This presentation shows how cross-linguistic comparison with multiple diachronic, parsed corpora can test and develop a detailed theory of linguistic variation and change in ways that data from a single corpus cannot. Using a diachronic parsed corpus of Icelandic (IcePaHC; Wallenberg et al. 2011) and four parsed diachronic corpora of English (Kroch et al. 2005, 2010; Kroch and Taylor 2000; Taylor et al. 2003), we present a comparison of long-term syntactic change in Icelandic and English of a type that was impossible to carry out before this kind of diachronic comparative data became available. In our case study, the same initial state of variation in embedded polar questions in Icelandic and English leads to two different outcomes in the two languages: replacement of one form by the other in Icelandic, and specialization of the two forms for different syntactic contexts in English. Furthermore, the change in English is slow enough that it is not observable at all without the remarkable time-depth and sample size afforded by the combined four corpora of historical English. The Icelandic and English data bear out Kroch (1994)’s prediction about the two possible outcomes of morphosyntactic variation, specialization and replacement, suggesting a particular theoretical direction in how variation must be represented in a Minimalist syntax.

The presentation will also include a systematic introduction to the methods the study used, geared towards practitioners and advanced students of syntax and language change who would like to make use of parsed corpora in their own research. The talk will introduce methods for searching multiple parsed corpora in a maximally efficient way, and making sure that queries are kept as controlled as possible across corpora of different languages. Finally, there will be a brief introduction to analyzing and displaying this type of data in R.

Our case study takes as its starting point the theory of “morphosyntactic doublets”, or “competing grammars”, as described in Kroch (1994), and applies it to the following variation found in English embedded polar questions:

1. John wondered whether Mary was coming to the party.
2. John wondered if Mary was coming to the party.

Interestingly, the same variation occurred in older stages of Icelandic:

3. vér vitum eigi, hvort vér tökum öndina
   We know not whether we take soul-the

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

Each case of variation arose from the same change, the innovation of a wh-word for polar questions, whether or hvort, from an earlier dual form (Bailey et al. 2012; Van Gelderen 2009). Kroch (1994) hypothesized that a morphological doublet like this one is diachronically unstable, and must result in either the replacement of one variant by the other, or the specialization of the two variants for different contexts such that they no longer compete for use (i.e. the speaker no longer chooses between them in any context).

The study shows that the prediction is correct, and more interestingly, that the same initial conditions can lead to replacement in Icelandic but specialization in English. Furthermore, the specialization change in English is shown to be so slow that the variation
observed in present-day English (1-2 above) must be the result of the same change, still in progress in the modern language. We are only able to observe a change this slow because all four parsed corpora of historical English, considered together, allow us to observe the change over a period of over 1000 years, and with a reasonable sample size of clauses throughout. Similarly, IcePaHC allows a comparison between Icelandic and English over an overlapping period of more than 850 years, allowing us to observe a different outcome of the same variation during roughly the same time-span.

Given that such results have only recently become possible, we would rightly expect new theoretical insights to emerge from this new source of data. This study suggests that linguists must be careful in analyzing apparent morphosyntactic “optionality” in present-day languages, and that the historical trajectory of any case of variation must be considered if we are to provide a truly explanatory analysis of any variable phenomenon. The talk will also reflect on the nature of morphosyntactic variation in a Minimalist syntax (Chomsky 1993, 2001, and subsequent), and suggest that all synchronic linguistic optionality may very well be cases of competing grammars, or “doublets” in the sense of Kroch (1994). If shown to be plausible, this hypothesis could affect much of the work done in synchronic syntactic theory.