A Gradient Grammar Approach to Concord Variation in Existential there+be Constructions

Robin Melnick
robin.melnick@sjsu.edu

INTRODUCTION

‘Gradient grammar’ (Bresnan & Hay 2006) here refers to statistical theories of language competence. A number of studies (Bod et al. 2003, Bresnan et al. 2005, Gries 2007, O’Connor et al. 2004, Snider & Zaenen 2006, Wasow & Arnold 2005, among others) have explored such probabilistic variation where speakers choose among grammatical options (dative and genitive alternations, complementizer optionality, and so on). English existential THERE+BE (‘ETB’, Crawford 2005) constructions, however, offer an opportunity to apply such stochastic analysis to an alternation where one of the forms is prescriptively ungrammatical: a choice between typical English number agreement and the otherwise proscribed discord that is, in fact, productive with ETBs in many dialects. Agreement-bending processes not peculiar to ETBs certainly contribute here—including agreement attraction (Bock & Miller 1991, Eberhard 1997), variant concord with collectives (Humphries and Bock 2005), dialectal variation with collectives (Bock et al. 2006), structural priming, and more—but prior ETB studies (Martinez Insua & Palacios Martinez 2003, Pietzsch 2005, Riordan 2007, and others) have also provided evidence that additional linguistic and processing elements affect ETB agreement in particular.

The present work hypothesizes that certain of these previously studied elements—including determiner type and distance between copula and head noun or determiner in the post-copular NP—can be subsumed under a broader and more significant factor that resonates with Hawkins’ principle of MINIMIZE DOMAINS (2001). Specifically, this study explores ‘cue distance’ (CD), defined here as the number of words from THERE+BE to whatever item first establishes SG vs. PL, independent of which syntactic element provides this ‘cue’ in a given utterance.

RESEARCH QUESTIONS

1. Is ‘cue distance’ (CD), as defined here, a statistically significant factor in ETB agreement?

2. Do previously observed variations in the relative influence on discord associated with different determiner types, in fact, simply correlate with the extent to which each of these determiner types varyingly contributes to increased CD?

METHOD

Using a combination of programmatic and hand coding, the Michigan Corpus of Academic Spoken English (MICASE) yielded 1563 tokens for there aren’t, there’re, there is(n’t) and there’s, after removing truncations, bare ETBs (no post-copular NP), locatives (What you find there is...’), and so on. To support the proposed analysis of the relationship between ‘cue distance’ and determiner type, the set of DET classification values included NONE, definite article, indefinite article, possessive, demonstrative, the particle ‘no’, cardinal number, and three further types of indefinite quantifiers: SG-QNT (a quantifier that requires a SG complement, e.g., ‘each’), its plural analog PL-QNT (e.g., ‘several’), and neutral-QNT (compatible with either SG or PL, e.g., ‘lots’). Finally, the parameter C-DA is the computed difference between ‘cue distance’ and determiner distance.

Prior studies have looked at distance to DET and distance to head N, but not this way they may interact to ultimately yield a minimum processing depth.

ETB discord is principally a phenomenon of PL-NP with contracted SG copula.
SUMMARY

The data suggest that a significant source of discord in ETBs is the extent to which determiners varyingly expand the minimum processing domain. This study provides evidence in support of the hypothesized significance of the notion of ‘cue distance’ and its role in promoting non-concord in English existential THERE+BE constructions. Further, ‘cue distance’ offers an explanation for the varying effects on concord/discord of different determiner types, as previously reported in multiple studies. Specifically, the data lend support to the suggestion that determiner-promoted discord stems, at least in part, from the extent to which a DET extends the distance that must be processed in order to firmly establish post-copular NP number.

The present study has not separated the potentially related effects of its multiple variables. A multivariate binary logistic regression would be a valuable next step.

SELECTED REFERENCES


ACKNOWLEDGEMENT

I am grateful to the following for their helpful input at various stages of this project: Liz Coppock, Chris Donlay, Stefan Frazier, Stefan Gries, Scott Grimm, Kevin Moore, Cathy O’Connor, John Ohala, Dan Silverman, Roula Svorou, Gregory Ward, Tom Wasow, and Annie Zaenen.