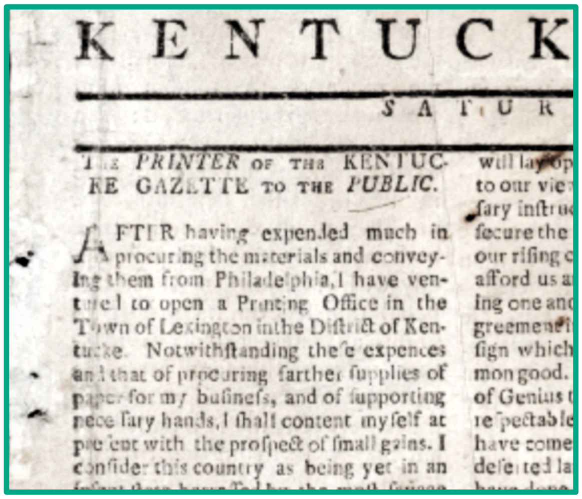


KENTUCK
SATUR

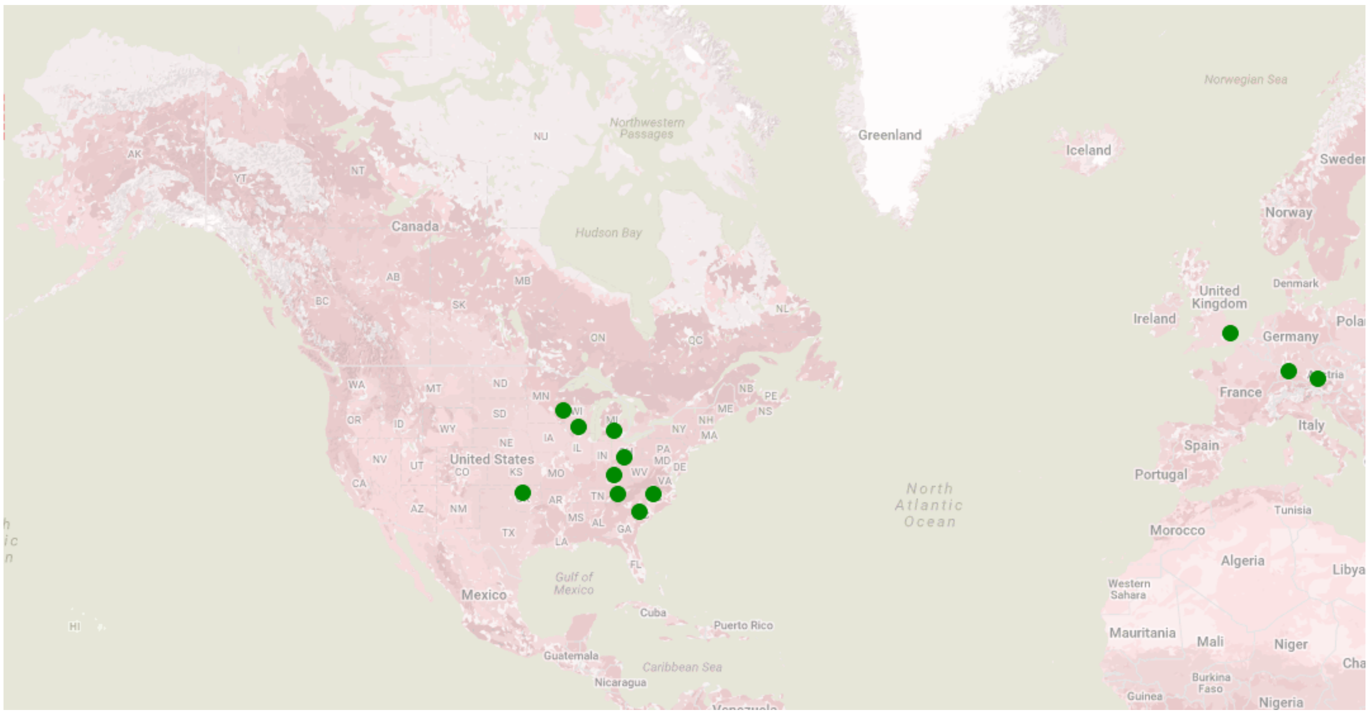
THE PRINTER OF THE KENTUCKE GAZETTE TO THE PUBLIC.

AFTER having expended much in procuring the materials and conveying them from Philadelphia, I have ventured to open a Printing Office in the Town of Lexington in the Diftrict of Kentucke. Notwithftanding thefe expences and that of procuring farther fupplies of paper for my bufinefs, and of fupporting neceffary hands, I fhall content myfelf at prefent with the profpect of fmall gains. I confider this country as being yet in an

will lay op
to our vie
fary inftru
fecure the
our rifing c
afford us a
ing one and
greement
fign which
mon good.
of Genius
refpectable
have come
defeated la

NARNiHS

North American Research Network in Historical Sociolinguistics




```
<s syn="NP"><w pos="PP" lemma="I">I</w></s>
<s syn="VP"><w pos="VBP" lemma="have">have</w>
    <s syn=VP"><w pos="VBN" lemma="venture">ventured</w>
        <s syn="S">
            <s syn="VP"><w pos="TO" lemma="to">to</w>
                <s syn="VP"><w pos="VB" lemma="open">open</w>
                    <s syn="NP"><w pos="DT" lemma="a">a</w><w pos="VBG" lemma="print">Printing</w><w pos="NN" lemma="office">Office</w></s>
                    <s syn="PP"><w pos="IN" lemma="in">in</w>
                        <s syn="NP">
                            <s syn="NP"><w pos="DT" lemma="the">the</w><w pos="NN" lemma="town">Town</w></s>
                            <s syn=PP"><w pos="IN" lemma="of">of</w>
                                <s syn="NP"><w pos="NP" lemma="Lexington">Lexington</w></s></s>
                                <s syn="PP"><w pos="IN" lemma="in">in</w>
                                    <s syn="NP"><w pos="DT" lemma="the">the</w><w pos="NN" lemma="district">Diftrict</w></s>
                                    <s syn="PP"><w pos="IN" lemma="of">of</w>
                                        <s syn="NP"><w pos="NP" lemma="Kentucky">Kentucke</w></s></s></s></s></s></s></s></s></s></s></s>
```

NARNiHS Inaugural Conference
21-23 July 2017
University of Kentucky | Lexington, KY, USA

Table of Contents

Facilities Guide

Welcome to the inaugural conference of the North American Research Network in Historical Sociolinguistics! All presentations at NARNiHS's inaugural conference will take place in Jacobs Science Building, Room 321. The hands-on workshop on Saturday afternoon and the poster session on Sunday afternoon will be in Room 221 of the same building. The street address of the building is:

680 Rose St.
Lexington, KY 40508

To get to Room 321 from the Rose Street entrance of the Jacobs Science Building, turn right towards the middle of the building. You'll find an atrium area with a large screen and staircases. Take the stairs to the third floor. If you wish to take the elevator, you'll find it in the hallway to the left of the main entrance. On the third floor, if you're facing the large screen in the atrium, the lecture hall, Room 321 will be directly to the left of the stairs. Room 221 is directly below Room 321.

Restrooms & Water Fountains

The nearest restrooms and water fountains are just around the corner from the lecture hall. When you exit the lecture hall, turn right and restrooms will be on the hallway to the right.

Book Displays

We are privileged to offer a display of books and journals of interest to historical sociolinguists and other materials for perusal. Materials and order forms are available for inspection from Peter Lang, the University of Chicago Press, John Benjamins Publishers, Routledge, and LINCOM. Please take a few moments to visit the book table at the front of the lecture hall during coffee breaks.

Let us Know You're Here!

There is no cost for this conference nor is registration required, but we'd like to know who's here and keep record of our expanding network. Please take a moment to sign in at the table in front of the entrance to the lecture hall.

Transportation

Around Campus

The University of Kentucky offers a free shuttle between campus buildings and parking lots. You'll find route maps and schedules here: http://www.uky.edu/transportation/buses-and-shuttles_campus-shuttles.

Parking

For information on visitor parking passes and campus parking lots, visit http://www.uky.edu/transportation/parking-info_visitor-parking.

Getting around Lexington

There are several amenities only a short walk from UK's campus. However, should you desire to venture further off campus, Lexington has a public transit system. You can find details about routes or download a full system map here: http://www.lextran.com/routes. (Note: the "detailed PDF schedule" of each document will contain fare information.)

Food and Beverages

Social Gathering at Michler's Kentucky Native Cafe

After the keynote presentation Friday night, we have planned to walk over as a group to Michler's Kentucky Native Cafe for food, drinks, and conversation. This plan is contingent on good weather, as the cafe features outdoor seating in a lovely shaded backyard. It's a little under a mile away from Jacobs Science Building and should be about a 15 minute walk. You are welcome to meet up with the group there if you would like to drive. More information about the cafe – including their menu, street address, and where to park – can be found at the following link: https://www.michlers.com/pages/cafe.

Coffee & Networking Breaks

The conference will include four coffee and networking breaks (two per day on Saturday and Sunday) where coffee, tea, water, and light snacks (including a few gluten-free and vegan options) will be served. The breaks are sponsored by John Benjamins Publishers.

Networking Lunches

A private dining area ideal for networking is reserved for us at the "Fresh Food Company @ The 90" for those who wish to continue conversations over lunch. Food is available for purchase at a reasonable price on location at the "Fresh Food Company @ The 90", but you should also feel free to bring your own lunch if you prefer – all conference attendees are welcome.

Evening Meals

Evening meals have not been organized, but we are leaving the option open for group events and will plan based on interest at the end of each day's sessions.

Other Dining Options

Information on dining out in Lexington can be found here: http://www.visitlex.com/restaurants/

Acknowledgements

Conference Committee

Mark Richard Lauersdorf, *University of Kentucky*
Joseph Salmons, *University of Wisconsin-Madison*
Fernando Tejedo-Herrero, *University of Wisconsin-Madison*
Donald Tuten, *Emory University*

Local Organizers

Mark Richard Lauersdorf, *University of Kentucky*
Abbey Thomas, *University of Kentucky alum; University of Texas at Arlington*
Kelly E. Wright, *University of Kentucky alum; University of Michigan*

Program Design

Abbey Thomas, *University of Kentucky alum; University of Texas at Arlington*

Sponsors

We gratefully acknowledge the financial support of the following sponsors:

- The Office of the Vice President for Research, University of Kentucky
- John Benjamins Publishing Company

Display copies of publications have been provided by:

- The University of Chicago Press
- Peter Lang International Academic Publishers

Informational and material support has been provided by:

- The Lexington Visitors Center
- The College of Arts & Sciences, University of Kentucky
- The Department of Chemistry, University of Kentucky

Conference Schedule

(All events are in Jacobs Science Building 321 unless otherwise noted.)

Friday, 21 July 2017	
7:00-8:30 PM	**Keynote Speaker: Stephan Elspaß** (Universität Salzburg) *The Present and Future of Historical Sociolinguistics*
9:00-11:00 PM	**Social Gathering at Michler's Kentucky Native Café**
Saturday, 22 July 2017	
Presentation Session 1	
9:30-10:00 AM	**Brian D. Joseph** (The Ohio State University) *Reconstructing historical sociolinguistic conditions from loanwords: The case of ERIC loans in the Balkans*
10:00-10:30 AM	**Uri Horesh** (University of Essex), **Enam Al-Wer** (University of Essex) *Dialect contact and change in an Arabic morpheme: Examining Jordanian and Palestinian dialects*
10:30-11:00 AM	**Coffee and Networking Break**
Presentation Session 2	
11:00-11:30 AM	**Samantha Litty** (University of Wisconsin-Madison) *Generational differences in Voice Onset Time and Final Obstruent Neutralization in Wisconsin German and English, 1863-2013*
11:30 AM-12:00 PM	**Fernando Tejedo-Herrero** (University of Wisconsin-Madison) *A failed attempt of standardization in the third-person clitic pronoun system in Spanish*
12:00-2:00 PM*	**Networking Lunch** *Location: Private dining area at "The Fresh Food Company @ The 90"
Presentation Session 3	
2:00-2:30 PM	**Nandi Sims** (The Ohio State University) *Regional Variation of Verbal -s in Earlier African American English*
2:30-3:00 PM	**Michael Montgomery** (University of South Carolina) *Documenting 'Earlier' American English*
3:00-3:30 PM	**Coffee and Networking Break**
3:30-6:00 PM*	**Kelly E. Wright** (University of Kentucky) *A Historical Sociolinguist's Digital Tools "Starter Kit"* *Location: Jacobs Science Building 221

Sunday, 23 July 2017	
Presentation Session 4	
9:30-10:00 AM	**Dennis R. Preston** (Oklahoma State University & Michigan State University Emeritus) *The History of Folk Linguistics and Historical Folk Linguistics*
10:00-10:30 AM	**Joseph Salmons** (University of Wisconsin-Madison) *Historical sociolinguistics and language shift: On verticalization*
10:30-11:00 AM	**Coffee and Networking Break**
Presentation Session 5	
11:00-11:30 AM	**Joshua R. Brown** (University of Wisconsin-Eau Claire) *The Relevance of Social and Cultural Histories in Understanding Heritage Language Shift*
11:30 AM-12:00 PM	**Benjamin E. Frey** (University of North Carolina at Chapel Hill) *Verticalization and Language Shift Among the Eastern Band of Cherokee Indians*
12:00-2:00 PM*	**Networking Lunch** *Location: Private dining area, "The Fresh Food Company @ the 90"
Presentation Session 6	
2:00-2:30 PM	**Jeffrey E. Davis** (University of Tennessee) *Historical Sociolinguistic Studies of North American Indian Sign Language*
2:30-3:00 PM	**Alina Ladygina** (University of Tübingen), **Igor Yanovich** (University of Tübingen) *The rise of the 19th century English progressive: variation between individual verbs*
3:00-4:00 PM	**Coffee and Networking Break**
4:00-6:00 PM*	**Poster/Software Session** *Location: Jacobs Science Building Room 221 *The LSA 2017 Summer Institute general poster session, held at the same time/place, will include posters on the topics of Corpus Linguistics, Dialectology, Historical Linguistics, and Sociolinguistics.*
	Phillip Barnett (University of Kentucky), **Taha Husain** (University of Kentucky) *Athenians talk like this, but Thessalians talk like this: What the Attic plays tell us about the sociolinguistics of Classical Greek discourse*
	Christopher Handy (Independent scholar), **Michael Litchard** (Independent scholar) *The Haks Language Modeling System: Examples from Buddhist Texts in Sanskrit, Tibetan and Chinese*

Contact Information for Participants

Enam Al-Wer – enama@essex.ac.uk

Phillip Barnett – phillip.barnett@uky.edu

Joshua R. Brown – brownjo@uwec.edu | http://www.joshuarbrown.com/

Jeffrey E. Davis – jdavis49@utk.edu

Stephan Elspaß – Stephan.Elspass@sbg.ac.at

Benjamin E. Frey – benfrey@email.unc.edu

Christopher Handy – handyca@mcmaster.ca

Uri Horesh – uri.horesh@essex.ac.uk

Taha Husain – taha.husain@uky.edu

Brian D. Joseph – joseph.1@osu.edu

Alina Ladygina – aladygina@yahoo.com

Mark Richard Lauersdorf – lauersdorf@uky.edu

Samantha Litty – slitty@wisc.edu

Michael Montgomery – ullans2000@gmail.com

Dennis R. Preston – preston@msu.edu

Joseph Salmons – jsalmons@wisc.edu

Nandi Sims – sims.408@osu.edu

Fernando Tejedo-Herrero – lftejedo@wisc.edu

Abbey Thomas – thomasabbey8@gmail.com

Donald Tuten – dtuten@emory.edu

Kelly E. Wright – kelly.wright@uky.edu

Igor Yanovich – igor.yanovich@uni-tuebingen.de

Abstracts

(Shown in the order of presentation.)

Keynote Speaker: Stephan Elspaß (Universität Salzburg)
The Present and Future of Historical Sociolinguistics

Historical sociolinguistics has developed into a vibrant field of research in historical linguistics in the new millennium. It has shifted the focus to the 'social' in the histories of languages, and it has enriched the study of historical linguistics by accounting for 'external' factors in the description and explanation of historical language variation and language change. Crucially, the (re)discovery of new (old) data, the advances in corpus linguistics and greater efforts to conduct comparative studies have sparked off a whole range of exciting research activities and new projects in the last two decades. Their findings help us to reconstruct historical language use and the social factors that had an effect on its variation and change, thus contributing to an understanding of language as a socio-cultural phenomenon. In effect, such findings sometimes cast a different light on traditional narratives of language histories and encourage alternative interpretations.

In the first part of my talk, I will try to map the field by giving a brief account of main topics and projects in present-day historical sociolinguistics. In the second part, the focus is on two core areas of research which not simply complement studies in historical linguistics, but may lead to findings that could not be achieved by other than historical-sociolinguistic approaches. As one such area, I will identify the study of standardisation processes, in particular viewed from a language history 'from below' perspective. Related to this is the study of the effects and limits of prescriptivism, in particular with respect to present-day variation. The examples and case studies are mainly from German; examples from other European languages will be referred to for comparative purposes.

Brian D. Joseph (The Ohio State University)
Reconstructing historical sociolinguistic conditions from loanwords: The case of ERIC loans in the Balkans

Loanwords typically, but not always, imply a social connection between peoples, in the form of face-to-face interaction and contact that make the loans possible. They thus offer a basis for looking back to earlier language and social states and of reconstructing the conditions under which the loans could have occurred. But usual typologies for loanwords, e.g. Bloomfield's cultural versus intimate loans or Hockett's need versus prestige borrowing, often miss the full picture of the earlier social conditions. Drawing on collaborative work with Victor Friedman (see, e.g., Friedman & Joseph 2014 (*IJSL*)), I propose here a new class of loanwords, ERIC loans — an acronym for loans that are "Essentially Rooted In Conversation" — that focus on the conversational interactions that speakers had and what such loans mean for us as historical sociolinguists. I illustrate the utility of these loans by bringing to the fore data from borrowing in the Balkans among the languages of the Balkan sprachbund.

References
Friedman, Victor A. and Brian D. Joseph. 2014. Lessons from Judezmo about the Balkan Sprachbund and contact linguistics. *International Journal of the Sociology of Language*, 226: 3-23.

Uri Horesh (University of Essex), Enam Al-Wer (University of Essex)
Dialect contact and change in an Arabic morpheme: Examining Jordanian and Palestinian dialects

Our data are derived from sociolinguistic interviews held in Amman, Sult and Ajlun in Jordan and Gaza, Jaffa, Nazareth, Umm al-Fahm, Kufur Yasif, Jerusalem and Ramallah in Palestine.

Generally speaking, in urban Palestinian dialects, the *-e* type suffix is the default, with lowering to *-a* occurring in certain phonological environments. Gaza is unique in its divergence from the general urban Palestinian pattern, in that, traditionally, there is only one variant, the open *-a*, akin to the situation in Cairo. However, as a result of dialect contact since the late 1940s, some speakers have a limited percentage of *-e* type variants as well.

Unlike the Palestinian case, in traditional Jordanian varieties *-a* type variants are the default, and *-e* type variants are phonologically marked. In this respect, these dialects are in line with those of the Horan region (spanning from the central Jordan to and including the south of Syria). However, in the dialect of Amman, the phonology of this suffix has undergone a structural change, essentially abandoning this pattern.

Phonetic differences are also attested. In northern Palestine (e.g., Nazareth), the *-e* type variants can be as high as *-ɪ*, similar to its realization in the dialects of neighboring Lebanon. In Amman, the capital of Jordan, where a new dialect has been forming over the last three generations, our data show that among the first generation of Palestinian migrants into the city, the high variant *-ɪ* is still intact. In contrast, the second and third generations of Palestinians living in Amman typically have the lower *-e/ɛ* variants. In fact, younger speakers often mimic and mock older speakers by using higher realizations of these vowels in an unnatural manner.

The case of Palestinians in Jordan (and in Amman in particular) illustrates how divergence in each of the traditional dialects—Palestinian and Jordanian—results in the two dialects meeting halfway, so to speak, resulting in a 'fudged form,' to borrow a term from Peter Trudgill's work. This leads us to a discussion of 'the bigger picture' in Arabic dialects. Eastern dialects of Arabic, notably in Iraq, are known as 'raising' dialects, i.e., the feminine ending in these varieties is categorically a higher, *-e* type vowel. In Saudi Arabia, some varieties which had traditionally been 'raising' are gradually reversing to a mixed system. The case of al-ʾAḥsāʾ in Eastern Arabia, currently studied by Moayyad Al Bohnayyah, is currently at a stage that very much resembles that of the Jordanian varieties.

As the factors that are involved in these changes are both phonological—and rather complex ones at that—and social (including ethnicity, age, migration status), we believe that only a combination of a historical analysis—via an apparent time lens—and a sociolinguistic approach, can provide a full account thereof.

Samantha Litty (University of Wisconsin-Madison)
Generational differences in Voice Onset Time and Final Obstruent Neutralization in Wisconsin German and English, 1863-2013

In this paper I compare voice onset time (VOT) and final obstruent neutralization (FON) data from audio recordings made 1946-2013 for speakers of German and English in Wisconsin. I focus on how the contrasts are exhibited in written sources from German immigrant writers and their descendants from 1863-1940. In written sources, a loss of distinction in how an author perceives an initial consonant might be shown in their spelling by writing *put* when "but" is intended. By using written sources from German immigrants to the Upper Midwest in the 19th century, combined with audio recordings made

from the 1940s to 2013, we can observe usage in individuals and at the community level over a long span of time to show the continual development of each feature.
Written sources were collected from five communities with a total of 98 authors. The documents, generally letters, were written by German immigrants and their descendants. Many of these writers were literate in German, but for those who have moved to the U.S. as adults, they were not subjected to the American education system and most had not received formal training in writing English. Recordings were made from both male and female speakers in German and English in the late 1940s and 2013.

I compare VOT and FON data by generation, first analyzing which authors exhibited vernacular data forms, then I focus on individual obstruent variation. Results show that while first and second generation German authors exhibit a variety of syllable initial vernacular forms, these level out by the third & fourth generation, where only /t^h/ vs. /t/ (transcription conventions show the distinction is maintained between [spread glottis] vs. unmarked [ ], where /t^h/ is the consonant in /t^hu:/ 'two' and /t/ is the consonant in /tu:/ 'do'). I find a variety of vernacular forms in data collected from both German and English speakers for syllable final forms, although specific vernacular forms are distinct in each language. In the first and second generations, German syllable final vernacular forms vary, but by the third generation only /s^h/ vs. /s/ (as in /s^hu:/ 'sue' vs. /su:/ 'zoo') vernacular forms are written. English syllable initial vernacular forms show /t^h/ vs. /t/ forms recede in written data by the third generation and /k^h/ vs. /k/ (as in /k^h.p^h/ 'cap' and /k.p^h/ 'gap') vernacular forms are present in the written data of second-third generation, but remain in spoken data from second-fourth generation.

Written sources can be used with audio recordings, bringing historical sociolinguistics together with sociophonetics and the analysis of generation with the linguistic data gives a more nuanced description of the development and progression of these features in the Upper Midwest.

Fernando Tejedo-Herrero (University of Wisconsin-Madison)
A failed attempt of standardization in the third-person clitic pronoun system in Spanish

Third-person clitic pronoun usage presents a great deal of dialectal variation in Spanish. Dialects in southern Spain and many regional lects in America exhibit a case-determined system (*le*(*s*) for reference in the dative case and *lo*(*s*), *la*(*s*) for reference in the accusative). In northern and central Spain, clitic selection is based on a semantic reference (*le*(*s*)*, la*(*s*) for masculine and feminine count-nouns and *lo(s)* for mass-noun referents). While the origin and development of these dialectal differences have been studied and explained as outcomes of a wide range of inter-dialectal mixing (Fernández-Ordóñez 1994, 2001; Tuten 2003), the selection of a pronoun system, its codification, and acceptance as the supraregional standard norm have received considerably less attention by scholars (Gómez Seibane 2013, Klein-Andreu 1992, 2000, Sáez Rivera 2008).

In this presentation I examine the interplay of variant dialect selection and ideological valuation in two specific time periods in the history of the Spanish clitic pronoun system. First, I review the social factors that contributed to the promotion of the referential clitic pronoun system, which favored *leísmo* and *laísmo* patterns of use, as the prestigious supraregional norm from the second half of 16th century and throughout the 17th century in peninsular Spanish. Second, I discuss the changes in the indexicality and language ideology (Eckert 2008, Michael 2015) that led the Academy at the end of the 18th century to proscribe the use of the referential clitic pronoun system in favor of either a case-determined (or etymological) system, or a hybrid system (*leísta* (+Human), but not *laísta*, nor *loísta*). Finally, this presentation discusses, in the context of the valuation of variants in the same dialect (i.e., north-central Castilian) in two different time periods, the configuration of supraregional standard norms and, in that process, the creation of dialectal norms.

Third-person clitic pronoun usage presents a great deal of dialectal variation in Spanish. Dialects in southern Spain and many regional lects in America exhibit a case-determined system (*le*(*s*) for reference in the dative case and *lo*(*s*), *la*(*s*) for reference in the accusative). In northern and central Spain, clitic selection is based on a semantic reference (*le*(*s*)*, la*(*s*) for masculine and feminine count-nouns and *lo(s)* for mass-noun referents). While the origin and development of these dialectal differences have been studied and explained as outcomes of a wide range of inter-dialectal mixing (Fernández-Ordóñez 1994, 2001; Tuten 2003), the selection of a pronoun system, its codification, and acceptance as the supraregional standard norm have received considerably less attention by scholars (Gómez Seibane 2013, Klein-Andreu 1992, 2000, Sáez Rivera 2008).

In this presentation I examine the interplay of variant dialect selection and ideological valuation in two specific time periods in the history of the Spanish clitic pronoun system. First, I review the social factors that contributed to the promotion of the referential clitic pronoun system, which favored *leísmo* and *laísmo* patterns of use, as the prestigious supraregional norm from the second half of 16th century and throughout the 17th century in peninsular Spanish. Second, I discuss the changes in the indexicality and language ideology (Eckert 2008, Michael 2015) that led the Academy at the end of the 18th century to proscribe the use of the referential clitic pronoun system in favor of either a case-determined (or etymological) system, or a hybrid system (*leísta* (+Human), but not *laísta*, nor *loísta*). Finally, this presentation discusses, in the context of the valuation of variants in the same dialect (i.e., north-central Castilian) in two different time periods, the configuration of supraregional standard norms and, in that process, the creation of dialectal norms.

Nandi Sims (The Ohio State University)
Regional Variation of Verbal -s in Earlier African American English

This paper explores the patterns of distribution of verbal –s in the present tense paradigm of the early African American Vernacular English (AAVE) represented in the written Federal Writer's Project ex-slave narratives. The widespread use of verbal –s with various subjects in early AAVE has led to a great deal of discussion and disagreement concerning the sources of the patterns of s-marking. One widely accepted view is that the verbal –s paradigm of earlier AAVE varieties reflects the Northern Subjects Rule (NSR) typically associated with the English varieties spoken in northern Britain (Poplack and Tagliamonte 2004; Walker and van Herk 2002).

Schneider showed that the frequency of s-marking in these same ex-slave narratives varied by state (1983, 1989). British sociohistorical migration data demonstrates that the area from the Appalachian to Ozark mountain ranges (henceforth known as the *mountain* region) experienced the highest percentage of migration from northern Britain and, thus the language varieties spoken there should have had the most influence of the NSR. Other regions (*gulf*, *upper-south*, and *lower-south*) experienced migration from East Anglia, Southern England, and other British colonies, such as Barbados, which were characterized by different s-marking patterns (Winford 2015).

This study tests variation in the early AAVE spoken in the four abovementioned regions using data from 134 of the FWP ex-slave narratives (1983, 1989). My dependent variable was the presence or lack of –s in tokens of regular verbs, *do*, *have*, and *go* in simple present and present perfect, present-reference constructions. In addition to the region, I coded each token for the type of subject (pronominal or non-pronominal), third singular (third singular or non-third singular), and adjacency (adjacency or non-adjacency between subject and verb) as per the NSR (Pietsch 2005). Other variables included preceding segment, following segment, and verbal aspect, shown by others to be relevant to s-marking. There was no significant difference in s-marking for type of subject, adjacency, or the interaction between the two either overall or within any of the regions. My results showed a significant difference only in the marking of verbal-s in non-third singular subjects between the regions. The two coastal regions had almost 100% marking of –s with both third-singular and non-third singular subjects. The

mountain and gulf regions showed almost 100% s-marking with third-singular subjects and only about 50% s-marking with non-third singular subjects.

None of the patterns of s-marking in this data set align with the distinctive NSR predictions. The high amount of s-marking in the coasts may reflect a different s-marking paradigm: that of Southern England, characterized by variable –s marking on all persons and numbers. Though further research needs to be done to compare these varieties of AAVE to antebellum varieties of Anglo, southern English to test this neo-Anglicist hypothesis.

References

Pietsch, Lukas. 2005. “Some do and some doesn’t”: Verbal concord variation in the north of the British Isles. In Bernd Kortmann, Tanja Hermann, Lukas Pietsch and Susanne Wagner (eds) *Agreement, Gender, Relative Clauses*: 125-210. De Gruyter.

Poplack, Shana and Sali Tagliamonte. 2004. Back to the present: Verbal -s in the (African American) English Diaspora. In R. Hickey (ed.) *The Legacy of Non-Standard Colonial English: The study of transported dialects*: 203-223. Cambridge: Cambridge University Press.

Schneider, Edgar W. 1983. The origin of verbal -s in Black English. *American Speech* 58(2):99- 113.

Schneider, Edgar W. 1989. *American Earlier Black English: Morphological and syntactic variables.*

Walker, James A. and Gerard Van Herk. 2002. "We labors under a great deal of disadvantiges”: Verbal -s in Early African American English. *Actes de l’ACL 2002/ 2002 CLA Proceedings*.

Winford, Donald. 2015. The Origins of African American Vernacular English: Beginnings. In Jennifer Bloomquist, Lisa J. Green, and Sonja L. Lanehart (eds). *The Oxford Handbook of African American Language.*

Michael Montgomery (University of South Carolina)
Documenting ‘Earlier’ American English

“The great art of the historical linguist is to make the best of this bad data – ‘bad’ in that it may be fragmentary, corrupted, or many times removed from the actual productions of native speakers” Labov 1972:100). What is called for is what Labov terms “methodological self-criticism [that] leads to a continual refinement of our methods” (1972:99).

1) The more documents we can amass and organize, the better.
2) The more meta-data on the authors of documents, the better.
3) The earlier the documents, the better.

Summing up, historical sociolinguistic research rests upon critical mass, on social and personal information on authors and authorship, and on historical depth. Optimally documents are numerous, informative, and old, and collectively these assumptions point us toward what I will call DOIs, i.e. = document of interest. However, they hardly help us find them, so researchers need ways to minimize our work. Two principles are crucial in pursuing DOIs.

Since DOIs should be close to speech, what does “speech-like” mean? Culpeper and Kytö. 2010, for example, take it as axiomatic that we should study dialogues, both those that are narrational (such as trial proceedings) and non-narrational (such as from plays). I argue that we should privilege not texts having the most features of interaction, but rather ones showing the least command of writing

conventions (capitalization, punctuation, and spelling), ones exhibiting the lack of literacy, that unambiguously have the absence of normative features.

Since DOIs are from writers of non-elite echelons of society, identifying the social station of writers may be more important than most other personal information, such as their names or locale. From the researcher's point of view, why a non-elite individual would write at all is the more important issue, because having an idea about this enables us to target DOIs in the first place. Therefore, we should target DOIs on the basis of the motivation to write. I propose that non-elite writers fall into one of three categories:

- lonely-hearts seeking to overcome separation from others (e.g. emigrant letters, Civil War letters)
- desperadoes seeking relief from privation, poverty, or oppression (e.g. African American letters in the Civil War, from Sierra Leone)
- functionaries seeking to fulfill an obligation of employment (e.g. plantation overseers, Indian traders)

Ascertaining nonliterate quality and social motivation are pragmatic issues, primarily because the time researchers have seeking DOIs in manuscript archives is normally quite finite. Library catalogues and finding aids tell us nothing about the literate quality of documents, whereas a researcher's quick visual inspection usually suffices to discover the presence or absence of punctuation and thus whether making a copy of a document is justified.

Library catalogues and finding aids infrequently provide information about writers to determine their non-elite status, if they identify such writers at all. To deal with time pressures, limitations of library records, the fact that DOIs are often buried within much larger collections, and other constraints, it is crucial to consult archival staff and social historians, who often retain a visual memory of documents for many years. In many cases my own own research has depended heavily on them.

Case Study of Verb-Phrase *have* Ellipsis.
Evidence from Irish emigrants (Montgomery 1997)
1) we Shoul^{d} not ∅ Lost one Town of them (Galphin, 15 October 1775)
2) my Journey there wou'd ∅ been but of little service (Croghan, 12 May 1765)

Is this tendency an isolated one geographically or structurally?

Evidence from *Corpus of American Civil War Letters* [10000+ letters & diaries, 5 000 000+ words] (Montgomery 2014, Montgomery, Ellis and Cooper 2014)
3) I *had* liked to *have* Shot a fisherman while Standing at the mouth of mitchael Sound. [John B. Lance, 10 November 1861, N.C.)
4) I waunt you to write your letters a litel plainer for I *had not a like to a* maid out half of your words. (T. Warrick, 30 April 1862, Ala.)
5) David goins is well I *had like to* forgot him. (W. H. Chapman, 10 May 1864, Tenn.)
6) I will tell you we had a time of it in the cold we *like to a* froze. (A. J. Spease, 6 Dec. 1863, N.C.)
7) I have bin quite sick ever sence last week but I *like to* ∅ died thursday night I was out of my mind nearly al night long and I had seven or eight fits. [J. W. Muire, August 1862, Va.)
8) i *liked to* for got to tell yoe Sarah C has gown to new barn (J. Hall, 13 April 1862, N.C.)

Reduction of *have* to *a* in CACWL:
9) She wood *a* ben Struck to *a* done Som good (G. Robertson [NC], 3 May 1863)

Ellipsis of *have* in CACWL after a modal = 34 (South = 10, Midwest = 8, Penn. = 3, Northeast = 11)

10) I would Ø answered your kind letter before now if I had of been able I have not been well since I was at Sis Loulies. (S. Buntin, 3 February 1861, Miss.)
11) you said that you I *would ∅ rote* befor if you thought i would answer it. (J. Linfor, 6 November, 1861, Ill.)
12) it appears you didnot get it as you *would ∅ said* something about it. (B. Hagenbuch, 19 February 1865, Penn.)
13) If the thing had ben atempted we proberly should have taken it but we *should ∅ lost* half our men and the place gained *would ∅ ben* of no imprtance to us. (M. Larry, 8 December 1863, Maine)

References

Culpeper, Jonathan, and Merja Kytö. 2010. *Early Modern English Dialogues*

Spoken Interaction as Writing

Labov, William. 1972. Some principles of linguistic methodology. *Language in Society* 1:97-120.

Montgomery, Michael. 1997. A Tale of Two Georges: The Language of Irish Indiaa Traders in Colonial North America. *Focus on Ireland*, ed. Jeffrey Kallen, 227-54. Amsterdam and Philadelphia: John Benjamins.

Montgomery, Michael. 2014. "Hain't We Got a Right to Say 'Ain't'?: The History of Negation Variants in Appalachian English." *Southern Journal of Linguistics* 38.

Montgomery, Michael, Michael E. Ellis, and Brandon Cooper. 2014. When did Southern American English Really Begin?: Testing Bailey's Hypothesis. *The Evolution of Englishes*, 331-48. Amsterdam: Benjamins.

Schneider, Edgar. 2002. Investigating variation and change in written documents. In *The Handbook of Language Variation and Change*, ed. J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 67-96. Oxford: Blackwell.

Wyld, Henry Cecil. 1936. *A history of modern colloquial English*. Oxford: Blackwell.

Kelly E. Wright (University of Kentucky)

A Historical Sociolinguist's Digital Tools "Starter Kit"

This workshop will present a baseline of Open Source digital tools for historical sociolinguistic investigation. I believe this will be possible, and most beneficial, if we are all working with a prepared dataset. The Parsed Corpus of Early English Correspondence (PCEEC) provides a rich, accessible data set for the kind of hands-on introductory workshop proposed here. I have prepared manageably sized and uniformly formatted sections of this dataset, which can be access here: http://www.uky.edu/~mrlaue2/narnihs2017/workshop.html.

We will explore various angles of inquiry by using the following tools on the PCEEC, providing a more faceted appreciation for the data. The workshop will begin with a Regular Expressions (RegEx) primer/refresher to enhance use of basic search strings in a text/code editor (e.g. BBEdit or Notepad++) or in the corpus analysis software, AntConc. This will be followed with a short introduction to the spelling normalization software VARD2. Following this, the group will experiment with social network visualizations using Gephi. If time allows, we will dig deeper into the Associated Information File, which contains various demographic information on the authors and readers of the PCEEC letters, and discuss approaches to documentation, preservation, and standardization of metadata.

These activities will illustrate that with some foreknowledge of a dataset, one can maximize search coverage and minimize noise in the results by using the special set of characters RegEx provides to help define search patterns. Changes/Variation in spelling and the estimations of naïve writers can thwart the accuracy of digital searches across a body of historical texts. VARD2 allows for variant spellings in the data to be gathered into a search using the current (or contemporary) standard. Using PCEEC node and edge lists, we will use Gephi to view several social networks, revealing the capacity for quantification and qualification of sociolinguistic interaction that such analysis adds to historical data.

Along with the PCEEC selection, I have made available to NARNiHS access to a full set of guided instructions, complete with screenshots. This will be useful for members who cannot attend the workshop or stay for its full duration, but are interested in the tools.

Dennis R. Preston (Oklahoma State University & Michigan State University Emeritus)
The History of Folk Linguistics and Historical Folk Linguistics

First I want to show what I consider to be the natural shared interests of the study of folk linguistics, language ideology (within the anthropological linguistics tradition), and language attitudes (within the social psychology of language tradition). I consider these interests to be so deeply shared that I have elsewhere referred to the conflation of them as "language regard," a conflation not only of shared research interests or goals but also of methodologies and interpretive and analytic tools.

Second, I want to place language regard research firmly within several sociolinguistic traditions: Hymesian ethnographic ones, Gumperzian interactive ones, and Labovian variationist ones.

Third, I want to distinguish and illustrate the difference in historical sociolinguistic approaches in general between 1) historical data that can be analyzed from a modern sociolinguistic perspective and 2) historical treatments of data that reflect sociolinguistic interest.

Finally, I will look at data of both these sorts with a focus on language regard issues.

Joseph Salmons (University of Wisconsin-Madison)
Historical sociolinguistics and language shift: On verticalization

This talk presents new support for a 'verticalization' view of language shift, where shift is driven by specific changes in community structure (Warren 1978). Until the late 19th century, strong 'horizontal' ties existed among economic, media, education, and religious organizations in Western communities. Over time, these have given way to 'vertical' ties which bring in English, as groups become more dependent on extra-community institutions. For example, in government, more powerful state boards of education bring English into schools; in the social sphere, many religious organizations centralize and merge, prompting shift to English; and in the media, small newspapers, including minority-language papers, merge or go under (Salmons 2005a, 2005b, Frey 2013).

I provide historical sociolinguistic comparative context on the learning of and shift to English in 19th / early 20th century immigrant groups in the Midwest. While Germans sometimes remained monolingual for two or more generations (Wilkerson & Salmons 2008, 2012, Frey 2013), other immigrant groups show a range of patterns, with Norwegians apparently learning English quickly but remaining bilingual, while Dutch and Danish immigrants seem to have shifted to English early. Other communities across the United States, like Hasidim or Old Order Amish, continue to transmit their languages today.

These case studies all support a verticalization account, and allow us to identify further correlates, e.g. community size, where larger communities can more easily create and maintain horizontal structures. For Germans, shifting to English involved dismantling an extensive German-language infrastructure; in other cases, less support correlates with faster shift. Successful language maintenance in some communities today reflects the development of robust horizontal structures and resistance to verticalization.

References

Frey, Benjamin E. 2013. *Towards a general theory of language shift: A case study in Wisconsin German and North Carolina Cherokee*. PhD dissertation, University of Wisconsin-Madison.

Salmons, Joseph. 2005a. "Community, Region and Language Shift in German-speaking Wis-consin." In *Regionalism in the Age of Globalism: Volume 1: Concepts of regionalism*, ed. Lothar Hönnighausen et al., 129–138. Madison: CSUMC.

Salmons, Joseph. 2005b. "The Role of Community and Regional Structure in Language Shift." In *Regionalism in the Age of Globalism* Volume 2: *Forms of regionalism*, ed. Lothar Hönnighausen, et al., 133–144. Madison: CSUMC.

Warren, Roland L. 1978. *The community in America*. 3rd edn. Chicago: Rand-McNally.

Wilkerson, Miranda & Joseph Salmons. 2008. 'Good old immigrants of yesteryear' who didn't learn English: Germans in Wisconsin. *American Speech* 83. 259-283.

Wilkerson, Miranda & Joseph Salmons. 2012. Linguistic Marginalities: Becoming American without Learning English. *Journal of Transnational American Studies* 4.2. acgcc_jtas_7115. http://www.escholarship.org/uc/item/5vn092kk .

Joshua R. Brown (University of Wisconsin-Eau Claire)
The Relevance of Social and Cultural Histories in Understanding Heritage Language Shift

A key task for historical sociolinguists is not to simply map the sociohistorical situation onto language use. It is imperative for both social and historical validity to reconstruct a comprehensive picture of the social context, in which the language varieties under investigation were used (Bergs 2005:8-21; Raumolin-Brunberg 1996:18). This point has been especially prominent in European studies within historical sociolinguistics (Horner & Rutten 2016). When looking at the negotiation of identities, especially when those identities are contested as in immigrant and language contact situations, extralinguistic factors that influence and result as part of that negotiation are necessary to capture fully the ecology of the language contact situation.

This presentation presents data from a seven-year critical ethnography in the Kishacoquillas "Big" Valley in Central Pennsylvania's Appalachian region. Big Valley is home to one of the most diverse and certainly the most compactly diverse Anabaptist (Amish and Mennonite) settlements in the world (Kauffman 1991). Originating with just a handful of Old Order Amish families from the same religious affiliation, today's settlement is home to as many as twelve distinct Anabaptist groups ranging from the ultraconservative Nebraska Old Order Amish to the most progressive Mennonite affiliations. Unlike many studies in historical sociolinguistics, which focus on investigating language from older written texts, this study focuses on investigating identity in language and relies largely on oral histories from the Valley's oldest residents, who were raised in Amish-Mennonite churches as children in the early- and mid-twentieth century. As a result, issues of change, beyond language behavior, and problems of memory need to be reconciled in constructing a comprehensive social history of the language shift that

occurred in the middle of the twentieth century among the then Amish-Mennonite congregations. Although written historical data carries with it many problems including authorship, data amount, availability etc., studies, which rely on memory have their own unique problems, even though the "shifters" were directly interviewed about the changes throughout their lifetimes.

Due in part to these problems of data collection, the social history of the language shift situation needs to be assessed to fully understand the processes underway at the time. This presentation aligns its findings with the theory of verticalization, modeled on the work of sociologist Roland Warren (1978) and applied to heritage language shift situations by Salmons (2005a,b), Lucht et al. (2011), Wilkerson & Salmons (2012), and Frey (2013). In this model, vertical structures index relationships that are at differing hierarchical levels, rather than previously forged local, horizontal relationships. In looking at the social and cultural changes underway in Big Valley in the early- and mid-twentieth century, vertical relationships became more pronounced and pervasive.

References

Bergs, Alexander. (2005). *Social Network Analysis and Historical Sociolinguistics*. Berlin: Mouton de Gruyter.

Frey, Benjamin E. (2013). *Toward a General Theory of Language Shift: A Case Study in Wisconsin German and North Carolina Cherokee*. Ph.D. Dissertation. Madison: University of Wisconsin-Madison.

Horner, Kristine, and Gijsbert Rutten. 2016. "Metalanguage: Social and Historical Perspectives on Germanic Languages in Europe." *Metalinguistic Perspectives on Germanic Languages European Case Studies from Past to Present*, ed. Kristine Horner and Gijsbert Rutten, 1-11. New York: Peter Lang.

Kauffman, S. Duane. (1991). *Mifflin County Amish and Mennonite Story, 1791-1991*. Mifflin County Mennonite Historical Society.

Lucht, Felecia, Benjamin Frey, and Joseph Salmons. (2011). A Tale of Three Cities: Urban and Rural Asymmetries in Language Shift. *Journal of Germanic Linguistics* 23: 371-398.

Raumolin-Brunberg, Helena. (1996). Historical Sociolinguistics. *Sociolinguistics and Language History: Studies Based on the Corpus of Early English Correspondence*, ed. Terttu Nevalainen and Helena Raumolin-Brunberg,11-37. Amsterdam: Rodopi.

Salmons, Joseph. (2005a). "The Role of Community and Regional Structure in Language Shift." *Regionalism in the Age of Globalism: Volume 1: Concepts of Regionalism*, ed. Lothar Hönnighausen, Marc Frey, James Peacock & Niklaus Steiner, 129-138. Madison: Center for the Study of Upper Midwestern Cultures.

Salmons, Joseph. (2005b). "Community, Region and Language Shift in German-speaking Wisconsin." *Regionalism in the Age of Globalism: Volume 2: Forms of Regionalism*, ed. Lothar Hönnighausen, Anke Ortlepp, James Peacock, Niklaus Steiner & Carrie Matthews, 133-144. Madison: Center for the Study of Upper Midwestern Cultures.

Warren, Roland. 1978. *The Community in America*. Chicago: Rand-McNally.

Wilkerson, Mirande E. and Joseph Salmons. (2012). Linguistic Marginalities: Becoming American without Learning English. *Journal of Transnational American Studies* 4 (2).

Benjamin E. Frey (University of North Carolina at Chapel Hill)
Verticalization and Language Shift Among the Eastern Band of Cherokee Indians

Many citizens of the Eastern Band of Cherokee Indians suffered under the assimilationist policies of federal boarding schools of the late 19th and early 20th centuries, yet a shift in home language use from Cherokee to English did not begin in earnest until the 1950s. The primary differentiating factor between the 20th century and other historical periods was the increasing degree of social contact that Eastern Band Cherokee citizens had with people from outside their communities, and the extent to which Cherokees became integrated into the larger American society during what Warren (1978) refers to as the Great Change. While previous studies (Salmons 2002, Lucht 2007, Bousquette & Ehresmann 2012 among others) have examined the correlation between social integration and language shift under the model of verticalization (Warren 1978), Eastern Cherokee represents the first case study in an indigenous context.

On the Qualla Boundary, the land base of the Eastern Band of Cherokee Indians, increased social contact with outsiders in multiple domains altered traditional community structures that had supported the widespread use of the Cherokee language. As those structures changed, domains of English use increased, while domains of Cherokee use decreased. Ultimately the number of domains in which people used English instead of Cherokee became so great that parents began raising children to exclusively in English. Secondary historical sources, oral interviews and census data suggest that community-wide shift to English began during the early 20th century – some 80 years after the founding of the local boarding school. This pattern correlates well with a verticalization-based model of language shift.

References

Bousquette, Joshua & Todd Ehresmann. 2010. West Frisian in Wisconsin: A

historical profile of immigrant language use in Randolph Township. *It Beaken*. 72(3). 247-276.

Lucht, Felecia. 2007. *Linguistic variation in a Wisconsin German community*. Madison, WI: University of Wisconsin-Madison dissertation.

Salmons, Joseph. 2002. The shift from German to English, World War I, and the German-language press in Wisconsin. In W.G. Rödel & Helmut Schmahl (eds.). *Menschen zwischen zwei Welten: Auswanderung, Ansiedlung, Akkulturation*. Trier: Wissenschaftlicher Verlag Trier. 179-193.

Warren, Roland L. 1978. *The community in America*. 3rd edn. Chicago: Rand-McNally.

Jeffrey E. Davis (University of Tennessee)
Historical Sociolinguistic Studies of North American Indian Sign Language

This presentation brings together historical linguistic research and contemporary sociolinguistic fieldwork documenting the use of indigenous sign language across major cultural areas of native North America. In contrast to urban societies, where sign language is used primarily by members of the Deaf community, in some indigenous communities, sign language has been used as an alternative to spoken language, and acquired as one of the native languages by both deaf and hearing community members. Up until modern times, a certain number of indigenous communities have maintained traditional ways of signing that are distinct from American Sign Language (ASL) that is used in Deaf communities of the U.S. and Canada. Although considered endangered today, indigenous sign language varieties are still being learned and used by some descendants of nomadic groups of the historic Great Plains cultural area; e.g., Blackfeet/Blackfoot, Crow, Mandan-Hidatsa, Nakota/Gros Ventre, Northern Cheyenne and

native communities bordering this geographic area. While Plains Indian Sign Language has been the best documented variety, distinct indigenous sign language varieties have also been documented for certain Native communities of Northwestern Canada (Inuit-Nunavut), the Southwestern United States (Navajo and Pueblo), and Mesoamerica (Maya of western Guatemala and Yucatán, Chiapas, and Oaxaca regions of Mexico).

Historical linguistic and sociolinguistic findings are presented from the presenter's ethnographic fieldwork and ongoing research of legacy materials from archival sources, encompassing documentary linguistic materials spanning three centuries (1800s to present). It has been well documented that a conventionalized and linguistically enriched sign language was used for a variety of discourse purposes across the major North American Indian cultural areas, encompassing twelve language families and members from forty spoken language groups. Sign language was once so prevalent and widespread across the expanse of Native North America that it was considered an "international auxiliary language" (i.e., a signed lingua franca shared between distinct linguistic groups). By all accounts, the signed lingua franca served numerous sociolinguistic purposes and discourse functions for many generations to an extent unparalleled by any currently or previously known case of an indigenous sign language. Among the communities where the signed lingua franca once flourished, it was considered a prestigious or high status language used by chiefs, elders, interpreters, and medicine men/women within and between North American Indian nations. Historically, sign language was used at varying levels of discourse, encompassing various sociolinguistic contexts and many discourse genres.

This presentation features the first fieldwork in over fifty years to focus on the linguistic status of American Indian Sign Language and historical sociolinguistic outcomes. It will address how indigenous sign language serves as an alternative to spoken language, how it is acquired as a first, second, or third language, how it is used among deaf and hearing tribal members, and how it is used internationally as type of lingua franca. It also takes into account language contact and historical change among endangered American indigenous signed and spoken languages, as well as contact between the indigenous sign language and urban sign language varieties used in American Deaf communities of the U.S. and Canada. In these multilingual contact situations code-switching is common-place, plus outcomes unique to the cross-modality nature of signed-spoken language contact, such as code-blending and the co-occurrence of signing and speaking. Evidence for historical relatedness in terms of lexical borrowing and grammatical similarity is presented, along with discussion of linguistic status, change, and resiliency. The aim is to engender a greater understanding of historical linguistic and sociolinguistic outcomes in signed language-spoken language contact situations.

Alina Ladygina (University of Tübingen), Igor Yanovich (University of Tübingen)
The rise of the 19th century English progressive: variation between individual verbs

Introduction. We examine the development of the English progressive for the 42 most frequent lexical verbs in the speech of the British Parliament throughout the 19th century. Using exploratory data analysis and mixed-effects regression modeling, we find that individual verbs show distinctive trajectories of change that cannot be obviously predicted by general linguistic features known to influence the construction. Previous studies would not detect this because our dataset is 2 orders of magnitude larger, allowing us to consider individual verbs in detail. We conclude that in the rise of the progressive, "each word has its own history".

Background. The progressive as it exists today is a relatively young feature of English. Although already established in Early Modern English, it only became obligatory in "semantically progressive" contexts in the 20th century [Kranich 2010] The 19th century was crucial for the process of the

progressive integration [Smitterberg 2005], with the overall frequency of the progressive greatly increasing throughout the period [Strang 1982], [Denison 1998: Sec. 3.3.3].

Motivation. Our aim is to see how the rise of the progressive plays out on the level of individual lexical verbs: which actual trajectories of change different verbs have. To be sure that we are dealing with true trajectories coming from a single speech community, as opposed to aggregated from heterogeneous usage by individuals who have never been in linguistic contact with each other, we use the debates of the British Parliament as our source. Though MPs come from different places and have different life histories, we can at least be sure that they were in regular linguistic contact with each other. The specificity of speech situations in the parliament helps, ensuring that different speakers in the corpus attempt to use roughly the same register in our texts. While parliamentary debates represent formal speech and thus could in principle show fossilized linguistic norms, we do observe clear changes in the usage throughout the 19th century, Fig. 1. Importantly for us, considerable variation by verb (gray) is clearly seen around the general trend (black).

Methods. The debates from the Hansard Archive from 1803 to 1889 were divided by calendar decade, and POS-annotated by TreeTagger [Schmidt 1994]. We identified 50 most frequent verbs over the whole period and extracted their finite progressives using regular expressions (N=89553 occurrences of the progressive in total, cf. N=53 for debates in [Smitterberg 2005]). The verbs for which a high proportion of non-progressive forms, e.g. predicative adjectival participles (That correction is wanting) was extracted, were discarded, leaving 42 lexical verbs for which the number of false positives was negligible. Instances of BE+going followed by t o were also excluded, as that combination usually is a futurate. Our metric was the number of BE+ing hits for the verb divided by the overall frequency of the verb in the corresponding time period. While this is only a rough proxy to the variationist variable we are interested in, namely the proportion of progressives in contexts where one would be possible, a proper delineation of that sociolinguistic variable would involve analyzing by hand millions of contexts to arrive at a comparably sized dataset.

Analysis. Our 42 verbs can be divided by trajectories they exhibit into those where the progressive's share increases (Fig. 2), decreases (Fig. 3) or stays largely stable (Fig. 4) throughout the period. The verbs within these three classes do not show obvious semantic or aspectual similarities. Conversely, verbs with similar features do not behave alike. E.g., the verbs of movements or speech demonstrate high variability in the relative frequency of progressives (Fig. 5-6). Interestingly, lexical aspect does not play a clear role either: even statives show a heterogeneous pattern. In general, they rarely occur in the progressive, but one of them – the verb suffer – shows a dramatic increase in relative frequency. The informal observations above are supported by regression modelling. We fit the following set of mixed-effects linear models1, all with fixed effect of time: 1) varying intercept for individual verb (AIC=-2053.6, logLik=1030.8); 2) the same plus varying slope for individual verb (AIC=-2132.0, logLik=1071.0); 3) var. intercept and slope for "trajectory class", according to the division into verbs in Fig. 2-4, plus var. intercept for individual verb (AIC=-2104.4, logLik= 1058.2); 4) var. intercept and slope for movement vs. speech (claimed to be leaders of change) vs. other verbs, plus var. intercept for ind. verb (AIC=-2049.6, logLik= 1030.8). Not surprisingly, Model 2, which assigns a "random" slope to each verb, is the best. But importantly, allowing slopes to vary by "trajectory class", Model 3, is clearly helpful, while doing the same for semantic classes, Model 4, does not improve even the baseline model that didn't have any varying slopes. We do not suggest that "trajectory classes" are a real phenomenon. Rather their relative statistical success illustrates that differences between individual verbs cannot be obviously traced to any general linguistic features.

Conclusion. The history of individual verbs in the rise of the progressive does not appear reducible to the influence of general linguistic factors. Individual word histories are significant. Future analyses should thus strive to distinguish true class effects from spurious ones driven by individual items (e.g. by suffer for statives.)

In our debates data, the membership of each verb in its "trajectory class" appears generally stable throughout the 19th century. But we do not know whether such membership is stable across different

genres and speech communities. It remains for future research to show whether individual verb histories that we see in the parliamentary debates generalize across text types and speech communities, or belong specifically to some of those, but not others.

1) Normally, logistic regression would be more appropriate for such counts data as ours. But in our case it is not, for two reasons: 1) we only observe the overall frequency of the lexical verb, not the £ of contexts where the progressive is possible, hence our counts do not come from a binomial distribution;
2) the trajectories in Fig. 6 obviously do not uniformly show the S-curve shape.

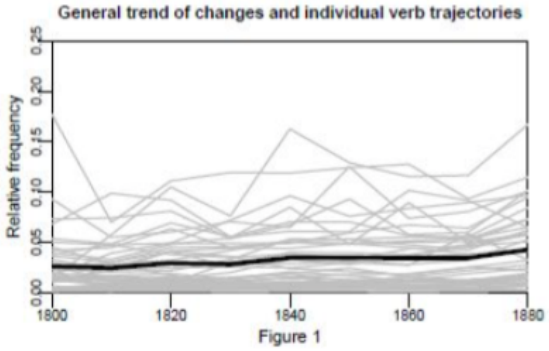


Figure 1

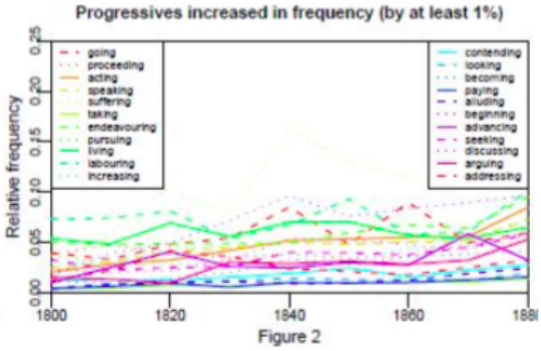


Figure 2

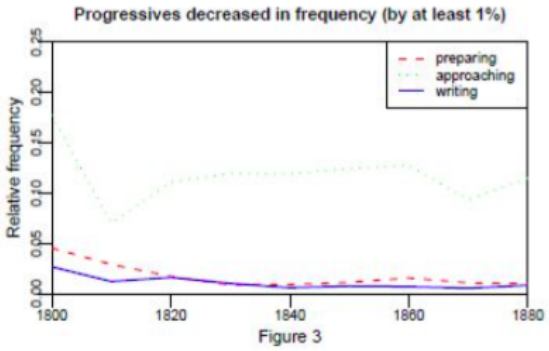


Figure 3

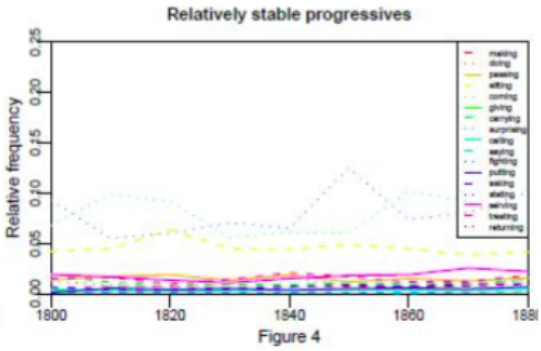


Figure 4

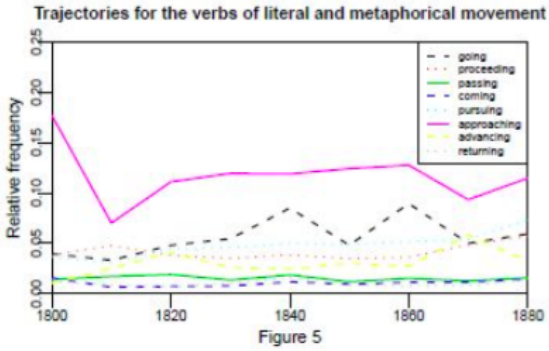


Figure 5

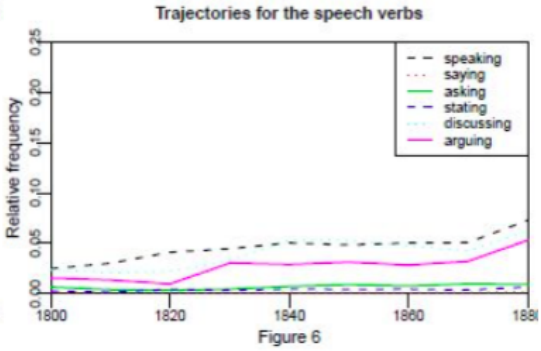


Figure 6

References

Hansard Parliamentary Debates (1st-3rd Series), February 1820 - August 1891 . http://www.hansard-archive.parliament.uk/

[Denison 1998] David Denison (1998). Syntax. In: The Cambridge History of the English Language, Vol IV: 1776-1997 , ed. Suzanne Romaine, CUP.

[Kranich 2010] Svenja Kranich (2010). The progressive in Modern English. A Corpus-Based Study of Grammaticalization and Related Changes. Issue 72 of Language and Computers - Studies in Practical Linguistics. Amsterdam: Rodopi.

[Schmidt 1994] Helmut Schmid (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees . Proceedings of International Conference on New Methods in Language Processing, Manchester, UK.

[Smitterberg 2005] Erik Smitterberg (2005). The progressive in 19th-century English: A process of integration. Issue 54 of Language and computers: studies in practical linguistics. Amsterdam - New York, NY, Rodopi, 2005

[Strang 1982] Barbara M.H. Strang (1982). Some aspects of the history of the be+ -ing construction. In: Language Form and Linguistic Variation: Papers dedicated to Angus McIntosh , ed. John Anderson, Amsterdam: John Benjamins.

Phillip Barnett (University of Kentucky), Taha Husain (University of Kentucky)

Athenians talk like this, but Thessalians talk like this: What the Attic plays tell us about the sociolinguistics of Classical Greek discourse

The Ancient Greek plays have been discussed by literary and linguistic scholars alike for mil- lennia. In the past decades, with the rise of sociolinguistics and historical sociolinguistics, the gaze of some researchers has become fixed on the social information embedded in these classical works. Aristophanes in particular has been studied for his wide array of dialect usage [1]. Much is known about the culture and social stratification of Ancient Greece, including some documentation of language attitudes [2]. We propose to construct a corpus of all of the complete attested Ancient Greek plays, in which we will tag all dialogue for speaker and listener social demographics—gender, age, social status, ethnicity.

As for the nature of our queries, we will focus first on discourse particles. There are several frequently occurring particles in Ancient Greek that serve various functions. One such particle, δέ is multifaceted in its usage and often serves as a simple continuative marker. It can also be used as a conjunction and in the common construction μέν...δέ, which roughly translates as 'on one hand...on the other hand.' The various and frequent nature of this particle in particular will be useful in our research as the authors of the plays may at some level—be it conscious or otherwise—use this discourse particle, and others like it, to indicate some social meaning.

We will query the corpus first for the frequency of such particles, focusing on the most common, and analyze the social information of each particle's most frequent users. At this point, we will analyze the function that the particles serve in the speech of the characters to determine if there is statistically significant variation therein. This will give us insight into the minds of the authors of the Ancient Greek plays. We will learn if these particles were, in the minds of the playwrights, used more or less by any particular social group or dialect.

Christopher Handy (Independent scholar), Michael Litchard (Independent scholar)

The Haks Language Modeling System: Examples from Buddhist Texts in Sanskrit, Tibetan and Chinese

We present a custom software package for language modeling, with specific case examples from Sanskrit, classical Tibetan and classical Chinese texts. Our system, which we call Haks, provides information about texts without requiring prior manual tagging of individual words. Haks treats texts as sets of connected syllables, termed n-grams, rather than as discrete words, and performs frequency analysis of n-grams in order to locate common words and phrases.

Haks is a practical method for extracting recurrent strings from digitized texts in cases where grammar, vocabulary and other information about the texts are partly or entirely unknown. Our method involves

building concordances of words and phrases from digitized input sets of texts, using a simple but effective pattern recognition algorithm. We demonstrate this process using texts from three major languages of the Buddhist tradition: Sanskrit, classical Tibetan and classical Chinese.

The above three languages of the Buddhist literary tradition lack word boundary delimiters in their traditional manuscripts. A person familiar with these languages can identify individual words in such manuscripts by means of context clues, but the texts have no spaces between words. Due to this lack of orthographic spacing, digital concordances for Sanskrit, Tibetan and Chinese often require some amount of manual part-of-speech tagging before performing further linguistic processing automatically. Numerous studies of Sanskrit, Tibetan and Chinese texts using hybrid methods (e.g., Huet 2006, Hackett 2000, Zhan et al. 2006) have yielded useful information, but are unable to deal with unknown texts without human intervention. These tools are therefore limited by the tagging efforts of human researchers. This limit is problematic in the large scale analysis of Buddhist texts, in which we frequently find that we have only partial information about the contents of a set of texts. Consequently, an ongoing issue in automated semantic analysis and translation projects for these languages is the problem of finding word boundaries. Haks overcomes this problem by treating texts as strings of syllables instead of discrete words.

Haks is a modular system for language analysis created in Haskell. It divides each text into individual syllables based on rules for that text's particular language. If the source language is unknown, Haks can also employ a generic division rule. After constructing these initial syllables, Haks then analyzes syllables in sets called n-grams, where 'n' represents the number of syllables in the set. By sorting these n-grams according to the number of times they appear, common words and phrases naturally bubble to the top, with n-grams below a threshold value being discarded. After building a database of frequently observed n-grams, the software can determine further complex relationships between individual n-grams as well as between sets of n-grams, allowing for genre classification and other kinds of semantic analysis.

This system works on any language, human or otherwise, since it does not require any knowledge about the rules or meaning of the source texts in order to find patterns within those texts. The example texts used in our demonstration come from free Internet databases, so that our results can be verified easily. We also provide source code for the project, to make this technique available to others. The modular nature of the system allows for features to be added easily for the analysis of particular aspects of specific languages, and we plan to continue developing this idea further to include other types of string analysis.

Thank you for attending!
We look forward to seeing you at the next event.