and linguistic diversity.

References

Boenigk, S. & A. Becker. 2016. Toward the importance of nonprofit brand equity. *Nonprofit Management and Leadership* 27(2). 181–198.

Hall-Lew, L., C. Cowie, S. J. McNulty, N. Markl, C. Lai, C. Llewellyn, B. Alex, N. Fang, Z. Elliott, and A. Klinger. 2021. The Lothian Diary Project: Investigating the Impact of the COVID-19 Pandemic on Edinburgh and Lothian Residents. *Journal of Open Humanities Data* 7(4):1–5. DOI: https://doi.org/10.5334/johd.25

Leemann, A., P. Jeszenszky, C. Steiner (2021). Linguistic fieldwork in a pandemic: Combining smartphone app recordings with Zoom videoconferencing. NWAV49, Austin (TX), October 21.

Sammut-Bonnici, T. 2015. Brand and branding. In Cooper, C. (ed), *Wiley Encyclopedia of Management*, Volume 12: Strategic Management. 3rd edition. 1–3.

Sneller, Betsy. 2022. (ed.) Covid-era sociolinguistics. Special issue, *Linguistics Vanguard.*

‘Americanness’ in British song lyrics: The case of intensifiers

Ayano WATANABE

There has been a widespread interest in ‘Americanness’ amongst British singers in the last few decades (e.g., Beal 2009; Simpson 1999; Trudgill 1983). Singing in American accents is popular with UK artists, as a recent article in *The Telegraph* entitled ‘Why you put on an American accent when you sing’ demonstrates. Whilst such articles lament the tendency for British singers to employ US accents in singing instead of using their UK accents, linguistic accounts have suggested a number of explanations for this singing style. Generally, such research has addressed the following questions, which are also the focus of my research:

1. How do British singers adopt American English in singing?
2. Why do British singers sing like this? Is this an attempt to imitate American songs?

My research departs from the study of accent and builds on Werner’s (2012) work which shows that the use of linguistic features which are often associated with American English usage is also found at a lexico-grammatical level. This paper will therefore scrutinize the frequency of lexico-grammatical variants associated with US English in contemporary British songs. The focus of the paper will be intensifiers such as *very* and *so*. Previous research (e.g., Aijmer 2018; Stratton 2018) has reported that certain intensifiers often appear much more readily in either British English corpora or US English corpora (e.g., *real* for ‘Americanness’ or *well* for ‘Britishness’), which makes these features ideal for the purpose of this study.

The paper will introduce two song lyrics corpora: British Popular Music Corpus of English (PMCE-UK) and American Popular Music Corpus of English (PMCE-US). Each corpus contains approximately 5,500 British/American singers’ songs (ca. 1,500,000 words) which appeared in the top 20 of Smith’s (2016) British charts and *Billboards*’ American charts from 1953-2009, respectively. A frequency comparison between British National Corpus (BNC) and Santa Barbara Corpus of Spoken American English (SBCSAE) permitted me to calculate the proportion of intensifiers indexing either American English or British English in PMCE-UK as well as in PMCE-US. The effects of musical genre as well as other factors such as the songwriter’s nationality and whether a song is a cover version were also examined.

In total, 1,700 intensifier tokens were extracted from PMCE-UK. The analysis demonstrates that British songs have a very high use of intensifiers associated with US English (e.g., *so*, *real*) (over 80%). Diachronic analysis revealed that proportion of ‘US’ intensifiers is stable, with a slight increase in the 1990s-2000s. Moreover, I found that the ‘US’ variants are almost categorical in hip hop and jazz and less frequent in other genres (e.g., pop), indicating that genres play an important role in the variable choice. By contrast, other factors were not very effective. It is important to note that these patterns are similar to those found in PMCE-US. Following Bell's (2001) *Referee Design*, the patterns would reflect the songwriter’s attempt to accommodate to American music, while other explanations (e.g., accommodation to the intended audience) are also possible.

References

Aijmer, Karin. 2018. "'That's Well Bad': Some New Intensifiers in Spoken British English." Pp. 60–95 in *Corpus approaches to contemporary British speech: Sociolinguistic studies of the spoken BNC2014*, edited by V. Brezina, R. Love, and K. Aijmer. New York: Routledge.

Beal, Joan C. 2009. "'You're Not from New York City, You're from Rotherham': Dialect and Identity in British Indie Music." *Journal of English Linguistics* 37(3): 223–40.

Bell, Allan. 2001. "Back in Style: Reworking Audience Design." Pp. 139–69 in *Style and sociolinguistic variation*, edited by P. Eckert and J. R. Rickford. Cambridge: Cambridge University Press.

Simpson, Paul. 1999. "Language, Culture and Identity: With (Another) Look at Accents in Pop and Rock Singing." *Multilingua* 18(4):343–67.

Smith, Nick. R. 2016. *British Top 20 Hits (1952-2015)*. CreateSpace Independent Publishing Platform.

Stratton, James. 2018. "The Use of the Adjective Intensifier *Well* in British English: A Case Study of The Inbetweeners." *English Studies* 99(8):793–816.

Trudgill, Peter. 1983. "Acts of Conflicting Identity. The Sociolinguistics of British Pop-Song Pronunciation." Pp. 141–60 in *On Dialect: Social and geographical perspectives*, edited by P. Trudgill. Oxford: Blackwell.

Werner, Valentin. 2012. "Love Is All around: A Corpus-Based Study of Pop Lyrics." *Corpora* 7(1):19–50.

Subjunctive II and diminutive as hedging phenomena in German in Austria Evidence from free conversation data and experimental settings

Anja Wittibschlager & Katharina Korecky-Kröll (University of Vienna)

Hedging is an interesting field of pragmatics that is characterized by various interfaces to other linguistic levels (e.g., semantics, vocabulary, syntax, morphology, see e.g., Schröder/Zimmer 1997, Reisigl/Wodak 2009). We focus on the pragmatics-morphology interface (e.g., Dressler/Merlini Barbaresi 1994) and investigate two phenomena that exhibit considerable variation not only as a function of the language variety or the participants' sociolinguistic characteristics investigated (e.g., gender), but also of the methodology employed, namely the subjunctive II and the diminutive. Both phenomena are often employed for mitigation (e.g., for being polite or for expressing one's uncertainty).

We investigate German in Austria, as Austrians are well known for their high dialect loyalty and dialect competence and for their frequent use of dialects (see Lenz 2019). In addition, the subjunctive II as well as the diminutive were identified as typical Austrian hedging strategies by previous studies on pragmatic strategies in the German-speaking area (Muhr 2008; Warga 2008). Therefore, Austria may be considered an "ideal sociolinguistic research laboratory" (Lenz 2018: 269) for investigating these two hedging phenomena: We find a broad inventory of different subjunctive II forms ranging from Standard German synthetic forms (e.g., *hätte* 'have-SUBJ2') or Standard German analytic forms (e.g., *würde sagen* 'would say') to colloquial analytic forms (e.g., *tät sagen* 'do-SUBJ2 say') or synthetic base dialect forms (e.g., *sogad* 'say-SUBJ2'), see Breuer/Wittibschlager (2020). The same holds for diminutives, which are characterized by a large range of suffixes (Standard *-chen* or *-lein*, colloquial *-(e)l*, dialect *-erl, -i, -lan, -le, -li*) that may be partially combined with umlaut (stem vowel change, e.g. *Häs-chen* 'hare-DIM', e.g., Korecky-Kröll/Dressler 2007).

We investigate language production data from 40 adult native speakers of German from five small rural locations belonging to the five main dialect regions of Austria (Central Bavarian, South-Central Bavarian, South Bavarian, Bavarian-Alemannic, Alemannic). Participants were part of two age groups (18-35 and 60+) as well as two educational backgrounds (+/- high school diploma) and were largely balanced for gender.

As subjunctive II forms as well as diminutives appear only in mid-to-low frequencies in free conversation data, it was necessary to use additional methods in order to elicit sufficient numbers of relevant data. Thus, each participant was recorded in four settings:

1) a more formal interview with an Austrian researcher,
2) a less formal free conversation with a friend from the same location,
3) a translation task from the local dialect into Standard German
4) a translation task from Standard German into the local dialect.

A main goal of this multi-method approach was to get insight into participants' individual vertical variety spectra by grasping high, intermediate and low varieties (i.e., from Standard German to the base dialect).

We will discuss effects of the dialect region, participants' gender and age group as well as individual preferences of participants followed by a critical discussion of the methods employed. We conclude that all methods have certain advantages and disadvantages, but that a multi-method approach of different spontaneous and experimental settings is most appropriate when investigating phenomena of mid-to-low frequencies in everyday speech.

References

Breuer, Ludwig Maximilian / Wittibschlager, Anja (2020): On the variation of the subjunctive II in Austria - Urban and rural analyses. *Linguistic Variation,* 20(1), 136–171. https://doi.org/10.1075/lv.19005.bre

Dressler, Wolfgang U. / Merlini Barbaresi, Lavinia (1994): Morphopragmatics. Diminutives and Intensifies in Italian, German, and Other Languages. Berlin: De Gruyter.

Korecky-Kröll, Katharina / Dressler, Wolfgang U. (2007): Diminutives and Hypocoristics in Austrian German (AG). In: Savickienė, Ineta / Dressler, Wolfgang U. (eds.): The Acquisition of Diminutives. A Cross-linguistic Perspective. Amsterdam: Benjamins, 207–230.

Lenz, Alexandra N. (2018): The Special Research Programme „German in Austria. Variation – Contact – Perception". In: Ammon, Ulrich / Costa, Marcella (eds.): Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language Choice in Tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l'Europe (Sociolinguistica 32), 269–277.

Lenz, Alexandra N. (2019): Bairisch und Alemannisch in Österreich. In: Herrgen, Joachim / Schmidt, Jürgen Erich (eds.): Deutsch. Language and Space, volume 4. Berlin, Boston: De Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft 30.4), 318–363.

Muhr, Rudolf (2008): The pragmatics of a pluricentric language: A comparison between Austrian German and German German. In: Schneider, Klaus P. / Barron, Anne (eds.): *Variational Pragmatics. A focus on regional varieties in pluricentric languages*. Amsterdam: Benjamins, 211-244

Reisigl, Martin / Wodak, Ruth (2009): The discourse-historical approach (DHA). In: Wodak, Ruth/Meyer, Michael (eds.): Methods of Critical Discourse Analysis. 2. ed. London: Sage, 87–121.

Schröder, Hartmut / Zimmer, Dagmar (1997): Hedging Research in Pragmatics: A Bibliographical Research Guide to Hedging. In: Markkanen, Raija / Schröder, Hartmut (eds.): Hedging and Discourse. Approaches to the Analysis of a Pragmatic Phenomenon in Academic Texts. Berlin: De Gruyter, 249-272.

Warga, Muriel (2008): Requesting in German as a pluricentric language. In: Schneider, Klaus P. / Barron, Anne (eds.): *Variational Pragmatics. A focus on regional varieties in pluricentric languages*. Amsterdam: Benjamins, 245-266.

Identity, language attitudes and language use in the Belarusian-Russian border region

Curt Woolhiser (Boston College)

When the Soviet Union collapsed in 1991, the Republic of Belarus became, virtually overnight, a fully independent state with a fully-fledged international border. The Belarusian language, despite enjoying a limited official status alongside Russian in the Soviet period, had become a de facto minority language by the 1980s, converging mainly toward Russian. In contrast to the other former Soviet republics, however, post-Soviet Belarus did not give precedence to the development of the "titular" national language, and instead, under the authoritarian Lukashenka regime, saw the continued expansion of the Russian language in most social domains. While Belarusian became relegated largely to the cultural sphere, with a minimal presence in education and government administration, researchers began to witness the emergence of distinct national variety of Russian in Belarus, differing from the Moscow-based standard Russian with respect to a number of variables (Mečkovskaja 2005, Norman 2008, Woolhiser 2014). Nonetheless, the linguistic and educational establishments in Russia and Belarus have to date largely resisted the concept of "Belarusian Russian" as a distinct non-dominant standard variety.

This study is based on an online survey to be administered in May-June 2022 focusing on awareness of, attitudes toward, and reported use of a number of phonological, morphological and lexical features characteristic of "Belarusian Russian" among students of local origin at higher educational institutions in four cities along the Belarusian-Russian border: Mahiliow (Rus. Mogilev) and Orsha on the Belarusian side, and Smolensk and Roslavl' on the Russian side. Data on informants' social identities and levels of identification with their regions and their respective national communities will be collected. It is predicted that there will be a significant border effect in terms of awareness of and reported use of typical features of "Belarusian Russian," despite the presence of a shared Belarusian-like dialectal substratum on both sides of the border. In the case of Belarusian respondents, we are likely to find that those who express strongly negative attitudes toward a potential Anschluss with Russia will be particularly conscious of, and perhaps most likely to report use of, features that distinguish "Belarusian Russian" from the Russian language of the metropole.

Durnovo, N. N., N. N. Sokolov, D. N. Ušakov. 1915. *Opyt dialektologičeskoj karty russkogo jazyka v Evrope s priloženiem očerka russkoj dialektologii*. Moscow.

Mečkovskaja, Nina B. 2005. Postsovetskij russkij jazyk: Novye čerty v sociolingvističeskom statuse. *Russian linguistics* 29:1, 49 – 70.

Norman, Boris. 2008. Russkij jazyk v Belarusi segodnja. *Die Welt der Slaven* LIII, 289-300.

Woolhiser, Curt. 2014. "The Russian Language in Belarus: Language Use, Speaker Identities and Metalinguistic Discourse." In L. Ryazanova (ed.), *The Russian language outside the nation: Speakers and identities.* Edinburgh: Edinburgh University Press.

Methods XVII Mainz
Conference Theme: (Dia)Lects in the 21st Century

Special thematic session proposal: "Borders, Dialects and Standard Varieties"

Convenors: Curt **Woolhiser** and Stefan **Dollinger**

Political borders have not figured prominently in dialect geography, which as a discipline developed in lockstep with the formation of European nation states. Dialect geography was therefore influenced by 19th-century nationalism and, in turn, 19th-century perceptions of nationality influenced dialectology and its perception of what constitutes "a language". If considered at all, dialect geography has, by and large, ruled out political borders as linguistically interesting.

While this perspective has changed incrementally since Kremer (1979), Chambers and Trudgill (1980), Auer, Kerswill and Hinskens (2005), today we see a refocussing on more essentialist interpretations of what makes a language (superordinate) and what a dialect (subordinate) in certain segments of the scholarly community via interpretations that are tied to the perceived or measured strength and relevance of political borders. We see this kind of new essentialism in statements such as, e.g. "two percent" of variation between standard varieties "hardly make a 'variety'" (Elspaß & Niehaus 2014: 54). Such statements stand in clear contrast to the dynamic nature of language attitudes and cognitive perceptions of what constitutes relevant difference or salience, which is decisive in distinguishing varieties, not absolute and categorical difference. As Hickey puts it, "the number [of differences] does not need to be great, and there are cases where single features are involved" (2017: 2) to discriminate one variety from another.

In this workshop at Methods in Dialectology we aim to test and elaborate the notions of linguistic autonomy and heteronomy (Chambers & Trudgill 1998) within the framework of pluricentric theory and its predictions (e.g. Muhr 2016, Clyne 1995, Auer 2005, Dollinger 2019b). By giving consideration to linguistic behaviour, linguistic attitudes, cognitive representations of identities and the dynamic nature of the interplay between these factors, this workshop brings data from seven settings to the table. The settings are the contexts of Belgium-The Netherlands, Scotland-England, Catalonia-Aragon, Austria-German, Belarus-Poland and Belarus-Russia and Canada-USA.

The contributions in this workshop explore the junction of traditional language conceptualizations, e.g. "Dutch", "English", "Spanish" or "German", as they morph into more recently moulded standard varieties, e.g. Netherlandic Dutch and Belgian Dutch, Scottish English and (Northern) English English, Catalan and Castilian Spanish, Austrian German and German German, Belarusian Russian and Russian Russian, Belarusian and Polish. The contributions explore diverging standard language dynamics (Grondelaers & Speelman), ongoing linguistic hegemonic regimes (De Ridder, Dollinger), dialectometric methods that

include social assessments (Valls; Llamas, Watt and Brown), or identify new divergences along, often very young, political borders (Konczewska, Woolhiser).

While concepts of 19th-century nation building no longer carry weight in academic circles, their traces can be found in linguistic conceptualizations of standard varieties (Dollinger 2019a: 23-76), that is e.g. Russian, not Belarusian Russian; German, not Austrian German, American English, not Canadian English and the like. In this context, the recent dissemination of quantitative and computational methods with algorithms that are, by and large, blind to social salience, seems to spur a new kind of unintended hegemonic linguistics in which traditional standards (Netherlandic Dutch, Castllian Spanish, Russian Russian) are unwittingly upheld and enforced by what were thought to be objective methods of description (e.g. Dollinger 2019a: 64-76; 2021: 139-59). The contributions in this special session address these methodological dilemmas either directly (Grondelaers & Speelman, Valls, Llamas et al.) or describe the descriptive and theoretical conundrums (Woolhiser, Konczewska, De Ridder, Dollinger) that reflect and lay bare increasing tension in contemporary sociolinguistic theory and practice. A tension that has considerable repercussions in applied linguistic and real-life situations for these varieties' speakers.

Auer, Peter. 2005. The construction of linguistic borders and the linguistic construction of borders. In: Filppula, Markku, Juhani Klemola, Marjatta Palandern and Esa Penttilä (eds). 2005. *Dialects Across Borders : Selected Papers from the 11th International Conference on Methods in Dialectology (Methods XI), Joensuu, August 2002,* 3-30. Amsterdam: Benjamins.

Auer, Peter, Frans Hinskens and Paul Kerswill (eds.). 2005. *Dialect Change: Convergence and Divergence in European Languages.* Cambridge: Cambridge University Press.

Chambers, J. K. and Peter Trudgill. 1980. *Dialectology*. 1st ed. Cambridge: Cambridge University Press.

Chambers, J. K. and Peter Trudgill. 1998. *Dialectology*. 2nd ed. Cambridge: Cambridge University Press.

Clyne, Michael. 1995. *The German language in a changing Europe*. Cambridge: Cambridge University Press.

Dollinger, Stefan. 2019a. *The Pluricentricity Debate: On Austrian German and other Germanic Standard Varieties*. London: Routledge.

Dollinger, Stefan. 2019b. Debunking "pluri-areality": on the pluricentric perspective of national varieties. *Journal of Linguistic Geography* 7(2): 98-112.

Dollinger, Stefan. 2021. [*Austrian German or German in Austria? Identities in the 21st Century] Österreichisches Deutsch oder Deutsch in Österreich? Identitäten im 21. Jahrhundert*. 3rd ed. Vienna, Hamburg: new academic press.

Hickey, Raymond. 2017. Areas, areal features and areality. In The Germanic languages and areal linguistics. In *The Cambridge Handbook of Areal Linguistics*, ed. by Raymond Hickey, 1-15. Cambridge: Cambridge University Press.

Kremer, Ludger. 1979. *Grenzmundarten und Mundartgrenzen*. Köln: Böhlau.

Muhr, Rudolf (ed.) 2016. *Pluricentric Languages and Non-Dominant Varieties Worldwide. Part I: Pluricentric Languages across Continents. Features and Usage.* Frankfurt: Peter Lang.

Muhr, Rudolf. 2005. Language attitudes and language conceptions in non-dominating varieties of pluricentric languages. In: Rudolf Muhr (ed.), *Standardvariationen und Sprachideologien in verschiedenen Sprachkulturen der Welt.* Vienna: Peter Lang Verlag, 11-20.

1. Measuring standard language dynamics in Belgian and Netherlandic Dutch. Big data meets experimental attitude research

Stefan Grondelaers (Radboud University Nijmegen) & Dirk Speelman (University of Leuven)

In this talk we use a computationally enriched experimental technique (Grondelaers et al. 2020) to visualise language ideology (change) in Belgian and Netherlandic Dutch. A better grasp of Belgian and Netherlandic value systems is essential for making (somewhat valid) predictions about how the European national varieties of Dutch differ and diverge.

In a free response task, 211 Belgian and 177 Netherlandic respondents returned three adjectives in reaction to a number of regional and ethnic accent varieties and (for Belgium) two supra-regional varieties, viz. Belgian Standard Dutch and a stigmatised colloquial variety dubbed "Tussentaal". Valence information (pertaining to the positive/negative character of the responses) and big data-based distributional analysis (to detect semantic similarity) were used to cluster returns into 11 positive and 11 negative evaluative dimensions. Correspondence analysis was employed to visualise the correlations between these evaluative dimensions and the investigated varieties.

Crucially, the resulting "perceptual maps" confirm the very different standard language dynamics previously observed for Belgian and Netherlandic Dutch. While Netherlandic Dutch has stabilised into a "relaxed", consensual standard, Belgian Dutch continues to be conditioned by (conservative) ideology and prestige considerations. A comparison of older and younger evaluations, however, demonstrates that the Belgian system is very much in motion...

Grondelaers Stefan, Dirk Speelman, Chloé Lybaert & Paul van Gent (2020). Getting a (big) data-based grip on ideological change. Evidence from Belgian Dutch. *Journal of Linguistic Geography* https://doi.org/10.1017/jlg.2020.2.

Grondelaers, Stefan, Roeland van Hout & Paul van Gent (2016). Destandardization is not destandardization: Revising standardness criteria in order to revisit standard language typologies in the Low Countries. In Chloé Lybaert, Steven Delarue, Anne-Sophie Ghyselen & Jacques Van Keymeulen (eds.), (*De)standardisation in Europe – qualitative and quantitative approaches*. Theme issue of *Taal & Tongval, Language variation in the Low Countries*, 68/2, 119-149.

2. The usage of Belgian Dutch in translated and non-translated (audiovisual) fiction today and its sociolinguistic implications

Reglindis De Ridder (University of Stockholm)

Dutch as it is used in Belgium (henceforth *Belgian Dutch**) has taken a different course from *Netherlandic Dutch***, despite massive language planning efforts in the second half of the 20th Century to prevent this (Jaspers and Van Hoof 2013). For a long time, Belgian Dutch was considered a deviation from 'proper' Dutch. The turn of the century, however, marked the official recognition of the Dutch language area as a pluricentric language area with two equal national varieties in Europe by the Dutch language planning body (Nederlandse Taalunie 2003). Nevertheless, Netherlandic Dutch is still the dominant variety in translated (audiovisual) fiction. This paper discusses the results of a sociolinguistic analysis of 290 children's programmes focussing on the usage of Belgian and Netherlandic Dutch (De Ridder 2020a) with a reception study into parents' opinion as regards their children's exposure to both varieties of Dutch (idem 2020b). It highlights differences between local and imported programmes and calls for further sociolinguistic research into the language used in different children's media and how it may

affect language development in children and language attitude. Children's television has been criticized for its lack of diversity, yet, linguistically, children's media can also be out of touch with reality.

* In English, often referred to as 'Flemish', however, the official term used by the Nederlandse Taalunie is 'Belgisch Nederlands'/'Belgian Dutch'.
** In Dutch, colloquially referred to as 'Hollands', however, the official term used by the *Nederlandse Taalunie* is 'Nederlands Nederlands'/'Netherlandic Dutch'.

De Ridder, Reglindis (2020a). Dutch national varieties in contact and in conflict. In: R. Muhr, J. A. Mas Castells & J. Rueter (eds.) *European Pluricentric Languages in Contact and Conflict*. Wien: Peter Lang Verlag, 65-80.
De Ridder, Reglindis (2020b). Linguistic diversity in audiovisual media for children in Belgium and Austria. In: R. Muhr and Thomas, J. (eds.) *Pluricentric Theory beyond Dominance and Non-dominance*. Graz/Berlin: PCL press, 121-136. Available from https://pcl-press.org/publications/pluricentric-theory-beyond-dominance-and-non-dominance-a-critical-view/ (Last accessed on 26 September 2021).
Nederlandse Taalunie. (2003). Rapport Variatie in het Nederlands: eenheid in verscheidenheid [Report on the variation within Dutch: unity in diversity]. Available from https://taalunie.org/publicaties/94/variatie-in-het-nederlands-eenheid-in-verscheidenheid (Last accessed on 26 September 2021).
Van Hoof, Sarah & Jaspers, Jürgen (2013). Hyperstandardisation in Flanders. Extreme enregisterment and its aftermath. *Pragmatics*, 23(2), pp.331-359.

3. Comparing human and computer classification of phonetic features in the Scottish/English border region

Carmen Llamas, Dominic Watt and Georgina Brown

The *Accent and Identity on the Scottish/English Border* (AISEB) project (UK ESRC RES-062-23-0525) examined the links between phonological variation and local, regional and national identities at the extreme ends of the political border between England and Scotland (Eyemouth and Berwick in the east; Gretna and Carlisle in the west). The border is said to coincide with the most tightly-concentrated bundle of dialect isoglosses in the English-speaking world, turning Scotland into a 'dialect island' (Aitken 1992). The border therefore represents a prime context for the investigation of language and identity. In this paper we will discuss one of the tests used as part of AISEB's speech perception strand. Under the researcher's supervision, participants were asked to classify short audio samples according to the perceived origin of the speaker as a way of gaining insights into the socio-geographical associations and relative salience of local pronunciation variants, to complement those obtained using the Social Category Association Test (SCAT) described in Llamas, Watt & MacFarlane (2016). As an additional, objective means of probing the notion of salience, we compare the findings for our human participants to those yielded by a prototype automated accent classification system, Y-ACCDIST (Brown & Wormald 2017).

Aitken, A.J. (1992). Scots. In McArthur, T. (ed.). *The Oxford Companion to the English Language*. Oxford: Oxford University Press, pp. 893-899.

Brown, G. & Wormald, J. (2017). Automatic sociophonetics: Exploring corpora with a forensic accent recognition system. *Journal of the Acoustical Society of America* 142 (1): 422-433.

4. Between dialectometry and sociolinguistics: the analysis of internal border effects

Esteve Valls (Universitat Internacional de Catalunya)

Following the realisation that the north-western dialect continuum of Catalan is splitting along the political border between Catalonia and Aragon, in Spain (as shown, for example, by Valls et al. 2013), this research upholds the view that internal borders should be incorporated into border studies, since they often trigger processes of linguistic convergence and divergence which alter the most common patterns of linguistic diffusion. The language change is analysed in apparent-time using a combination of dialectometric techniques that constitutes an innovation within the field of border effects, and which, in the specific case of Catalan in Aragon, illustrates the usefulness of dialectometry in detecting processes of structural hybridisation in areas where the vitality of the language is most seriously undermined. Lastly, this investigation evinces the need to further develop a form of social dialectometry that not only answers sociolinguistic questions, but also makes it possible to objectively evaluate the social motivations fuelling the ongoing changes —an attempt to bring dialectometry and sociolinguistics closer together that we explore by using generalised additive mixed-effects regression modelling, in line with Wieling et al. (2011).

Valls, Esteve; Wieling, Martijn & Nerbonne, John. (2013). Linguistic advergence and divergence in north-western Catalan: A dialectometric investigation of dialect leveling and border effects. *Literary and Linguistic Computing* 28(1).

Wieling, Martijn; Nerbonne, John & Baayen, Harald. (2011). Quantitative Social Dialectology: Explaining Linguistic Variation Socially and Geographically. *PLoS ONE*, 6(9).

5. Cross-Border Language Contacts in the Polish-Belarusian Border Region in the 21st Century

Katarzyna Konczewska (Polish Academy of Science)

This paper focuses on border effects involving varieties of two closely related Slavic languages in contact, Polish and Belarusian, in the little-studied northern part of the contemporary Polish-Belarusian border region. The preliminary results of research allow us to posit the prevalence of productive bilingualism with diglossia in this area and to define the sociolinguistic situation as exoglossic, unbalanced, and four-component. The material for the study was collected by the author in 2015-2019 during dialectological interviews in communities on both sides of the Polish-Belarusian border, which it the lateral dialectology area of the Balto-Slavic contact zone.

Microarea studies are key regarding linguistic contact in the transitory zones since they allow, e.g. for the tracing of directions, the depth of local dialect infiltration.
The microarea under investigation is characterized by its relative inaccessibility, as well as heterogeneity of local residents in terms of national identity and religious affiliation. At present, the autochthonous local population is comprised of both Eastern Orthodox and Roman Catholics, mainly Poles and Belarusians in terms of national self-identification, who are the descendants of peasants and the petty gentry. A unique feature of the area under investigation is that for more than five hundred years it was an integral unit within various state formations; it was divided by a political border only in 1948. Uneven settlement processes due to landscape features, as well as historical and political factors influenced the formation of specific, multicomponent sociolinguistic situations on each side of the border.
In this presentation I will examine the pluricentric languages common in the area, as well as the linguistic codes used by their native speakers. While the theory and methodology of research on language and dialect contact in border regions have been addressed in the scholarly literature (e.g. Woolhiser 2005), Konczewska (2021) has shown that the peculiarities of the formation and development of the area under investigation would benefit from a more individualized approach.
Hypotheses concerning the course of linguistic contacts in peripheral areas are the key elements in the research of linguistic contact in the greater Baltic area. The verification of such hypotheses will optimize research quality and make new knowledge available. In these studies I strive to go pass beyond the research models of traditional linguistics, taking the work of ethnographers and ethnohistorians into account as well.

Konczewska, Katarzyna. 2021. Polacy i język polski na Grodzieńszczyźnie. Kraków: Instytut Języka Polskiego PAN.
Woolhiser, Curt. 2005, *Political Borders and Dialect Convergence/Divergence in Europe*. In: P. Auer, F. Hinskens, P. Kerswill (eds.), *Dialect Change: Convergence and Divergence in European Languages*. Cambridge: Cambridge University Press, 236-262.

6. Modelling standard varieties: epistemological considerations, “fail-safes”, and the peculiar case of German

Stefan Dollinger (UBC Vancouver)

This paper puts at its centre the German “Pluricentricity Debate” (Dollinger 2019), which explores the question whether the pluricentric view of languages is still adequate today. This debate is important, as recent critics have re-introduced the counter term of “pluri-areality” (Scheuringer 1996) and German dialectology has seen the branding of pluricentricity as an outdated model that is hampered by national limitations (e.g. Elspaß and Niehaus 2014, Herrgen 2015, Langer 2021, Koppensteiner & Lenz 2021). The pluricentric perspective of German – one language, several national standards – is, in German linguistics, now questioned more than at any point since Clyne’s (1984) landmark publication.

The debate affords the opportunity to inquire how German – and any other codified language – should be modelled in the 21st century and allows conclusions about gaps in English, Dutch and other varieties of comparable social use. To that purpose, a comparative view is taken in this meta study that contrasts the sociolinguistic situations, linguistic behaviours, attitudes and perceptions in German with other Germanic varieties. Although philology-specific concepts do have their place, it will be shown that "pluri-areality" represents no such case, leaving pluricentricity as the most appropriate theory to date, a concept that abides by the epistemological principle of hypothesis testing (Popper 1966).

Clyne, Michael G. 1984. *Language and Society in the German-speaking Countries*. New York : Cambridge University Press.

Dollinger, Stefan. 2019. *The Pluricentricity Debate: On Austrian German and other Germanic Standard Varieties*. London: Routledge

Elspaß, Stephan and Konstantin Niehaus. 2014. The standardization of a modern pluriareal language. Concepts and corpus designs for German and beyond. *Orð og tunga* 16: 47–67.

Herrgen, Joachim. 2015. Entnationalisierung des Standards. Eine perzeptionslinguistische Untersuchung zur deutschen Standardsprache in Deutschland, Österreich und der Schweiz. In: Lenz & Glauninger (2015), 139-164.

Koppensteiner, Wolfgang & Alexandra Lenz. 2021. Standard(s) aus der Perspektive von „Nicht-LinguistInnen" in Österreich. In Hoffmeister et al. (2021), 391-416.

Langer, Nils. 2021. Review of Dollinger 2019, with rejoinder. *Zeitschrift für Rezensionen zur Germanistischen Sprachwissenschaft*. Online First. https://www.academia.edu/45577124/

Popper, Karl R. 1966. *Logik der Forschung*. Zweite, erw. Auflage. Tübingen: Mohr.

7. Identity, language attitudes and language use in the Belarusian-Russian border region

Curt Woolhiser (Boston College)

When the Soviet Union collapsed in 1991, the Republic of Belarus, overnight, became a fully independent state with a fully-fledged international border. The Belarusian language, despite enjoying a limited official status alongside Russian in the Soviet period, had become a de facto minority language by the 1980s, converging mainly toward Russian. In contrast to the other former Soviet republics, however, post-Soviet Belarus did not give precedence to the development of the "titular" national language, and instead, under the authoritarian Lukashenka regime, saw the continued expansion of the Russian language in most social domains. While Belarusian became relegated largely to the cultural sphere, with a minimal presence in education and government administration, researchers began to witness the emergence of distinct national variety of Russian in Belarus, differing from the Moscow-based standard Russian with respect to a number of variables (Mečkovskaja 2005, Norman 2008, Woolhiser 2014). Nonetheless, the linguistic and educational establishments in Russia and Belarus have to date largely resisted the concept of Belarusian Russian as a non-dominant variety.

This study is based on an online survey to be administered from March to May 2022 on awareness of, attitudes toward, and reported use of a number of phonological, morphological

and lexical features characteristic of “Belarusian Russian” among students of local origin at higher educational institutions in four cities along the Belarusian-Russian border region: Mahiliow (Rus. Mogilev) and Orsha on the Belarusian side, and Smolensk and Roslavl’ on the Russian side. Data on informants’ social identities and levels of identification with their regions and with their respective national communities will be collected. It is predicted that there will be a significant border effect in terms of awareness of and reported use of typical features of “Belarusian Russian,” despite the presence of a shared Belarusian-like dialectal substratum on both sides of the border. In the case of Belarusian respondents, we are likely to find that those who express strongly negative attitudes toward a potential Anschluss with Russia will be particularly conscious of, and perhaps most likely to report use of, features that distinguish “Belarusian Russian” from the Russian language of the metropole.

Durnovo, N. N., N. N. Sokolov, D. N. Ušakov. 1915. *Opyt dialektologičeskoj karty russkogo jazyka v Evrope s priloženiem očerka russkoj dialektologii.* Moscow.

Mečkovskaja, Nina B. 2005. Postsovetskij russkij jazyk: Novye čerty v sociolingvističeskom statuse. *Russian linguistics* 29:1, 49 – 70.

Norman, Boris. 2008. Russkij jazyk v Belarusi segodnja. *Die Welt der Slaven* LIII, 289-300.

Woolhiser, Curt. 2014. “The Russian Language in Belarus: Language Use, Speaker Identities and Metalinguistic Discourse.” In L. Ryazanova (ed.), *The Russian language outside the nation: Speakers and identities.* Edinburgh: Edinburgh University Press.

Convergence and divergence of tone paradigms across Tai dialects in the 21st century

Chingduang Yurayong[1,2], Yuttaporn Naksuk[1] & Saknarin Pimvunkum[1]
[1]Mahidol University, [2]University of Helsinki

Having dispersed from the Proto-Tai-Kadai language spoken approximately 5000 years ago in Southern China (Ostapirat 2005: 126), the Tai branch and its linguistic populations have spread outside China towards India and Mainland Southeast Asia. Previous studies have proposed a taxonomical structure of the Tai branch with three major groups: 1) Northern Tai; 2) Central Tai; and 3) Southwestern Tai (Chamberlain 1975; Li 1977; Luo 1997). However, the subgrouping of Southwestern Tai in particular remains disputable. By only applying a conventional comparative method of historical linguistics, common patterns and clustering of phonological innovations do not give a straightforward dialect classification. This is very likely due to later language contact resulting from migration and relocation of the linguistic populations, and causing convergence among Tai dialects spoken in adjacent areas (Pittayaporn 2009: 298). Such a scenario is particularly common in Laos and Thailand where people were forced to migrate to new places, particularly during the war times when many Lao speakers fled to different parts of Thailand, creating new diaspora communities. As the resettlement of Lao speakers concerned mostly two generations upwards, the younger generation of dialect speakers already show a sign for shifting towards a national language or regional dialect of their current location (Akkharawatthanakun 2003).

In the present study, we conduct a large-scale investigation of over two hundred Tai dialects (Black Tai, Lue, Shan, Lao, Thai, etc.), focusing on tone paradigm. The question concerns how tones D (close syllables *-p*, *-t*, *-k, -ʔ*) rhyme with tones A-B-C (open syllables *-V*, *-N*). Our main goal is to identify change in progress, concerning rhyming patterns in tone paradigm of dialects which may have diverged from their proto-systems as we have arrived in the 21st century. The data collected from various grammatical and phonological descriptions of Tai dialects in Southeast Asia is organised according to whether tones D match with tones A, B and/or C in given dialects. This information is then processed by a Neighbor-Net algorithm (Bryant & Moulton 2004), which produces a network diagram showing the distance and clustering among tone profiles of each dialect under investigation.

The preliminary interpretation of the generated network diagram identifies several dialect clusters: 1) Lao proper; 2) Northern Thai; 3) Central Thai; and 4) Southern Thai. By capturing dialects which do not belong to a cluster of their own, we look further into their current speaking areas on the map and migration history of their speaking populations. A number of cases clearly point to a scenario where a given dialect has shifted its rhyming pattern of tones D, converging with a local dialect of the new settlement area in line with sociocultural assimilation. This phenomenon is manifest of language shift in progress which has taken and is taking place in many areas of Thailand where a younger generation of Lao dialect speakers in particular is shifting towards a national language, Standard Thai.

References

Akkharawatthanakun, Phinnarat. (2003). *Tone change: A case study of the Lao language* (Doctoral dissertation). Bangkok: Chulalongkorn University.

Bryant, David, & Moulton, Vincent. (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, *21*(2), 255-265.

Chamberlain, James R. (1975). A new look at the history and classification of the Tai languages. *Studies in Tai linguistics in honor of William J. Gedney*, 49-66.

Li, Fang-Kuei. (1977). *A handbook of comparative Tai*. Manoa, HI: University Press of Hawai'i.

Luo, Yongxian. (1997). *The subgroup structure of the Tai languages: A historical-comparative study* (Journal of Chinese Linguistics, Monograph Series 12). Hong Kong: The Chinese University of Hong Kong Press.

Ostapirat, Weera. (2005). Notes on phonological correspondences and vocabulary distribution. *The peopling of East Asia: putting together archaeology, linguistics and genetics*, 107-131.

Pittayaporn, Pittayawat. (2009). *The phonology of Proto-Tai* (Doctoral dissertation). Ithaca, NY: Cornell University.

Measuring language attitudes towards ethnolectal features in Swiss-German-speaking children: A mixed-methods approach

Melanie Röthlisberger (University of Zürich, Switzerland), Eline Zenner (KU Leuven, Belgium), Laura Rosseel (Vrije Universiteit Brussel, Belgium)

Background: This paper reports on a mixed-methods project that investigates how children aged 6 to 12 use and evaluate ethnolectal features in Swiss-German. Features of ethnolectal Swiss-German (e.g. example 1) are typically discussed in connection with adolescents with a migrant background (e.g., Tissot et al. 2011) and in the way these features are appropriated by non-migrant adolescents for stylistic purposes (Auer 2002; Schmid 2017). Little is known about the way adolescents acquire the usage and the social meaning attached to these forms.

(1) Omission of prepositions, articles, pronouns or auxiliaries, as in *Chani bleistift?* <Can I pen?> for *Chani en bleistift ha?* <Can I have a pen?>

Aim: This project investigates Swiss-German preadolescents' production, perception and evaluation of ethnolectal features through a mixed method approach, (a) taking stock of the linguistic repertoire available to children (production study) and, (b) measuring the social meanings attached to these features (evaluation study).

Production study: To identify the linguistic features of interest, we investigate anecdotal claims about the usage of ethnolectal features as defined by Auer (2002, see also Tissot et al. 2011) using spoken data collected through the diapix task (Baker & Hazan 2011) and free storytelling. For the production-oriented part, two studies were conducted in and around the city of Winterthur in one urban and one rural primary school with children aged 6-12. Preliminary analysis of the sampled speech indicates that the use of ethnolectal features is already common among children under 12. This is the case in both areas; this usage, however, seems to be restricted to migrant children in the rural area while also being used among non-migrant children in the urban areas.

Evaluation study: The social evaluation children attach to these features is assessed in a second, experimental step that includes a visually enriched and child-friendly version of the matched guise technique and a language awareness test (N=86). Results show that the youngest children have no preference for either the Swiss German or the ethnolectal guise and seem to lack awareness of the ethnolectal features. However, that awareness develops with age and so does a more positive overall evaluation of the Swiss German guise compared to the ethnolectal guise. The latter is furthermore increasingly associated with social meanings of 'non-Swissness'. Interestingly, social meanings of urbanity, which have been reported in adult populations, do not (yet?) appear in the evaluations held by pre-adolescents.

Implications: Obtaining a better understanding of the acquisition trajectory of Swiss-German ethnolectal features, more generally informs us about innovation and language change (cf. Cheshire et al. 2011) and about ongoing restructuring processes in Swiss-German (see, e.g. Leemann et al. 2014).

References

Auer, Peter. 2002. "Türkenslang": Ein jugendsprachlicher Ethnolekt des Deutschen und seine Transformationen. In Annelies Häcki Buhofer (ed.), *Spracherwerb und Lebensalter*, 255–264.

Tübingen/Basel: Francke.
Baker, Rachel & Valerie Hazan. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43(3). 761–770. doi:10.3758/s13428-011-0075-y.
Cheshire, Jenny, Paul Kerswil, Sue Fox & Eivind Torgersen. 2011. Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics* 15(2). 151–196.
Leemann, Adrian, Marie-José Kolly, Iwar Werlen, David Britain & Dieter Studer-Joho. 2014. The diffusion of /l/-vocalization in Swiss German. *Language Variation and Change* 26(2). 191–218.
Schmid, Stephan. 2017. Differenzierungsprozesse im Sprachgebrauch von Jugendlichen in der Deutschschweiz: zur sozialen Interpretation von ethnolektalen Sprechweisen in Schweizer Medien. *Bulletin Suisse de Linguistique Appliquée (Vals-Asla)* Nspécial(t. 1). 105–116.
Tissot, Fabienne, Stephan Schmid & Esther Galliker. 2011. Ethnolektales Schweizerdeutsch: Soziophonetische und morphosyntaktische Merkmale sowie ihre dynamische Verwendung in ethnolektalen Sprechweisen. In Elvira Glaser & Natascha Frey (eds.), *Dynamic des Wandels - Wandel und Variation: Akten des 3. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD)*, 319–344. Stuttgart: Steiner.
Zenner, Eline, Rosseel, Laura, & Dirk Speelman. 2020. Starman or Sterrenman: an acquisitional perspective on the social meaning of English in Flanders. *International Journal of Bilingualism* 25(3): 568-591.

Intra- and interindividual variation on the lexical level - Analyses across Austria

Theresa Ziegler, Jan Höll & Amelie Dorn
(University of Vienna)

The present talk discusses so-called 'Austriacisms' (Austrian 'shibboleths'), whose role for linguistic and social identity is highly debated in meta-language discourse (cf. Wodak et al. 2009). mainly concern the lexical domain, and less often also grammatical and phonetic levels (Wiesinger 2015, Auer 2014). A set of Austriacisms related to food even enjoy official protection under the Austrian EU Accession Treaty (e.g., Erdäpfel 'potatoes', Fisolen 'green beans'). Despite their sociolinguistic and ideological importance, Austriacisms have to-date not been the subject of any significant large-scale studies. Our contribution addresses this research gap, presenting results from large-scale surveys carried out across Austria. Importantly, our analyses of the collected survey data implement an integrated approach in which investigations of linguistic behavior and cognitive associations are juxtaposed.

Our empirical analyses are based on two data sources: First, the "conversation corpus" features the language data of 150 speakers (from 13 rural locations of Austria), elicited in two settings (Lenz 2018): a (more formal) researcher-led interview (by an unfamiliar academic explorer) and a (more informal) conversation setting "among friends". The aim of these two settings was to elicit different registers of the individual repertoires of linguistic variation, which enables both inter-individual and intra-individual comparisons. These data will be the basis for our analyses, which will focus on the use of lexical Austriacisms. Second, we present analyses based on nation-wide written surveys (approx. 750 participants). The questionnaire concentrated on aspects of enregisterment (cf. Agha 2007) and thus on the link between linguistic elements and social identities.

Our quantitative and qualitative analyses will provide answers to the following research questions: What role do Austriacisms play in the language behaviour of Austrian speakers? Which lexical phenomena show what kind of variation on the areal-horizontal and vertical-social dimension of variation? Which social values are attributed to which variants and which enregisterment processes (cf. Auer 2014) are these attributions based on?

The results of our analyses provide evidence for the following hypotheses: We anticipate salient inter-regional differences with regard to lexis across Austria (as we have found to be the case with regard to phonetic, morphological and syntactic variation). We also expect that inter-regional differences on the basis of lexical variation (mainly between the Bavarian and Alemannic areas of Austria) occur across the entire dialect/standard axis, including "intermediate" registers. With regard to those (mainly) lexical features which are highlighted as "Austrian peculiarities" in public and private discourses on Austrian language and which bear highly social values (revealed by our questionnaire), we expect only low usage frequencies in our "conversation corpus" (interviews and conversations among friends).

References

Agha, Asif. 2007. Language and Social Relations. Cambridge: Cambridge University Press.

Auer, Peter. 2014. Enregistering pluricentric German. In Soares da Silva, Augusto (ed.). Pluricentricity. Language Variation and Sociocognitive Dimensions (Applications of Cognitive Linguistics 24). Berlin & Boston: de Gruyter Mouton, 19–48.

Lenz, Alexandra N. 2018. The Special Research Programme ‘German in Austria. Variation – Contact – Perception‘. In Ammon, Ulrich & Marcella Costa (eds). Sprachwahl im Tourismus – mit Schwerpunkt Europa. Language Choice in Tourism – Focus on Europe. Choix de langues dans le tourisme – focus sur l’Europe (Yearbook Sociolinguistica 32). Berlin & Boston: de Gruyter, 269–277.

Wiesinger, Peter. 2015. Das österreichische Deutsch in der globalisierten Umwelt: Wandlungen durch bundesdeutsche Einflüsse. In Lenz, Alexandra N., Timo Ahlers & Manfred M. Glauninger (eds.). Dimensionen des Deutschen in Österreich. Variation und Varietäten im sozialen Kontext (Schriften zur deutschen Sprache in Österreich 42). Frankfurt am Main: Lang, 91–122.

Wodak, Ruth, Rudolf De Cillia, Martin Reisigl & Karin Liebhart. 1999. The discursive construction of national identities. Edinburgh: Edinburgh University Press.