

Direct compositionality in English why-stripping: A corpus-based approach

Jong-Bok Kim & Joanna Nykiel
jongbok@khu.ac.kr, joanna.nykiel@us.edu.pl

Kyung Hee U., Seoul & U. of Silesia

Corpus-based and Experimental Approaches to Ellipsis
The University of Kentucky
July 29, 2017

Striping and *Why*-stripping

- Why-stripping in English consists of the adverbial *why* with a focal-stressed non-wh remnant. The remnant can have either a correlate or no overt correlate in the antecedent clause (contra Yoshida et al. 2015)
 - 'Daddy, will you buy me a violin?' Sonya said. 'Why **violin**?' (COCA 1995 FIC)
 - "Good morning to you, too," she says, and turns to face you. "I know what you're doing," you say, and she smiles at this. "I'm making you breakfast. Why **so suspicious**?" (COCA 2015 FIC)

Key research question

- How do we account for the semantically propositional character of what appears to be syntactically less than sentential structures?

Two main approaches to clausal ellipsis

- Movement and PF-deletion approach (Ross 1969, Merchant 2004, Yoshida et al. 2015):
 - An ellipsis site has internally structured material through the derivation, and PF deletion renders some of it unpronounced. The meaning composition is dependent upon the derivational source.
- Direct Interpretation approach (Ginzburg and Sag 2000, Culicover and Jackendoff 2005, Jacobson 2016):
 - There is no syntactic structure at the ellipsis site other than the remnant. The interpretation depends on the surrounding discourse.

Movement and deletion approach

- Yoshida et al.'s (2015) example
 - (2) A: John was eating natto.
B: Why NATTO (and not another food)?
- Ortega-Santos et al. (2014) and Yoshida et al. (2015) argue that Why-stripping in English involves a base generation of *why* in the Spec of CP with movement of the focused remnant to the specifier position of a Focus projection, and clausal ellipsis of the remaining parts:
 - (3) $[_{CP1} \text{Why} [_{CP2} \text{NATTO} [_{TP} \text{he was eating } t_{\text{natto}}]]]]?$

Connectivity effects: Preposition selection

- The movement plus ellipsis account is mainly motivated by connectivity effects (Yoshida et al. 2015: 331–337)
- The preposition is selected by the matrix verb
 - (4) Matrix *why*-stripping
 - A: John relies on Mary.
 - B: Why ON/*OF MARY (but not others)?
 - (5) Embedded *why*-stripping
 - John relies on Mary, but I don't understand why ON/*OF MARY.

Connectivity effects: Preposition omission and stranding

- A PP remnant can optionally omit the preposition in both matrix and embedded *why*-stripping
 - (6) Matrix *why*-stripping

A: John was talking to Mary.
B: Why (to) Mary?
 - (7) Embedded *why*-stripping

John was talking to Mary, but I don't understand why (to) Mary.
- This could follow from the P-stranding generalization (Merchant 2001, 2004), since either the focused PP *to Mary* or the NP *Mary* (stranding *to*) can undergo movement followed by ellipsis of the remaining clause (but see Levin 1982 for counter-evidence from English)

Connectivity effects: Voice mismatch

- Unlike VP-ellipsis, sluicing disallows voice mismatch:
 - (8) a. The janitor must remove the trash whenever it is apparent that it should be [removed]. (VP-ellipsis)
 - b. Someone must remove the trash, and it is apparent *who by/*by who [~~the trash must be removed~~]. (Sluicing)
- Why-stripping also disallows voice mismatch.
 - (9) A: Max brought the roses. B: Why Max?
 - (10) A: Max brought the roses. B: *Why by Max?
 - (11) *Max brought the roses, but I don't understand why by Max.

Connectivity effects: Binding effects

- The R-expression *John* in the remnant can be coindexed with the subject of the antecedent clause under the same conditions that its counterpart embedded in a full clause could be

- (12) a. A: He_i is selling all of these pictures. B: *Why (even) PICTURES OF JOHN_i?
- b. A: His_i mother is selling all of these pictures. B: Why PICTURES OF JOHN_i?
- (13) a. *He_i is selling pictures of John_i.
- b. His_i mother is selling pictures of John_i.

Online corpora used in the research

- COCA (Corpus of Contemporary American English): 520 million words of text with equally divided among spoken, fiction, popular magazines, newspapers, and academic texts. It includes 20 million words each year from 1990-2015

Search methods

- search methods: simple string searches with some regular expressions (only matrix clause why-stripping examples with the period)

(14) a. `why * ?`: 1285 tokens

b. `why * * ?`: 1150 tokens (but removed 888 instances of `why v * ?` for examples like *why should I?*)

c. `why * * * ?`: not used but instead searched `why XP ?`

- tried to use more complex searches when necessary:

(15) `why [nn*]|[v*]|[j*]|[r*] ?, why det|art * ?`

Island insensitivity

- If *why*-stripping involves movement plus deletion, then we expect it to be subject to constraints on movement. Instead, island insensitivity is a typical property (Yoshida et al. 2015 also notes this)

- (16) a. Well, I think what's happening is both risky and immoral.
Why **immoral**? (COCA 2009 SPOK)
- b. A pack of lean dogs, all different colors, loped across the street far ahead of her. Why **dogs**? (COCA 1992 FIC)

Lack of case-matching effects

- If the non-wh remnant is repeated from the antecedent (Yoshida et al. 2015), then case-matching effects are expected. But this expectation is unsupported
 - (17) Of course this is a war against Islam. They say that Saddam is a dictator, but why **him**? (COCA 2003 SPOK)
 - (18) A: You, the boy, Carrier, and two carpenters. Go in, negotiate a purchase. Tell them we've no interest in them beyond buying what we need.
 B: But why **us**? (COCA 2004 FIC)

Non-identity with the correlate: Preposition mismatch

- The movement-ellipsis account predicts the preposition identity between the remnant and its correlate (Yoshida et al. 2015), but examples violating this morphosyntactic identity constraint can be observed.

- (19) a. John relies on Mary, but I don't understand why/how come **on/*of Mary**.
- b. Robin's body had been found at the golf course just a half mile from her home under some leaves. Why **on the golf course**? (COCA 2005 SPOK)

An implicit correlate

- The remnant has no overt correlate, but is linked to one of the linguistic expressions in the context.
(20) a. I'm not into all that X-Files bullshit... but it was a psychic connection. Why **to me**? (COCA 2002 FIC)
b. It asks the questions. What has happened? Why has it happened? Why **to me**? (COCA 2010 ACAD)

How to account for this sprouting why-stripping?

- The movement-deletion approach might be able to account for such cases, but our corpus search show us that in many authentic uses, the remnant does not have its overt correlate.
- The correlate is cued by many different ways: syntactic, semantic, and contextual information.
- In many cases, context clues are the key rather than the syntactic or semantic (mutual entailment) identity condition (Merchant 2002).

Relationship with antecedent

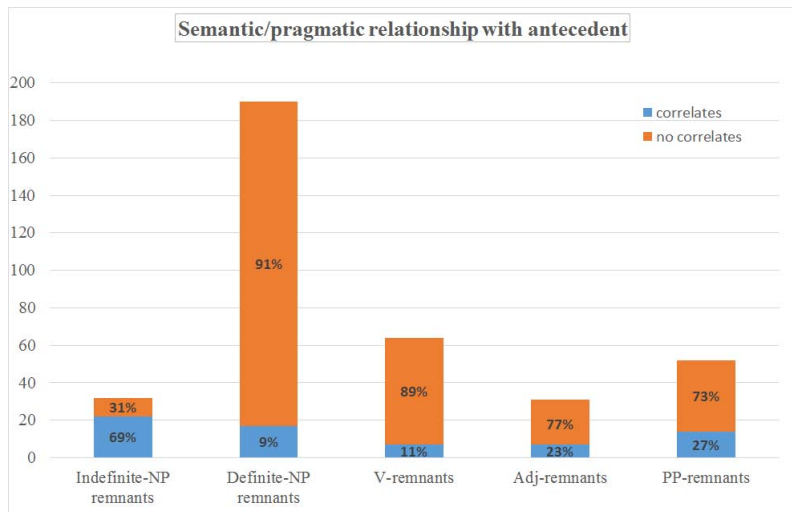


Figure: Relationship between remnant and antecedent

Indefinite NP remnants

- Indefinite-NP remnants: 10 out of 32 (31.2%) have no (exactly matching) correlate:

- (21) a. Derrick 'Starfire' Dukes, who'd had a few professional boxing matches, but made his living mostly as a professional wrestler. **Why a wrestler?** (COCA 1994 SPOK)
- b. Ms-GAMMON: Jason was a member of the DARE program. He didn't do drugs and he didn't normally drink. This particular day he did, though. RIVERA: Any idea why he would do that, though? **Why a gun?** (COCA 1996 SPOK)

Definite NP remnants

- For definite-NP remnants 173 out of 190 (91%) have no correlates. 17 have a correlate, and of these 6 (35.3%) show syntactic category mismatch.

- (22) a. Her pastor was leaving town. And soon. She frowned.
Why the rush? (COCA 2011 FIC)
- b. We were surprised at how healthy the reef was, Smith says. Why **the surprise?** (COCA 2009 MAG)

V-Remnants

- V-remnants: 57 out of 64 (89%) have no correlates

(23) a. Heaven forbid the voters should elect anyone else.

They've always voted for a Magetry . Why **change**?
(COCA 2002 FIC)

b. I want to turn it around now . Why **wait**? (COCA 2004
SPOK)

AP remnants

- With the single Adj, the remnant typically has a correlate, but with the AP remnant, more than 70% have no correlate.
- (24) a. In Italy, Lieutenant Robert Martin's hand-me down P-51 Mustang, 'Queen Cole,' got a fresh coat of red paint on the tail section. **Why red?**
- b. The Sterling Tigers begin the season ranked No. 1 despite finishing 12-16 last year . Why **so high?** (2003 NEWS)

PP remnants

- With the PP remnants, of total 52 tokens, 38 (73%) have no correlates
 - (25) a. "Maybe they wish to sponsor your spaceship-" "Why should they do that? Why now? Why **in person**? (COCA 2000 FIC)
 - b. This is her first introduction to the American people, right? Why now? Why **in this manner**? (COCA 2011 SPOK)

No category identity with the correlate

- The remnant and the correlate can mismatch in terms of syntactic category
- (26) a. What are these white men so angry about? ... Why **the anger**? (COCA 2010 SPOK)
- b. But you changed the policy, saying that a gun used once in a crime should be destroyed, not recycled. Why **the change**?

Cued from synonymous words

- The remnant and the correlate are semantically synonymous

(27) a. Well, why are tomatoes so complex? ... the tomatoes have all these different things about them. Why **so complicated**?

b. We've got to get over to your apartment fast. Why **the hurry**? (COCA 2008 FIC)
- Semantically related: subset/superset relationships

(28) And as I understand it, Nicole, that's what drew you into the movement when you were 9 years old. Why **at that age**? (COCA 1999 SPOK)

Cued from pragmatic situations

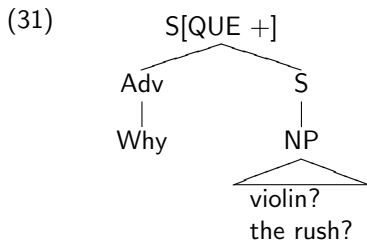
- The remnant can have a semantic/pragmatic relationship with the antecedent, with no clear correlate (as is typical of Bare Argument Ellipsis, see Culicover and Jackendoff 2005)
- (29) a. In fact, they reviewed and approved our press release. So we're a bit scratching our heads to figure out why **this**. (COCA 2004 FIC)
- b. But to actually go overseas and, you know, it's an investment of time. There is safety issues as well. Why **that**?
- (30) a. She raised her fist and punched the air. "Why **the anger**?" (COCA 2004 FIC)
- b. So I let him babble on about strange happenings and dramatic rescues. Why **argue**? (COCA 2008 FIC)

Summary

Authentic uses of the why-stripping construction tell us that there need not be a (morpho-)syntactically identical correlate for the remnant. In many cases, the remnant has no linguistically overt correlate. It is cued from the context.

A Direct Interpretation approach

- The complete syntax of the fragmental remnant in Why-stripping is just the categorial phrase projection of the fragment itself (Ginzburg and Sag 2000, Culicover and Jackendoff 2005, Sag and Nykiel 2011, Kim 2015, Jacobson 2016, among others)
- The focus marking expression *why* combines with an S projected from any NSU remnant in Why-stripping as long as it functions as a SAL-UTT (Salient Utterance or Focus Establishing Constituent).



Constructional licensing the fragment utterance

- There is no syntactic structure at the ellipsis site and fragments can serve as the sole daughter of an S-node, directly generated from the constructional constraints.

(32) Head-Fragment Construction

$$\left[\begin{array}{cc} \text{SYN} & \text{S} \\ \text{DGB} & \left[\text{SAL-UTT} \left[\begin{array}{cc} \text{SYN} & [\text{CAT } \boxed{1}] \\ \text{SEM} & [\text{INDEX } i] \end{array} \right] \right] \end{array} \right] \Rightarrow \left[\begin{array}{cc} \text{SYN} & [\text{CAT } \boxed{1}] \\ \text{SEM} & [\text{INDEX } i] \end{array} \right]$$

DGB and SAL-UTT

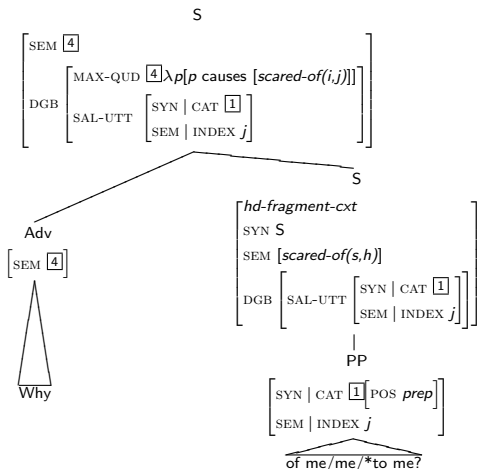
- The construction allows a fragment to be projected to a sentential expression while it is functioning as the SAL-UTT (salient utterance) in the discourse
- DGB (Dialogue-Game-Board): records where the contextual parameters are anchored and where there is a record of who said what to whom, and what/who they were referring to (Ginzburg 2012)

Why and QUD

- Typical examples: overt matching correlate
(33) I finally said, "You. I'm scared of you." Why of me? (COCA 2011 FIC)
- As stated, uttering the Why-stripping construction in the dialogue introduces the information about QUD (Question-Under-Discussion) as well as SAL-UTT.
- The QUD evoked is that the speaker is scared of the hearer, and the hearer is asking the reason for this.

A generated structure for (34)

(34)



Connectivity effects: Case matching and mismatching

- The Head-Fragment Construction ensures the `CAT` and `INDEX` values of the fragment are identical with those of the `SAL-UTT`. This salient information is related to an element in the `MAX-QUD`.
- This allows flexibility in the case matching between the remnant and its correlate. The `CASE` value is determined constructionally (Kim 2016), which allows the case mismatch between the remnant and its correlate (data repeated here). This predication is different from movement-deletion approaches (Yoshida et al. 2015).

(35) Of course this is a war against Islam. They say that Saddam is a dictator, but why **him**? (COCA 2003 SPOK)

Island repair

- Island repair is problematic for the deletion-based approach, and *why*-stripping, like other clausal ellipses, can repair island violations (Yoshida et al. 2015 assumes a PF deletion for violating examples, à la Merchant 2001).
- By contrast, our DI approach avoids this problem: remnants are directly generated, and no island-sensitive operations are thus involved. Island sensitivity, if observed, is likely due to processing difficulty (Culicover and Jackendoff 2005, Goldberg 2013, Schmeh et al. 2015)

With an implicit correlate

- The remnant has no overt correlate, but is linked to one of the linguistic expressions in the context (data repeated)
- (36) a. I'm not into all that X-Files bullshit... but it was a psychic connection. Why **to me**? (COCA 2000 FIC)
- b. It asks the questions. What has happened? Why has it happened? Why **to me**? (COCA 2010 ACAD)

context updating analysis

- two major types of null complements in English, definite and indefinite null complements (Ruppenhofer and Michaelis 2014):

- (37) a. John loves to read [e]. (indefinite null instantiation (INI))
b. We arrived [e] at 8 pm. (definite null instantiation (DNI))

Null instantiation

- Incorporating this idea within the type feature system (where types are in italics), we can introduce two signs *overt* and *ini*, the latter of which can be resolved to a covert argument or an instance of NI (null instantiation).

(38) Lexical entry for *happen*:

$$\left[\begin{array}{l} \text{FORM } \langle \textit{happen} \rangle \\ \text{ARG-ST } \langle \text{NP}_{x'} (\text{PP}_y) \rangle \\ \text{SYN } \left[\begin{array}{l} \text{SUBJ } \langle \text{NP}[\textit{overt}] \rangle \\ \text{COMPS } \langle \text{PP} \left[\begin{array}{l} \textit{ni} \\ \text{PFORM } \textit{to} \end{array} \right] \rangle \end{array} \right] \\ \text{SEM } \textit{happen}(x, y) \end{array} \right]$$

Activation

- Uttering the sentence *Why has it happened? Why to me?* would then update the DGB as following, triggered from the verb *happen*

$$(39) \left[\begin{array}{c} \text{DGB} \\ \text{SAL-UTT} \\ \text{SEM} \end{array} \left[\begin{array}{c} \text{MAX-QUD } \boxed{4} \lambda p [p \text{ causes } [\textit{happen-to}(i,j)]] \\ \text{SYN PP} \left[\begin{array}{c} \textit{ni} \\ \text{PFORM } \textit{to} \\ \text{INDEX } j \end{array} \right] \\ \text{INDEX } j \end{array} \right] \right]$$

- The question remains what ensures the presence of the preposition in the sprouting *why*-stripping (e.g., *Why to Mary* but not **Why Mary?*)

A processing account

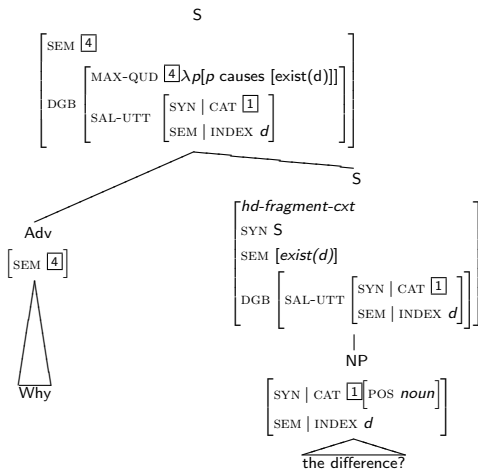
- Full Instantiation Constraint (FIC): The syntactic information (e.g., case features) not available at surface but updated in the DGB needs to be fully specified in the subsequent syntax (Kim 2015).
- In the course of processing the antecedent, the lexical entry with the implicit argument is activated in memory. By virtue of this activation, the semantic and syntactic role of *to me* became available. In this sense, indirect licensing make use of an antecedent S as well as the lexical information of *happen* (for a similar idea, see Culicover and Jackendoff 2005:260).

Correlate cued by the context

- Example:
 - (40) Japan came in first at 61 mbps; the average speed for all 30 nations considered was 9 mbps. Why the difference? (COCA 2008 MAG)
- The MAX-QUD introduced here is from the context: there is no linguistic correlate here. The QUD could vary but it is linked to the question *why the difference*.
 - (41) a. ... why the difference exists
 - b. ... why the difference occurs between the two
 - c. ... why the difference happens
 - d.

A generated structure

(42)



Concluding remarks

- The DI approach we adopt here introduces no additional syntax: fragments are mapped into non-sentential utterances and induce sentential interpretations from the enriched discourse.
- Why-stripping is simply the projection of combining the focus marking *why* with such a fragment.
- With a system that represents clear discourse structures with the information about salient utterances and QUD, we have straightforward mapping relations from fragments (*why*-stripping) to propositional meaning.

Selected references

- Culicover, Peter and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford University Press, Oxford.
- Ginzburg, Jonathan. 2012. *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.
- Ginzburg, Jonathan and Ivan A. Sag. 2000. *Interrogative investigations*. Stanford: CSLI Publications.
- Jacobson, Pauline. 2016. The short answer: implications for direct compositionality (and vice versa). *Language*, 92:2: 331-375.
- Kim, Jong-Bok. 2016. *The Syntactic Structures of Korean: A Construction Grammar Perspective*. Cambridge University Press.
- Kim, Jong-Bok. 2015. Syntactic and semantic identity in Korean sluicing: A direct interpretation approach. *Lingua* 66, 260–293.
- Merchant, Jason. 2004. Fragments and ellipsis. *Linguistics and Philosophy* 27, 661–738.
- Nykiel, Joanna. 2017. Preposition stranding and ellipsis alternation. *English Language and Linguistics*. 21(1): 27–45.

Selected references

- Ortega-Santos I, Yoshida M, Nakao C. 2014. On ellipsis structures involving a wh-remnant and a non-wh-remnant simultaneously. *Lingua*. 138: 55-85.
- Phillips, Colin and Dan Parker. 2014. The psycholinguistics of ellipsis. *Lingua* 151, 78–95
- Ross, John Robert. 1969. Guess who? In Robert I. Binnick, Alice Davison, Georgia M. Green, and Jerry L. Morgan (eds.), *Proceedings of the Fifth annual meeting of the Chicago Linguistics Society*, 252–286. Chicago, Illinois.
- Ruppenhofer, Josef and Laura A. Michaelis. 2014. Frames and the interpretation of omitted arguments in English. In Katz Bourns, Stacey and Lindsay L. Myers (eds.), *Perspectives on Linguistic Structure and Context: Studies in honor of Knud Lambrecht*, 57–86. John Benjamins.
- Yoshida, Masaya, Chinzuru Nakao, and Ivan Ortega-Santos. 2015. The syntax of *Why*-stripping. *Natural Language and Linguistic Theory*. 33(1): 323-370.