ISIS: It's Not Disfluent, but How Do We Know That?

ELIZABETH COPPOCK*, JASON BRENIER*[†], LAURA STAUM*,
    and LAURA MICHAELIS[†]
Stanford University, *University of Colorado, Boulder*[†]

1.   Introduction

One of the most robust and exceptionless rules of English grammar is that there can only be one tensed verb per clause. When an English speaker utters two tensed verbs in succession, he or she is generally not speaking fluently; the second verb can be interpreted as a repair of the first, and the first can be interpreted as a false start.

A possible exception to this rule concerns the present tense form *is*. It seems that when English speakers use *is* twice in a row in examples like the following, they are speaking perfectly fluently.

(1) But the thing *is is* that I'm naturally thin...

Although this conclusion is intuitively clear to native speakers who have encountered the phenomenon, it is challenging to rule out the possibility that these examples are mere repetition disfluencies. In this paper, our goal will be to empirically support the intuition that spoken examples of this putative exception, which we will refer to as "ISIS", following Zwicky (2002), don't "sound" disfluent, i.e., they don't have the acoustic properties of disfluencies.

The first person to publish an observation of this special property of *is* was Dwight Bolinger, in his 1987 article entitled "The remarkable double IS." He thought that the phenomenon was "not more than two or three decades old," despite an isolated example from a letter written by Charles Darwin that he cites:

(2) My excuse and reason is, is the different way all the Wedgwoods view the
        subject from what you and my sister do.

McConvell (1988) says: "As I heard more and more examples in natural conversation in different places from English speakers with widely differing dialect backgrounds, I became aware that something systematic was going on.

Once I had 'tuned in' to the phenomenon, I began hearing politicians and businessmen on TV and radio, and journalists in prepared (and presumably scripted) TV commentary doing it... when I finally received reports of high school and college students in Australia and New Zealand writing double copulas, I realised that the construction was becoming well-established in some speakers" (p. 287).

Andersen (2002) agrees that what she calls the "double copula construction" "seems to be a recent phenomenon", and argues explicitly for its legitimacy as a part of English syntax: "Although this may look like a mere spelling mistake, or in the case of speech, a hesitational feature, several facts suggest that it is neither" (p. 43). She adds, "the repeated instance of *is* is not as haphazard and random as spelling mistakes or hesitational features" (p. 45). Her first argument is that the double copula occurs not only in speech but also in writing. Examples of written tokens are in (3) and (4).

(3) "The really sad thing is," she finally said, "is that no one believed you back then, did they?" [http://journals.aol.com/delela1/Metamorphosis]
(4) And the best part is, is that whoever believes in him is his child. [http://anointedyouth.org/info/wijesus.htm]

Her second argument is that it tends to occur in particular constructions. This argument, interpreted literally, depends on its conclusion, but what she seems to mean is that the copula is "systematically" repeated after certain nouns, such as *issue* and *point* (p. 46).

We can add to Andersen's points that we often find *is*-doubling before short, easy-to-process clauses where we wouldn't expect disfluencies. Consider example (1), repeated below, which is from the Fisher corpus. The constituent that follows contains a first person pronoun (highly accessible) subject, and very little else:

(5) But the thing *is is* that I'm naturally thin...

It has been shown in work on sentence comprehension that low grammatical weight makes constituents easier to plan and process (Gibson 2002, Arnold, Wasow, Losongco, and Ginstrom 2000). We would not expect this environment to produce an unusually large number of disfluencies, so the large number of *is is* sequences in this environment requires some other explanation.

McConvell (2005) argues, as we do in this paper, that ISIS lacks the acoustic trappings of a disfluency. One of his arguments is based on prominence: the first *is* (BE1) is more prominent than the second (BE2), whereas neither of Levelt and Cutler's (1983) repair types, *marked* or *unmarked*, would be expected to have that prominence distribution. In a marked repair, BE2 tends to be more prominent than BE1; in an unmarked repair, BE1 and BE2 would be equally prominent. He also suggests that ISIS lacks some of the phonetic effects in the reparandum that Shriberg (2001) found in disfluencies.

Other authors have taken this conclusion for granted. Some analyses propose that ISIS is licensed indirectly through other constructions (McConvell 1988, Tuggy 1996, Brenier and Michaelis 2005). Others attempt to derive ISIS directly from grammatical principles, with only historical connections to other constructions (Andersen 2002, Shapiro and Haley 2002, Massam 1999). What all of these accounts have in common is that they presuppose that ISIS is a grammatical construction of English.

The only disfluency analysis of this phenomenon that we have encountered is in the Penn Treebank (Marcus *et al.* 1994), where it is annotated using the disfluency tag –DFL-, and BE1 is treated as an "edited" portion, as shown in Figure (6).

(6) Treebank analysis of ISIS

```
(TOP (S (INTJ (UH Well))
        (, ,)
        (NP-SBJ (DT the)
                (NN thing))
        (EDITED (RM (-DFL- \[))
                (VP-UNF (VBZ is))
                (, ,)
                (IP (-DFL- \+)))
        (VP (VBZ is)
            (RS (-DFL- \]))
            (SBAR-PRD (IN that)
                      (S (NP-SBJ (PRP I))
                         (VP (VBP live)
                             (PP-LOC (IN in)
                                     (NP (DT a)
                                         (NN dorm)))))))
        (. .)
        (-DFL- E_S)))
```

Perhaps we find this analysis in the Treebank only because the annotators are faced with the difficult task of giving these examples a syntactic analysis, and the disfluency analysis is simpler. Since the disfluency analysis is simpler, it is challenging to rule out.

2.   Fisher Study

The intuition that we seek to support empirically in the current study is that ISIS doesn't sound like a disfluency. Our main prediction is that *is is* sequences in the syntactic environments where ISIS is found will not exhibit the acoustic properties of disfluencies.

In order to test our prediction, we must know where to locate ISIS. While many syntactic analyses of ISIS have been proposed, there is no definitive

analysis that enumerates the environments that license a double *is*. In the literature, practically the only environment where authors present examples as ISIS is in the introduction of assertions. For example, *you're always late* is being asserted in *The problem is (is) that you're always late*. In fact, the only exception to this rule is when the following constituent is a question, as in, *The question is, is do we have enough time?* Authors never present ISIS in predicative sentences (e.g. *John is (is) happy* or *John is (is) in Paris*). Our intuitions suggest that this is not an accidental gap; we would go so far as to star the following sentence, pronounced with typical ISIS prosody:

(7) *The thing IS, is downstairs.

Nor is it found in environments where *is* functions as an auxiliary:

(8) *The thing IS, is going to fall apart.

Thus our more specific hypothesis is that *is is* sequences that introduce assertions will sound more fluent than those in other environments.

2.1.   General methods

Our sample consisted of *is is* sequences in Part 1 of the Fisher English Training Speech corpus (Cieri *et al*. 2004). This corpus consists of conversational telephone speech from a diverse set of speakers, with full conversations up to 10 minutes long. 60% of the *is is* sequences were randomly selected for coding.

We coded each *is is* sequence using Praat, with a textgrid for each token. An example textgrid is shown in (9). The waveform and spectrograph of the utterance are shown on the first two tiers, and a transcript of the utterance is shown on the third tier. The following tiers contain our hand annotations:

RATING: a subjective rating 1-7 of how fluent the example sounds, with 1 meaning "definitely a disfluency", 7 meaning "definitely not a disfluency"
LABEL: syntactic properties of the NP preceding BE1 (usually the subject of the sentence.[1]) We recorded whether this NP was headed by a *wh*-word (+/-wh), and what its syntactic function was in the surrounding clause.
COUNTERWEIGHT: the syntactic type of the largest constituent following BE2. Example values on this tier were: 'cl' (finite clause), 'np', 'ap', and 'pp'

On the lower two tiers are word and phone alignments, obtained by time-aligning the transcripts using the Sonic continuous speech recognizer.[2] This enabled us to extract phonetic properties of specific phones.
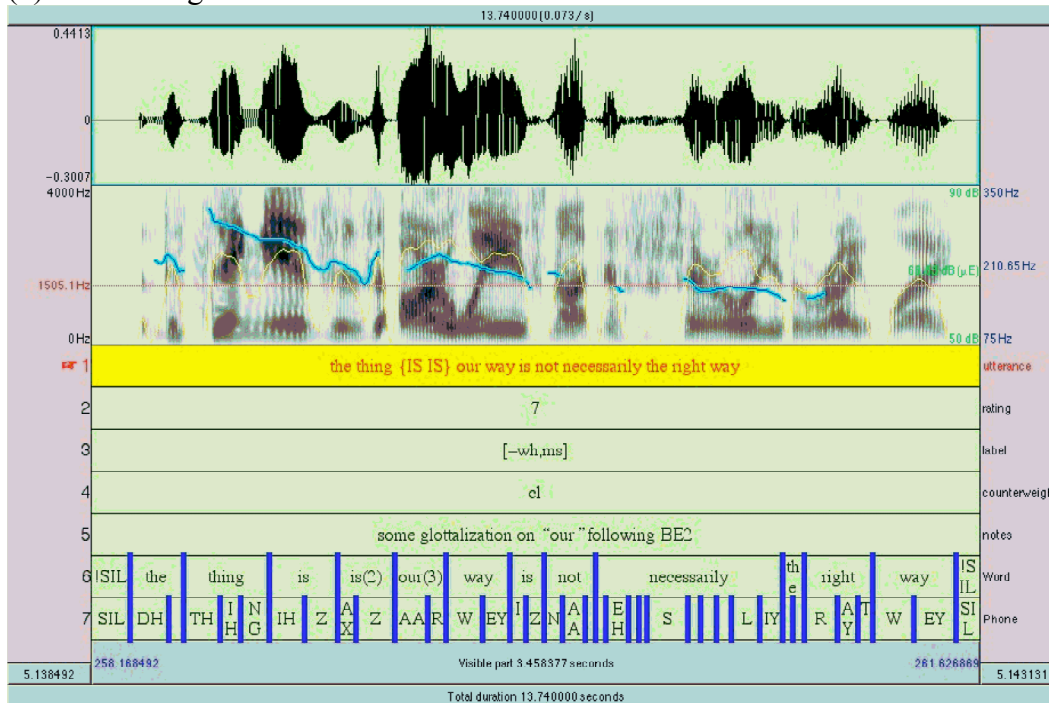
---

[1] To be precise, the Label tier contains syntactic properties of the NP preceding or including BE1; the relevant NP includes BE1 in the case of pseudoclefts whose subject NP ends in *is*, as in *[What it is], is a computer program*.
[2] These alignments were made possible by help from Bryan Pellom.

(9) Praat Textgrid

```
13.740000 [0.073/s]
0.4413

0

-0.3007
4000 Hz                                                        90 dB  350 Hz

1505.1 Hz                                                              210.65 Hz

0 Hz                                                            50 dB  75 Hz
```

1	the thing {IS IS} our way is not necessarily the right way	utterance
2	7	rating
3	[−wh,ms]	label
4	cl	counterweight
5	some glottalization on "our" following BE2	notes

Word row: SIL | the | thing | is | is(2) | our(3) | way | is | not | necessarily | the | right | way | IS IL SIL

Phone row: SIL | DH | TH I H N G | IH | Z | A X | Z | AA R | W | EY | I Z | N | A A | E H | S | L | IY | R | A T Y | W | EY | IS IL SIL

```
5.138492   258.168492        Visible part 3.458377 seconds        261.626869   5.143131
                        Total duration 13.740000 seconds
```

    In the analyses that we will present in the following sections, we only included those examples that have a subject NP directly preceding the *is is* sequence; we will call these *post-subject doubles*. Included in this category are the following examples, which do have a subject NP right before the *is is* (the subject NP is shown in brackets):

(10)   [the bad thing] is is that I smoke
(11)   [one of them] is is really uh overweight

We *excluded* examples that have inversion as in (10), along with pseudocleft examples as in (11), and examples such as (12), called "hypotactic apposition" by Brenier and Michaelis (2005):

(10) [ ]is is it spring there?
(11) [what that is] is we gotta...
(12)   that's [what smoking is] is it puts...

We also defined several functional environments, based on the syntactic category of the phrase following BE2:

ASSERTIVE: the *is is* sequence precedes a declarative clause, for example, *The problem is is that you're always late*. The Assertive environment is where we

expect to find tokens of ISIS, primarily.

AUXILIARY: the *is is sequence* functions as an auxiliary verb, as in, for example, *John is is singing*.

EQUATIVE*: the *is is* sequence equates the subject and a following referential expression. Due to a lack of foresight on our part during the coding process, this category does include some predicative sentences, because it includes all examples in which the following constituent is an NP, whether the NP is definite (in which case the example is equative) or indefinite (in which case the example would probably be predicative).

PREDICATIVE: the *is is* sequence is followed by a predicate, either an adjective phrase or a prepositional phrase.

We used these environments as levels of the independent variable in the analyses that follow, expecting the Assertive environment to pattern against the others.

2.2. Subjective fluency ratings

Indeed, when we subjectively rated each example for fluency on a 1-7 scale, and then analyzed our ratings by their environment, in all but the Assertive environment, our subjective fluency ratings tended to be quite low; histograms are shown in (13). This suggests that *is is* sequences preceding a declarative clause (in the Assertive environment) sound more fluent to the naked ear.

(13) Histogram of ratings across environments
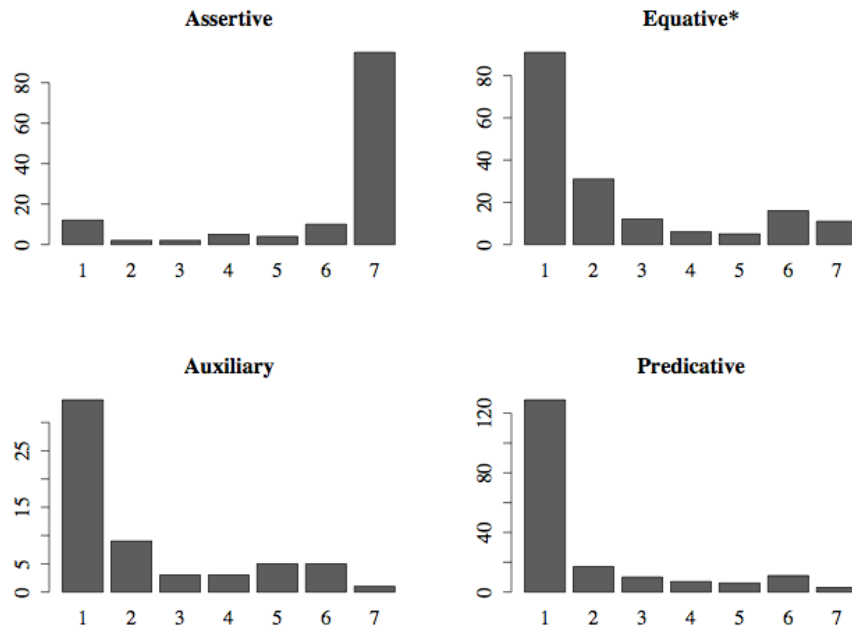
These ratings provide additional motivation for our choice of environments to compare; they suggest that the Assertive environment contains a disproportionate

number of grammatical tokens, i.e., tokens of ISIS. If it is true that sequences that sound fluent perceptually will be acoustically distinguishable from disfluent sequences, then the Assertive environment should contain a disproportionate number of *is is* sequences that have the acoustic properties of fluent speech.

2.3.   Pauses

A more objective measure of fluency is the presence of planning pauses between the repeated words. To the extent that repetition disfluencies signal difficulty in processes of linguistic production such as sentence planning and word retrieval, we expect to find speakers slowing down when they utter them, in order to give themselves time to overcome the difficulty.

Researchers on disfluencies have always found that pauses are associated with repetition disfluencies, although the location of the pause can vary. Shriberg (1995), following Hieke (1981), defines two different types of repetition disfluencies, both of which have pauses. *Prospective* repetition disfluencies are ones which anticipate an upcoming pause, and function as a way of stalling for time. *Retrospective* ones come after a pause, functioning to "smooth over the break," providing a transition to fluent speech (Dickerson 1971). In both types, pauses surround the repeated word by definition. Clark and Wasow's (1998) "Commit-and-Restore" model of disfluency production also motivates repetitions on the basis of pauses: the reparandum (e.g., BE1) makes a preliminary commitment to an upcoming constituent, and the repair (e.g. BE2) restores continuity to the delivery of the consituent, after a suspension in the flow of speech. Shriberg (2001) notes that "disfluency is often indicated by unfilled pauses in the editing phrase" as predicted by Levelt's (1989) model (p. 164).

If examples of what we believe to be the ISIS construction were really examples of repetition disfluencies, then repeated *is* before assertions should be just as likely to exhibit pauses as repeated *is* elsewhere. To the contrary, we predict that we will find fewer planning pauses where ISIS is licensed, and more pauses where it is not. We therefore investigated the number of pauses in *is is* sequences in assertive environments, and we compared these with pauses in *is is* sequences in predicative, equative, and auxiliary environments.

To our surprise, the speech recognizer that was used in time-aligning the word and phone transcripts identified very few pauses surrounding BE2 anywhere; less than 20% of the examples in any environment had a pause either preceding or following BE2. This is unexpected under the assumption that most of these *is is* sequences are disfluent, and that disfluencies contain planning pauses.

In order to determine whether these low numbers were due to errors made by the speech recognizer, we listened to some examples that we judged disfluent, in which no pauses were detected. Our first impression was that the speech recognizer was indeed failing to detect pauses, because to our ears, there *was* silence between BE1 and BE2. But upon closer inspection, we found that what we heard as silence was actually a voiceless portion of the final segment (/z/) of BE1. In other words, we heard a break in voicing between BE1 and BE2 as a pause.

With this in mind, we set out to measure breaks in voicing longer than Sonic's cutoff for identifying a pause, which is 175 milliseconds. This is below the 200 ms that are generally considered necessary for planning (Goldman-Eisler 1968; cited in Shriberg 2001). With this as the measure, *is is* sequences in the Assertive environment stood out dramatically from the others, as shown in (14).

(14)   Distribution of Breaks in Voicing > 175 ms between BE1 and BE2
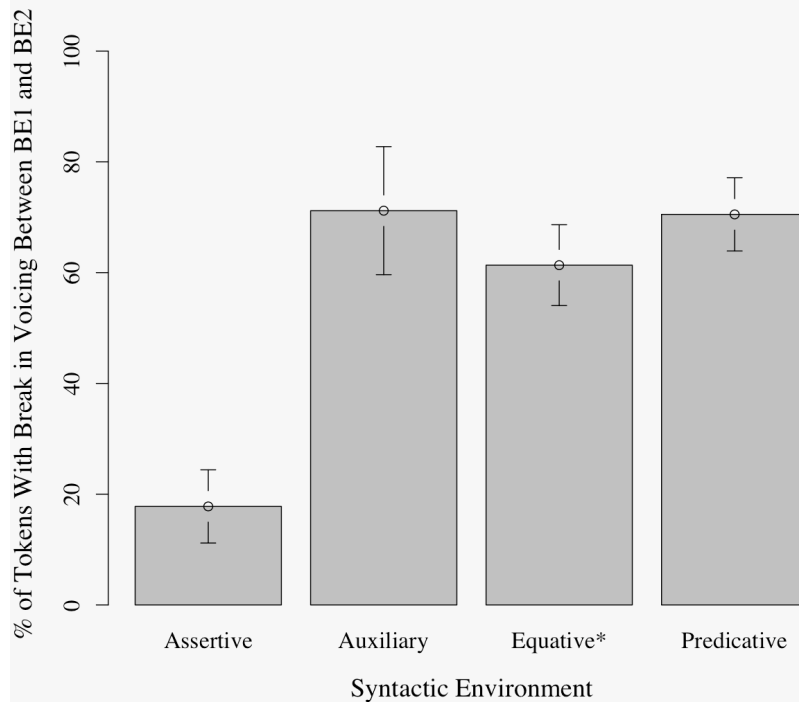
The percentage of breaks in voicing in the Assertive environment is clearly well below the number of breaks in voicing in the other environments, as we hypothesized.

This cannot be explained on the basis of differences between the environments in the grammatical complexity of the following constituent, because declarative clauses tend to be longer than the types of phrases represented in the other three categories; if grammatical complexity were playing a role, we would expect an effect in the opposite direction: more complex planning units such as assertive clauses should be more likely to be preceded by a break.

Nor can it be explained on the basis of the presence of a disproportionate number of "prospective" repetition disfluencies in the Assertive environment. Recall Hieke's (1981) categorization of repetition disfluencies into "prospective" and "retrospective": the latter have a pause between the reparandum and the repair, but the former have a pause after the repair. One could possibly imagine that there happens to be a large number of prospective-type repetition disfluencies in the Assertive environment, for some reason. This idea is not supported by the

number of pauses directly following BE2, of which there are very few in any of the environments, nor is it supported by the number of breaks in voicing greater than 175 ms. Only around 30% of the tokens in any environment have a break in voicing greater than 175 ms, and there is no significant difference between the environments with respect to this variable. When we look at the total number of pauses surrounding BE2, either before or after it, we find that the Assertive environment remains clearly distinct from the others, as shown in (15).

(15) Distribution of Breaks in Voicing > 175 ms surrounding BE2
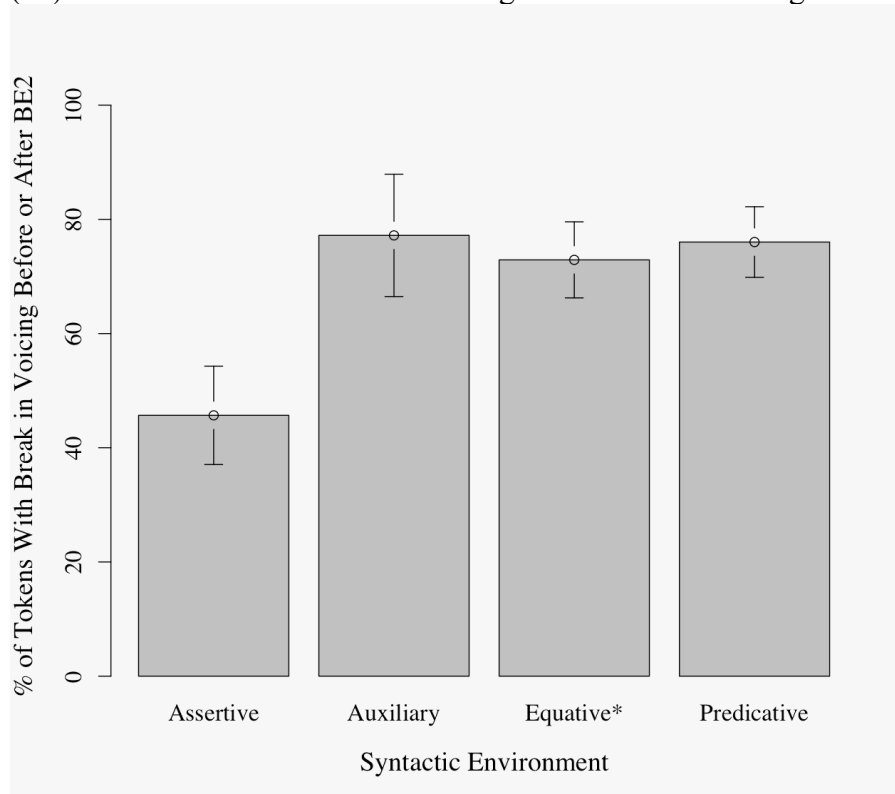
Thus, the Assertive environment contains significantly fewer breaks overall. This supports our hypothesis that *is is* sequences in the Assertive environment sound more fluent.

3.   Conclusion

This study has provided quantitative support for the observation that the ISIS construction doesn't sound like a disfluency. Sequences of double *is* that are followed by assertions have fewer breaks in voicing than double *is* sequences in other environments. This can be explained by the idea that both instances of *is* tend to be grammatically licensed in the Assertive environment, but not elsewhere. This in turn can be explained by the idea that there is such a construction as ISIS, licensed primarily in the introduction of assertions, and that it makes up a sizeable portion of the tokens in the Assertive environment.

This result confirms a subjective impression given by Andersen (2002): "According to my observations, the two copulas are generally produced in a rapid sequence, and there is usually no pause between them or any other sign of hesitation on the part of the speaker, although there may be a pause immediately following the double copula, or after the complementiser *that*" (p. 45). She goes so far as to deny that the comma typically inserted between BE1 and BE2 signals a prosodic break: "Internet users commonly insert a comma between the two tokens of *is* in the double copula construction. In fact, this is more common than not, as it happens in more than 70 per cent of the cases. However, it seems unlikely that this is done in order to represent a pause; rather, a more plausible explanation may be that the writer uses the comma as a way of preventing the erroneous interpretation that the double copula is a spelling mistake" (p. 56).

This lack of break is also interesting in light of Brenier & Michaelis's prosodic optimization theory of ISIS's origin. According to that theory, ISIS is a way of satisfying multiple constraints: that there be a prosodic break after BE1, and that the the VP be uninterrupted. If ISIS satisfies both of these constraints, then there is a prosodic break after BE1; this would lead us to expect a fair number of pauses there. The fact that we don't find them there doesn't mean that prosodic optimization wasn't one of the initial motivations for the development of this construction, but it suggests that the prosodic boundary is eroding. Bolinger (1987) suggested this with regard to the use of the comma; he writes that "the disjunction (signaled by the comma ...) has tended to disappear" (p. 39).

One of the lessons we can draw from this is that when speech recognizers do not find pauses as heard by humans, a break in voicing might be a good way to automatically capture the percept of a pause. This is a technique that could be applied more generally in order to identify disfluencies. Distinguishing ISIS tokens from disfluency tokens is a very important task for engineering applications and scientific studies that require estimating speakers' disfluency rates.

This study sets the stage for investigating quite a few remaining questions, such as the historical origin, social distribution and social meaning of this construction; the way that ISIS is related to other constructions of English (e.g. Hypotactic Apposition and Pseudocleft); and the grammatical principles from which ISIS is derived. It remains to be understood what could license two finite verbs in a row, how many arguments BE1 and BE2 are taking, and whether BE1 and BE2 are even verbs. An alternative hypothesis, which is mentioned by both Massam and McConvell, and one for which there is a certain amount of evidence, is that BE2 is a focus marker grammaticalized from the copula, hence some kind of adjunct taking no arguments. A monovalent analysis for BE1 is explored in Brenier and Michaelis (2005). The jury is still out on what is going on syntactically, but such investigations may proceed on firmer ground, given our investigations here.

Another question that we have not answered is whether ISIS is an "amalgam." The relationship between speech errors and conventionalized amalgams is one

that Michaelis (this volume) addresses, and resolving this question would constitute an important contribution to this line of inquiry. It is important to note that the question of whether ISIS examples are fluent or disfluent is different from the question of whether ISIS is a conventionalized amalgam. Non-conventionalized amalgams do not always sound disfluent; consider *How soon before midnight will they meet?* which has been discussed on LanguageLog.com, a possible amalgam between *How soon will they meet?* and *How long before midnight will they meet?* It is easy to imagine this being delivered perfectly fluently. Thus, even though we have given evidence that ISIS tokens do not sound disfluent, it is an open question whether they are conventionalized, if amalgams.

We look forward to investigating these questions in future research, and hope to understand this puzzling construction one day. To paraphrase Bob Dylan:

> *ISIS, oh, ISIS, you mystical child.*
> *What drives me to you is, is what drives me insane.*

References

Andersen, Gisle. 2002. Corpora and the double copula. In Leiv Egil Breivik and Angela Hasselgren eds., *From the COLT's mouth... and others': Language corpora studies: In honour of Anna-Brita Stenstrom*, 43—58. Amsterdam: Rodopi

Ariel, Mira. 1990. Accessing Noun Phrase antecedents. London: Routledge.

Ariel, Mira. 2001. Accessibility Theory: An Overview. In *Text representation: Linguistic and psycholinguistic aspects*, edited by T. Sanders and al. Amsterdam: Benjamins.

Brenier, Jason and Laura Michaelis. 2005. Optimization via syntactic amalgam: Syntax-prosody mismatch and copula doubling. *Corpus Linguistics and Linguistic Theory* 11:45—88. Walter de Gruyter.

Bolinger, Dwight L. 1987. The Remarkable Double *Is*. *English Today* 9:39—40.

Cieri, C., Graff, D., Kimball, O., Miller, D., and Walker, K. 2004. Fisher English Training Speech Part 1 Transcripts. Philadelphia: Linguistic Data Consortium.

Clark, H. H. and Wasow, T. 1998. Repeating words in spontaneous speech. *Cognitive Psychology* 37:201–242.

Gibson, Edward. 2002. Linguistic Complexity in Sentence Comprehension. The Encyclopedia of Cognitive Science. MacMillan.

Goldman-Eisler, F. 1968. *Psycholinguistics: Experiments in Spontaneous Speech*. New York: Academic Press.

Hieke, A. 1981. A content-processing view of hesitation phenomena. *Language and Speech* 24(2):147–160.

Jaeger, T. Florian and Thomas Wasow. 2005. Processing as a Source of

Accessibility Effects on Variation. In *Proceedings of the 31st Meeting of the Berkeley Linguistics Society*.

Jehn, Richard Douglas. 1979. That's something that I wouldn't want to account for, is a sentence like this. *Calgary WPL* 5:51— 62.

Levelt, W. and Cutler, A. 1983. Prosodic marking in speech repair. *Journal of Semantics* 2(2):205–217.

Markus, M., Kim, G., Marcinkiewicz, M.A., MacIntyre, R., Bies, A., Ferguson, M., Katz, K. and Schasberger, B. 1994. The Penn Treebank: Annotating predicate argument structure. *ARPA Human Language Technology Workshop* 114—119.

Massam, Diane. 1999. Thing is constructions: the thing is, is what's the right analysis? *English Language and Linguistics* 3(2):335— 52.

McConvell, Patrick. 1988. To *be* or double *be*? Current changes in the English copula. *Australian Journal of Linguistics* 8:287—305.

McConvell, Patrick. 2004. Catastrophic change in current English: Emergent Double-be's and Free-be's.

Ross-Hagebaum, Sebastian. 2004. "And that's my big area of interest in linguistics is discourse" — The forms and functions of the English *that's X is Y*-construction. Paper presented at BLS 2004.

Shapiro, Michael & Michael C. Haley. 2002. The reduplicative copula *is is*. *American Speech* 77(3):305—312.

Shriberg, Elizabeth. 1995. Acoustic properties of disfluent repetitions. *Proc. of the International Congress of Phonetic Sciences* .

Shriberg, Elizabeth. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31(1):153—169.

Tuggy, David. 1996. The thing is is that people talk that way. In Eugene Casad (ed.), *Cognitive linguistics in the redwoods: The expansion of a new paradigm in linguistics,* 713—52. Berlin: Mouton de Gruyter.

Zwicky, Arnold. 2002. Seeds of Variation and Change. Handout from NWAV 31.

Elizabeth Coppock, Laura Staum, Jason Brenier
Department of Linguistics
Margaret Jacks Hall, Building 460
Stanford, CA 94305

Laura Michaelis
Department of Linguistics
295 UCB
Boulder, CO 80309

{coppock,lstaum,brenier}@stanford.edu,
michaeli@colorado.edu