On pronominalization and ellipsis in clausal idioms*

Nick Huang & Gesoel Mendes

University of Maryland, College Park

1. Introduction

Idiom chunks can participate in anaphoric relations, as many have observed (Nunberg, Sag, and Wasow 1994, Bruening 2015, Rottman and Yoshida 2013 etc.). However, to the best of our knowledge, the theoretical implications of this fact have not been properly appreciated.

In this paper, building on existing observations, we present a novel paradigm that shows that pronominalization and VP ellipsis in clausal idioms are not as free as some would expect. Namely, pronoun subjects and VP ellipsis appear to have to co-occur to preserve the interpretation of a clausal idiom.

We argue that this paradigm lends new support to (i) the idea that pronouns can be derived from full noun phrases (e.g. Elbourne 2001, Postal 1966), (ii) an identity condition on ellipsis that takes into account the content of a syntactic chunk larger than the elided constituent (e.g. Gengel 2007, *pace* Merchant 2001), and (iii) the idea that idiomatic interpretation is sensitive to pragmatic inferences about remnants of ellipsis.

This paper is organized as follows. In Section 2, we present the co-occurrence restriction, constrasting it with pronominalization and VP ellipsis in non-idiomatic contexts. In Section 3, we discuss implications for theories of pronouns and ellipsis and give a pragmatic analysis of the co-occurrence restriction. We conclude in Section 4.

2. A co-occurrence restriction for pronoun subjects and VP ellipsis in clausal idioms

The dialog in (1), where Ben denies Alex's claim, shows that in general, pronominalized subjects and VP ellipsis (VPE) can freely occur, when antecedents are available.

*Special thanks to Howard Lasnik, Alexander Williams, Benjamin Bruening, Jason Merchant, Dan Goodhue, Maria Polinsky, Kyle Johnson, members of the UMD S-Lab community, and NELS participants and reviewers for comments. This material is based upon work supported by the National Science Foundation under Grant No. 1449815 and CAPES 0652-14-8.

(1) Alex: The green balloon hit the ceiling.
Ben: No, . . .

a.	. . . the green balloon didn't hit the ceiling.	[−Pron., −VPE]
b.	. . . it didn't.	[+Pron., +VPE]
c.	. . . it didn't hit the ceiling. (It hit the lamp.)	[+Pron., −VPE]
d.	. . . The green balloon didn't. (The blue one did.)	[−Pron., +VPE]

The examples in (2) show that this flexibility disappears in the context of clausal idioms. That Ben's response in (2a) has an idiomatic interpretation is not surprising, since the whole idiom is present. Curiously, (2b) shows that neither a subject pronoun nor VPE cancels the idiomatic interpretation, even though no part of the idiom is present on the surface. Even more interestingly, in (2c) and (2d), when parts of the idiom are repeated, the idiomatic interpretation is harder to get.

(2) Alex: When the news got out, the shit hit the fan.
Ben: No, ...

a.	... the shit didn't hit the fan.	[−Pron., −VPE]
b.	... it didn't.	[+Pron., +VPE]
c.	... #it didn't hit the fan.	[+Pron., −VPE]
d.	... #the shit didn't.	[−Pron., +VPE]

Other clausal idioms, like *the cat is out of the bag*, show the same restriction. The restriction can be stated as the generalization in (3).

(3) **Generalization:** in dialog environments, the idiomatic interpretation of clausal idioms is best preserved when the whole idiom is repeated or when it is entirely omitted under anaphoric relations.

3. Implications for theories of pronouns and ellipsis

As pointed out in the introduction, it is not novel to observe that a sentence like (2b) retains an idiomatic interpretation even though no part of the idiom is present on the surface. This paper's contribution lies in addressing *how* the idiomatic interpretation comes to be available and what implications this fact has for our theories for pronouns and ellipsis.

3.1 Assumptions about idioms

To account for the fact that the idiomatic interpretation is available in (2b), we adopt the idea that the entire idiom is present in the syntax (Katz and Postal 1963, Fraser 1970, among many others). In this view, an idiom is a pairing of special meaning and a full fledged syntactic structure with lexical items like what is illustrated in (4).

(4) $[_{vP}$ the shit $[_{VP}$ hit the fan$]]$

This analysis explains why synonym replacement, e.g. *the excrement reached the ventilator* instead of *the shit hit the fan*, results in the loss of an idiomatic interpretation. One might therefore expect pronominalization and VPE, which also affect the realization of lexical items, to have a similar impact. Seen in this light, it becomes puzzling why the generalization in (3) should hold.

3.2 Hidden syntactic material in VP ellipsis and anaphoric pronouns

The nature of anaphoric pronouns and VP ellipsis has been a point of debate since the early days of generative grammar. The point of contention that is most relevant to us is whether VPE and anaphoric pronouns are derived from full-fledged constituents with internal syntactic structure or are base-generated anaphoric elements that are interpreted mainly via discursive devices (or surface vs. deep anaphora, to use Hankamer and Sag's (1976) typology).

(5) *Approaches to anaphoric pronouns*

a. **Derivational approaches:** anaphoric pronouns are remnants of partially unpronounced NPs/DPs with regular syntactic structure (e.g. Postal 1966, Elbourne 2001);[1]

b. **Interpretive/base-generation approaches:** anaphoric pronouns are base generated elements whose interpretation is established in the discourse and constrained by syntax/semantics/pragmatics (e.g. Jackendoff 1969, Bresnan 1970, Lasnik 1976, Heim 1982)

(6) *Approaches to VP ellipsis*

a. **Derivational approaches:** elided VPs are unpronounced VPs with regular syntactic structure (e.g. Chomsky 1955, Ross 1967, Merchant 2001)

b. **Interpretive/base-generation approaches:** elided VPs are phonologically null pro-forms (e.g. Bach and Partee 1980, Hardt 1993, Lobeck 1995).

The fact that the idiomatic interpretation can be preserved under pronominalization and VP ellipsis can be straightforwardly captured by derivational analyses, in which the idiom is present in earlier steps of the derivation. It is not immediately clear how an interpretive analysis can capture these facts (but see Bruening 2015 and section 3.3.1).

More specifically, we adopt Elbourne's (2001) analysis, in which pronouns can be derived via NP ellipsis. In (2b)/(7), the pronoun *it* is derived from *the shit*. Because the idiom is present at some point in the derivation (7b), the idiomatic interpretation is preserved.

[1]Derivational approaches also include analyses where pronouns are derived from full-fledged NPs/DPs via substitution (e.g. Chomsky 1955, Lees and Klima 1963, Hornstein 2007).

(7) a. Alex: When the news got out, the shit hit the fan.
Ben: No, it didn't. [+Pron., +VPE] (=(2b))

b. [$_{DP}$ the ~~shit~~] didn't [$_{VP}$ ~~hit the fan~~]. = *it didn't*. (strikethrough = elided)

3.3 Ruling out alternative analyses for the pronoun subject

3.3.1 Base generation

An alternative analysis is to claim that the syntax of the pronoun subject are identical to those of regular (unbound) pronouns. Specifically, in a sentence like (2b), the pronoun is base-generated, not derived.

However, if the pronoun were base-generated, the idiom would be absent in the syntax: there is no instance of *the shit*. All else being equal, there should not be an idiomatic interpretation.

One solution is to claim that some mechanism identifies the pronoun with the antecedent idiom subject, so that an idiomatic interpretation becomes available. As suggested implicitly in the literature, this mechanism might be the same one that is used for interpreting pronouns in non-idiomatic contexts.

This analysis provides a unified syntactic analysis of pronouns in idiomatic and non-idiomatic contexts. However, it does not capture a major difference in the interpretation of pronouns between these two contexts: pronouns are clearly referential in non-idiomatic contexts, setting aside the cases of bound pronouns and expletive pronouns. In contrast, in a clausal idiom context, positively identifying a referent for the idiom subject is often difficult. If the subject does not refer, then nor does the pronoun.

For instance, although *the shit hit the fan* seems to mean "a (serious) problem appears" or "chaos breaks out," (8) shows neither *the shit* nor *it* denotes the same referent as *the problem* or *the chaos*.

(8) The shit hit the fan. We suspect that {#the shit/#it/the problem/the chaos} was caused by Mary.

3.3.2 An expletive pronoun account

For the sake of argument, suppose that clausal idioms enter the derivation as *v*Ps. In (9), *v*P is elided and an expletive *it* inserted to satisfy the EPP.

(9) [$_{TP}$ it$_{exp}$ didn't [$_{vP}$ ~~the shit hit the fan~~]].

This analysis preserves the idiomatic reading, not least because expletive *it* is not referential. However, it incorrectly predicts that expletive *it* should be available with clausal idioms with plural subjects (10). In reality, however, the idiomatic reading is available only when the subject of Ben's response is the plural pronoun *they*.

(10) Alex: The chickens have come home to roost.
Ben: Yes, I'm afraid {#it has/they have}. (note: the plural *they have* retains the idiomatic reading)

3.4 A weaker e-GIVENness condition on ellipsis

In this section, we discuss implications for theories of ellipsis licensing. There is a consensus that an elided constituent must be identical in some sense to its antecedent. Less clear is how this identity condition should be stated. Here, we show that the availability of an idiomatic reading in (2b) bears on this question. We give an argument in support of proposals like Gengel 2007, in which identity between an elided constituent and its antecedent is necessary but not sufficient for licensing ellipsis, *pace* accounts like Merchant 2001 (see also Fox and Lasnik 2003 for independent arguments).

We begin with a review of Merchant's proposal (2001), given in (11).

(11) a. *Focus condition on VP ellipsis* (Merchant 2001:26 ex. 43)
A VP α can be deleted only if α is e-GIVEN.

b. *e*-GIVEN*ness* (ibid. ex. 42)
An expression E counts as e-GIVEN iff E has a salient antecedent A and, modulo $\exists$-type shifting,
(i) A entails F-clo(E), and
(ii) E entails F-clo(A)

c. *F-clo(sure)* (ibid. p. 14 ex. 8)
The F-closure of α, written F-clo(α), is the result of replacing F-marked parts of α with $\exists$-bound variables of the appropriate type (modulo $\exists$-type shifting).

As these definitions show, Merchant's definition of e-GIVENness relies on the elided constituent and its antecedent entailing each other, which requires computing their denotations. However, doing so is difficult, if not impossible, for clausal idioms, on the assumption that proper subparts of an idiom lack denotations of their own. After all, these constituents do not clearly refer to entities or events in a discourse. For example, it is unclear what kind of denotation the VP *hit the fan* might have in *the shit hit the fan*. The same with *shit*, in *the shit*. If a denotation is unavailable, neither *hit the fan* nor *shit* can be said to be e-GIVEN, even when an antecedent is available. We would then incorrectly rule out VPE and NPE, since e-GIVENness is a necessary condition on ellipsis.

Our solution is to weaken the e-GIVENness condition to the one in (12), in effect following Gengel's proposal (2007:229). (12) states that the mutual entailment essential for e-GIVENness can be computed for a constituent *containing* the elided constituent. In clausal idioms, this "containing" constituent is presumably the idiom itself, which has a coherent denotation. A schematic for the dialog in (2) is presented in (13).

(12) *Proposed e-*GIVEN*ness condition*
XP is e-GIVEN iff XP is (reflexively) dominated by a constituent YP_E, such that YP_E has a salient antecedent YP_A, and modulo ∃-type shifting,

a. YP_E entails F-clo(YP_A), and
b. YP_A entails F-clo(YP_E).

(13) a. $\overbrace{\text{[[the shit] [}_{\text{VP}}\text{ hit the fan]]}}^{YP_A}$

b. [NEG $\overbrace{\text{[[the ~~shit~~] [}_{\text{VP}}\text{ ~~hit the fan~~]]}}^{YP_E}$] = *it didn't*. (strikethrough = elided)

An alternative solution for the above problem might be to posit that constituents that lack denotations of their own can mutually entail. We argue that this analysis overgenerates. If mutual entailment were possible under such circumstances, we would incorrectly predict that we can elide parts of any clausal idiom as long there is another salient clausal idiom whose constituents also lack denotations of their own, regardless of whether the idioms are identical. For example, *the shit* and *hit the fan* should be able to license ellipsis of *chickens* and *come home to roost*, to the extent that these constituents all lack denotations of their own in a clausal idiom context. This prediction is clearly not borne out, as (14) shows.

(14) Context: a criminal conspiracy between Alex and Ben has been exposed. The police have found out about their crimes and are now coming to arrest them.

Alex: The shit has hit the fan!
Ben: #They have! (meaning "The chickens have come home to roost," with *they* being derived from ellipsis of the NP *chickens* (antecedent: *shit*) and the VP *come home to roost* being elided (antecedent: *hit the fan*).)

3.5 Idioms meets Pragmatics

Our analysis so far predicts subject pronominalization and VP ellipsis can occur freely without affecting idiomatic interpretation, contrary to fact, as we showed before:

(15) Alex: When the news got out, <u>the shit hit the fan</u>.
Ben: No, ...

a. ... <u>the shit</u> didn't <u>hit the fan</u>. [−Pron., −VPE]
b. ... it didn't. [+Pron., +VPE]
c. ... #it didn't <u>hit the fan</u>. [+Pron., −VPE]
d. ... #<u>the shit</u> didn't. [−Pron., +VPE]

Since ellipsis and pronominalization can indeed preserve the idiomatic interpretation, we believe what disrupts the idiomatic interpretation in the above paradigm is neither VP ellipsis nor NP ellipsis, at least not directly. We suggest that there is a pragmatic explanation

for why they must co-occur in (2b)/(15b), based on inferences made about the remnants of VP and NP ellipsis when they do not co-occur.

A speaker who repeats the VP while pronominalizing the subject (15c), or repeats the subject while eliding the VP (15d), leads a hearer to infer that the speaker intends to contrast the repeated element with another entity or predicate. For instance, *it didn't hit the fan* suggests that *the shit* did something else. Conversely, *The shit didn't* implies that something else hit the fan. Since *the shit* does not refer, nor does *hit the fan*, there are no plausible contextual alternatives for the hearer to consider.

We would like to suggest that the same rationale can be applied to other types of idioms in similar contexts. Consider the following examples with a VP idiom:

(16) Alex: Did John kick the bucket last night?
Ben: No, . . .

a. . . . he didn't kick the bucket.
b. . . . he didn't.
c. . . . #he didn't kick it.

In (16a) the whole idiom is repeated and thus no emphasis or contrast is inferred. In (16b), the whole idiom is elided, hence again nothing special is inferred about sub-parts of the idiom to the detriment of others. Finally, in (16c), only a part of the idiom is repeated, thus inviting that inference. We suggest this affects the availability of idiomatic interpretation.

The decision on whether to elide parts of a sentence has implications on how sentences are interpreted and idioms seem particularly sensitive to this decision. We believe this to be a promising line of research for both idioms and pragmatic aspects of ellipsis. We leave a full investigation about these effects and how exactly they arise for future work.

4. Conclusion

In the context of clausal idioms, pronominalization and ellipsis can apply without affecting idiomatic interpretation. Making standard assumptions about idiomatic interpretations and ellipsis, we argued that this fact is theoretically significant, providing new evidence for the claims that (i) some pronouns can be derived from full noun phrases and that (ii) the identity condition on ellipsis must be able to refer to a constituent that properly contains the one that is elided. Finally, we suggested a way to deal with cases where idiomatic interpretation is hard to get when only pronominalization or VP ellipsis obtains. Specifically, if just a subportion of a clausal idiom is elided, either for pronominalization or VP ellipsis, the remnant portion receives an contrastive interpretation which disrupts the idiomatic interpretation.

References

Bach, Emmon, and Barbara H. Partee. 1980. Anaphora and semantic structure. In *Papers from the Parasession on Pronouns and Anaphora, Proceedings of the Chicago*

Linguistic Society, ed. Jody Kreiman and Almerindo E. Ojeda, 1–28. Chicago, Illinois: University of Chicago.

Bresnan, Joan. 1970. An argument against pronominalization. *Linguistic Inquiry* 1:122–123.

Bruening, Benjamin. 2015. Idioms, anaphora, and movement diagnostics. Ms., University of Delaware.

Chomsky, Noam. 1955. The logical structure of linguistic theory. Harvard/MIT. [Published in part, 1975, New York: Plenum].

Elbourne, Paul. 2001. E-type anaphora as NP-deletion. *Natural Language Semantics* 9:241–288.

Fox, Danny, and Howard Lasnik. 2003. Successive-cyclic movement and island repair: The difference between sluicing and VP-ellipsis. *Linguistic Inquiry* 34:143–154.

Fraser, Bruce. 1970. Idioms within a transformational grammar. *Foundations of Language* 6:22–42.

Gengel, Kirsten. 2007. Focus and ellipsis: A generative analysis of pseudogapping and other elliptical structures. Doctoral dissertation, University of Stuttgart.

Hankamer, Jorge, and Ivan Sag. 1976. Deep and surface anaphora. *Linguistic Inquiry* 7:391–428.

Hardt, Dan. 1993. Verb phrase ellipsis: Form, meaning and processing. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.

Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Doctoral dissertation, University of Massachusetts, Amherst.

Hornstein, Norbert. 2007. Pronouns in a minimalist setting. In *The copy theory of movement*, ed. Norbert Cover and Jairo Nunes, 351–385. Amsterdam: John Benjamins.

Jackendoff, Ray. 1969. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.

Katz, Jerrold J., and Paul Postal. 1963. Semantic interpretation of idioms and sentences containing them. *Quarterly Progress Report of the MIT Research Laboratory of Electronics* 70:275–282.

Lasnik, Howard. 1976. Remarks on coreference. *Linguistic Analysis* 2:1–22.

Lees, Robert B., and Edward S. Klima. 1963. Rules for English pronominalization. *Language* 39:17–28.

Lobeck, Anne. 1995. *Ellipsis: Functional heads, licensing and identification*. New York: Oxford University Press.

Merchant, Jason. 2001. *The syntax of silence*. Oxford: Oxford University Press.

Nunberg, Geoffrey, Ivan A Sag, and Thomas Wasow. 1994. Idioms. *Language* 70:491–538.

Postal, Paul. 1966. On so-called pronouns in English. In *Modern studies in English*, ed. David Reibel and Sanford Schane, 201–223. Englewood Cliffs, NJ: Prentice-Hall.

Ross, John Robert. 1967. Constraints on variables in syntax. PhD dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Rottman, Isaac, and Masaya Yoshida. 2013. Sluicing, idioms, and island repair. *Linguistic Inquiry* 44:651–668.

Nick Huang & Gesoel Mendes

On pronominalization and ellipsis in clausal idioms

znhuang@umd.edu, gmendes@umd.edu