Optional to and prosody
Thomas Wasow, Rebecca Greene (Stanford University), & Roger Levy (University of California, San Diego) wasow@stanford.edu

Stress sensitivity; Syntactic reduction; Corpus analysis; American English

Numerous studies have investigated factors influencing the presence or absence of that, at the beginning of complement or relative clauses. These include the predictability of the clause in its context, the complexity of the clause, and various properties of the clause-internal subject NP, plus such factors as modality (speech vs. writing) and speech rate (e.g. [1], [2], [4], [5]).

In contrast, the factors influencing where the infinitival to will occur when it is optional remain almost entirely unexplored. One environment where to is optional is what we call the do be construction (DBC), exemplified in (1). DBC is characterized by a relative clause in the subject containing some form of do, followed by a copula interpreted as an assertion of identity, which is in turn followed by a verb phrase. The post-copula verb (PCV) is usually in base form (uninflected, without to), but infinitival to also occurs frequently. The forms with and without to appear semantically indistinguishable.

We analyzed over 1000 DBC examples from the spoken portion of the Corpus of Contemporary American English. Several factors behave as we expected, based on the analogy with optional that: sentences with longer subjects or longer post-copula VPs have higher rates of to, and to is more likely when the do in the subject is in infinitival form than when it is finite or base (arguably due to priming). But frequency of the post-copula verb is positively correlated with the occurrence of to, contrary to what accounts in terms of predictability would lead one to expect.

Considering prosody as a factor in the distribution of optional to helps to explain this surprising correlation. Specifically, avoidance of stress clash and stress lapse – that is, two adjacent syllables that are both stressed or unstressed ([3]) strongly influences whether to is used. Infinitival to is unstressed; so insertion of to would always either prevent clash (if the adjacent syllables are both stressed) or cause lapse (if either adjacent syllable is unstressed). The rate of to use is almost three times as high in environments where it prevents clash as in environments where it causes lapse. Frequent verbs tend to be monosyllabic and hence initially stressed, and the DBC copula carries some stress (as evidenced by the fact that it never contracts). Thus, in the absence of adverbs, use of to with frequent PCVs often prevents stress clash. When an adverb intervenes between the copula and the PCV, to is extremely rare; examination reveals that its use in these cases would usually create lapse rather than prevent clash. Once we included stress in our models of to use, the expected negative correlation with PCV frequency emerged.

Our study of optional to supports the processing accounts of the distribution of optional that put forward by Jaeger and others. In mixed models of our DBC data, prosody consistently ranks among the strongest predictors of to. This suggests that other studies examining the factors influencing syntactic alternations should be including prosody as well.

(1) a. What the former colonial subject has done is (to) beat the master at his own game.
   b. All they can do is (to) say we’ll try to enforce the no-fly zone.
   c. The best thing to do was simply (to) remind the reader repeatedly.

References