

# ISIS: It's not a disfluency, but how do we know that?

Jason Brenier\*<sup>†</sup>, Elizabeth Coppock<sup>†</sup>,  
Laura Michaelis\*, and Laura Staum<sup>†</sup>

\*University of Colorado at Boulder

<sup>†</sup>Stanford University

# What is ISIS?

- Linguists have observed the emergence of a nonstandard presentational construction: **ISIS**  
(Zwicky 2002, McConvell 1988, Tuggy 1996, Massam 1999).

**The thing** is is **our way is not necessarily the right way.**

**SETUP**                      **COUNTERWEIGHT**

...**the problem** is is **that every time they do this,...**



This has been assumed to be different from repetition disfluencies containing an *is is* sequence:

One of them is is really uh overweight.



# What does ISIS look like?

- Utterances identified as ISIS in the literature typically have the following syntactic/lexical characteristics:
  - Definite **subject NP** with a propositional headword, e.g., *thing, problem, question, point*.
  - Two contiguous finite forms of the copula: *BE1* and *BE2*.
  - A clausal complement.

The **thing** *is is* our way is not necessarily the right way.

# What does ISIS sound like?

- Utterances identified as ISIS in the literature typically have the following prosodic characteristics (Brenier and Michaelis 2005):
  - BE1 is typically more prosodically prominent than the subject NP, but BE2 is unaccented.
  - No intonational break between BE2 and the clausal complement.

...the problem is is that every time they do this,...



# Possible Analyses of ISIS

- Non-disfluency (constructional) analyses
  - Some analyses propose that ISIS is licensed through other constructions (e.g. as an amalgam)
  - Others attempt to derive ISIS directly from grammatical principles (only historical connections to other constructions)
- Disfluency analyses
  - Most non-standard *is is* sequences annotated in Penn Treebank as disfluencies

# Amalgam analyses

- Brenier and Michaelis 2005
  - **ISIS** is a syntactic amalgam that repairs prosodic defects of Simplex:
    - The problem is // he has to leave (Simplex)
    - The problem is // is he has to leave (ISIS).
- Tuggy 1996
  - **The 2-be construction** is multiply licensed by other constructions with similar functions and forms
- McConvell 1988
  - **Double copula** examples are a blend between the two possible syntactic structures of Simplex

# Non-amalgam analyses

- Anderson 2002
  - The string *is is that* has been reanalyzed as generalized focus construction (rather than one specific to pseudoclefts)
- Shapiro and Haley 2002
  - The second *is* is a proclitic on complementizer *that*, strengthening the meaning of identity between the constituents
- Massam 1999
  - **T-i constructions** are specificational pseudoclefts without the wh-word

# Our Question

Are we justified in ruling out a disfluency analysis  
of ISIS?




# Written Examples

- ISIS examples appear in writing (formal and informal)
  - “The really sad thing is,” she finally said, “is that no one believed you back then, did they?”  
[<http://journals.aol.com/delela1/Metamorphosis>]
  - And the best part is, is that whoever believes in him is his child.  
[<http://anointedyouth.org/info/wijesus.htm>]

# Consistent Doublers

- Some individual speakers produce lots of examples very consistently (also pointed out by McConvell 1988)
- These from one speaker over a 4-hour period:
  - *You know what's funny is, is when I was a kid, soccer was a girlie sport.*
  - *But the thing is, is they wear their emotions on their sleeve.*
  - *The thing is, is I don't smoke.*
  - *My philosophy is is that...*

# Easy Processing Contexts

- ISIS often comes before short, easy-to-process clauses that should not produce disfluencies
  - But...but the thing is is that I'm naturally thin 

# Intuition

- ISIS doesn't *sound* like a disfluency!
  - noted by McConvell 1988, 2004

# Our Study

Investigating the acoustic  
properties of ISIS

# Method

- Collected all sequences of *is is* in the Fisher corpus
- Labeled **syntactic** features of the constructions and automatically extracted **phonetic** properties of the copulas
- Correlated **ISIS-like** syntactic properties with phonetic properties and compared these to known properties of repetition disfluencies

# The corpus

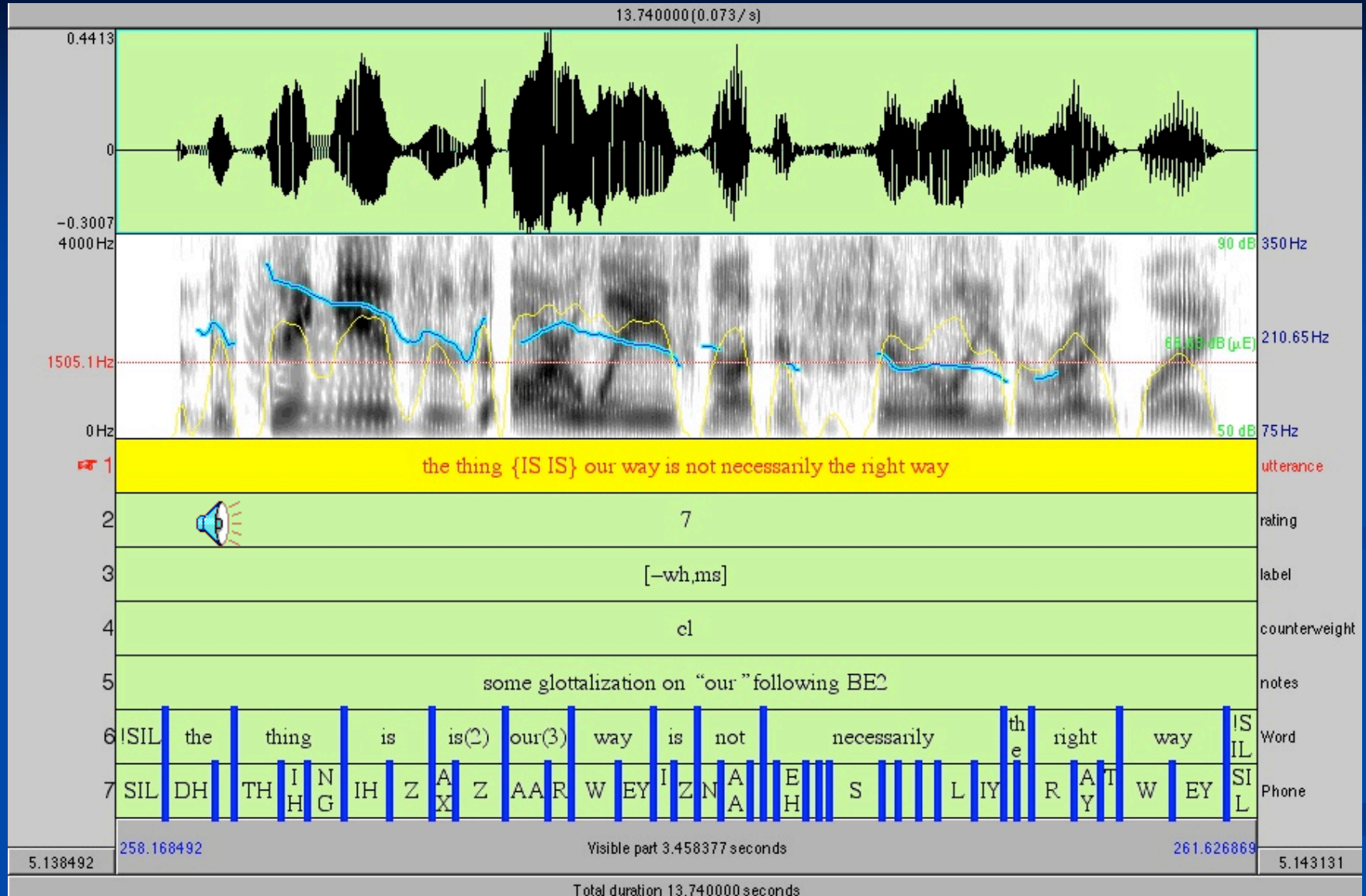
- Fisher English Training Speech Part 1
  - Corpus of telephone speech from LDC ([www ldc upenn edu](http://www ldc upenn edu), LDC2004S13)
  - Spontaneous conversational speech
  - Diverse set of speakers
  - Full conversations up to 10 minutes long
- Time-aligned transcripts (using Sonic continuous speech recognizer)
- Randomly selected 60% of *is is* sequences in Fisher Part 1 for coding

# Annotation






We coded each *is is* sequence in the corpus for:

- Syntactic function of **setup NP** (NP preceding or including BE1)
  - *matrix subject, matrix object, embedded subject, predicate nominal*
- Type of **counterweight** (constituent following BE2)
  - *finite clause, NP, wh-NP, AdjP, participle, inf. clause*





# Selecting Examples to Consider

- We are only interested in examples in which there is a **subject NP** preceding the *is is* sequence (**post-subject doubles**):
  - Including: *[the bad thing] is is that I smoke* 
  - Including: *[one of them] is is really uh overweight* 
  - Excluding: *[ ]is... is it spring there?* 
  - Excluding: *[what that is] is we gotta...* 
  - Excluding: *that's [what smoking is] is it puts...* 

# ISIS vs. Disfluencies

Among post-subject doubles, we compared 2 classes:

1. **ISIS-like** examples
  - Proposition-denoting **headwords**
    - *thing, problem, issue, reason, question, etc.*
  - Clausal (propositional) **counterweights**
    - ... is that S.
2. **Non-ISIS-like** examples
  - NEITHER feature




Examples with only ONE feature (headword or clause) were excluded from the analyses.

# ISIS-like vs. Non-ISIS-like

Features examined:

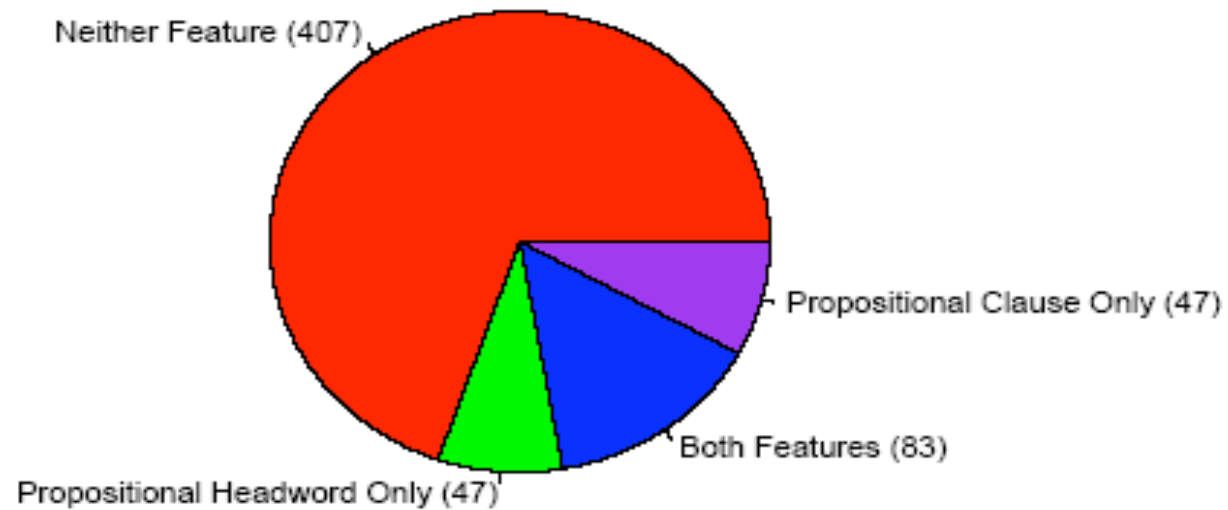
1. **Propositional headword** (*thing, point, etc.*)
2. **Clausal counterweight**

Examples:

- **ISIS-like:** [*the bad thing*] is is [*that I smoke*] 
- **Non-ISIS-like:** [*one of them*] is is [*really uh overweight*] 
- **Excluded:** [*the other point*] is is [*kind of inertia*] 

# Identifying ISIS

ISIS-like tokens in the Fisher sample



# Hypothesis 1

Prosodic and phonetic differences exist between:

- sequences with proposition-denoting headwords and clausal counterweights (**ISIS-like** sequences)
- and
- **non-ISIS-like** *is is* sequences following an NP subject (which we assume are disfluent).

# Testing Hypothesis 1

In **ISIS-like** sequences, BE1 > BE2 in:

- **mean pitch** ( $p < 0.005$ )
- **duration** ( $p < 0.001$ )
- **voicing** ( $p < 0.001$ )
- **vowel fullness** (frontness and height) ( $p < 0.001$ )
- **intensity** ( $p < 0.001$ )

In **non-ISIS-like** sequences, BE1 < BE2.

# Conclusion 1

- ☞ In our sample, *ISIS-like* examples differ significantly in their acoustic characteristics from *non-ISIS-like* examples. This confirms Hypothesis 1.



# Hypothesis 2

**ISIS-like** sequences will not have the properties of well-studied repetition disfluencies involving other high-frequency function words (e.g. *the*).

# Types of Repetition Disfluencies (Hieke, 1981)

- *Prospective*: The repair (R2) functions as a “pause device”, a way of stalling for time
  - In this case, we expect lengthening on R2, possibly followed by a pause.
- *Retrospective*: R2 functions to make a smooth transition to following material after a pause
  - In this case, there should be a pause right before R2



# Acoustic properties of repeated *the* (Shriberg 1995)

## Prospective

- := Pause after R2
- Optional pause between R1 & R2
- R1 shorter than R2
- Decrease in pitch from R1 to R2

## Retrospective

- := No pause after R2
- := Pause between R1 & R2
- R1 longer than R2
- Little change in pitch between R1 and R2



# Acoustic properties of ISIS

- Not Prospective
  - Rare pause after BE2 (2.4% of tokens have one)
  - BE1 longer than BE2 ( $p < 0.005$ )
- Not Retrospective
  - Rare pause between BE1 & BE2 (4.8% have one)
  - Decrease in pitch from BE1 to BE2 ( $p < 0.001$ )

☞ ISIS resembles neither disfluency type

# Conclusion 2

- ☞ Sequences in propositional environments differ prosodically from both prospective and retrospective repetition disfluencies. This confirms Hypothesis 2.

# Main Conclusions

- ISIS-like *is is* sequences (those in propositional environments) have a reliably distinctive prosodic pattern.
- Moreover, this prosodic pattern is unlike that of any known disfluency.

# Implications

- ☞ Acoustics can be used to investigate grammaticality status.
- ☞ Suprasegmental cues could be used in speech applications for determining whether ambiguous strings are disfluent or not.
- ☞ Ruling out a disfluency analysis of ISIS is a prerequisite for developing a syntactic analysis of this puzzling and unique construction.

# Remaining Questions

- What is the historical origin of ISIS?
- What is ISIS's social distribution and meaning?
- How is ISIS related to other constructions of English (e.g. Hyp. App., Pseudocleft)?
- What are the grammatical principles from which ISIS is derived?
  - What licenses two finite verbs in a row?
  - How many arguments do BE1 and BE2 take?
  - Are BE1 and BE2 even verbs?



**Thank you!**

# Acknowledgements

- Thanks to Arnold Zwicky for inspiration, Patrick McConvell for discussion, and Bryan Pellom for generous help with aligning the Fisher corpus.

# References

- Andersen, Gisle. 2002. Corpora and the double copula. In Leiv Egil Breivik & Angela Hasselgren (eds.), *From the COLT's mouth... and others': Language corpora studies: In honour of Anna-Brita Stenstrom* (Amsterdam: Rodopi), 43—58.
- Brenier, Jason and Laura Michaelis. 2005. Optimization via syntactic amalgam: Syntax-prosody mismatch and copula doubling. *Corpus Linguistics and Linguistic Theory* 1-1, 45-88. Walter de Gruyter.
- Bolinger, Dwight L. 1987. The remarkable double *Is*. *English Today* 9.39—40.
- Clark, H. H. and Wasow, T. 1998. Repeating words in spontaneous speech. *Cognitive Psychology*, 37:201—242.
- Hieke, A. 1981. A content-processing view of hesitation phenomena. *Language and Speech*, 24(2):147—160.
- 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.
- 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. [Powerpoint presentation](#) (with sound files!) or [PDF](#).
- 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, E. 1995. Acoustic properties of disfluent repetitions. *Proc. of the International Congress of Phonetic Sciences*.
- 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* (Berlin: Mouton de Gruyter), 713—52.
- Zwicky, Arnold. 2002. [Seeds of Variation and Change](#). Handout from NWAV 31.