Anchoring language evolution in the real world

May 24, 2019

Abstract

Language is a complex dynamic system, changing in response to internal and external influences. It captures and reveals so much of our human nature, and encapsulates our entire evolution. Internal pressure from within a specific community leads to the adaptation of the language to reflect the community's collective personality and cultural preferences. External influences lead to changes in language that would reflect the new status quo brought about by these events, or their long term effects. In this course we will read and discuss papers that investigate some of these phenomena, and help us look at language in a new way.

1 Introduction

Internal influences (although probably misnamed) reflect pressures on the language brought by speakers imposing structure or particular pronunciation conventions to make language more efficient. This has happened and is happening continuously. On the large historical scale, this is how languages we speak today have evolved and separated from their common ancestors. But we see this phenomenon happening currently as well: whether it is the vernacular used within specific communities, or the language used in the various social media (e.g. Twitter, use of emojis, etc.)

External influences on language change are events outside of language that get reflected in language eventually. Such changes could be the birth of new words, changes in word senses, or the disappearance of words from everyday vocabulary.

Changes in the real world can be abrupt, or gradual with a long term effect on society. For example, the industrial revolution has had considerable impact on society, through an increase in technological development, the creation of new types of jobs, and the urbanization of a large segment of the population over a large period of time, i.a. . These changes must have had a strong influence on language, with new words or word senses to reflect the new technologies and the new kinds of jobs. The urbanization and decrease in interaction with nature lead to the decline in usage of terms related to nature. The loss of such words is not an abstract problem, but it entails a loss of culture and of experience. As a contemporary example, the increase in social networking and

virtual interactions lead to increased frequency and impact of relations fostered by social networks, with new types of relations appearing, such as *frenemy*.

Language is also shaped by the communication of information with a particular purpose, like persuasion or manipulation. This is reflected in the different manners of news reporting, particularly for organizations that consider news a form of entertainment. The sentiment profile of news is skewed towards negative stories, reflecting – or maybe, adapting to – the population's negativity bias. An informal observation in the way news have been reported over the past couple of decades seem to indicate that there is an escalation in the degree of alarmism or urgency implied, as people become inured to an existing level of negativity after a period of exposure.

We do not need to look at large amounts of text for evidence of important events and their influence on language – some words encapsulate entire stories. The word *rival* for example, tells us that competition for natural resources – in this case water – has been such an important part of human history that a word expressing this concept is part of our basic vocabulary. The word *calendar* tells us about socially important rituals and customs. In the olden days of the Roman empire, the priests' shouting (calandum, from a Proto-Indo European root *kele* – to shout) announced the phases of the moon that were important for agricultural purposes or celebrations. These became the days (calendae) when payments were due, and were written down in a *calendarium* which was then mostly an accounting book.

External influences are events outside of language, that get reflected in language eventually. Under this group I put events that lead to the birth of new words (I omit imports from other languages because of economic/social/political factors), by making a new form of an existing word, or combine existing words, to reflect the new concept:

old example: rival – peoples/tribes competing for the same water resource

new example: frenemy – a friend who is actually an enemy

and their disappearance (see the Horologicon for some examples).

Both the appearance of a new word or some disappearances can have (causal) roots in the real world. For the disappearance for instance – see the Vocabularist article about nature words disappearing (The vanishing words we need to save `http://www.bbc.com/culture/story/20151126-robert-macfarlane-on-the-wild-words-were-losing`) – this is a reflection of social and behavioural shifts in the real world: words denoting specific nature phenomena or objects are lots because people don't use them anymore, because they don't need them, because most of them live in cities. The appearance of words like "frenemy" reflects the more elaborate social relations, and possibly the influence of social networks in personal connections, some of which may not be genuine.

Word etymologies reveal (more or less explicitly) how a word was formed and its relation to the words it was derived from. Based on this we can trace its story and possibly match it with some external evidence that reveal what events made

it necessary. As mentioned before, this continues to happen, possibly at a more rapid pace because of the wider and faster reach of a potential communication event.

Question to think about (among many others):

What are the consequences (or causes) for the occurrence of a new word or word sense: how does it change a language ontology? where does it fit? why did a new sense appear? what kind of "hole" does it fill?

Does it make sense to compare a historic language (Latin, in particular) with modern English? apart from technological terminology, how has the language changed? What are the domains, how detailed are they? Nature may be more detailed in the olden days.

How word associations change to show how words (that still have the same sense) have shifted. E.g.: bread – associated mostly with other foods, now is associated with carbs and gluten, which show the change in perspective in part of the society.

2 Topics and paper list

A more comprehensive list of papers as well as instructions are posted on the course website (`https://www.cl.uni-heidelberg.de/courses/ss19/LC/material/LC_paperlist.html`).

Hannes Wettig, Suvi Hiltunen, Roman Yangarber (2011)
MDL-based Models for Alignment of Etymological Data
`https://www.aclweb.org/anthology/R11-1016`

Hannes Wettig, Roman Yangarber (2011)
Probabilistic Models for Alignment of Etymological Data
`https://www.aclweb.org/anthology/W11-4634`

Javad Nouri, Jukka Siren, Jukka Corander, Roman Yangarber (2016)
From alignment of etymological data to phylogenetic inference via population genetics
`https://www.aclweb.org/anthology/W16-1905`

Jouna Pyysalo (2017)
Proto-Indo-European Lexicon: The Generative Etymological Dictionary of Indo-European Languages
`https://www.aclweb.org/anthology/W17-0234`

Hannes Wettig, Kirill Reshetnikov, Roman Yangarber (2012)
Using context and phonetic features in models of etymological sound change
`https://www.aclweb.org/anthology/W12-0215`

Alina Maria Ciobanu, Liviu P. Dinu (2018)

Simulating Language Evolution: a Tool for Historical Linguistics
https://www.aclweb.org/anthology/C18-2015

Alina Maria Ciobanu, Liviu P. Dinu (2015)
Automatic Discrimination between Cognates and Borrowings
https://www.aclweb.org/anthology/P15-2071

Ryo Nagata (2014)
Language Family Relationship Preserved in Non-native English
https://www.aclweb.org/anthology/C14-1183

Jack Grieve (2018)
Natural selection in the modern English lexicon
http://evolang.org/torun/proceedings/paperpdfs/Evolang_12_paper_171.pdf

SIMON KIRBY, JENNIFER CULBERTSON, AND MARIEKE SCHOUWSTRA (2018)
The origins of word order universals: evidence from corpus statistics and silent gesture
http://evolang.org/torun/proceedings/paperpdfs/Evolang_12_paper_70.pdf

Michael Pleyer (2018)
Language evolution and the emergence of constructions in interaction
http://evolang.org/torun/proceedings/paperpdfs/Evolang_12_paper_154.pdf

Christine Cuskley (2018)
Alien symbols for alien language: Iterated learning in a unique, novel signal space
http://evolang.org/torun/proceedings/paperpdfs/Evolang_12_paper_42.pdf

(the next two papers are to be presented together, but count as 1)
Andres Karjus, Richard A. Blythe, Simon Kirby and Kenny Smith (2018) Identifying linguistic selection and innovation while controlling for cultural drift http://evolang.org/torun/proceedings/paperpdfs/Evolang_12_paper_117.pdf
Christopher A. Ahern, Mitchell G. Newberry, Robin Clark, Joshua B. Plotkin (2016)
Evolutionary forces in language change
https://arxiv.org/abs/1608.00938

Hamilton, W. L., Leskovec, J., Jurafsky, D. (2016)
Cultural Shift or Linguistic Drift? Comparing Two Computational Measures of Semantic Change

https://www.aclweb.org/anthology/D16-1229

X. Wang and A. McCallum (2006)
Topics over time: A non-Markov continuous-time model of topical trends
https://www.cics.umass.edu/~mccallum/papers/tot-kdd06s.pdf

David Hall, Daniel Jurafsky, and Christopher D. Manning (2008)
Studying the history of ideas using topic models
https://www.aclweb.org/anthology/D08-1038

Pankaj Gupta, Subburam Rajaram, Hinrich Schütze, Bernt Andrassy (2018)
Deep Temporal-Recurrent-Replicated-Softmax for Topical Trends over Time
https://www.aclweb.org/anthology/N18-1098

Ian Stewart and Jacob Eisenstein (2018)
Making "fetch" happen: The influence of social and linguistic context on the success of lexical innovations
https://arxiv.org/abs/1709.00345

Jacob Eisenstein (2013)
Phonological factors in social media writing.
https://www.aclweb.org/anthology/W13-1102

Khan, F., Bowers, J., Frontini, F. (2017)
Situating Word Senses in their Historical Context with Linked Data
https://www.aclweb.org/anthology/W17-6916
Eger, S., Mehler, A. (2016)
On the Linearity of Semantic Change: Investigating Meaning Variation via Dynamic Graph Models
https://www.aclweb.org/anthology/P16-2009

Koper, M., Schulte im Walde, S. (2018)
Analogies in Complex Verb Meaning Shifts: the Effect of Affect in Semantic Similarity Models
https://www.aclweb.org/anthology/N18-2024

Ramiro, C., Srinivasan, M., Malt, B.C. and Xu, Y. (2018)
Algorithms in the historical emergence of word senses
http://www.cs.toronto.edu/~yangxu/ramiro_et_al_2018_sense_emergence.pdf

Kemp, C., Xu, Y., and Regier, T. (2018)
Semantic typology and efficient communication
http://www.cs.toronto.edu/~yangxu/kemp_xu_regier_2017_ec.pdf

Xu, Y., Malt, B.C., and Srinivasan, M. (2017)

Evolution of word meanings through metaphorical mapping: Systematicity over the past millennium
http://www.cs.toronto.edu/~yangxu/xu_malt_srinivasan_2017_metamap.pdf

Cibelli, E., Xu, Y., Austerweil, J.L., Griffiths, T.L. and Regier T. (2016)
The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color
http://www.cs.toronto.edu/~yangxu/cibelli_xu_austerweil_griffiths_regier_2016_whorfcolor.pdf

Xu, Y., Regier, T. and Malt, B.C. (2016)
Historical semantic chaining and efficient communication: The case of container names
http://www.cs.toronto.edu/~yangxu/xu_regier_malt_2016_historical_chaining.pdf

Peter D Turney, Saif M Mohammad (2018)
The natural selection of words: Finding the features of fitness
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0211512

(either this paper or the next, but not both)
Mitra, S., Mitra, R., Maity, S., Riedl, M., Biemann, C., Goyal, P., Mukherjee, A., (2015)
An automatic approach to identify word sense changes in text media across timescale
http://dx.doi.org/10.1017/S135132491500011X

Sunny Mitra, Ritwik Mitra, Martin Riedl, Chris Biemann, Animesh Mukherjee, Pawan Goyal (2014)
Thats sick dude!: Automatic identification of word sense change across different timescales
https://www.aclweb.org/anthology/P14-1096

William L. Hamilton, Jure Leskovec, Dan Jurafsky (2016)
Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change
https://www.aclweb.org/anthology/P16-1141

David John Britain (2009)
One foot in the grave? Dialect death, dialect contact, and dialect birth in England
https://www.researchgate.net/profile/David_Britain/publication/237971543_One_foot_in_the_grave_Dialect_death_dialect_contact_and_dialect_birth_in_England/links/57d1c1bf08ae6399a38b8646.pdf

David John Britain (2008)
When Is a Change Not a Change? A Case Study on the Dialect Origins of New Zealand English
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.165.1860

C. Douglas Chrtien (1962)
The Mathematical Models of Glottochronology
https://www.jstor.org/stable/411186

Isidore Dyen, Joseph B. Kruskal and Paul Black (2010)
An Indoeuropean Classification: A Lexicostatistical Experiment
http://cysouw.de/home/2010_quantitative_historical_linguistics_files/cysouwBIBQUANTHISTLING.pdf

3 Grading

Your grade for the course will depend on three components: participation (20%), presentation (35%) and project (45%).

3.1 Participation

Be active and ask questions in class, send in answers to the questions sent out by the discussion leader for the upcoming class.

3.2 Presentation

Lead the discussion (with or without slides, it is up to you) for the paper you chose, a week in advance send out 3 questions to your colleagues (cc me)

3.3 Project

There will be a common project proposal, but if you have an idea you want to pursue, come talk to me and if it fits the topic, you're good to go!

Project (with more details on the project page):

- choose 3 words (I suggest English, but it can be any language as long as you have some corpora from at least 2 different time frames (a distance more than 50 years).
- Research their etymology and establish the relation between them and their source (or whether they themselves are a basic word – e.g. like mother and father in English).
- Using the different time frame corpora, obtain their collocations or larger contexts, and analyze their senses.

- Establish what senses they had in each time frame, and analyze the relation between their basic sense and the others (could be metaphoric expansions, or generalizations, etc.)
- Determine (using clustering for example, or some other automatic method) the closest words to each sense in each time frame, and their associated contexts. Compare them with contemporary senses.

4 Contact

You can always email me (nastase@cl.uni-heidelberg.de), and if you would like to meet, we will arrange a mutually acceptable time.

5 Additional reading

Serious reading Numerous textbooks on historical linguistics, but this one seems to have a practical side which suits us:

Lyle Campbell, 2013
Historical Linguistics

Fun reads while also being very informative

Here is a selection of fun and very instructive books about various issues related to language change (not to be read for the class, but for your own pleasure):

The horse, the wheel and language – David Anthony

Etymologicon – Mark Forsyth

Horologicon – Mark Forsyth

The secret life of words – Henry Hitchings

Salt, a world history – Mark Kurlansky

`www.etymonline.com`

BBC, The vocabularist (do a search with these words, and enjoy the results)

Writing and Script: A Very Short Introduction – Andrew Robinson