

**Conceptual effects on agreement:
A corpus study of mismatch in English copular constructions***

Cecily Jill Duffield

Department of Linguistics, University of Colorado at Boulder

1. Introduction

The phenomenon of subject-verb agreement plays a key role in both linguistic theory (e.g., Corbett, 2006) and psycholinguistic research (e.g., Eberhard, Cutting & Bock, 2005). While a vast amount of research of subject-verb agreement has been done, we still have an incomplete understanding of how speakers actually produce it. Specifically, the exact role of syntactic and conceptual information in the production of agreement is unclear. While experimental work is usually based observations of naturally occurring errors, they do not, for the most part, draw on the quantitative, distributional evidence that is now available from spoken language corpora. Relatively few psycholinguistic experiments have been based upon corpus studies of how speakers produce mismatch in spontaneous, communicative contexts (e.g., Lorimor, 2007). This study aims to address that issue by providing a large-scale corpus investigation of mismatch in English subject-verb agreement. Results suggest that speakers are more likely to produce subject-verb agreement mismatch in English copular constructions displaying a particular set of characteristics than in other sentence types, and that the tendency to produce such mismatch is the result of conceptual information available during the production process.

2. Subject-verb agreement and agreement mismatch in English

Subject-verb agreement is a relationship in which features of certain elements in the utterance are dependent upon the features of other elements in a particular grammatical domain. In English subject-verb agreement, the verb takes on person, number and/or gender features of the subject noun phrase. This study focuses on the number feature, which in English has a value of either singular or plural. Examples of singular (SG) and plural (PL) subject-verb agreement are shown in utterances (1-2) taken from the Switchboard Corpus (Godfrey, Holliman & McDaniel, 1992).

1. *One of the biggest problems-SG now is-SG, like, paralegals and stuff.*
2. *The benefits-PL were-PL just amazing.*

Because the verb usually matches the syntactic form of the subject noun phrase, it is often assumed that speakers draw primarily upon syntactic information when producing agreement (e.g., Bock & Miller, 1991:83; Bock et al., 2006:99; Vigliocco & Franck, 2001:370). Furthermore, when speakers produce agreement errors—what I will refer to as *agreement mismatch*—the source of the agreement mismatch is often syntactic. For example, the syntactic features of a noun embedded inside a subject noun phrase may induce mismatch, as in *The key-SG to the cabinets-PL are-PL on the table* (Bock & Miller, 1991).

However, there is also evidence that semantic, or conceptual, information can cause mismatch, as seen in (3-4).

* I would like to thank Bhuvana Narasimhan, Lise Menn, Laura Michaelis-Cummings, Al Kim, and Alice Healy, Gary McClelland, Susan Brown, Will Corvey, Steve Duman, David Harper, Alison Hilger, Les Sikos, and Michael Thomas, Michael Barlow, Kay Bock, Greville Corbett, Suzanne Kemmer, Colin Phillips, Michael Tanenhaus, and John Trueswell, all of whom have graciously given constructive feedback during the development of this project. All errors are of course my own.

3. *The label on the bottles-SG are-PL* peeling off.
4. *Forty minutes-PL is-SG* too long to wait in the doctor's office.

In example (3), the plural verb *are* indicates the semantic plurality of many labels on many bottles, despite the syntactic singularity of *label*. In (4), the verb *is* reflects the speaker's construal of *Forty minutes* as a chunk of time, rather than a collection of individual units. Thus, while it is clear that both syntactic and conceptual features contribute to mismatch, the exact nature of these influences has yet to be fully understood.

3. A new type of mismatch

New data from this corpus study present a pattern that has been previously unexamined in experimental work, but may contribute to the understanding of how conceptual features influence the production of agreement mismatch. These new data are exemplified by the following attested examples of subject-verb agreement mismatch (from the Switchboard Corpus (Godfrey et al., 1992)):

5. *The only thing-SG* we've taken back recently *are-PL* plants.
6. *The other one-SG* is...*are-PL* the Saturns.
7. *Repeating patterns-PL is-SG* what you have to check for when you buy your paper.

In each of these examples, the verb does not match the grammatical features of the subject noun, but neither does it seem to reflect the semantic number of the lexical items in subject (i.e., the semantic number of those items outside of the full context of the utterance). Perhaps more perplexing is that the mismatch in most tokens cannot be explained by the interference of a noun embedded in the subject noun phrase. There are, however, several features in these tokens that are of interest:

- *Semantically light subjects*. The subject head nouns are typically items such as *thing*, or *one*, which can refer to a range of items, or are lexical nouns that name general, abstract entities (e.g., *solution*, *pattern*, *rate*, *issue*).
- *Verbs number features matching those of the predicate nominal*. In the examples above, the verbs each agree with the number features of the following noun (e.g., *are...plants*).
- *Equative constructions*. The structure in which agreement mismatch occurs is the equative construction. This construction performs a particular communicative function: to equate two NPs or other constituents (Birner, Kaplan & Ward, 2007).
- *Modified subjects*. In many cases, the subject noun is modified by a relative clause or prepositional phrase, which may either affect the semantic construal of the subject referent, or contribute to interference in agreement production by increasing linear distance between the subject and verb.

Unlike cases of reference transfer (Pollard and Sag, 1994:85) and "pancake" sentences (Corbett, 2006:150), changing the verb does not give a different construal of the subject noun phrase. For example, changing the verb from *are* to *is* in (9) above, resulting in, *The only thing we've taken back recently IS plants* does not change the propositional content of the utterance.

These examples suggest that speakers produce agreement mismatch such that the verb agrees with a predicate nominal when there is a conceptual (equative) link between the subject and the predicate nominal and a number of constraints combine to promote the predicate nominal as the controller of the agreement features to be marked on the verb. When the subject and predicate nominal are in an equative relationship, but differ in terms of grammatical number, the speaker may mark the verb with the features of the predicate nominal. This seems particularly likely when the subject is semantically light and the predicate nominal relatively more concrete, and when a subject modifier adds some linear

distance between the subject head noun and the verb. To demonstrate that the presence of such features predict when a speaker is significantly more likely to produce mismatch, the following corpus study was conducted.

4. The Corpus study

The data for this study were collected from the treebanked portion of the Switchboard corpus (Godfrey et al 1992). The data set was created by collecting all tokens containing 3rd person present forms of the verbs BE (*is, are*), DO (*does, do*), and HAVE (*has, have*), along with the 3rd person past forms of BE (*was, were*) in the main clause of the sentence. The total number of tokens collected was 4964. Of these, 84 (approx. 2%) contained agreement mismatch, while 4610 (approx. 98%) did not. Each token was hand-annotated for agreement mismatch as well as for the four hallmark properties of the proposed construction: a semantically light subject head noun (limited to *thing(s)* or *one(s)*); a (prepositional phrase or relative clause) subject modifier; the presence of a copula as the main verb (as a proxy for the equative construction); and a post-verbal nominal whose number did not match that of the subject head noun.

To determine whether or not each of these properties for which the data were annotated had a significant impact on rates of agreement mismatch, a multiple logistic regression analysis was performed using SAS statistical software, in which each of the properties were treated as independent predictor variables for the likelihood of agreement mismatch (the dependent variable). In addition, an interaction between the presence of a copula main verb and the mismatching post-verbal nominal was included to see if the effect of one variable depended upon the other—in other words, to show that it is not merely the presence of a mismatching object that interferes with subject-verb agreement, but that only within the copular construction are such errors are likely to occur. Results are presented in Table 1.

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Estimate	Odds Ratio	Standard Error	Wald χ^2	Pr > ChiSq
Intercept	1	-4.8260	0.008	0.2867	283.2570	<0.0001
Light subject (<i>thing(s)</i> or <i>one(s)</i>)	1	-0.7222	0.486	0.4008	3.2479	0.0715
Subject modifier (PP or RC)	1	1.6077	4.991	0.2644	36.9684	<0.0001
Equative (Main copula verb)	1	-0.0871	0.917	0.3335	0.0682	0.7940
Mismatch post- verbal nom	1	0.2578	1.294	0.6466	0.1589	0.6902
Equative * Mismatching post-verbal nom	1	3.1549	23.451	0.7019	20.2041	<0.0001

Table 1: Results from the statistical analysis of distributional patterns seen in the corpus.

The overall model was shown to significantly predict the likelihood of agreement mismatch ($Wald \chi^2 207.886, p < 0.0001$). The presence of a semantically light subject trended toward, but did not reach, significance ($p=0.0715$). This may be due to the limited definition of “semantically light subject” as the lexical item *thing(s)* or *one(s)*, while many of the errors observed occurred with such abstract subjects as *cuts, issues, problems, solutions* or headless relative clauses (e.g., *What I want...*). Recoding the subject type may result in an observed significance of this feature in predicting agreement mismatch. The presence of a subject modifier, however, was a significant predictor of agreement mismatch, with speakers 4.99 times more likely to produce mismatch when the modifier is present ($Wald \chi^2 36.968, p < 0.0001$). There was no main effect of the copular construction ($p=0.7940$), or of the mismatching post-verbal nominal ($p=0.6902$). Thus, the presence of either feature alone has no effect on the likelihood of a speaker producing agreement mismatch. The interaction between these two features is, however, highly significant. The interaction can be interpreted as showing that when producing a copula construction and a mismatching predicate nominal, speakers are 23.45 times more likely to produce agreement mismatch than when only one of the features is present ($Wald \chi^2 20.204, p < 0.0001$). The presence of *both* features is necessary for there to be a significant increase in the likelihood of mismatch.

5. Conclusion

Previous research has demonstrated that while subject-verb agreement is primarily a syntactic phenomenon, semantic information may influence speakers to produce mismatch. This corpus study provides evidence that conceptual features play a stronger role in the production of subject-verb agreement than has previous experimental work would suggest. Constraints involving the conceptual relationship between the subject and predicate nominal, the modification of the subject, and the mismatching number of the predicate nominal combine to allow the predicate nominal features to interfere at the time when the speaker encodes agreement information on the verb. While of course needing to be investigated experimentally, this processing explanation can be supported by the statistical significance of the distributional patterns shown in this study. Such evidence is just as important as distributional patterns of mismatch elicited in an experimental setting in informing theories of agreement production, as it reflects speakers’ processing in a real-world setting.

Selected References

- Birner, B. J., Kaplan J. P., & Ward, G. (2007). Functional compositionality and the interaction of discourse constraints. *Language, 83*(2) 317-343.
- Bock, K., & Miller, C. A. (1991). Broken Agreement. *Cognitive Psychology, 23*(1), 45-93.
- Bock, K., Cutler, A., Eberhard, K. M., Butterfield, S., Cutting, J. C., & Humphreys, K. R. (2006). Number Agreement in British and American English: Disagreeing to Agree Collectively. *Language, 82*(1), 64-113.
- Corbett, G. G. (2006). *Agreement*. Cambridge, Cambridge University Press.
- Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: number agreement in sentence production. *Psychological Review, 112*(3), 531-59.
- Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone Speech Corpus for Research and Development. Proceedings of ICASSP-92 (pp. 517-520). San Francisco.
- Lorimor, H. (2007). *Conjunctions and grammatical agreement*. (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign, Urbana, Illinois.
- Pollard, C., & Sag, I. A. (1994). *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- Vigliocco, G., & Franck, J. (2001). When Sex Affects Syntax: Contextual Influences in Sentence Production. *Journal of Memory and Language, 45*(3), 368-390.