Diachronic Language Models

Lecture 2 - Speed and Types of Linguistic Change

Natural Language Processing Group

Dr. Wei Zhao 24/10/2023

Organization

- How many students haven't blocked time slots
  - Seats are still available
  - Take a third slot, but the audience has the freedom to attend or not
  - Happy with topics?

	Enter your name in Presenters.			
Date	**Topics**	**Disciplines**	**Presenters**	**References (All students should re Please write down both the pointe**
17/10/2023	Language change and language models	CL	Wei Zhao	L1: Revisiting the five foundational problems
24/10/2023	Speed of linguistic change	HL	Wei Zhao	A3: Sociolinguistic typology and the speed o
31/10/2023	Grammaticalization	HL	Melis Çelikkol Lydia Körber	C1: Introduction to Grammaticalization C2: What is it then, this Grammaticalization?
07/11/2023	Syntactic change	HL		
14/11/2023	No seminar (GaML 2023)			
21/11/2023	Large language models	NLP	Ke Ren, Wenzhuo Chen	
28/11/2023	Diachronic language models	NLP		
05/12/2023	The science of prompt engineering for large language models	NLP	Veerav Chebrolu	(Veerav) M1 Pre-train, Prompt, and Predict:
12/12/2023	Guest lecture on semantic change	CL	Dominik Schlechtweg	No need
19/12/2023	Semantic change detection	CL	Katharina Altrichter, Blanca Birn	
Winter Break				
09/01/2024	Syntactic change detection	CL	Maya Arseven Hiu Lam Choy	Maya: G2 Detecting Syntactic Change Using
16/01/2024	Temporal machine translation	NLP	Long Kim, Atila Marte	I1 A Machine Translation Approach for Mode
23/01/2024	Temporal text summarization	NLP	Yu-Chuan Cheng, Xin	(Xinyu)H3 An Evaluation Corpus For Tempo
30/01/2024	Reconstruction of historical texts	NLP	Lisanne Rüh Tim Kolber	J3: Machine Learning for Ancient Languages
06/02/2024	Temporal misinformation detection	CSS	Geng Zhao, Amir Ghadanfar, Dana Simedrea (?)	

Organization

- Write paper review and prepare two questions
  - Please do both
  - Submit to us via email in the format: diaclms-02-weizhao.pdf
  - If you haven't submitted, please do it soon.

	Enter your name in Presenters.			
Date	**Topics**	**Disciplines**	**Presenters**	**References (All students should re Please write down both the pointe**
17/10/2023	Language change and language models	CL	Wei Zhao	L1: Revisiting the five foundational problems
24/10/2023	Speed of linguistic change	HL	Wei Zhao	A3: Sociolinguistic typology and the speed o
31/10/2023	Grammaticalization	HL	Melis Çelikkol Lydia Körber	C1: Introduction to Grammaticalization C2: What is it then, this Grammaticalization?
07/11/2023	Syntactic change	HL		
14/11/2023	No seminar (GaML 2023)			
21/11/2023	Large language models	NLP	Ke Ren, Wenzhuo Chen	
28/11/2023	Diachronic language models	NLP		
05/12/2023	The science of prompt engineering for large language models	NLP	Veerav Chebrolu	(Veerav) M1 Pre-train, Prompt, and Predict:
12/12/2023	Guest lecture on semantic change	CL	Dominik Schlechtweg	No need
19/12/2023	Semantic change detection	CL	Katharina Altrichter, Blanca Birn	
Winter Break				
09/01/2024	Syntactic change detection	CL	Maya Arseven Hiu Lam Choy	Maya: G2 Detecting Syntactic Change Using
16/01/2024	Temporal machine translation	NLP	Long Kim, Atila Marte	I1 A Machine Translation Approach for Mode
23/01/2024	Temporal text summarization	NLP	Yu-Chuan Cheng, Xin	(Xinyu)H3 An Evaluation Corpus For Tempo
30/01/2024	Reconstruction of historical texts	NLP	Lisanne Rüh Tim Kolber	J3: Machine Learning for Ancient Languages
06/02/2024	Temporal misinformation detection	CSS	Geng Zhao, Amir Ghadanfar, Dana Simedrea (?)	

Organization

- 6 LP and 8 LP
  - For students with 8 LP, we expect you are advanced students
  - Your presentation be more substantial than those with 6 LP.
    - 6 LP: present what is outlined in the paper
    - 8 LP: present what is outlined, and your own opinions & thoughts supported by existing resources (you find yourself on the Internet).
      - Optional: resolve your own questions.
      - Optional: present & discuss 2 related papers.
  - Please mention LP on the front page of your presentation.

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Recap

Foundations of diachronic language models

- Language Change: a subject of study in linguistic change over time
- Language Modeling
  - Embedding: a computer algorithm that translates dictionary to coordinates in vector space
  - Generative: a time-series-ish algorithm to generate human-like text

Language Embedding

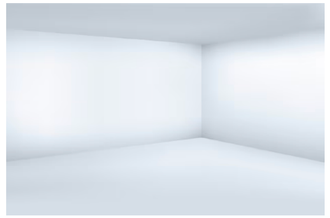
How it works

Where should coordinates be positioned in space?

- Semantic constraints: semantic relations (WordNet)
  - X ∼ Y if they are synonyms on WordNet
- Syntactic constraints: grammatical functions (syntactic tree)
  - X ∼ Y if they share a grammatical role (e.g., word class)
- Morphological constraints: inflection forms
  - X ∼ Y if they share a word ending
- Contextual constraints: You shall know a word by the company it keeps.
  - X ∼ Y if their companies are similar.

Lexicons

→



Coordinates in space

Language Change

Weinreich, Labov & Herzog (1968)

Inform an empirical theory of language change that should address the following problems:

- Constraints problem: discovering the laws of change
- Transition problem: tracking previous changes
- Embedding problem: identifying change causes
- Evaluation problem: evaluating change effects
- Actuation problem: understanding why a change occurs at a particular time

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Motivation

What is speed of change

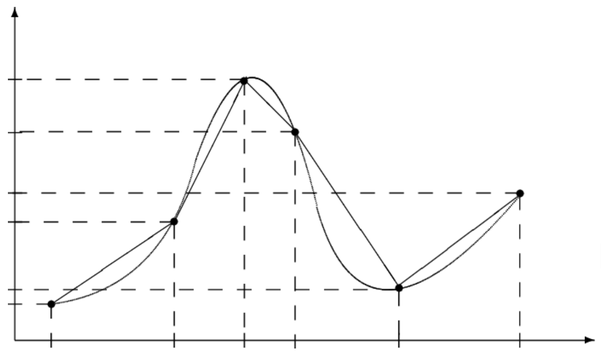
It refers to the rate at which a language undergoes changes over time, particularly to deal with how slowly or rapidly changes in words, sounds and grammar occur over time.

Motivation

What is speed of change

"Language is not some gradually and imperceptibly changing object which smoothly floats through time and space [...]. Rather, the transmission of language is discontinuous, and a language is recreated by each child on the basis of the speech data it hears."

—Paul Kiparsky 1968



Language development is a piecewise-linear function

Motivation

What is speed of change

“Historians have always known that languages sometimes undergo a period of rapid change, then settle into relative stasis. [...] Changes sometimes take place ‘catastrophically’, and [...] grammars change abruptly.”

—David Lightfoot 1999

Motivation

Why are we learning this

- If LMs understand the speed of language change, then
  - they can tell when a change occurs
  - they can distinguish between gradual and rapid linguistic changes
  - they can predict ongoing changes
- Break down these abilities into languages, linguistic features, dialects, mediums, registers, etc.

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Synopsis on

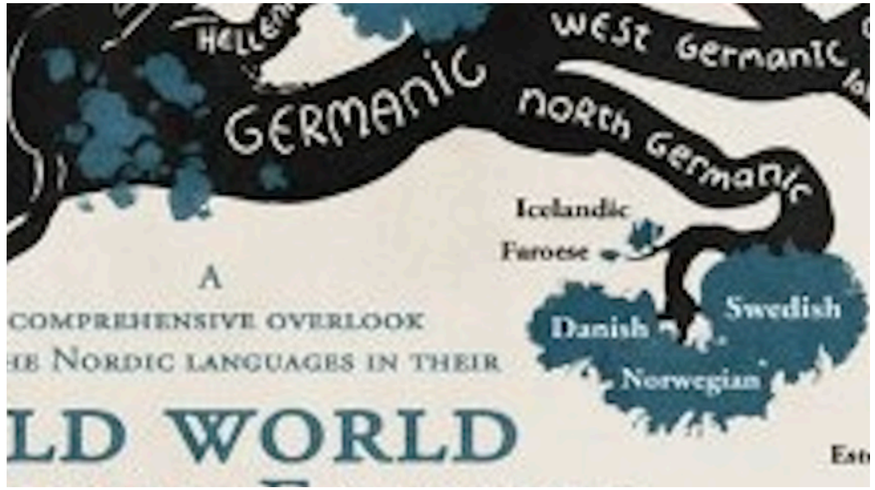
“Sociolinguistic Typology and the Speed of Linguistic Change”

Observations

Rate of language change is not constant

Not constant across languages:

- More changes in Danish than Icelandic, though they descend from Old Norse.
  - Danish influenced by Low German
  - Icelandic (also Faroese) are linguistically stable

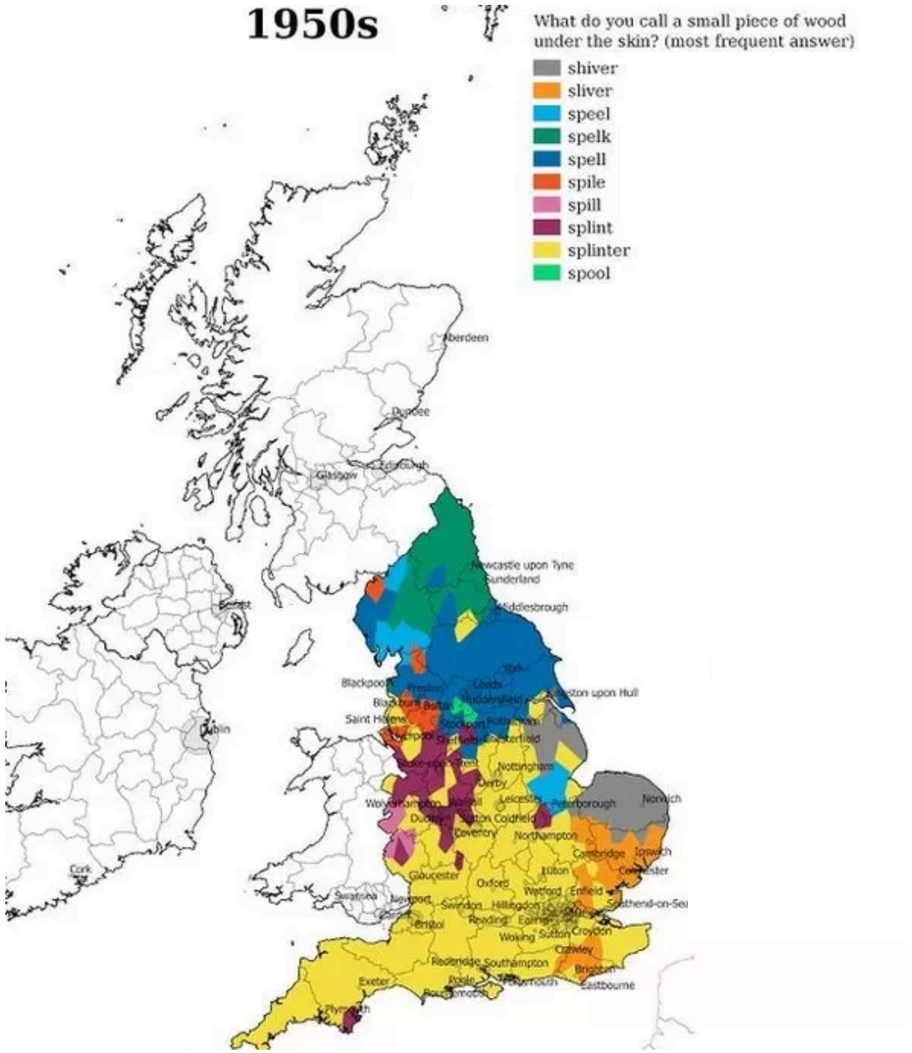


Observations

Rate of language change is not constant

Not constant across dialects:

- More innovative are the southeastern English dialects than the northeastern



https://www.foolproof-proofreading.com/etymology/

Observations

Rate of language change is not constant

Not constant across time periods:

- Conservative and innovative periods in English

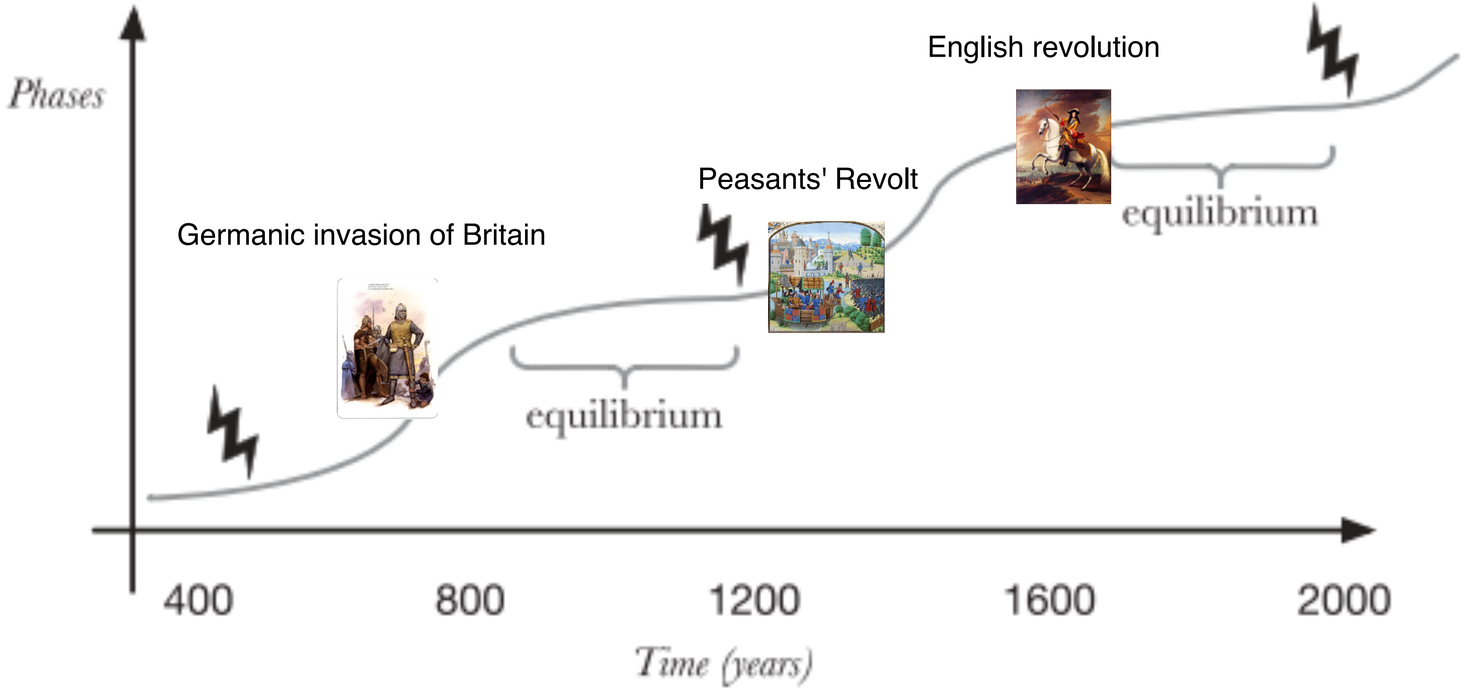


Figure 3: History of English as punctuated equilibria (modified from Bergs 2005: 54).

Modified from https://shorturl.at/gM124

Observations

Rate of language change is not constant

Not constant across linguistic features:

- Syntactic and morphological changes become stable from 16th century.

> As I remember, Adam, it was upon this fashion bequeathed me by will but poor a thousand crowns; and, as thou sayest, charged my brother on his blessing, to breed me well; and there begins my sadness. My brother Jacques he keeps at school, and report speaks goldenly of his profit; for my part, he keeps me rustically at home, or, to speak more properly, stays me here at home unkept; for call you that keeping for a gentleman of my birth, that differs not from the stalling of an ox?

Shakespeare, 1600

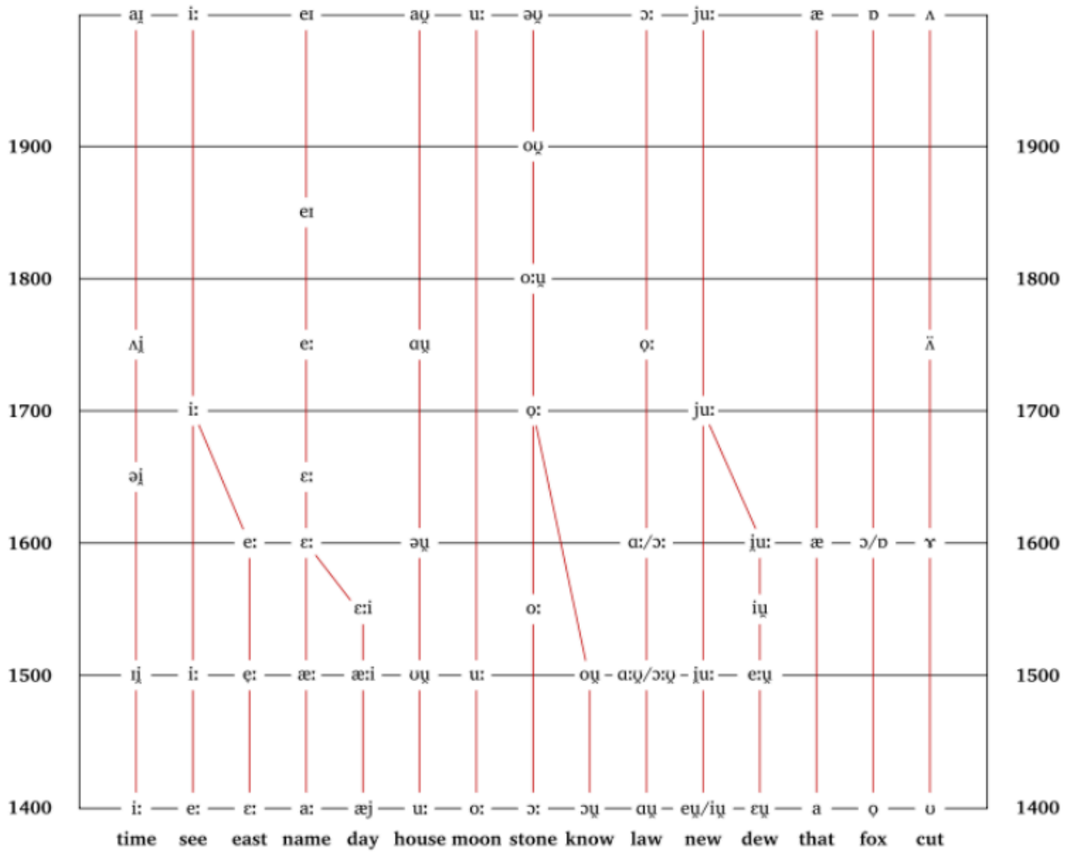
- Some words are currently not in use, but the grammar and inflection system are the same as we use today

Observations

Rate of language change is not constant

Not constant across linguistic features

- Phonological change is rapid
  - Great Vowel Shift 1400-1700
  - Continued sound changes afterwards via assimilation, dissimilation, etc.



Great Vowel Shift (1400-1700)

https://en.wikipedia.org/wiki/Great_Vowel_Shift

Observations

Rate of language change is not constant

Not constant across linguistic features

- The speed of semantic change varies from one word to another

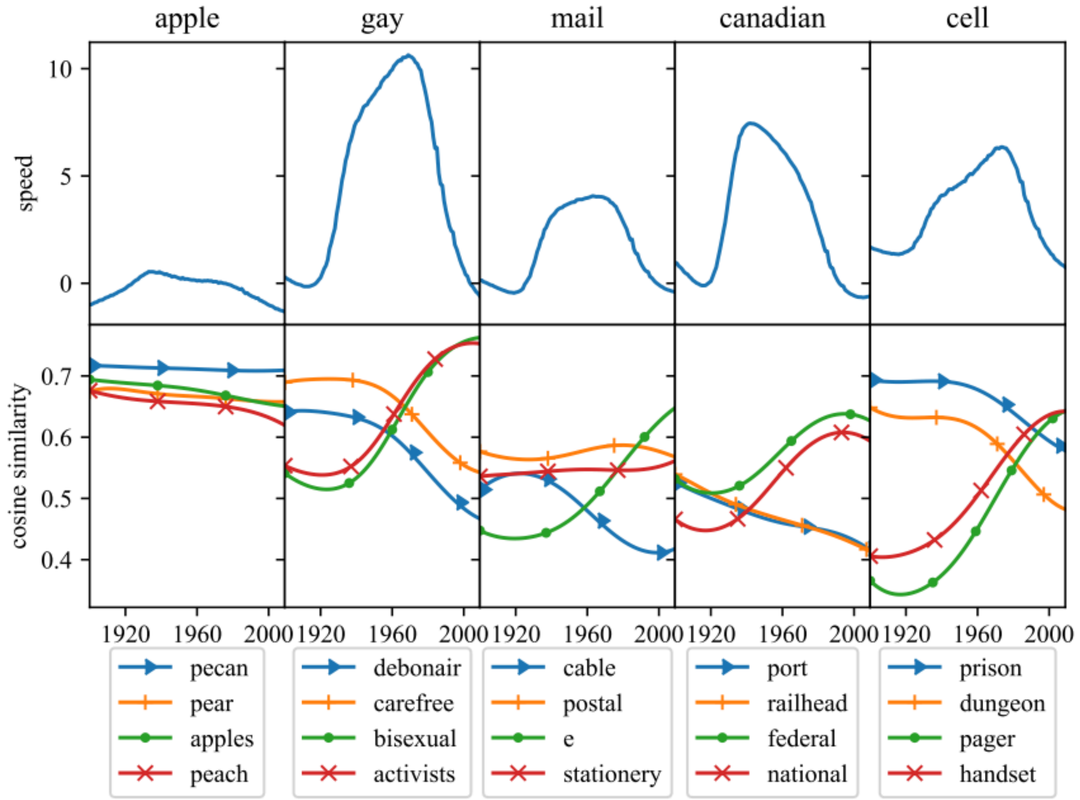


Figure 4: Speed and nearest neighbors over time of selected words. The top graphs as the speed at which a word changes usage according to our model. The bottom graphs are selected nearest neighbors for those words. Each of the chosen nearest neighbors appear as a top 10 nearest neighbor to the word at some year.

http://aclanthology.lst.uni-saarland.de/N18-1044.pdf

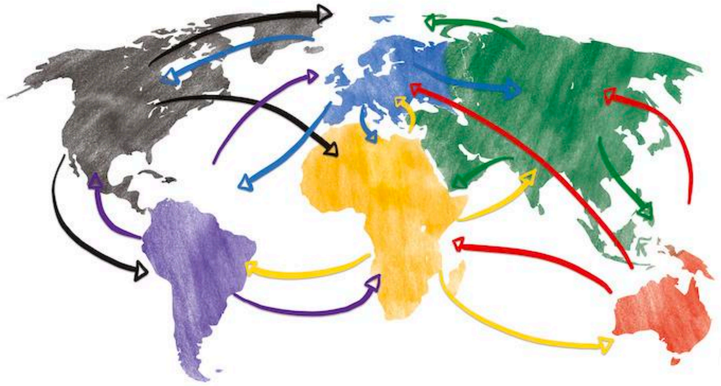
Explanations

Why rate of language change is not constant

Two sociological factors cause non-constant rate of change:

- The degree of contact between speech communities
  - Geographical isolation
- The social stability/instability
  - Germanic invasion of Britain
  - Peasants' Revolt
  - English revolution

Questions



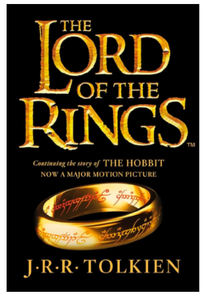
- Any other factors impact the speed of change
  - Internationalization/globalization/technical progress
    - Speed of internationalization may control the speed of change
      - The speed may be controlled by politics: Conservative countries may ban innovative technologies, while some countries are more innovative.
      - Perhaps noticeable changes are in word meaning and vocabulary (via semantic change and neologizing)

Questions

- Any other factors impact the speed of change
  - The social stability/instability (recent catastrophes like the Ukraine War)

As a result of Russia's invasion of Ukraine there have been a few lexical, orthographic, and semantic changes of note taking place in the Ukrainian language. They propagate alongside the flood of information, memes, and propaganda flowing over Ukrainian social media, primarily on telegram and Meta platforms. Some are more widespread than others, some may not last, but it's curious to look at how war can change the perception of one's neighbor in such a short period of time, with the language following along in changes in attitude.

New synonyms for "russia"/"moscovia" (sic) are: "katsapstan", "orcostan", "mordor". The latter synonyms are derived from the widespread Tolkeinian references to the invading army as a horde of orcs due to the poorly coordinated human wave attacks, slaughtering of civilians, and general disorder characteristic of the russian army. The army is also frequently referred to as "орда", the Mongol horde which caused a great deal of destruction in the region in the past.



https://spiegelmock.com/2022/05/01/war-and-ukrainian-linguistic-changes/

Questions

- Factors do not impact language change
  - Events that are not countrywide or international in scope, and that last within a short period of time.

Questions

- What are eroding languages?
  - Endangered languages that are at risk of disappearing
    - Globalization
    - Decreased speakers

https://www.statista.com/chart/26867/endangered-threatened-languages-by-world-region/

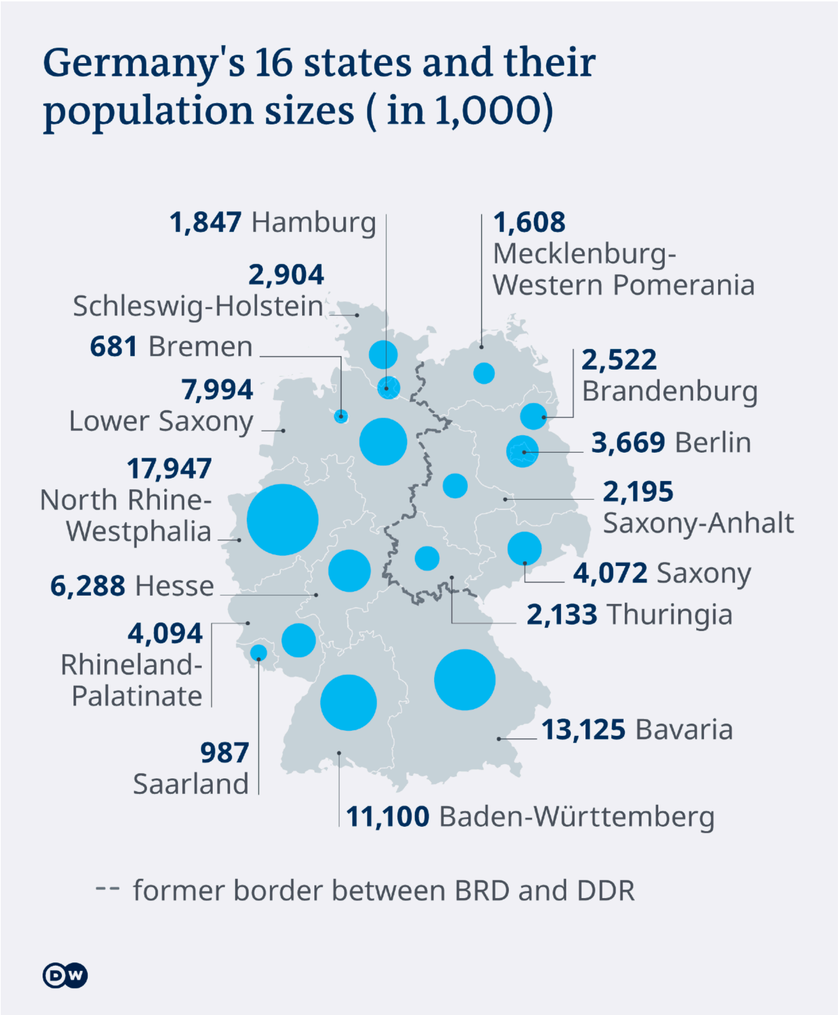
Questions

- Any counterexamples: languages that changed a lot despite being isolated?
  - Yes, but they do not change as rapid as international changes.
  - Isolated languages still evolve over time, influenced by only domestic language contact, linguistically and culturally.
    - Icelanders from other regions move to Reykjavik

https://www.worldatlas.com/maps/iceland

Questions

- Interactions between the degree of contact and social stability?
  - Yes. Big cities were hit badly during the Covid-19 pandemic, people may consider moving to rural regions for isolation from the disease.
  - Any linguistic changes from this?



Set up the scene:

We will look into changes in various linguistic features and the rate of change for each of them.

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Phonological Change

Overview

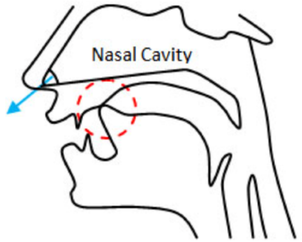
- Change affects various phonology aspects, including tone, stress, and syllable structure
  - Tone: Old Chinese (4 tones), Cantonese (6 tones)
  - Stress: record, present
  - Syllable structure: diversary (Donald Trump): diversity + lottery

If two sounds are similar, then one sound will likely take the features of the neighboring sound when speaking fast.

Phonological Change

Assimilation

- hand [hænd] + bag [bæg] → handbag [hæmbæg]
  - [b]: a bilabial sound
  - [n]: nasal sound
  - Nasal assimilation: [n] takes up the bilabial sound ([n] → [m]); [d] is omitted



- cup [kʌp] + board [bɔːrd] → cupboard [kʌbərd]
  - Regressive assimilation: [p] is subsumed into [b]
  - Schwa insertion: replacing long vowel with a schwa to make speech fluent

Phonological Change

Epenthesis

- This idea is good → This idea-r-is good
  - is: [ɪz]
  - idea: [aɪˈdɪə]
- Epenthesis breaks up a sequence of vowels or consonants to avoid a pause.

Phonological Change

Great Vowel Shift in English

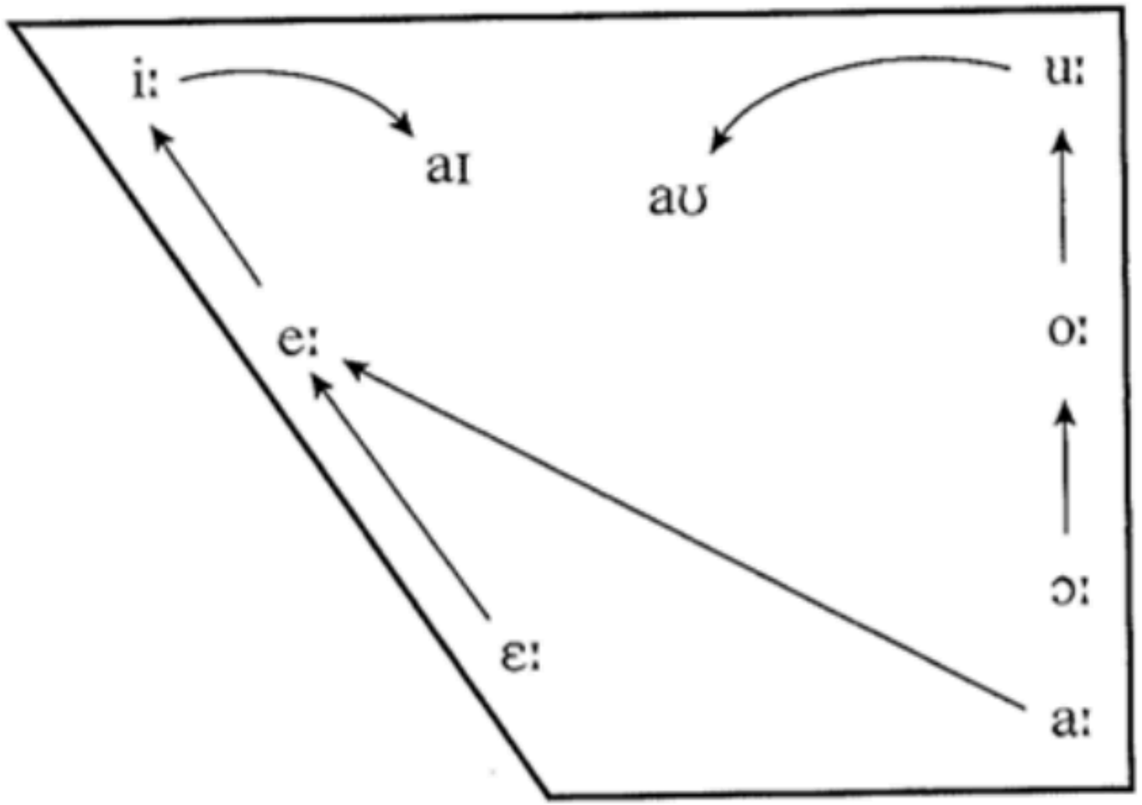
- Seven long vowels undergo changes between 1400-1600

Middle English (c. 1100–1450)		Early Modern English (c. 1450–1600)	Example			
Middle English		**Modern English**	**Middle English**		**Modern English**	
[iː]	→	[aɪ]	[miːs]	→	[maɪs]	mice
[uː]	→	[aʊ]	[muːs]	→	[maʊs]	mouse
[eː]	→	[iː]	[geːs]	→	[giːs]	geese
[oː]	→	[uː]	[goːs]	→	[guːs]	goose
[ɛː]	→	[eː]	[brɛːken]	→	[breːk]	break
[ɔː]	→	[oː]	[brɔːken]	→	[broːk]	broke
[aː]	→	[eː]	[naːmə]	→	[neːm]	name

https://shorturl.at/qxBG9

Phonological Change

Great Vowel Shift in English



https://shorturl.at/qxBG9

Phonological Change

Speed of change on the loss of word-final stop fortition

- Word-final stops with a consonant sound [b], [d], [g] block the airflow, making it difficult to speak fast.
- In late Middle English, word-final stops become less pronounced and turn into a glottal stop.
  - dog [dɒg] → [dɒʔ]

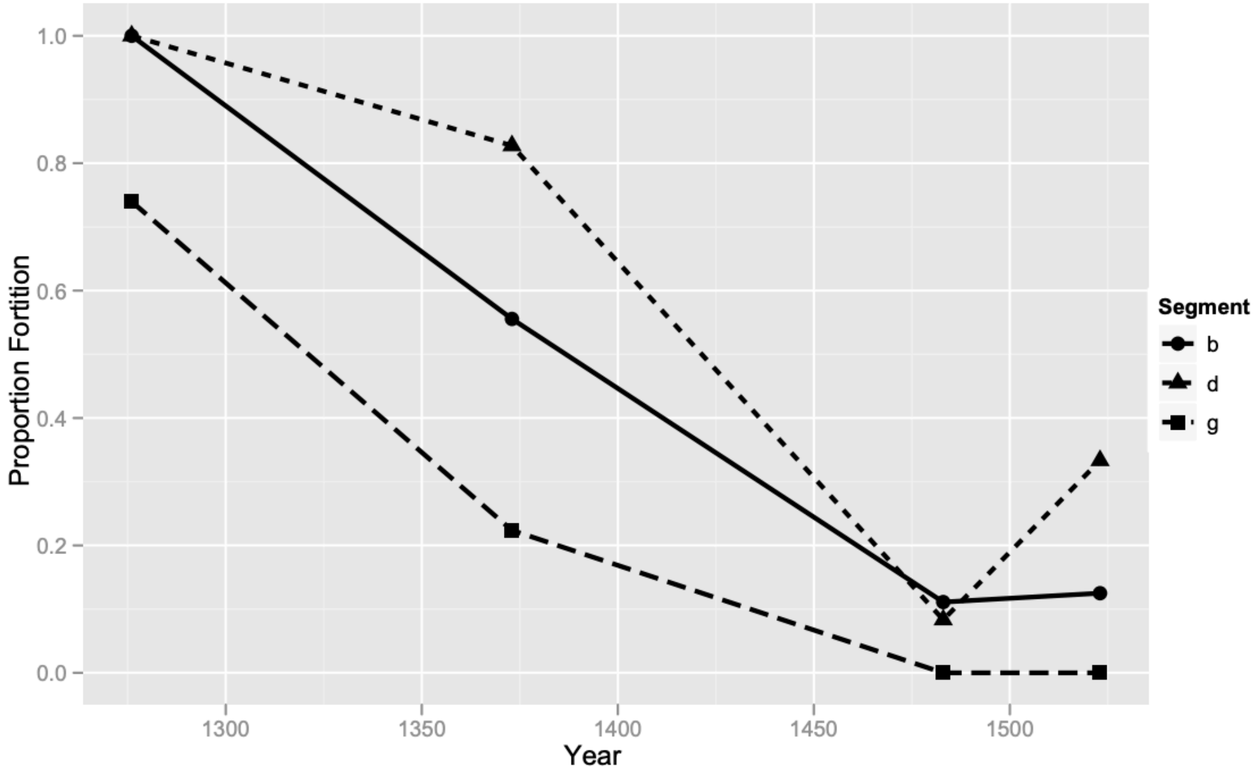


Figure 1: Data from Glaser
Segment:Year Interaction Not Significant

Phonological Rule Change: The Constant Rate Effect

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Morphological Change

Overview

- Morphological change occurs within morphological processes.
  - Word formation: derivation, compounding, etc.
  - Inflection: tense, mood, gender, case, person, number, etc.
    - Derivation: causing meaning change, word class change (verb to noun)
    - Inflection: no meaning change, no word class change

Morphological Change

Inflection on case marking

Table 7.30 Old English case affixes

	Masculine	*Neuter*	*Feminine*
Singular			
	hund 'dog'	dēor 'animal'	gief 'gift'
Nominative	hund	dēor	gief-u
Accusative	hund	dēor	gief-e
Genitive	hund-es	dēor-es	gief-e
Dative	hund-e	dēor-e	gief-e
Plural			
Nominative	hund-as	dēor	gief-a
Accusative	hund-as	dēor	gief-a
Genitive	hund-a	dēor-a	gief-a
Dative	hund-um	dēor-um	gief-um

https://shorturl.at/qxBG9

Morphological Change

Loss of inflection on case marking

Table 7.31 The loss of case affixes through sound change (in English *hound*)

	Old English	*Middle English (e = [ə])*	*Modern English*
Singular			
Nominative	hund	hund	hound
Accusative	hund	hund	hound
Genitive	hund-es	hund-(e)s	hound's
Dative	hund-e	hund-(e)	hound
Plural			
Nominative	hund-as	hund-(e)s	hounds
Accusative	hund-as	hund-(e)s	hounds
Genitive	hund-a	hund-(e)	hounds'
Dative	hund-um	hund-(e)	hounds

https://shorturl.at/qxBG9

Morphological Change

Loss of inflection on case marking

- Loss of case system causes the use of prepositions
  - “to” for the dative
    - Mathis hat mir das Papier geschickt (mir: dative case)
    - Mathis submitted the paper review to me (to me: dative phrase)
  - “of” for the genitive
    - Die Farbe der Blumen gefällt mir (der Blumen: genitive case)
    - I like the color of the flowers (of the flowers: genitive phrase)
  - Leading to restrictions on word order

Morphological Change

Loss of word formation on derivational morphemes

- Old English used to have a suffix -u to derive nouns from adjectives, but this was completely lost.

eald ("old") + *-u* → *ieldu* ("age")

hāliġ ("holy, sacred; pious") + *-u* → *hāligu* ("holiness")

hāl ("sound, healthy, intact") + *-u* → *hǣlu* ("wholeness, health")

hāt ("hot") + *-u* → *hǣtu* ("heat, warmth")

mennisc ("human, natural, humane") + *-u* → *menniscu* ("humanity")

miċel ("big, large; great") + *-u* → *miċelu* ("greatness, size")

- It has been replaced by -ness in Modern English

https://en.wiktionary.org/wiki/-u#Old_English

Morphological Change

Gain of word formation on derivational morphemes via grammaticalization

- Grammaticalization is a process for a word to gain a grammatical function.
  - Lexical morpheme → grammatical morpheme that often changes word class

Old English word		*Modern English Suffix*
hād	'state, condition, rank'	-hood (childhood)
dōm	'condition, power'	-dom (freedom)
(ge-)līc	'similar, equal, like'	-ly (fatherly)

https://shorturl.at/qxBG9

  - lexical *do* → grammatical *do* (auxiliary verb, e.g., to create a question structure)
    - I do my paperwork → do you like football?

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Lexical Semantic Change

Example

Definitionen von Oxford Languages · Weitere Informationen

nice

/nʌɪs/

adjective

adjective: **nice**; comparative adjective: **nicer**; superlative adjective: **nicest**

1. giving pleasure or satisfaction; pleasant or attractive.
   "we had a nice time"

   Ähnlich: enjoyable | pleasant | pleasurable | agreeable | delightful | satisfying

   - (of a person) good-natured; kind.
     "he's a really nice guy"

     Ähnlich: pleasant | likeable | agreeable | personable | charming | delightful

Positive connotations

Lexical Semantic Change

Example

"And up they goon, and doun agayn anon, Til that the millere sporned at a stoon, And doun he fil bakward upon his wyf, That wiste no thyng of this ***nyce*** *stryf;"*

—Chaucer: The Reeve's Tale at 14th century

Scene: An intense conflict between a man and two students. The students were in relationships with the man's wife and his daughter. The man strongly disagreed and engaged in a flight with the students.

nyce stryf → <u>nice</u> <u>strife</u>: <u>enjoyable</u> <u>conflict/diagreement</u>

Oxymoronic

Negative connotations

Nice: foolish

Lexical Semantic Change

Example

c1300–1617	**nice**, adj. & adv. Of a person: foolish, silly, simple; ignorant. Obsolete.
a1387–1665	**nice**, adj. & adv. Of conduct, behaviour, etc.: characterized by or encouraging wantonness or lasci
c1387–	**nice**, adj. & adv. Precise or particular in matters of reputation or conduct; scrupulous, punctilious
a1393–1450	**nice**, n.[1] A foolish or simple person; a fool.
a1393–1657	**nice**, adj. & adv. Of an action, utterance, etc.: displaying foolishness or silliness; absurd, senseless.
a1393–1605	**nice**, adj. & adv. Of a person: wanton, dissolute, lascivious. Obsolete.
a1393–1703	**nice**, adj. & adv. In early use: faint-hearted, timorous, cowardly, unmanly. Later also: effeminate. Obsolete.
1395–1771	**nice**, adj. & adv. Of dress: extravagant, showy, ostentatious. Also in extended use. Obsolete.

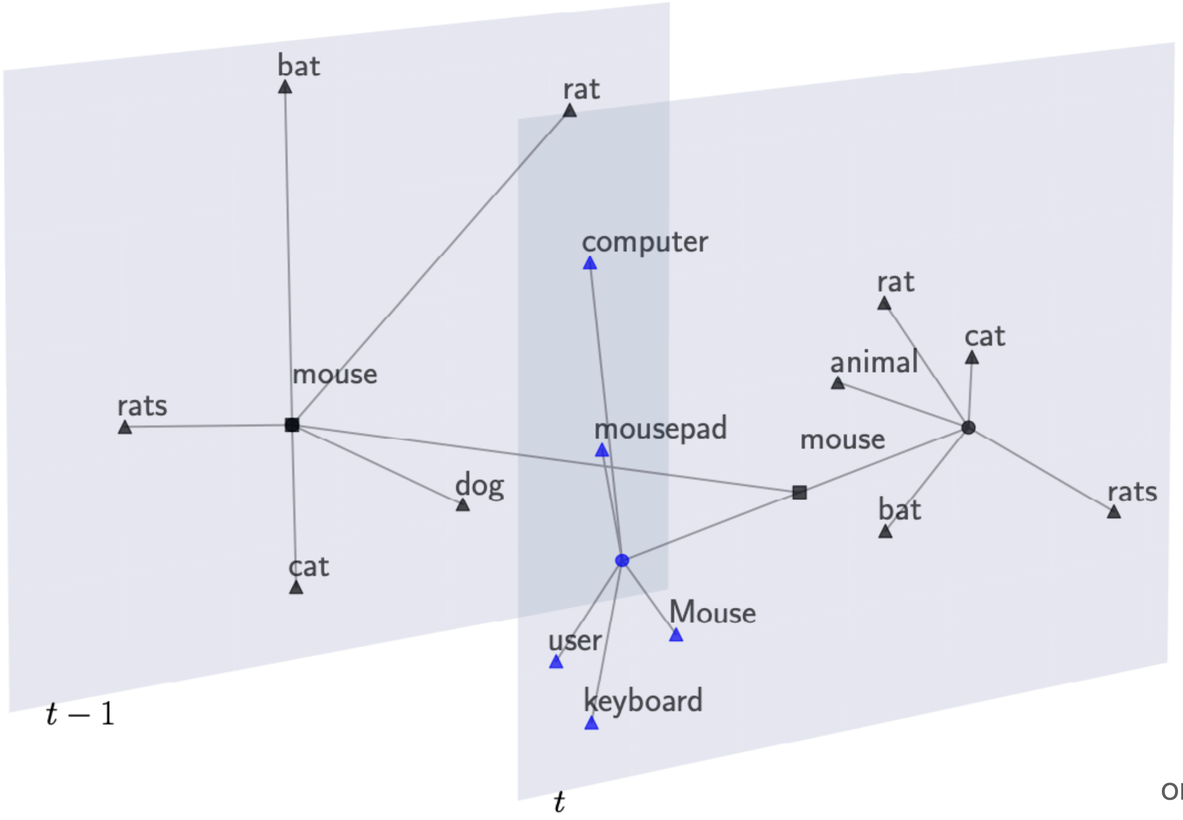
1958–	**to make nice**, in nice, adj. & adv. North American colloquial. to make nice (also nice-nice): to be pleasant or polite, esp. in an expedient or hypocritical way; to smooth things over.
1970–	**nice one**, in nice, adj. & adv. colloquial. nice one: expressing approval or congratulations for something done well. In later use also ironically.
1988–	**nice**, v. transitive. To be (excessively) nice or polite to (someone); to persuade by being nice.
1989–	**nice-but-dim**, adj. Designating a person regarded as foolish, incompetent, and ingenuous, while also being utterly good-natured and unmalicious.

https://www.oed.com/

Lexical Semantic Change

Broadening

- A word gains a new meaning while retaining the original one.

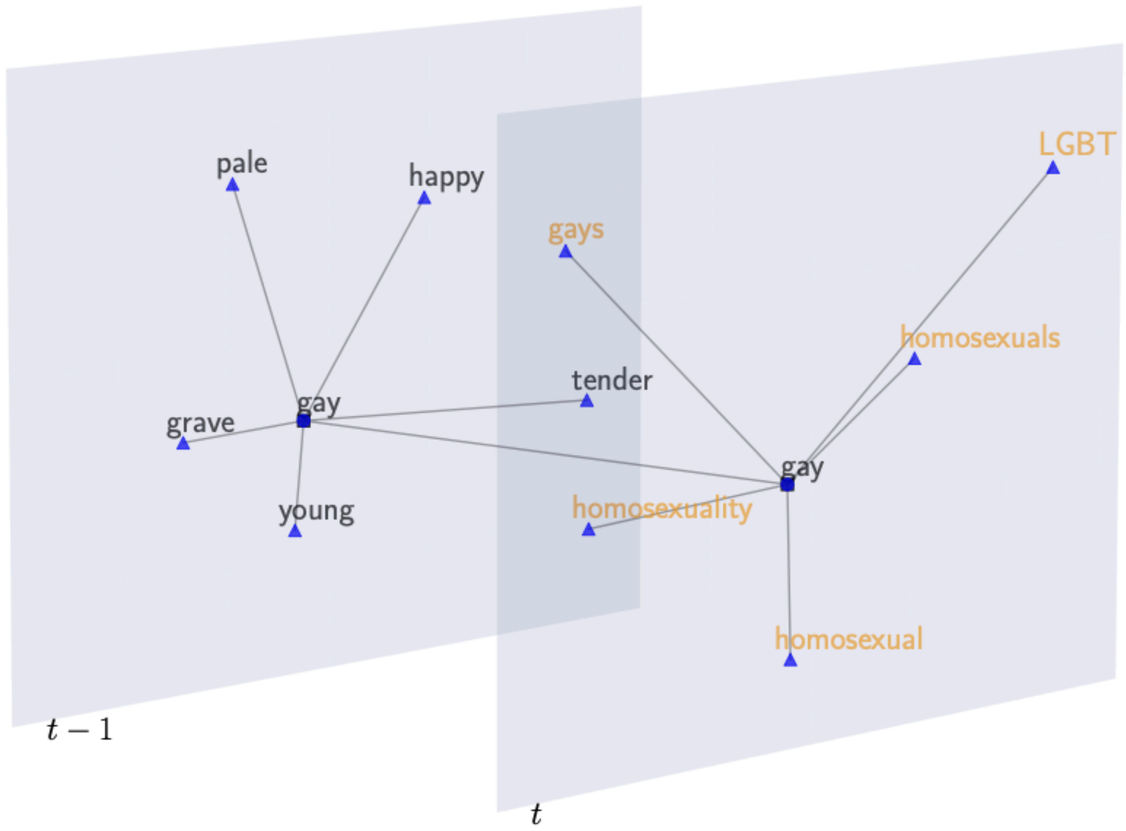


ongoing work

Lexical Semantic Change

Shifts

- A word loses a meaning that it had previously, and gains a new meaning.

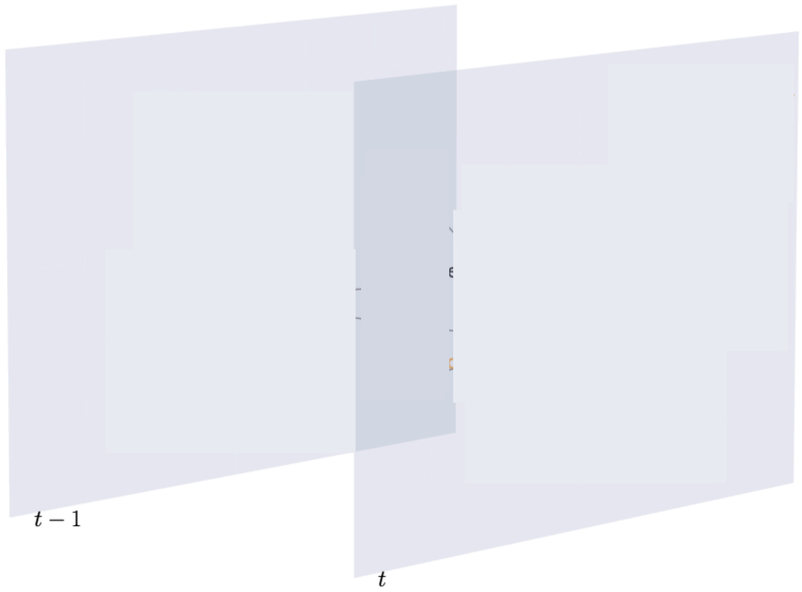


ongoing work

Lexical Semantic Change

Narrowing

- A word's meaning become more specific over time.
  - Meat: any kind of food -> animal flesh
  - Girl: young person of either gender -> female



Lexical Semantic Change

Lexical diffusion on semantic & sound change

- A linguistic change may initially appear in a few words, and then spread throughout the vocabulary.
  - Change in stress pattern of words
  - In the 15th-16th, 3 words (rebel, outlaw, and record) began stressing the first syllable when used as nouns. This shift in stress has been spreading since then.

(graph from O'Grady *et al* 2001)

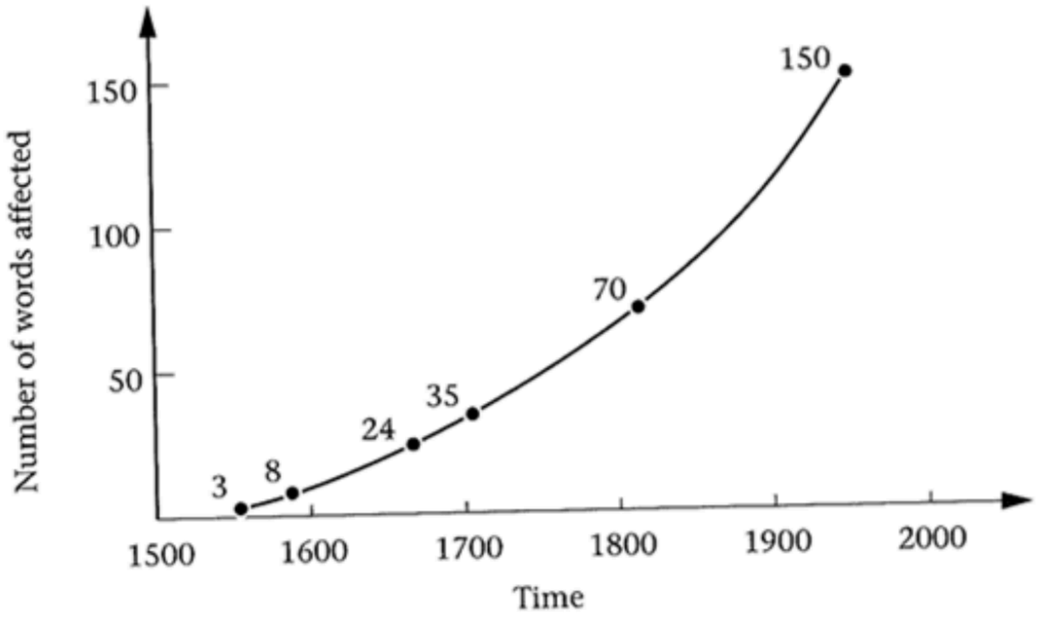
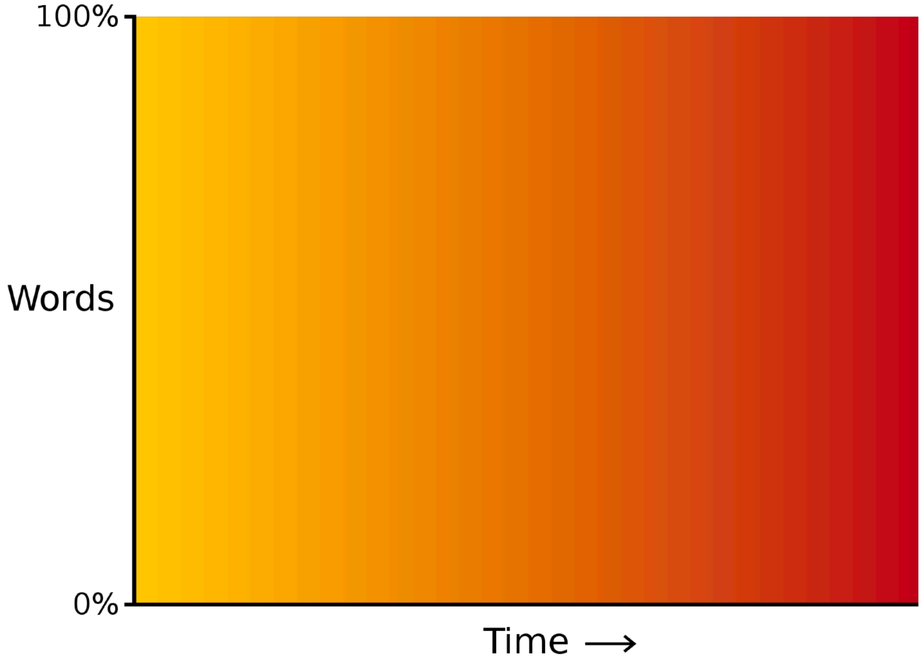


Figure 7.11 Diffusion of stress shift in English

Lexical Semantic Change

Uniform lexical diffusion

- Some changes apply immediately equally to all words.
- Cuban Spanish weakens the pronunciation of consonants:
  - Syllable-final [s] weakens to [h] for all words immediately.



https://en.wikipedia.org/wiki/Lexical_diffusion

	Spanish Spanish	Cuban Spanish	
es - ti - lo	[estilo]	[ehtilo]	“type”

Outline

- Recap on language change and language modeling
- Speed of change
  - Motivation
  - Paper Synopsis
    - Observations
    - Explanations
    - Questions
- Types of change
  - Phonological change
  - Morphological change
  - Semantic change
  - Syntactic change

Syntactic Change

Word order in OE

- Old English (OE) is flexible in word order
- SVO is found in simple transitive clauses:

Hē geseah þone mann

He saw the man

Syntactic Change

Word order in OE

- When the clause began with an element like þa ("then"), the verb would precede the subject, resulting in VSO:

þa	sende	sē	cyning	þone disc
then	sent	the	king	the dish

"Then the king sent the dish."

Syntactic Change

Word order in OE

- When the object is a pronoun, the word order is SOV:

Hēo hine lærde

She him saved

"She saved him."

Syntactic Change

Word order from OV to VO

...þæt ic **ðas boc** of Ledenum gereorde to Engliscre spræce **awende**.
...that I this book from Latin language to English tongue translate
"...that I translate this book from the Latin language to the English tongue."
(AHTh, I, pref, 6; van Kemenade 1987: 16)

...þæt he **his stefne** up **ahof**.
...that he his voice up raised
"...that he raised up his voice."
(Bede 154.28)

...forþon of Breotone **nædran** on scippe **lædde wæron**.
...because from Britain adders on ships brought were
"...because vipers were brought on ships from Britain."
(Bede 30.1-2; Pintzuk 1991: 117)

Ian Roberts, Diachronic Syntax, Oxford University Press, 2007

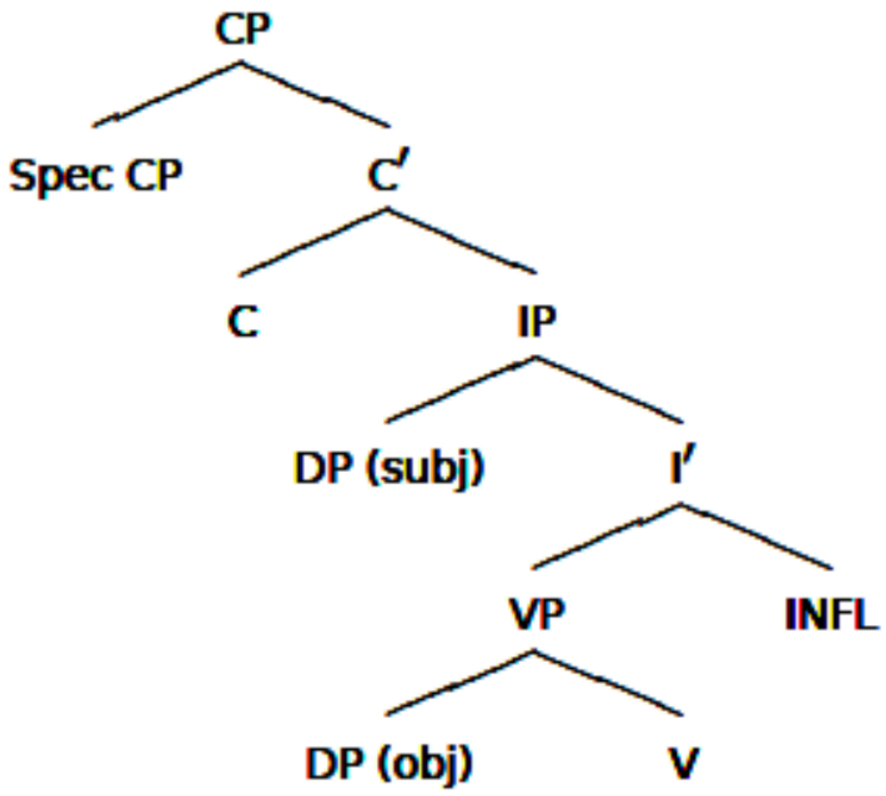
Syntactic Change

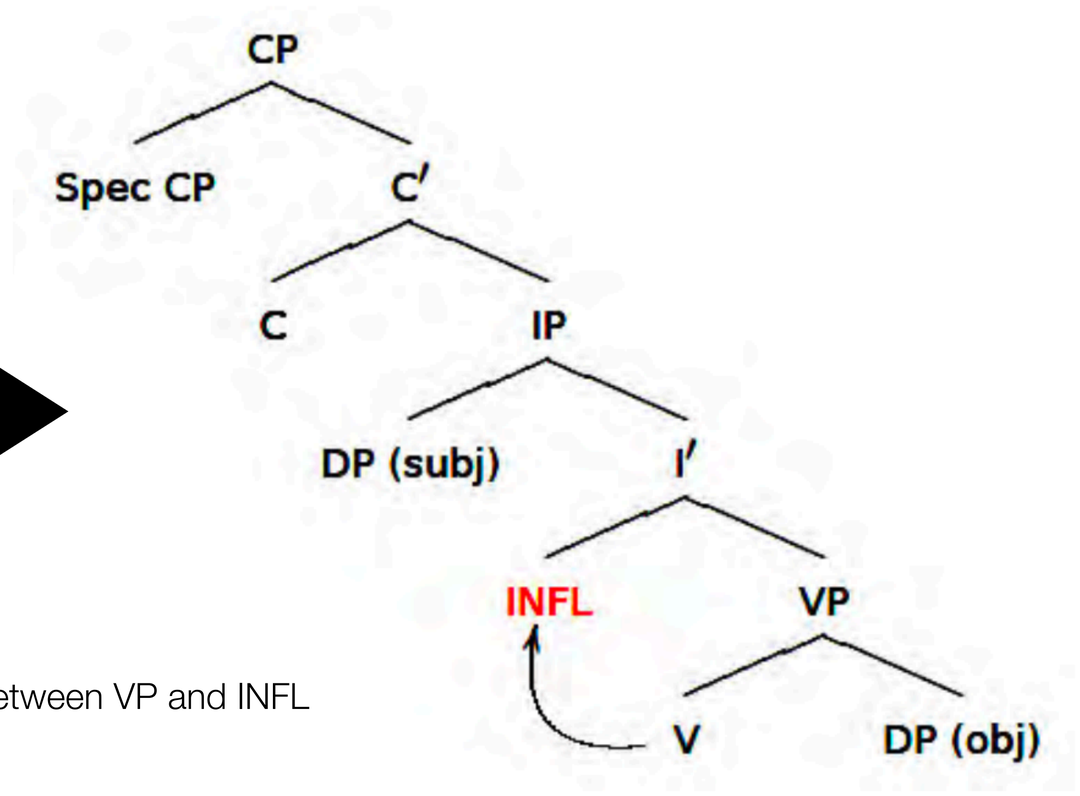
Word order from OV to VO

- Pintzuk (1995) argues that the change from OV to VO arises from the change in the order of INFL and its complement.
  - INFL starts at the final position but shifts to the middle over time. This change results in the shift from OV to VO.

Syntactic Change

Change in generative grammar structure





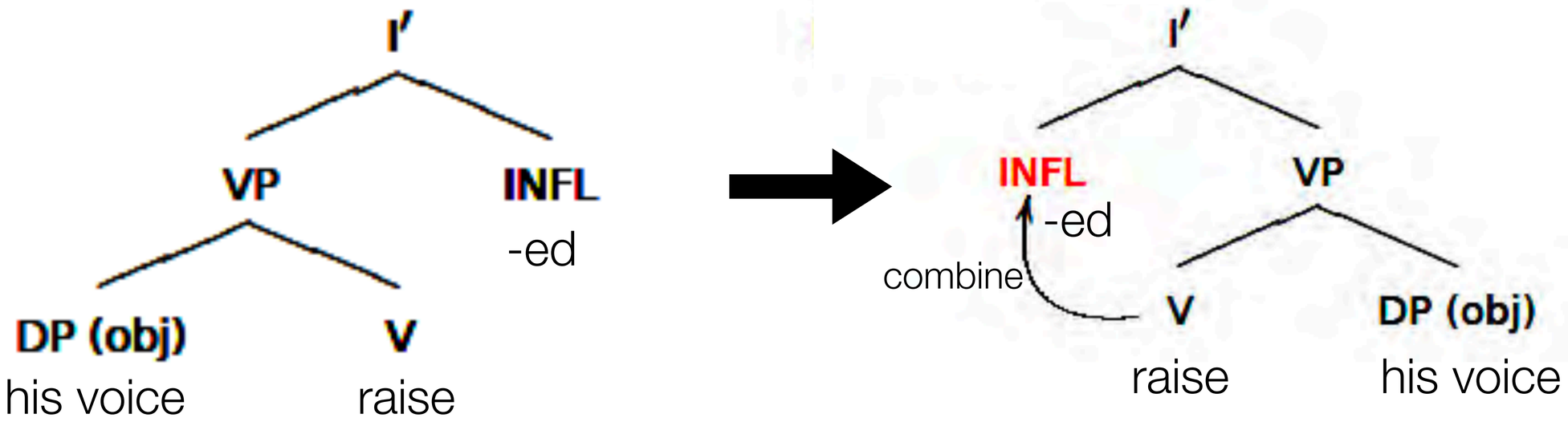
Switch between VP and INFL

https://shorturl.at/uwY02

Syntactic Change

Word order from OV to VO

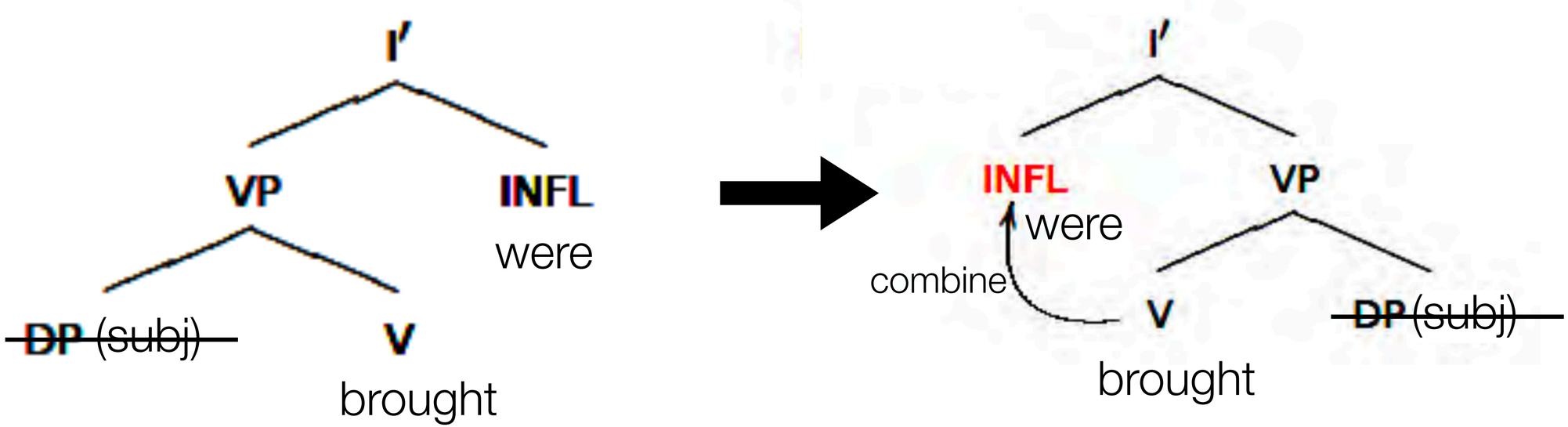
...þæt he **his stefne** up **ahof**.
...that he his voice up raised
"...that he raised up his voice."
(Bede 154.28)



Syntactic Change

Word order from OV to VO

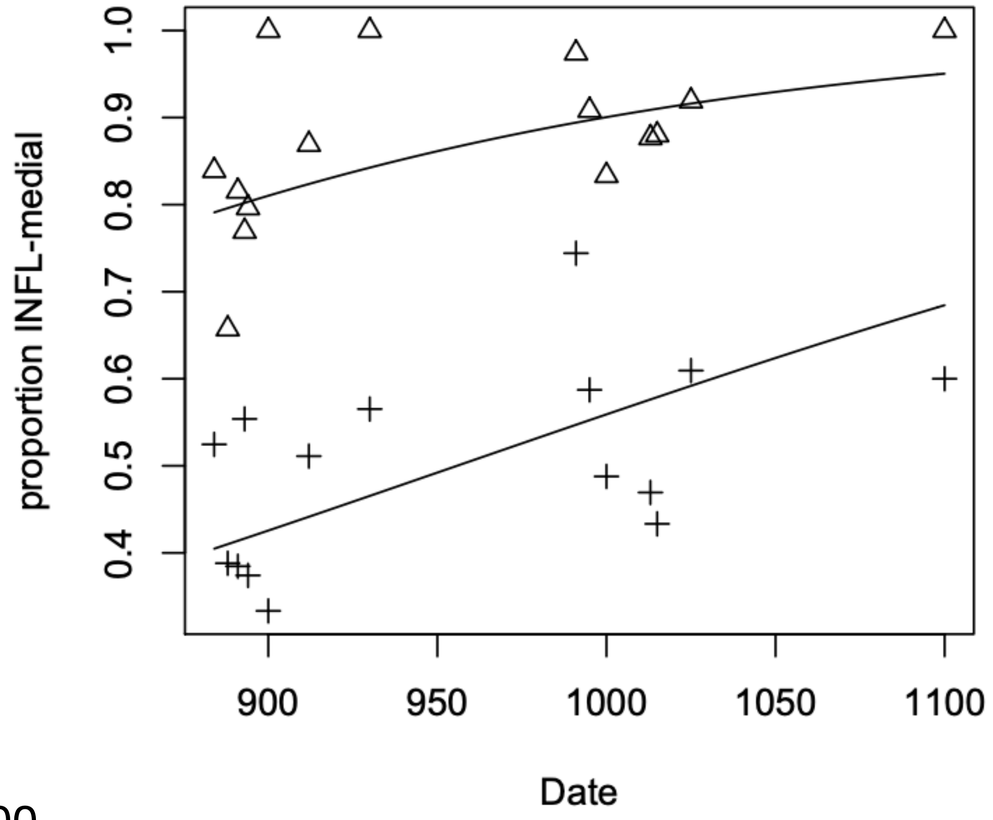
...forþon of Breotone **nædran** on scippe **lædde wæron**.
...because from Britain adders on ships brought were
"...because vipers were brought on ships from Britain."
(Bede 30.1-2; Pintzuk 1991: 117)



Syntactic Change

Change from OV to VO vs. Change in INFL

Year	*1000*	*1200*	*1300*	*1400*	*1500*
OV %	53	53	40	14	2
VO %	47	47	60	86	98



https://shorturl.at/uwY02

- Triangles – main clauses; crosses – subordinate clauses
  - INFL change is slow from 900 to 1000
  - INFL-middle: main clauses > subordinate clauses (frequency)
  - #high-freq INFL-middle = #low-freq INFL-middle from 900-1100
  - These results align with half-OV-half-VO results in the table from 1000-1200

Take-aways

Things to remember

- Rate of language change is not constant
  - different from one language to another
  - different across dialects
  - different across time periods
  - different across linguistic features
- What causes the difference
  - Geographical isolation
  - Social stability/instability