An Intelligent Survey System for Measuring Dialect Grammaticality

Grammaticality judgment tasks, in which informants are asked to rate the acceptability of sentences that encapsulate a particular grammatical construction, are a well-established means of collecting data for syntactic analysis (Schütze 2016). Elicitation of native speaker intuitions is essential in the study of dialect syntax, as non-Standard morphosyntactic features do not necessarily occur frequently in spontaneous speech; they may also be rare or absent in existing corpora (Cornips & Poletto 2005). As an extension of this approach, researchers seeking to collect data on dialect syntax make use of a variety of techniques adapted from dialectology and sociolinguistics, such as written questionnaires and in-person interviews with dialect speakers (Cornips & Jongenburger 2001). These methods have the potential to provide specific responses, allowing researchers to elicit particular features, as well as ask follow-up questions and seek clarification. At the same time, fieldwork is costly in terms of time and money.

The advent of readily available Internet survey systems means that native speaker assessments of grammaticality may be collected not just in experimental or face-to-face fieldwork settings, but also online. Internet-based surveys have the advantage of allowing researchers to reach large numbers of participants with a small initial time investment, and they are appealing to users who may be used to completing surveys or questionnaires as a form of entertainment. A number of systems have been developed specifically for language judgement tasks: MiniJudge is designed for small-scale experiments focussed on binary factors (Myers 2007); comparable systems such as Turkolizer and Turktools use Amazon's Mechanical Turk to reach a large number of participants (Gibson, Piantadosi & Fedorenko 2011; Erlewine & Kotek 2016). These systems provide a number of useful features, such as randomisation of sentence presentation, inclusion of distractor sentences, and complex statistical analysis.

Regardless of the means of collection, however, grammaticality judgment tasks are problematic. Because 'naïve' speakers are unable to access and report their implicit knowledge of language, when asked for *grammaticality judgments*, informants rate sentences according to whether they are *acceptable* (Schütze & Sprouse 2013). These acceptability judgments reflect a variety of factors other than grammaticality, such as pragmatic plausibility; moreover, written forms may also be particularly subject to the influence of prescriptive norms that do not reflect actual use (Cornips & Jongenburger 2001)

Grammaticality judgements are also problematic because speakers may accept or reject the same sentence for different *grammatical* reasons. For instance, a speaker might reject (1) because they cannot use an embedded passive without *to be*; because they cannot use this construction with *want* (but might with *need*); because they can use *want* only with animate subjects; because they allow such constructions only with a present participle (*mowing*); or because they favour the passive participle form *mowed*.

(1) The lawn wants mown

For a questionnaire with a fixed set of questions, relationships amongst the grammatical features of different sentences raise an additional problem. Where a speaker finds (1) grammatical, judgments about the sentences in (2) may be superfluous; where a speaker finds (1) ungrammatical, judgments on sentences in (2) may be required to provide insight into the reasons for the unacceptability of (1).

(2) a. The lawn wants to be mown
b. The lawn needs mown
c. The cat wants fed
d. The lawn wants mowing
e. The lawn wants mown

Such considerations become especially important in the context of dialect syntax, where judgments may be influenced by a complex set of (potentially non-binary) factors. Researchers are also likely to be interested in whether individual speakers allow both Standard and non-Standard forms, which has implications for how we view variation within a single grammar.

We therefore propose a dynamic, computer-based system for the collection of grammaticality judgments. This system allows linguists with no computing background to annotate strings with specific linguistic features; the features are recorded to a *knowledge graph*, a computational graph structure that maps connections between entities (Soares et al. 2018). Strings are thus presented to participants for judgment in a traditional way, but each has underlying encoding that incorporates analysis at the level of features rather than just strings, allowing integration of hypotheses about these features into the system.

As a test case we have looked at the verbs *need*, *want*, or *like* followed directly by a passive participle:

(3) a. The windows need washed (cf. Standard English: *The windows need to be washed*)
b. The cat wants fed
c. The kitten likes cuddled

This construction is attested in Scottish English, but theoretical analysis has been largely based on data from speakers in the American midlands (e.g. Murray, Frazer & Simon 1996; Murray & Simon 1999; Edelstein 2014). These previous studies make a number of claims, most notably:

- On a global level, the construction exemplified in (3) is more often used with *need* than *want*, and more often with *want* than *like.* Moreover, for any individual speaker, use of this construction with *like* will mean they can use it with *want*, and with *want* will mean they can use it with *need.*
- Some speakers allow *want* and *like* in this construction with animate subjects, where they would be disallowed in the Standard form, e.g. *my hair wants cut*.
- Some speakers have only the non-Standard construction, indicating that it is not simply an elliptical form of the Standard equivalent with *to be*.

This construction therefore makes an ideal candidate for the testing of pre-existing predictions in a novel linguistic context (i.e. Scotland). As with other non-Standard dialect forms, we may also predict differences between speakers who use this construction exclusively and those who have both Standard and non-Standard forms, raising the possibility of cross-dialectal influence. Our approach to collection of data for this construction has the following advantages:

- Annotation of linguistic features means the questionnaire can be randomised so that each informant may be presented with different sentences comprising the same features. For instance, to test the use of definite, inanimate subjects with *want*, one speaker might be given *my hair wants cut*, while another might be asked to judge *her house wants painted*. This randomisation potentially reduces effects otherwise attributable to the semantics or pragmatics pertaining to specific lexical items.
- Randomisation also avoids repetition of overly similar sentences, while still investigating subtly different constructions; for instance, a speaker asked to give a judgment on *her house wants painted* might then be presented with a different subject-verb pair for an inanimate subject in the Standard construction, e.g. *his car wants to be washed* rather than *her house wants to be painted*.
- Incorporation of hypotheses into the system allows the researcher to specify follow-up questions in the manner of those in (2), or for the system to generate new hypotheses based on observed patterns in respondent answers.

This system can complement traditional fieldwork methods, confirming and expanding data collection. It goes beyond previous grammaticality judgment survey systems by encoding linguistic knowledge to allow researchers to better interpret implicit knowledge from survey data and to validate hypotheses.

References

Cornips, Leonie & Willy Jongenburger. 2001. Elicitation techniques in a Dutch syntactic atlas project. *Linguistics in the Netherlands 2001* (November).

Cornips, Leonie & Cecilia Poletto. 2005. On standardising syntactic elicitation techniques (part 1). *Lingua* 115(7). 939–957.

Edelstein, Elspeth. 2014. This Syntax Needs Studied. In Raffaella Zanuttini and Laurence Horn (eds.), *Micro-Syntactic Variation in North American English*. Oxford: Oxford University Press. 242–268.

Erlewine, Michael Yoshitaka & Hadas Kotek. 2016. A streamlined approach to online linguistic surveys. *Natural Language and Linguistic Theory* 34(2). Springer Science+Business Media Dordrecht. 481–495.

Gibson, Edward, Steve Piantadosi & Kristina Fedorenko. 2011. Using mechanical turk to obtain and analyze English acceptability judgments. *Linguistics and Language Compass* 5(8). 509–524.

Murray, Thomas E, Timothy C Frazer & Beth Lee Simon. 1996. Need + Past Participle in American English. *American Speech* 71(3). [Duke University Press, American Dialect Society]. 255–271.

Murray, Thomas E & Beth Lee Simon. 1999. Want + Past Participle in American English. *American Speech* 74(2). [Duke University Press, American Dialect Society]. 140–164.

Myers, James. 2007. MiniJudge: Software for small-scale experimental syntax. *International Journal of Computational Linguistics and Chinese Language Processing* 12(2). 175–194.

Schütze, Carson T. 2016. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Berlin: Language Science Press.

Schütze, Carson T & Jon Sprouse. 2013. Judgment data. In Robert J Podesva & Devyani Sharma (eds.), *Research Methods in Linguistics*, 27–50. New York: Cambridge University Press.

Soares, Ricardo, Elspeth Edelstein, Jeff Z Pan, & Adam Wyner. 2018. Knowledge Driven Intelligent Survey Systems for Linguists. In *Proceedings of the Joint International Semantic Technology Conference* (JIST2018). Cham: Springer.