Cross-linguistic Corpora and the Theory of Language Change

Joel C. Wallenberg
Newcastle University
joel.wallenberg@ncl.ac.uk

June 28, 2013

Introduction

Variation in grammar is often described as falling into one of two categories.

1. Competing Grammars
   - Typically leads to language change via the replacement of one grammatical process by another.
   - Competition is parameterized in some fashion, as in competing flavors of the same functional head (Kroch, 1994).
2. Optionality (within a grammar?)
   - Diachronically stable variation between grammatical processes.

Introduction

Hypothesis: all variation, including grammatical optionality is formally the same as competing grammars, with the following consequences:

- Variation (apparent optionality) can be expected to resolve in either replacement of one form by another, or specialization for different functions in use.
- The difference between language change and stable variation/optionality must be explained by some mechanism outside of the grammar itself.
- Occasional exceptions are possible, depending on the mathematical character of some extragrammatical dimension with which the variation interacts.
    - Partial specialization of variants along a continuous dimension.

Outline

Blocking and Contrast

"Blocking Effect"

- General cognitive pressure against two forms existing for one function ("doublet") (e.g. morphosyntactic doublets as in Kroch 1994).

  {*dived*, *dove*} (dive-PST)
  {*jimmies*, *sprinkles*} (candy topping)

"Principle of Contrast"

- A strategy that children use in acquiring language: assume that two forms have two meanings (or uses)(Clark, 1987, 1990, *inter alia)*.
- Children hypothesize that novel words also refer to novel objects (as in Markman and Wachtel, 1988, among many other replications of the effect).

Blocking and Contrast

- A doublet is two variants competing for finite resources, as in e.g. biological evolution.
  - Instead of competing for something like food, they are competing for use (time in the mouths/brains of speakers)
  - Selection operates on the number of times a variant is **heard** (and accurately analyzed) by an acquirer.
- Either one variant has an advantage, and so replaces the other (following a logistic function; Nowak, 2006).
- Or neither variant has an advantage (or much of one), in which case random walk behaviour ensues.
- But in linguistic doublets, random walk cannot persist indefinitely because of the acquisition pressure of the Principle of Contrast.

The Principle of Contrast

- A strategy that children use in acquiring language: assume that two forms have two meanings (or uses).
  - Synonyms should only be acquired as a last resort.
- Demonstrated many times, in experiments like Markman and Wachtel (1988).
  1. 20 children
  2. 6 pairs of one familiar item (banana, cow, cup, plate, saw, spoon) and one unfamiliar item (cherry pitter, odd shaped wicker container, lemon wedgepress, radish rosette maker, studfinder, tongs).
  3. **Control**: "Show me one"
  4. **Test**: "Show me the X" (X = nonsense syllable)
- Control children pick the unfamiliar object at chance levels, but test children choose unfamiliar objects significantly higher than chance.

Competing Grammars

Competing Grammars, general form: 2 variants are available to a speaker with overlapping functions (e.g. the same meaning), and can't both be used at the same time.

- E.g. two featural versions of the same syntactic head.
- E.g. two different output mappings for the same phonological input.
- E.g. two different spell-outs of a morpheme.

- Necessary for the description of **any** linguistic change in a categorical dimension.
  - E.g. word-order parameters (Pintzuk, 1991; Santorini, 1992); a phonological rule like German final stop devoicing (Fruehwald et al., 2010).
  - In any such case, a speaker in the middle of the change in progress (code-)switches between categorical variants.

Blocking and Contrast

The possible historical outcomes of doublets (Competing Grammars), driven by selection and the Principle of Contrast are:

- Replacement of one by the other.
- Specialization of the two forms to different functions or meaning.

Proposal: every case of categorical linguistic variation or optionality can be reduced to competing grammars, leading to one of these two outcomes.
This simplifies the grammatical architecture necessary to account for both optionality and language change.

Example: English "Topicalization"

- Prince (1985, 1998, 1999): felicitous in two English discourse contexts, both of which require a certain type of contrast to appear on the fronted XP.

  (1) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And the third she'll feed junk food.

  (2) She was here two years. [checking transcript] Five semesters she was here.
  (Prince, 1999, 8,9)

- However, it is **never** obligatory.

Example: English Topicalization

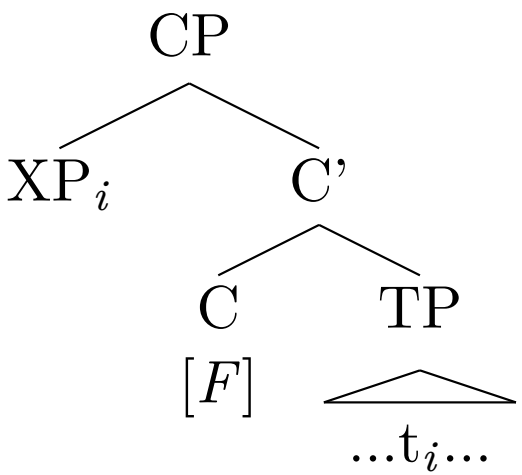
- As long as the accent pattern is kept constant, both orders are felicitous:

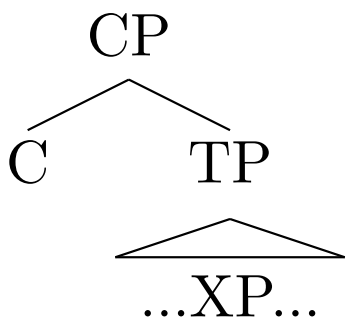
  (3) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And **the third** she'll feed **junk food**.

  (4) She's going to use three groups of mice. One, she'll feed them mouse chow, just the regular stuff they make for mice. Another she'll feed them veggies. And she'll feed **the third junk food**.

Topicalization in Minimalism

- Move is triggered by the feature content of some head.
- Given "Merge...preempts Move" (Chomsky, 2000), a feature cannot encode optional movement.
- Therefore, optional movement must involve a choice (for the **Numeration**) between two variants of a functional head, out of an inventory of possible heads:





- This is the core case of morphosyntactic doublet (i.e. competing heads) described in Kroch (1994).

A Minimalist Hypothesis

Given that:

- these mechanics are necessary to encode syntactic optionality in a Minimalist system,
- the same mechanics are necessary to describe a change in progress

Then, the system is simplest if no more machinery is added to deal with optionality/variation.

A Minimalist Hypothesis

- **Prediction:** every case of syntactic optionality or variation is one of the following:
  1. A replacement change in progress (outright competition going to completion).
  2. A specialization change in progress (specialization for different functions going to completion).
  3. **The only real case of diachronically stable variation/optionality:** variants have partially specialized along a continuous (or ordinal) dimension, e.g. style, prosodic weight.
- If categorical variants specialize along a categorical dimension, complete specialization should eventually result.
- If categorical variants specialize along a continuous or ordinal dimension, then complete specialization can **never** result (but replacement can still be arrested).

Case Study: Embedded Polar Questions

A quantitative study of embedded *yes/no*-questions in English and Icelandic, comparing the use of *whether* vs. *if*, and *hvort* vs *ef* found specialization in English, and replacement in Icelandic (Bailey, Wallenberg, & van der Wurff 2012).

(5) John wondered whether Mary was coming to the party.

(6) John wondered if Mary was coming to the party.

This variation does not exist in modern Icelandic, but it did in earlier Icelandic.

Origins of *whether* / *if* Variation

- The *if*-questions are older, as they occur throughout Germanic.
- *whether* had an old meaning as a dual *wh*-pronoun ("which of two"), from the Proto-Germanic class of duals.
- The *whether*-questions came from a very early reanalysis (possibly proto-Northwest Germanic).
- Icelandic once had this variation but no longer does, whereas English shows variation throughout its history (up to PDE).
- **Hypothesis:** Both are grammar competition (i.e. doublets); the Icelandic case is one of **replacement**, whereas the English case is one of **specialization**. (These are the only two possible outcomes.)

Two meanings of *whether*

- Van Gelderen (2009): question-meaning of *whether* came from the older dual-meaning of *whether*.

dual-meaning:

(7) hwæðer ðara twegra dyde ðæs fæder willan
whether of-the two did the father's will
"Which of the two did the will of his father?"
(*West Saxon Gospels*, from *York-Toronto-Helskinki Corpus of Old English Prose*; Taylor et al. 2003)

question-meaning:

(8) cweþe ge, la leof, hwæðer he sylf Crist sy
say you EXCL dear whether he self Christ is
"Please say whether he is Christ himself"
(*Ælfric's Catholic Homilies*, YCOE)

Context for Reanalysis

- Context where a child might make a mistake, misinterpreting the dual-meaning of *whether*, and creating the question-meaning of *whether*.

  Disjunctive Yes/No Questions:

  (9) I asked whether John wants tea or coffee.

  I asked which of the two he wants, A or B.

  $\longrightarrow$

  I asked does he want A, or B?

Can we find this disjunction context in Old English?

(10) he gecyðde hwæðer he mænde, ðe ðæs
he revealed which/whether he meant either the
modes foster ðe ðæs lichoman
mind's nourishment or the body's
"he revealed which/whether he meant nourishment for the mind or for the body"
(*Cura Pastoralis*, date: 9th c.YCOE)

Can we find this disjunction context in Old English?

(11) hwæðer ys mare ðe ðæt gold ðe ðæt templ
which/whether is bigger either the gold or the
ðe ðæt gold gehalgaþ
temple which gold makes-holy
"Which/whether is more important the gold or the temple that makes the gold holy?"
(*West Saxon Gospels*, date: pre11th c., YCOE)

Can we find this disjunction context in Old English?

(12) and hire axode of hwilcere þeode hi wære and
and her asked from which people she was and
hwæder hi wære Cristen and frig oðde
which/whether she was Christian and free or
þeowa
servant
"and asked her which people she was from and which/whether she was Christian and free, or a servant."
(*Life of Saint Margaret*, YCOE)

A Quantitative Study

- A quantitative study of embedded *yes/no*-questions in English and Icelandic, comparing the use of *whether* vs. *if*, and *hvort* vs *ef*.
- **Result 1:** A strong predictor of *whether* vs. *if* in both languages is the presence/absence of a disjunction (i.e. *or*, *eða*) in the clause, with *whether* being favoured in the disjunction case more than in the simple case in both languages, across their whole histories.
- This is a remarkably long-lasting "persistence" effect of the original reanalysis environment (cf. *have* vs, *have got* study by Shawn Noble, reported in Kroch 1989, cf. also Labov 1989).

Case Study: Embedded Polar Questions

In all stages of English and in historical Icelandic, a disjunction favors *whether*.

English

Disjunction Context:

(13) I wonder {**whether**, if} John or Bill is bringing coffee.

(14) I wonder {**whether**, if} John is bringing tea or coffee.

(15) I wonder {whether, if} John is bringing tea or not.

Simple Context:

(16) I wonder {whether, **if**} Bill is bringing coffee.

Example: Embedded Polar Questions

Disjunction Context:

(17) eftir því **hvort** maður vill heitt eða kalt
according it-DAT whether man wants hot or cold
"According to whether one wants hot or cold"
(*Sagan Öll*, date: 1985, from IcePaHC)

Simple Context, (older) Icelandic:

(18) vér vitum eigi, **hvort** vér tökum öndina
We know not whether we take soul-the

(19) og spurðu, **ef** hann væri Kristur
and asked if he were Christ
(*Icelandic Homilies*, date: 1150, from IcePaHC)

A Quantitative Study

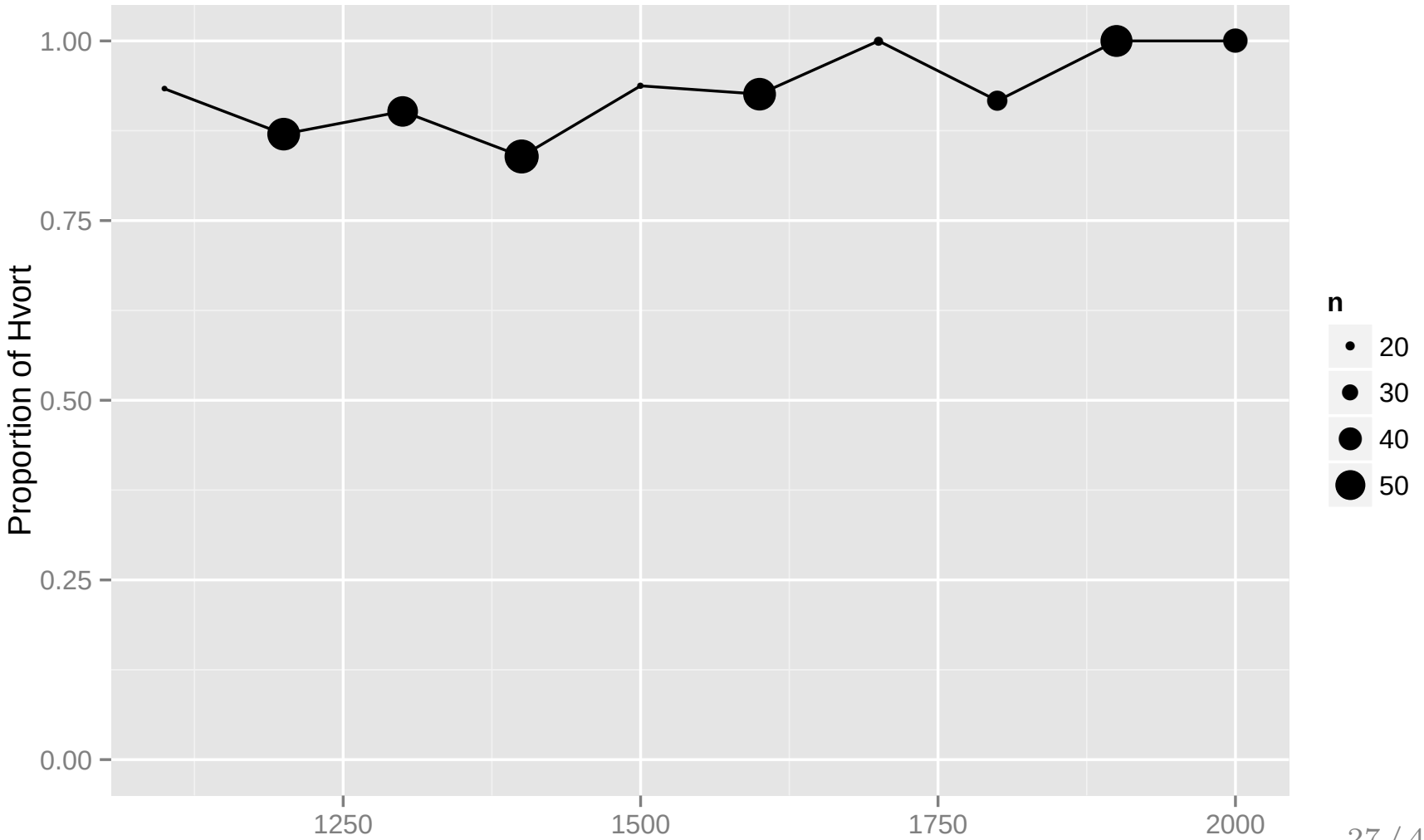
- **Result 2:** The *whether* structure completely replaces the *if* structure in the Icelandic case, but not in the English case.
- If the two possible outcomes of a morphosyntactic doublet are replacement or specialization (Kroch, 1994), Icelandic shows the former and English shows the latter.
- We propose that replacement must occur when there is some selectional advantage to one of the variants (in Darwinian terms, where reproduction = learning).
- Specialization must occur when there is no selectional advantage to one of the variants.
- **Experimental Infrastructure:** accurate parsed diachronic corpora: YCOE (Taylor et al., 2003), PPCME2 (Kroch and Taylor, 2000), PPCEME (Kroch et al., 2005), PPCMBE (Kroch et al., 2010), and IcePaHC (Wallenberg et al., 2011).

English *whether* vs. *if* Questions, N = 1929 clauses

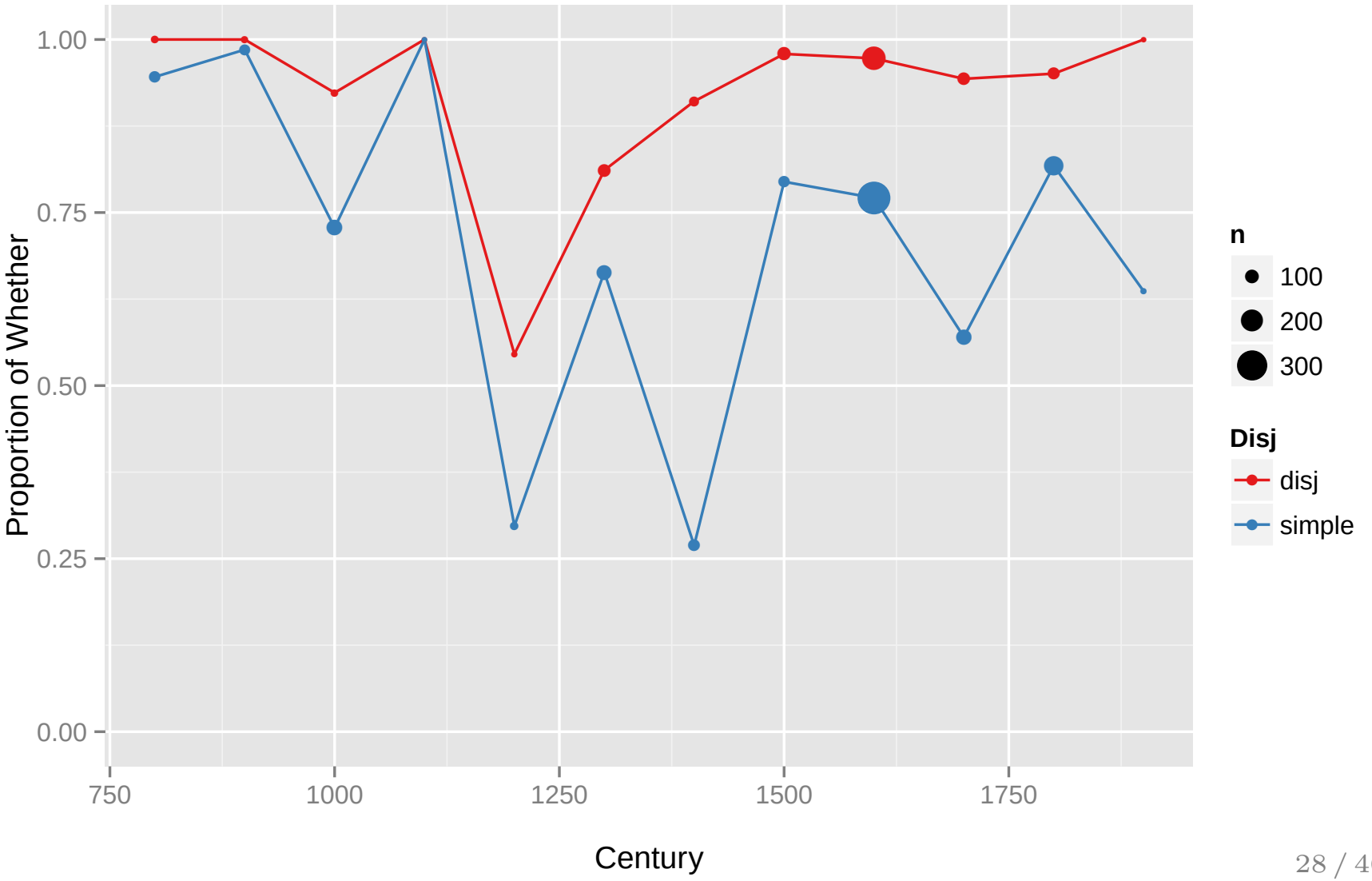
Parsed Corpora: YCOE, PPCME2, PPCEME, PPCMBE

Icelandic *hvort* vs. *ef* Questions, N = 397 clauses

IcePaHC 0.9 (Wallenberg, AK Ingason, EF Sigurðsson, & E Rögnvaldsson 2011)



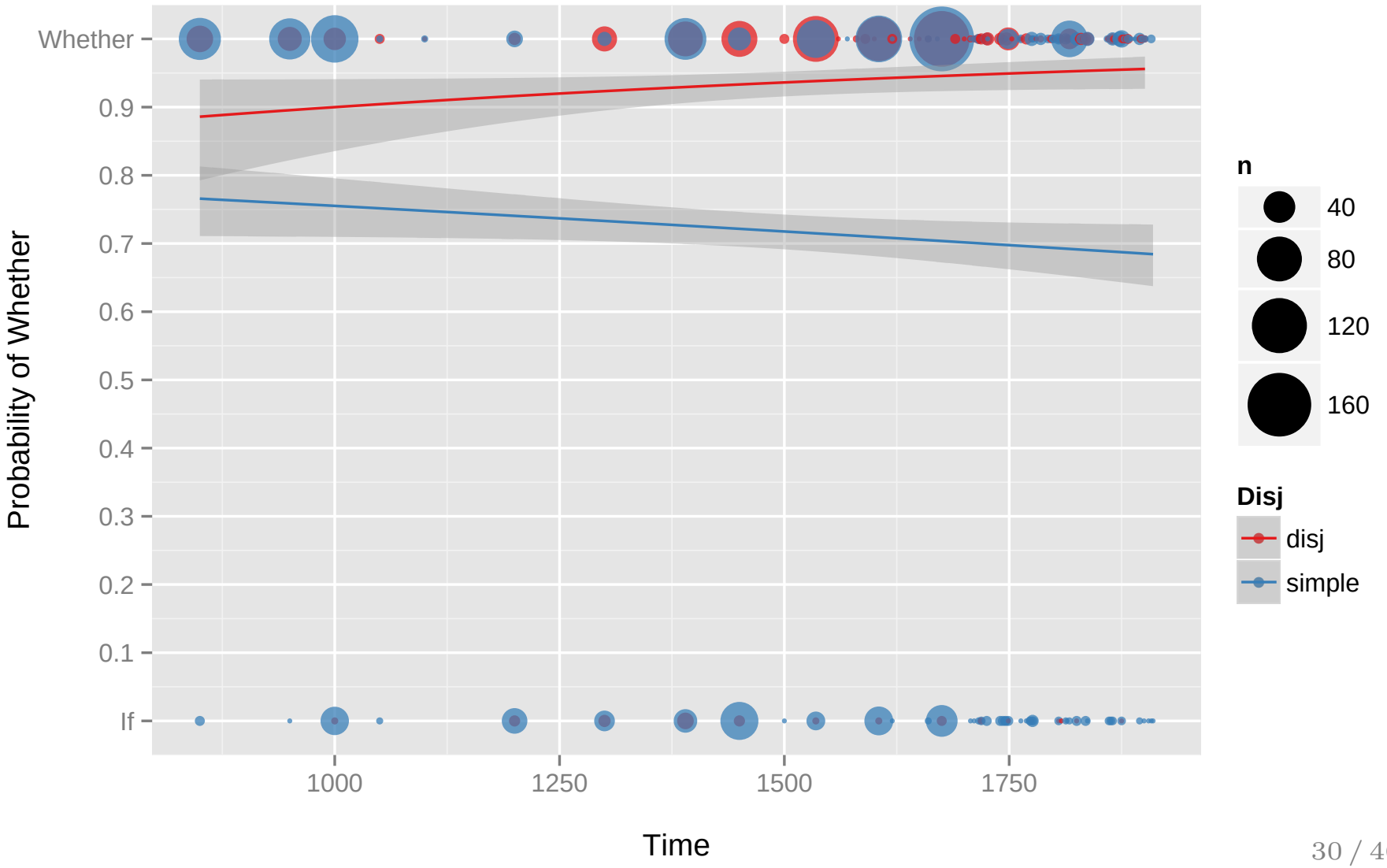
Specialization in English (N = 1929 clauses)



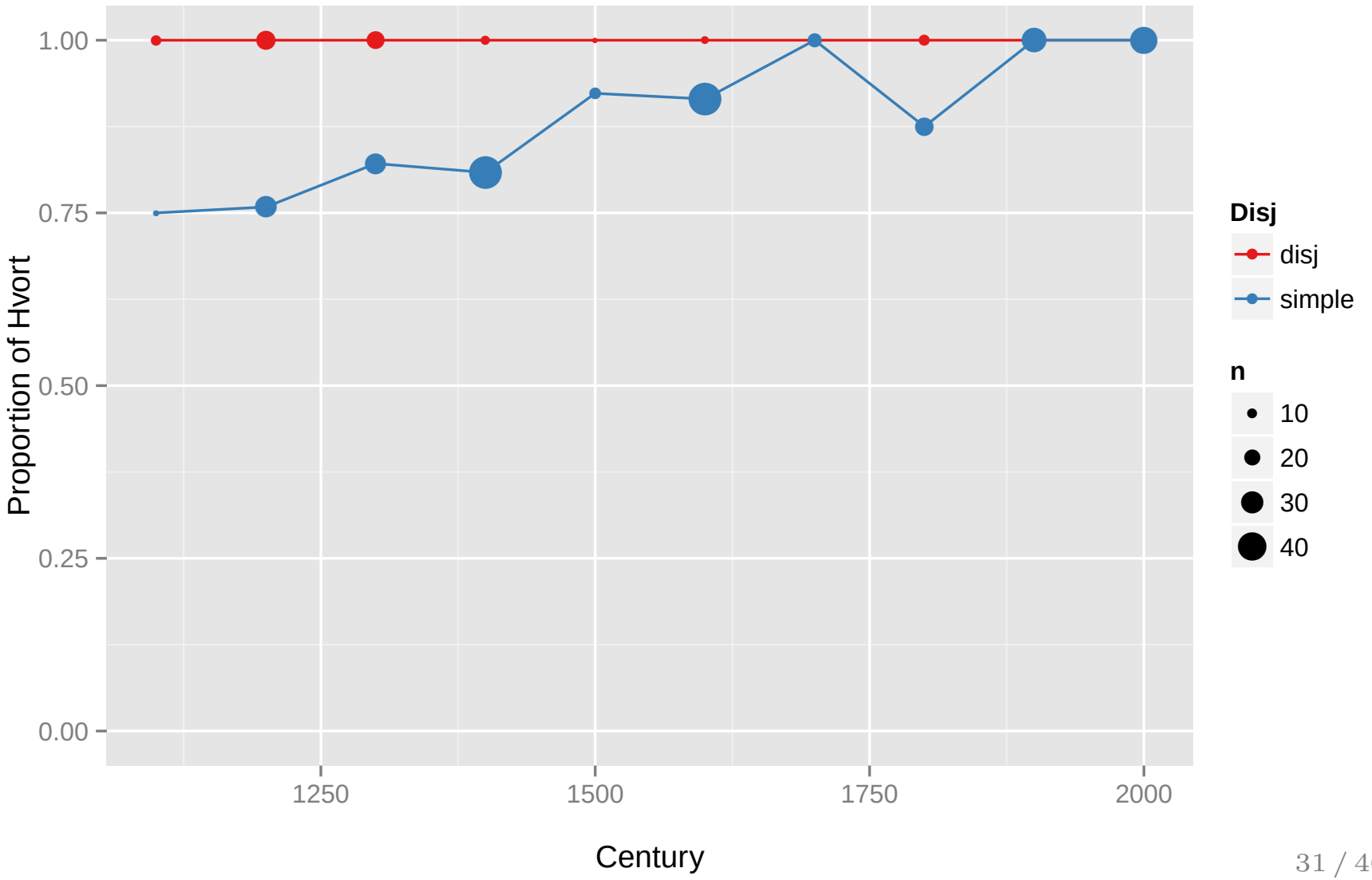
	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			1928	1928.3	
Disj	1	152.667	1927	1775.7	< 2e-16
Time	1	1.480	1926	1774.2	0.224
Disj:Time	1	5.401	1925	1768.8	**0.0201**

- A model without an interaction between Disjunction and Time fits significantly worse.
- Note that there is no clear effect of Time on *whether* use in general; the interesting effect is an interaction between Time, Disjunction, and *whether* use.
- In other words, *whether* is not in decline, being replaced by *if*, but rather they are diverging from each other in use, specializing for the two contexts.

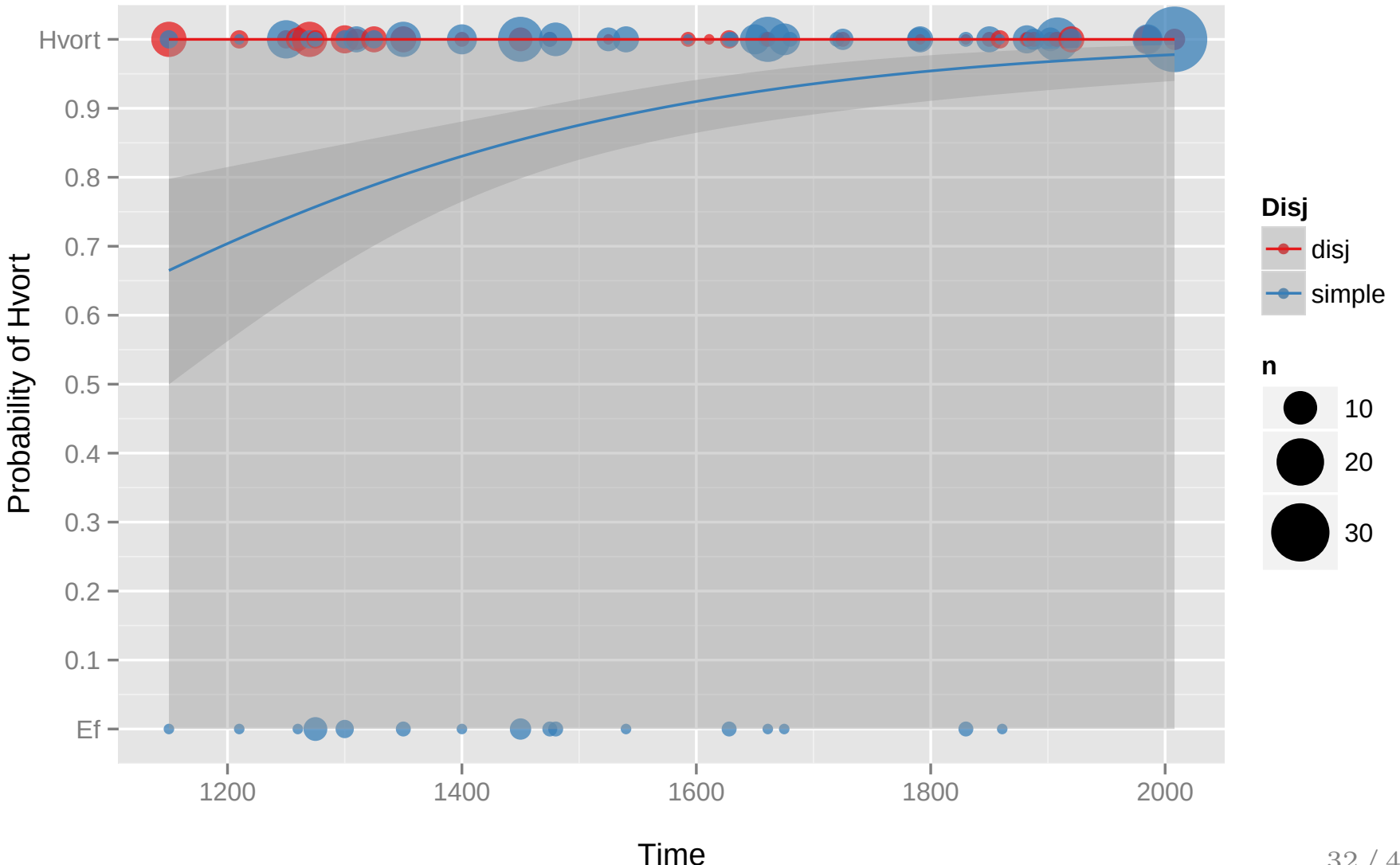
English, Logistic Model, N = 1929



Replacement in Icelandic (N = 397 clauses)



Icelandic, Logistic Model, N = 397



An Evolutionary Process

- The Blocking Effect is reducible to Darwinian selection plus the Principle of Contrast.
- A doublet resolves in replacement when one form has a selectional advantage.
- A doublet resolves in specialization when neither form has a selectional advantage (or a very small one).
- Unlike biology, the Principle of Contrast is built into acquisition and prevents random walk.
- In biology, a selectional advantage is a higher probability of reproduction.
- In language change, a selectional advantage is a higher probability of a child hearing and acquiring a particular structure.

Icelandic and English

- There must be some selectional advantage in the Icelandic case that is not present in the English case.
- Icelandic has retained the two function of *hvort* to the present day, much longer than English retained the two functions of *whether*.
- If the reanalysis we propose continues to occur over the history of Icelandic during acquisition, then the learner will consistently overestimate the amount of question *hvort* in the primary linguistic data.

Icelandic and English

- The child occasionally mis-analyses dual-*hvort* as question-*hvort*.
- This provides an advantage to *hvort* over *ef* because *hvort* reproduces slightly more often (in the child's learning) than *ef*.
- In English, dual-*whether* is lost much earlier, and so the system tends towards very gradual specialization after that point, due to the Principle of Contrast's pressure in acquisition.
- The difference between the languages could be due to the timing of an overlapping change in English: *which* and *what* taking over the function of dual-*whether*.

Continuous Dimensions

Hypothesis: Stable variation, i.e. optionality, results from categorical variants specializing along a continuous dimension.

There are many possible continuous dimensions, including language internal dimensions like

- weight (word length)
- prosodic accent (number of aligned prosodic peaks, degree of stress clash between two positions)

and language external dimensions like

- style
- speech rate

How is Topicalization different from *whether/if*?

- Is the frequency stable over time? Probably, at least since Late Middle English (Speyer, 2010).
- Is it specialized for different styles? Not that we know of.
- Is it sensitive to prosody? Definitely (Speyer, 2010).

(20) The first she'll feed mouse chow, the second she'll feed veggies, and **the third** she'll feed **junk food**.

(21) ? The first Anders will feed, the second Joel will feed, and the third Wim will feed.

(22) ?? Joel Anders will pay, Jill Wim will pay, and Ann Maggie will pay.

If you would like to find out more of this extension of the theory, come to DiGS 15 in Ottawa!

Conclusions

- Within syntax, only one formal account of optionality/variation is available, the same one that accounts for language change: Competing Grammars.
- This results in replacement, specialization, or stable variation (true optionality) in exceptional cases.
- Replacement occurred in the *whether/if* variation in Icelandic due to a selectional advantage.
- In English, since the variation could be mapped onto a categorical domain of specialization, it was.
- Stable variation is (only) the result of mapping categorical variation onto a continuous dimension of specialization.

Conclusions

- All categorical variation/optionality/change = Blocking Effect, Competing Grammars
- Blocking Effect = Darwinian selection, Principle of Contrast (and a domain of specialization)
- Thus, all categorical variation/optionality/change is reduced to interactions of Competing Grammars, Darwinian selection, Principle of Contrast

Methods, Step-by-Step

1. CorpusSearch coding queries. (Plus some checking of the codes by hand.)
2. Extract a file containing only the codes.
3. Import into R.
4. Take the relevant subset of codes for analysis.
5. Statistics, plots, etc., in a fairly painless way.

References I

Bailey, Laura, Joel C. Wallenberg, and Wim van der Wurff. 2012. Embedded yes/no questions: reanalysis and replacement. Paper presented at Presented at the 2012 Annual Meeting of the Linguistics Association of Great Britain (LAGB), University of Salford.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of howard lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–155. Cambridge, Mass.: MIT Press.

Clark, Eve. 1987. The Principle of Contrast: A constraint on language acquisition. In *Mechanisms of language acqusition*, ed. Brian MacWhinney, The 20th Annual Carnegie Symposium on Cognition. Lawrence Erlbaum Associates, Inc.

References II

Clark, Eve. 1990. On the pragmatics of contrast. *Journal of Child Language* 17:417–431.

Fruehwald, Josef, Jonathan Gress-Wright, and Joel C. Wallenberg. 2010. Phonological Rule Change: The Constant Rate Effect. In *Proceedings of 40th Meeting of the Northeast Linguistic Society (NELS) 40*.

Kroch, Anthony, Beatrice Santorini, and Lauren Delfs. 2005. Penn-helsinki Parsed Corpus of Early Modern English. Size 1.8 Million Words.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

References III

Kroch, Anthony S. 1994. Morphosyntactic variation. In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, ed. K. Beals et al et al.

Kroch, Anthony S., Beatrice Santorini, and Ariel Diertani. 2010. Penn Parsed Corpus of Modern British English. Size ∼ 950000 words.

Kroch, Anthony S., and Ann Taylor. 2000. Penn-Helsinki Parsed Corpus of Middle English. CD-ROM. Second Edition. Size: 1.3 million words.

Labov, William. 1989. The child as linguistic historian. *Language Variation and Change* 1:85–97.

References IV

Markman, Ellen M., and Gwyn F. Wachtel. 1988. Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology* 20:121–157.

Nowak, Martin A. 2006. *Evolutionary dynamics: exploring the equations of life*. Cambridge, MA: Belknap Press of Harvard University Press.

Pintzuk, Susan. 1991. Phrase structures in competition: Variation and change in Old English word order. Doctoral Dissertation, University of Pennsylvania.

Prince, Ellen. 1985. Fancy syntax and shared knowledge. *Journal of Pragmatics* 9:65–81.

Prince, Ellen. 1998. On the limits of syntax, with reference to left-dislocation and topicalization. *Syntax and semantics* 281–302.

References V

Prince, Ellen. 1999. How not to mark topics: 'Topicalization' in English and Yiddish. In *Texas linguistics forum*, chapter 8. University of Texas, Austin: Citeseer.

Santorini, Beatrice. 1992. Variation and Change in Yiddish Subordinate Clause Word Order. *Natural Language & Linguistic Theory* 10:595–640.

Speyer, Augustin. 2010. *Topicalization and stress clash avoidance in the history of english*. Topics in English Linguistics. Berlin/New York: Mouton de Gruyter.

Taylor, Ann, Anthony Warner, Susan Pintzuk, and Frank Beths. 2003. The York-Toronto-Helsinki Parsed Corpus of Old English Prose.

References VI

Van Gelderen, Elly. 2009. Renewal in the left periphery: economy and the complementiser layer. *Transactions of the Philological Society* 107:131–195.

Wallenberg, Joel C., Anton K. Ingason, Einar F. Sigurðsson, and Eiríkur Rögnvaldsson. 2011. Icelandic Parsed Historical Corpus (IcePaHC). Version 0.9. Size: 1 million words. URL `http://www.linguist.is/icelandic_treebank`.