A Corpus Study of Pseudogapping and Its Theoretical Consequences

Philip Miller

This paper presents the results of a large-scale investigation of the use of NP-remnant pseudogapping (PG) in the COCA corpus. Discourse conditions on the use of noncomparative and comparative PG are discussed. It is shown that the data raise problems for mainstream generative analyses involving remnant-raising and an alternative interpretive analysis is suggested. The question of whether PG is a subcase of VPE is discussed.

Keywords: pseudogapping, ellipsis, corpus analysis, antecedent mismatches, discourse pragmatics, usage preferences

1 Introduction

Pseudogapping (PG) is a construction similar to Verb Phrase Ellipsis (VPE) in that it is characterized by an ellipsis behind an auxiliary. But, contrary to VPE, the auxiliary is followed by a complement (the 'remnant'), which corresponds to a complement of the antecedent, as illustrated in (1).[1] It appears both in comparative (cf. (1b,c)) and noncomparative (cf. (1a,d)) structures. As these examples show, the ellipted material does not necessarily form a constituent, nor even a continuous subsequence of the antecedent.

(1) a. "It doesn't bother me," I said untruthfully. "Well, it does ~~bother~~ me," he growled, and I let it rest. (Fic)
b. We'll let you know if it deals with the heat and humidity as well as it did the frigid slop. (Mag)
c. [...] all treat him with deference due a social superior, as they do his wife, (Acad)
d. [...] the whole room seemed like a great relief to me and I knew it must ~~seem like a great relief~~ to him, too. (Fic)

Following Kuno 1981, most syntacticians working in transformational frameworks (e.g. Jayaseelan 1990, Lasnik 1999, Gengel 2013) have claimed that PG is a subcase of VPE, where the

I would like to thank Peter Culicover, Marianne Desmets, Jim Donaldson, Gabriel Flambard, Barbara Hemforth, Anne Hertz, Philip Hofmeister, Laura Kertz, Bob Levine, Jason Merchant, Alain Rouveret and Saghie Sharifzadeh, and Gary Thoms for comments and discussion. I would also like to thank Christopher Piñón and an anonymous reviewer for detailed comments on the initial version. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

[1] To clarify the intended interpretation of the examples, the antecedent is underlined, the pre-elliptical auxiliary is double underlined and the remnant is wavy underlined; in some examples, the ellipted material is struck out in the putative ellipsis site. This is not intended to represent a syntactic analysis. Unless otherwise mentioned, all of the examples cited in the paper are taken from the COCA (Corpus of Contemporary American English, http://corpus.byu.edu/coca/, see Davies 2008-), a large corpus of American English (450 million words), evenly divided into 5 registers (Acad(emic), Fic(tion), Mag(azine), News(paper), Spok(en)), with approximately 90 million words each.

EISS 10

Empirical Issues in Syntax and Semantics 10, ed. Christopher Piñón, 73–90
http://www.cssp.cnrs.fr/eiss10/
© 2014 Philip Miller

remnant is moved out of the VP before deletion of the VP under identity by VPE.[2] In Miller 1990, a nontransformational approach, I proposed an interpretive analysis, extending to PG the VPE-as-proform analysis of Schachter 1978 and Hardt 1993. This line of analysis has been explored further by Hoeksema 2006, as well as by Kubota and Levine 2014, who develop a Hybrid TLCG analysis, which is similar in spirit to the analysis suggested here.

In what follows, I will provide corpus data on pseudogapping, based on an extensive corpus investigation of the COCA, which resulted in over 1700 occurrences of PG. I will first briefly discuss the way the data were collected and classified and set out some of the central properties of the actual usage of PG, including discussion of the discourse conditions under which it is felicitous. I will then address the relevance of the corpus data for the various proposed syntactic analyses of PG, suggesting that they favor an interpretive analysis rather than a syntactic analysis in terms of ellipsis. In particular, I will show that putative remnant movement does not respect island constraints or connectivity. In a final section, I will provide some preliminary discussion of what the corpus data bring to bear on the question of whether PG and VPE are the same construction, suggesting that the differences between them might be explained away as resulting from independent discourse factors.

2 Collecting the Data

The COCA is tagged for parts of speech but it is not parsed, so that one cannot directly search for specific syntactic structures such as PG. Heuristic strategies had to be set up in order to find them. For this study, I concentrated on the central case of PG, namely, PG with NP-remnants. Since what characterizes this case is the presence of an NP complement after an auxiliary, strategies were devised to detect such configurations. Specifically, the following sequences were systematically checked: (i) auxiliaries followed by an object personal pronoun (except for *you* which does not have a distinct object form and consequently leads to massive noise from Subject-Auxiliary Inversion);[3] (ii) auxiliaries followed by *you* and either punctuation or *too*, *as well*, or *at all*; (iii) auxiliaries preceded by a subject pronoun and followed by *you*;[4] (iv) auxiliaries preceded by a subject pronoun or a noun and followed by an article, a determiner (quantificational or otherwise), an adjective, a noun, or a possessive (dependent or independent).[5]

Because these searches lead to a lot of noise, especially in the case of *do*, among which actual examples of PG were selected manually, I am certain to have missed some examples that could have been found by these strategies, due to lapses of attention. There are also certainly examples of PG which could not be found by these searches (e.g. because of unusual NP patterns or unusually placed adverbs or parentheticals), though they can be assumed to be relatively

[2]More recent analyses involve LF-copying and other variants. Distinctions between these various analyses will not be relevant here and I will henceforth subsume them all under the term 'deletion'.

[3]*It* shares this property with *you*, but, as is well-known, *it* does not occur as a remnant in PG because it cannot be stressed. This property was confirmed by searches on *it* of the type described in (ii) for *you*.

[4]Because 86% of pronominal remnants other than *you* in our data are followed by punctuation or *too/as well/at all* and 91% of PG cases in our data have a pronoun subject, one can assume that the great majority of *you* remnants were found by combining both of these search strategies.

[5]The auxiliaries *have* and *be* raise specific problems. Because of their very frequent use with a wide variety of NP complements, it was impossible to search for general PG patterns. I have conducted searches for comparative pseudogappings for these auxiliaries, but have not included them in the numerical findings presented here, as they would have skewed the proportions of comparative and noncomparative PG.

rare.[6] Overall, I believe that the sample of 1415 occurrences that I have compiled contains the great majority of the cases of NP-remnant PG in the COCA. However it is clear that pronominal subjects and remnants are less likely to have been missed than full NPs. Consequently, except for a possible minor bias in favor of pronominal subjects and remnants, the data can be assumed to be relatively representative of the COCA overall and, presumably, of current American usage.

3 Central Properties of NP-Remnant PG

A first observation that should be made concerns the distribution of NP-remnant PG in comparative and noncomparative structures. The present study found that 96.7% of occurrences were comparative. Table 1 provides a breakdown by register. The large sample studied here thus exhibits an even stronger bias than previous studies by Hoeksema 2006 (87% of the PGs in his corpus of 227 occurrences were comparative) and Sharifzadeh 2012 (90% were comparatives). Furthermore, it is clear that this difference in frequency of occurrence correlates with other features which differentiate comparative and noncomparative PG, which will now be discussed.

3.1 Noncomparative NP-Remnant PG

The COCA data provide 47 occurrences of noncomparative PG. Given the limited number of occurrences, it was tempting to combine these data with the 37 occurrences of noncomparative PG collected by Levin 1986, leading to a total of 84 cases. In the numerical analyses for noncomparative PG, I will systematically provide two figures, separated by a slash. The first of these indicates the figure for the COCA, the second for the combined COCA and Levin data.

Noncomparative PG is typical of the spoken register The breakdown by register provided in Table 1 is misleading since it does not take into account the fact that most noncomparative PGs outside the spoken register occur in reported speech. Spoken and reported speech combined contain 87.2%/91.7% of the occurrences of noncomparative PG while only 12.8%/8.3% of cases of noncomparative PG occur in narrative segments of fiction and non-fiction prose.[7]

The subject of the PG is almost always a personal pronoun Table 2 provides a breakdown of noncomparative PG by subject of the pre-elliptical auxiliary. There is a striking dominance of pronominal subjects and especially of *it*, which by itself accounts for about half of the occurrences (we will see that this is in stark contrast with the comparative PGs, where 89% of the subjects are pronouns). Pronouns, and especially *it*, are known to require a very highly accessible antecedent (cf. Ariel 1990). In the single case found with an NP subject (a case of nonfiction prose), given in (2), the referent of *Mom* is highly accessible in the context of the topic of the

[6] This was corroborated by a series of searches likely to uncover comparative PGs which could not be found by the above strategies and which led to no new occurrences.

[7] It should be noted that noncomparative PG is apparently less acceptable than comparative PG in general. Hoeksema 2006 reports an acceptability study on PG and finds an average acceptability of 8.4 out of 10 for comparative NP-remnant PG and of 4.6 for noncomparative coordinate cases. Ongoing acceptability experiments I am conducting corroborate this finding, with a greater loss of acceptability for full NP subjects than for pronominal subjects (as expected given the following discussion). The reasons for the lesser acceptability of noncomparative PG are as yet unknown, but it might simply be an artefact of the written presentation of the stimuli in the acceptability experiments. It may be that subjects have trouble imagining appropriate intonation patterns (which would make the sentences completely acceptable) and thus find the sentences unnatural. Checking this idea would require acoustic stimuli.

Table 1
Comparative vs. noncomparative NP-remnant PG by register in the COCA

Register	Total PG	Noncomp	Comp
Acad	146	0 (0%)	146 (100%)
Fic	430	18 (4.2%)	412 (95.8%)
Mag	414	10 (2.4%)	404 (97.6%)
News	224	1 (0.4%)	223 (99.6%)
Spok	201	18 (9%)	183 (91%)
Total	1415	47 (3.3%)	1368 (96.7%)

Table 2
Subjects in noncomparative NP-remnant PG

Subject	he	I	it	she	they	we	you	Total pro	NP
COCA	5 (10.6%)	7 (14.9%)	22 (46.8%)	2 (4.3%)	4 (8.5%)	0 (0%)	6 (12.8%)	46 (97.9%)	1 (2.1%)
COCA+Levin	7 (8.3%)	12 (14.3%)	44 (52.4%)	3 (3.6%)	7 (8.3%)	0 (0%)	10 (19.9%)	83 (98.8%)	1 (1.2%)

'family as a unit' especially given the immediately previous mention of *Dad*.[8]

(2) That notion is naively anachronistic in an age when the family as a unit of cultural consumption hardly exists: Dad watches ESPN, Mom does Lifetime, Little Bro works his Game Boy Advance, and Kid Sis is a Powerpuff Girl. (Mag)

It is important to keep in mind that these results are not as spectacular as they might seem, given that the noncomparative PGs are mostly spoken register or reported speech. Indeed, Francis et al. 1999 found that 91% of the subjects (out of 31,021 declarative sentences) in a part of the Switchboard Corpus (a corpus of telephone conversations) were pronominal. Thus, though the strictly spoken data from the present corpus of PG has 100% pronominal subjects, this is only 9% more than what was was found for subjects of declaratives overall in the Switchboard study.

Beyond being pronominal, the subject of the PG is, in general, coreferent with the subject of the antecedent clause. This is true in 38 out of 47 cases in the COCA data and 73 out of 84 in the COCA+Levin. Among the cases which are not coreferent, all (except for (2) cited above) fall into one of two patterns. The first, noted by Levin 1986, is the mirror pattern, as in (3a), where the referents of the subject and object of the antecedent clause appear in reverse order in the PG (4/5 cases). The second is the parallel pattern, as in (3b), where the subject and the dependent possessive in the object of the antecedent clause are coreferent, and the same is true of the PG, except that the object is reduced to an independent possessive, typically resulting in an *I—mine* combination (4/5 cases). These two configurations have in common that they make the referents in contrast more accessible (there are only two in the mirror pattern, and the referents are linked pairwise by parallel possessive relations in the parallel pattern) as opposed to four

[8]Jim Donaldson (p.c.) points out that (2) might in fact not be a case of PG at all, but rather a use of main verb *do* (as made clear by the fact that *watches* can be replaced by *does* in the antecedent: *Dad does EPSN, Mom does Lifetime*). If that is the correct analysis for this example, the corpus exhibits 100% pronominal subjects.

Table 3
Remnants in noncomparative NP-remnant PG

Remnant	COCA	COCA+Levin	Remnant	COCA	COCA+Levin
me	17 (36.2%)	33 (39.3%)	NP.dem	1 (2.1%)	4 (4.8%)
mine	10 (21.3%)	12 (14.3%)	NP.Ø	3 (6.4%)	5 (6%)
you/yours	4 (8.5%)	10 (11.9%)	NP.some	1 (2.1%)	2 (2.4%)
Other pron	4 (8.5%)	5 (6%)	NP.the	2 (4.3%)	4 (4.8%)
Total pron	35 (74.5%)	60 (71.4%)	NP.their	1 (2.1%)	1 (1.2%)
N.prop	2 (4.3%)	3 (3.6%)	NP.other	0 (0%)	1 (1.2%)
NP.a	2 (4.3%)	2 (2.4%)	Total NP	4 (25.5%)	7 (28.6%)
NP.any	0 (0%)	2 (2.4%)	Pron+Ana NP	41 (87.2%)	71 (84.5%)

independent referents in (2).

(3) a. I ain't scared of your gun. I got a gun, too. I can shoot you before you can me, (Spok)
b. Yes, you my [=might, PhM] love your baby and your toddler to death—I did mine—but that doesn't mean to say a child can fulfill all the needs of an adult. (Spok)

Noncomparative PG remnants are typically pronominal and/or anaphoric and form a contrastive focus with the corresponding complement of the antecedent Table 3 provides a breakdown of the remnants of noncomparative NP-remnant PG. 74.5%/71.4% of the remnant objects are pronominal, among which the first person singular *me* and *mine* are highly dominant. This finding is much more surprising than the case of subjects, as Francis et al. 1999 found only 34% pronominal objects in their sample of the Switchboard. Beyond this, it turns out that among the 24 cases of NP remnants, 10 are anaphoric, being headed by pro-N *one*, a Ø noun, or a repeated head noun. In all, 87.2%/84.5% of remnants are thus anaphoric. Furthermore, the nonanaphoric NP remnants are highly accessible in the discourse context.

One of the central characteristics of noncomparative PG is that it has to have a contrast between the remnant and the corresponding complement of the antecedent.[9] This correlates with the preponderance of first person singular remnants (and to a lesser extent of second person remnants), since contrast with the speaker (and to a lesser extent the addressee) are typically very relevant to her/him. More generally, for NP remnants, it is usually the case that the object of the antecedent has an obvious hyperonym within which it forms a contrastive pair of hyponyms with the remnant (e.g. two TV channels as in (2) cited above, or a contrastive pair of politicians, or a contrastive pair of electronic appliances, viz. a Web-TV vs. a big-screen PC) so that evocation of the object of the antecedent makes the remnant inferrable. We will see below that all these properties are in stark contrast with those of comparative PG.

Discourse conditions on noncomparative PG As was just mentioned, PG always involves contrastive objects, as illustrated in (4a). I will call these cases Obj-choice, following the terminology of Miller and Pullum 2014 (i.e. the symmetric of Subj-choice VPE—as in *He knows the answer and she does too*, but with contrasting objects rather than subjects). It can also involve

[9]One example found in the COCA violates this restriction and is presumably a speech error: "NOVAK: Doesn't worry you? DALEY: No, it doesn't me at all." (Spok).

double contrast (viz. the 'mirror' and 'parallel' cases discussed above), as in (4b), and thus be both Subj-choice and Obj-choice. It can furthermore involve secondary Aux-choice, as in (4c), which has a contrast between *can* and *will*, and (4d), which has a contrast in polarity ((3a) illustrates a case where there is a triple contrast on subjects, objects, and auxiliaries). PG cannot be simply Aux-choice, as shown by the infelicity of the variant of (4d) given in (4e). Only VPE is acceptable in that case, as in (4f) (cf. also footnote 9).

(4) a. Your weight affects your voice. It does mine, anyway. (Mag)
b. [...] we want to treat your POWs with dignity and we hope that you do ours as well. (Spok)
c. I can't read most of these lyrics, but I will some. (Spok)
d. "It doesn't bother me," I said untruthfully. "Well, it does me," he growled, and I let it rest. (Fic)
e. "It doesn't bother me," I said untruthfully. #"It does you," he growled.
f. "It doesn't bother me," I said untruthfully. "It does," he growled.

These considerations make it possible to tentatively propose the following discourse conditions on noncomparative PG:

Type 1: Object choice

Formal characteristics: The subject of the antecedent is identical to that of the PG construction but the object is distinct, and stressed if it is a pronoun.

Discourse requirement: Both the referent of the remnant and a particular open proposition $p(x)$ must be highly salient in the discourse context, and the point of the utterance containing the PG must be limited to identifying something or someone satisfying $p(x)$ and such that it forms a contrastive focus with the referent of the correspondent of the remnant in the antecedent.

Type 2: Subject and object choice

Formal characteristics: The subject and object of the antecedent are distinct from those of the PG construction, and both are stressed if they are pronouns.

Discourse requirement: Both the referents of the remnant and subject and a particular doubly open proposition $p(x, y)$ must be highly salient in the discourse context, and the point of the utterance containing the PG must be limited to identifying a pair satisfying $p(x, y)$ and such that they form a pair of contrastive foci with the referents of the correspondent of the remnant and the subject of the antecedent.

Noncomparative PG is typically endophoric Miller and Pullum 2014 discuss Subj-choice exophoric VPE and argue that it is exceedingly rare (as opposed to Aux-choice VPE) because the nonlinguistic context is usually unable to make open propositions salient. Since the discourse conditions on PG that have just been proposed also require a salient open proposition, we can expect exophoric PG to be exceedingly rare as well. One example of this type was found in the COCA, which might be considered to be a speech error. The context is obviously a cooking demonstration. Presumably Maggipinto's intention is to convey the fact that you can *use* tangerines in the recipe as well. The question is the extent to which the previous discourse and extralinguistic context can make salient the open proposition 'you can use x in this recipe'.

(5) Ms-MAGGIPINTO: It looks very glamorous, doesn't it? CURRY: It does. Now, it has tangerines? Ms-MAGGIPINTO: Those are mandarin oranges. CURRY: OK, mandarin oranges. Ms-MAGGIPINTO: Yes, you can tangerines, though. CURRY: OK, and you have what, beets? Ms-MAGGIPINTO: Those are beets. (Spok)

3.2 Comparative NP-Remnant PG

Register As shown in Table 1 above, the 1368 examples of comparative NP-remnant PG occur in all registers but are especially frequent in fiction and magazines. They typically do not exhibit the somewhat marked status of noncomparative PG (this corroborated by their much greater frequency of occurrence in the corpus and by the acceptability experiments mentioned in footnote 7). The comparative cases can be divided into two broad classes, those involving comparison of degrees as in (6a) and those involving comparison of manner (6b) or factual identity (6c) (the difference between the latter two is not always easy to establish, as shown by (6d)).

(6) a. It hurt me as much as it did her. (SPOK)
b. You must treat him as you would me. (FIC)
c. Let Thomas' confirmation serve to remind you, as it did me, that the fight is far from over. (MAG)
d. You're not much better, Noah. I took you in, just as I did him. (FIC)

Subjects of comparative NP-remnant PG Table 4 provides basic data on subjects in comparative PG. Several properties are worth noting. Pronominal subjects continue to dominate strongly. More specifically, in the spoken register, 97.8% of the subjects are pronominal, a figure very similar to that of the noncomparatives, and again significantly higher than that of Francis et al. 1999. The average for the written registers is 89.8%. Biber et al. 1999 provide comparable data for the overall use of pronouns vs. full NPs in written registers, finding 65% pronominal subjects in fiction, 25% in newspapers, and 20% in academic. Thus the presence of pronouns as subjects in written register PG is massively more frequent than would be expected, even more so than for noncomparative PG. On the other hand, the pronouns found most frequently are not the same. Whereas *it* was the subject of almost 50% of noncomparative PGs, it represents only 19.7% of comparatives. What dominates here is the use of the pronoun *you* as a generic, as illustrated in (7). Notice that in this type of example, the remnant is an indefinite NP with determiner *a(n)*, Ø, or *any*. This configuration is very frequent, especially in the magazine register and the fact/manner comparatives and, by itself, it accounts for close to 25% of the comparative PG data.

(7) a. "Treat a loan from a relative or friend as you would any business loan," advises Baltimore financial planner Jay Perry. (Mag)
b. Use the same criteria to select a healthy nut tree as you would a fruit tree. (Mag)

Remnants of comparative NP-remnant PG Table 5 provides information on remnants in comparative PG. The most immediately striking property of these data is the contrast in the proportion of pronoun vs. NP remnants, namely, approximately 13.5% vs. 86.5% (as opposed to 74.5% vs. 25.5% for noncomparatives), so that we observe a complete reversal of proportions. Even in the spoken register, we find only 23.6% of pronominal remnants, as opposed to 76.4% of full NP remnants. More specifically, indefinite NPs (in particular NP.a, NP.Ø and NP.any) are very

Table 4
Subjects in comparative NP-remnant PG

I	you	it	they	other pronoun	total pronoun	full NP	total
124	310	269	219	321	1243	125	1368
9.1%	22.7%	19.7%	16%	23.5%	90.9%	9.1%	100%

Table 5
Remnants in comparative NP-remnant PG

me/mine	you/yours	other pron	total pron	N.Prop	NP.a	NP.Ø	NP.any	NP.the	other NP	total NP
84	30	70	184	59	429	226	109	218	143	1184
6.1%	2.2%	5.1%	13.5%	4.3%	31.4%	16.5%	8%	15.9%	10.5%	86.5%

highly represented, especially in the fact/manner comparatives. As mentioned above and illustrated in (7), these often combine with generic *you* subjects, producing comparisons with generic situations. On the other hand, definite NP remnants are more frequent in the degree comparatives, because there is a greater tendency to compare degrees between uniquely identifiable entities. Among the pronominal remnants, the first person singular is dominant (almost 50% of the occurrences), showing a similar tendency to the noncomparatives.

Discourse conditions on comparative NP-remnant PG Though the comparative remnant and the complement of the antecedent are often cohyponyms of a hyperonym which is made easily accessible by the mention of the latter (as was the case in examples (7), where the hyperonyms are *loan* and *tree*) this need not at all be the case: remnants can represent entirely new and unpredictable information, as in (8), contrary to what we saw for noncomparatives.

(8) a. Feeding Las Vegas' one-armed bandits does as much for your biceps as it does your bank account. (Mag)
b. But the minor league instructors [...] came back talking breathlessly about the kid the way a tourist would the Grand Canyon. (Mag)

These examples make clear that in comparative PG, contrary to noncomparative, there does not have to be any kind of contrast between the referents of the subject and remnant of the PG clause and their correspondent in the antecedent (though, of course, they cannot be coreferent). Rather, they must simply be comparable.

Another difference between noncomparative and comparative PG concerns contrasting subjects. As mentioned above, in the noncomparative case these typically involve mirror or parallel configurations, which make the contrasting referents more accessible (cf. (3) above). There is no such constraint on comparative PG: the corpus contains 176 examples of comparative PG with different subjects out of 1368 (12.9%), only one of which is a mirror configuration. The greater flexibility of comparative PG in this respect is probably a consequence of the parallelism imposed by the very nature of the comparative construction, which helps make clear how the subjects and objects of the antecedent and elliptical clauses align. The ease of processing linked to this obvious parallelism may explain why PG is much more acceptable and frequent in comparatives.

To conclude, I should mention the existence of a subtype of comparative PG which, to my knowledge, has never been pointed out, namely, cases where the antecedent has no overt object corresponding to the remnant. There are 13 occurences of this type in the corpus, two of which are given in (9). In (9a) the antecedent has a null anaphor as object, the reference of which is provided explicitly in the discourse context. This is the most usual case, as opposed to (9b) in which the antecedent can simply be considered intransitive.

(9) a. Let the peas dry on a tray in the house for a few days, then store and label in airtight jars out of direct sunlight as you would any other bean or pea. (Mag)
b. My echoes are no longer tormentors but friends, and when one of them dies (as, inevitably, they have begun to) I mourn a little, as I would a sister. (Fic)

4 Some Consequences for Syntactic Analyses of PG

4.1 Raising of Remnant Analyses

As mentioned in the introduction, most syntacticians working in the Principles and Parameters and Minimalist frameworks have proposed that PG is a subcase of VPE where the remnant has been moved out of the VP before ellipsis. This kind of analysis was first advocated by Kuno 1981 and was developed by Jayaseelan 1990, who proposed that the remnant was affected by Heavy NP Shift (HNPS). Lasnik 1999 provides convincing criticism of this analysis, centering on the fact that pronouns make excellent PG remnants whereas they do not undergo HNPS. He proposes instead that the remnant raises by *A*-movement to spec of Agr_o. This analysis makes four central predictions. The first is that remnants must be direct objects; the second is that remnant movement should obey constraints on *A*-movement, in particular island constraints; the third is that the form of the remnant, in particular its case (or preposition, if it is a PP), should be identical to that which it would have in a nonelliptical context; the fourth is that there should be an appropriate syntactic antecedent in the context. I will discuss these in turn.

Remnants must be direct objects Citing Levin 1986, Lasnik notes a potential objection to this constraint, namely, that objects of prepositions can be PG remnants ('deprepositionalized' PG in Levin's terms). He proposes that this is made possible by reanalysis and that there should be a "consistent correlation between pseudogapping and pseudopassive" (Lasnik 1999:145). He also notes that more extreme cases of reanalysis, such as *take advantage of*, also support PG.

The data from the present corpus make these proposals very difficult to sustain as there are numerous cases where reanalysis is not supported by any of the usual criteria, in particular the possibility of a prepositional passive. This is illustrated in (10).

(10) a. In other words, walk into a seafood market as you would a fresh flower market, with your eyes, nose, ears and hands all on full alert. (News) [Compare: #The market was walked into.]
b. It [= the wind] blows through me as it would an abandoned house. (Mag) [Compare: #The house was blown through.]
c. For example, if people wish I would sound like I used to sound, then it says more about them than it does me. (Spok) [Compare: #I was said something about.]

All in all, there are 115 'deprepositionalized' remnants in the corpus (i.e. 8.1%), many of which

are more or less implausible candidates for reanalysis.[10] Furthermore, in some cases, as in (10c), there is an unconstrained direct object between the V and the preposition making reanalysis completely impossible.

Beyond these cases, there are numerous examples where the remnant corresponds to a complement of an embedded verb with no plausible reanalysis possible, as illustrated in (11):

(11) a. [...] he could no more imagine himself contradicting the man striding on ahead of him than he could his grandfather (Fic)
b. Health care providers acknowledge they may have a much harder time stopping the regulations than they would any bills in Congress. (News)
c. [...the custodians] will instead seek to acquire it [= Rodchenko's monochrome] as they would any other picture lest their collections otherwise betray a gap in [...] (Acad)

Finally, searching for adjective initial NP remnants showed that there were in fact 13 PGs with copular verbs and predicative AP remnants among the comparative PGs, as illustrated in (12). (Levin 1986 suggests that these are acceptable, contra Lasnik 1999:142.)

(12) a. Tortilla soup tastes as good cold as it does hot. (Mag)
b. [...] which sounds as seductive as it does sincere. (Mag)
c. Ang Lee seemed as embarrassed as he did thrilled to be named best director. (Spok)

Remnant movement and island constraints A central prediction of any analysis based on movement of the remnant is that usual island constraints should be respected. In the case of Lasnik's proposed *A*-movement analysis, *A*-movement constraints should be obeyed. More recently, Gengel 2013 has argued that *A-bar* movement is involved, specifically Focus Movement. Under this analysis, *A-bar*-movement constraints should be respected. However, once again, the data collected show that there are numerous instances where this is not the case.[11] The examples in (13) illustrate this for various types of islands. Each example is followed by variants which show that *A*-movement and/or *A-bar*-movement of the remnant are impossible.[12]

(13) a. According to current ideas, the frothiness of space retards the arrival of a burst's highest-energy photons more than it does the lowest-energy photons. (Mag) [Complex NP Constraint (CNPC): *the photons which it retards the arrival of; *The photon was retarded the arrival of.]

[10]Besides the three cited in (10), the following can be found: *agree with*, *appear to*, *ask for*, *bite into*, *care about* (2 cases), *cling to*, *commiserate with*, *cram into*, *depend on*, *do for*, *eat at*, *feel about*, *feel like*, *fly into*, *flirt with*, *forget about*, *get in and out of*, *go after*, *go beyond*, *go through*, *go into and go through*, *go to*, *grab onto*, *interfere in*, *hold onto* (2 cases), *jump into*, *knock on*, *know about*, *look like*, *make of*, *move on in*, *move through*, *know one's way around*, *plan around*, *play at*, *pore over*, *react to*, *relate to*, *revolve around*, *ride at*, *settle over*, *share with*, *sit at*, *sit on*, *stand behind*, *stick in*, *tee up for*, *throw NP around*, *tinker with*, *walk into*, *work at*, *work for*, *work with* (2 cases), and *zero in on*. It should also be noted that the frequency of deprepositionalized PG in the corpus seems much higher than that of prepositional passives. Though I have been unable to find data on present-day English, Seoane Posse 1999 finds that 2.3% of passives are prepositional in the period 1640-1710.

[11]Culicover and Jackendoff 2005:274-5 note similar problems for raising analyses of gapping.

[12]I have used relative clauses to illustrate *A-bar*-movement as it is known that they are less degraded by island violations than *wh*-questions. Some kinds islands are known to be less unacceptable or to lead to variable judgments, specifically the types illustrated in (13e) and (13f).

b. [...] the voting preferences of black women much more closely approximated the pattern of black men than they did white women. (Acad) [CNPC: *people who they approximated the pattern of; *Those people were approximated the pattern of.]

c. Meeting with the committee members was perhaps the most important thing I would do during my transition, and they would examine what I wore as intensely as anything else—as they would any woman who met with them, it occurred to me. (Mag) [*Wh*-island: *a woman who you would examine what wore; *She was examined what wore.]

d. Bring the same kind of carry-ons (diapers, medications, toys, etc.) when traveling by train as you would by air; you're allowed two per person. (Mag) [Adjunct island: *a means of transport which I brought the same kinds of carry-ons when traveling by; *The train was brought the carry-ons when traveling by.

e. [...] we'll walk with them and we'll make them exercise like you would a boxer on a treadmill, [...] (Spok) [Object of causative or perception verb island: %someone who you will make exercise; *A*-movement is possible: He was made to exercise.]

f. give these smaller newspapers your best efforts, of course, just as you would a large-circulation national magazine. (Acad) [First object of double object construction island: %something which she gave her best efforts; *A*-movement OK: The newspapers were given ...]

All in all, I have gathered 60 cases of island violations of these types (of which 20 CNPC violations), so that such violations appear in approximately 4.2% of the occurrences in the corpus. This seems too great a rate to make it plausible to consider them as speech errors. All the more so that they clearly do not have the flavor of island violations. My feeling is that they range between perfectly acceptable and slightly sloppy, rather than ill-formed. It should be recalled that all of the examples cited, and most of those collected, have in fact appeared in print. The variants with true island violations given after the examples are far more unacceptable (most of them are simply unintelligible). Thus, any theory claiming that these cases of PG violate the same constraints as those which rule out the variants would be confronted with a major problem in trying to explain why there is such a great difference in acceptability between them.

Nonconnectivity in remnant form One of the most intuitively appealing predictions made by syntactic theories of *A-bar* movement is that the moved element will show up with the same form as it would have had without movement, namely, so-called 'connectivity' effects: a prepositional or case-marked complement is realized with the usual preposition or case when it is fronted. In the case of PG, theories involving *A-bar* movement of the remnant thus predict that the remnant should exhibit connectivity effects.

In this connection, consider (14a):

(14) a. Ask Doll, who spoke as much about his schoolboy career ending as he did of the season in general: "I don't want it to end." (News) [Compare: He spoke as much about his career as he did about/#to Peter.]

b. He cared as much about his career as he did about/#of/#for the season in general.

In (14a), the remnant appears with the preposition *of*, whereas the correspondent of the remnant in the antecedent has *about*. To begin with, it should be noted that it is not possible in general

to switch prepositions in this manner without affecting acceptability. This is shown by the unacceptability of the *of* variant in (14b) where *care* has been substituted for *speak*. I tentatively suggest (as I did in Miller 1990:(38) for a similar invented example) that acceptability depends on a combination of two factors, namely the semantic relation of the remnant to the predicate and the possible subcat frames of the antecedent verb. Specifically, if a given verb has two subcat frames which allow for syntactically distinct objects, a discrepancy between the two frames, in the antecedent and PG clause respectively, will be all the more acceptable that the semantic relation of the complements in question to the verb is similar. For instance, *speak* allows PP[about], PP[of], and PP[to] complements. In the case of the first two of these, the semantic relation is very close, whereas it is completely different for the third. This explains that the combination of PP[about] and PP[of] in the attested example (14a) is acceptable, whereas the combination of PP[about] and PP[to] in the variant is unacceptable. In the case of (14b), *care* does not allow a PP[of] complement, so that the syntactic side of the constraint makes it infelicitous. It does allow a PP[for] complement, but not with the same semantic relationship as its *about* complement, so that, in this case, the semantic side of the constraint makes it infelicitous.

It is unclear how the differences in acceptability presented in (14) can be accounted for under a remnant raising analysis. Consider the putative structures for the variants of (14a):

(15) a. he [$_{\text{VP}}$ spoke t_i] [$_{\text{PP}_i}$ about the season in general] (antecedent of (14a))
b. he [$_{\text{VP}}$ spoke t_i] [$_{\text{PP}_i}$ of the season in general] (elliptical clause of (14a))
c. he [$_{\text{VP}}$ spoke t_i] [$_{\text{PP}_i}$ to the public in general] (elliptical clause of (14a))

If the VP in (15b) is to be deleted under identity with that of (15a), then the content of the trace t_i must be identical in both VPs. As a consequence, the trace must not contain any information about the specific preposition occurring in the raised remnant. But if that is true, how can one rule out deleting the VP in (15c) under identity with that of (15a), and thereby obtaining the unacceptable variant of (14a) with *to*?

In a similar vein, consider (16), in which the antecedent has *give* with a [—NP NP] complementation, whereas the elliptical clause has a [—NP PP[TO]] complementation. Remnant raising leads to non-identical VPs as shown in (17a) and (17b), so that it is unclear how deletion under identity could apply. On the other hand, the proposal sketched here potentially explains the acceptability of (16), as the remnant and its correspondent in the antecedent are both possible complements of *give* and have the same semantic relation to it: they are both recipients.

(16) "[…] It's hard enough to take two hours out of my day to put out a legal fire" — much less give the matter the same attention he would to something that's actually going to generate some cash for the company. (Mag) [Compare: he gave the matter the same attention he would give to something that's actually going to …]

(17) a. [$_{\text{VP}}$ give [$_{\text{DP}}$ the matter] [attention]] ⇒ [$_{\text{VP}}$ give [$_{\text{DP}}$ t_i] [attention]] [$_{\text{DP}_i}$ the matter] (Antecedent of (16))
b. [$_{\text{VP}}$ give [attention] [$_{\text{PP}}$ to the matter]] ⇒ [$_{\text{VP}}$ give [attention] [$_{\text{PP}}$ t_i]] [$_{\text{PP}_i}$ to the matter] (Elliptical clause of (16))

It should be emphasized again that it is not an option to consider that the category of the trace can be ignored. Indeed, this would incorrectly predict that (18b) should be able to mean (18a).

(18) a. He kicked Jay more than he kicked at Joe.

b. #He [$_{VP}$ kicked [$_{NP}$ *t*]] [$_{NP}$ Jay] more than he did [$_{VP}$ ~~kicked [$_{PP}$ *t*]~~] [$_{PP}$ at Joe].

It thus seems that analyses based on remnant raising feeding deletion under identity either undergenerate by predicting that examples like (14a) and (16) are ungrammatical, or overgenerate by predicting that (18b) and the unacceptable variants of (14) are well-formed.[13]

Absence of an appropriate linguistic antecedent Beyond the island and connectivity problems, the deletion under identity approach relies on the hypothesis that there is an appropriate linguistic antecedent in the context. It turns out, however, that this is not always the case. As with VPE and other elliptical phenomena, it is necessary to follow Cornish 1999 in making a difference between the antecedent-trigger (the segment of text that allows one to recover an appropriate antecedent) and the antecedent itself, which Cornish defines as the discourse-model representation making interpretation of the ellipsis possible, which is inferred from the antecedent-trigger. Consider the following examples:

(19) a. "The kids, we, all come here together," said Tommy Foday, a double amputee who at 50 was the oldest of the group. They all called him Pa Tommy, just as they would any village elder in Sierra Leone. (News)

b. Type in your PIN, just hit those buttons like you would a phone. (Spok)

c. EPA urged the Corps "to work directly with the affected communities as well as seek professional assistance in this matter as they would any other environmental issue." (Acad)

In (19a), one might first think that the antecedent is 'call *x* Pa Tommy', but of course this makes no sense at all. The actual antecedent is 'call *x* Pa *y*', where *y* is *x*'s first name. The antecedent clause explicitly provides an antecedent-trigger, which is underlined, but the actual antecedent is obtained by non-grammatical inference. Similarly, in (19b) the intended interpretation is not that one 'hits a phone', but rather that you should use the system in the same way you would use a phone, namely, hitting the *buttons* of the phone. Finally, in (19c), clearly the intended interpretation is 'the Corps would act with respect to any other environmental issue by working directly with the communities affected by the issue and seeking professional advice on the issue'. The intended antecedent is of course easy to infer from the previous clause, but is nowhere present in the appropriate syntactic form.

To the extent that cases like these are grammatical, and I do not see any reason to consider that they aren't, they raise tremendous problems for any analysis that requires any form of syntactic and/or semantic identity. They seem to require an analysis of the type proposed in Miller 1990 and Hardt 1993 where the auxiliary is treated as a proform, and general proform resolution processes find an appropriate antecedent. Of course, as I noted at the time, such an approach considerably overgenerates. It is beyond the scope of this paper to investigate

[13]In the corpus one finds six cases where an accusative remnant shows up instead of an expected possessive, as in *The music gets you feeling good, you start cracking some drinks, pretty soon there's some girls there and the music, it sets your soul on fire. It does me, anyway.* (Mag) [Compare: *It does mine, anyway]*. These seem at first sight to be instances of violations of connectivity. However, cases like these appear to be all the more acceptable that, in the event denoted by the PG clause, affecting the possessor results in a situation similar to affecting the possessed (thus, setting someone's soul on fire, in the intended metaphorical sense, is essentially the same thing as setting the person on fire). As pointed out by an anonymous reviewer, this actually suggests an alternative analysis, where one assumes that the nonelliptical variant of the PG clause is simply *it sets me on fire*, removing the apparent connectivity violation.

precisely how this can be dealt with, but I suggest that an appropriate solution might be found by a combination of (i) semantic conditions of the type sketched above in the discussion of remnant nonconnectivity; and (ii) processing constraints involving preferences for parallelism (cf. Dubey et al. 2005 and Frazier and Clifton 2001).[14]

4.2 Base-generated Interpretive Analyses

Kubota and Levine 2014 develop a Hybrid TLCG analysis of PG, arguing that it allows a synthesis of the transformational and nontransformational approaches. In a nutshell, the flexibility of constituency in the system allows analyses where both the syntax and semantics of the ellipted material are explicitly represented as a constituent in the derivation of the antecedent clause (this is even possible when the antecedent forms a discontinuous sequence, as in (16) above). The difficulty for this approach, however, is to avoid overgeneration by constraining flexibility (so that not just anything can be ellipted) without at the same time ceasing to generate well-formed cases. It remains to be seen how far in this direction the analysis can be extended. I believe, for instance, that the system will have trouble making the relevant differences among the cases of connectivity violation discussed above, cf. (14)-(18). It will also have a problem with the cases without appropriate syntactic antecedents. A detailed discussion goes beyond the scope of this paper. However, the categorial approach is superior to the remnant raising approach because it is at least capable of producing all of the acceptable sequences. The resulting overgeneration could then be reigned in by semantic and processing constraints of the type sketched section.[15]

5 Is Pseudogapping a Subcase of VPE?

Some authors (e.g. Levin 1986, Hoeksema 2006) have argued that PG and VPE exhibit too many differences to make it reasonable to analyze them as instances of the same construction. While it is clearly true that they cannot be exactly the same construction, it is not clear that there is a greater difference between PG and VPE than between the different subconstructions traditionally grouped together under the VPE label. Given the similarities between PG and VPE, the question arises of the extent to which it might be possible to consider them as subconstructions of a more general VPE construction, explaining away the differences on the basis of a better understanding of their discourse uses. In this section, I will address some of the most commonly noted differences between PG and VPE, but will only be able to provide a very preliminary and somewhat speculative discussion, in particular because the details of the usage of the sub-

[14]Since writing the initially submitted version of this paper, Thoms (to appear), a new remnant raising analysis, has come to my attention. Taking Thoms' analysis fully into account would have required major revisions of the present paper, which were not possible. It should be noted, however, that one of the significant advances of the paper is that it provides distinct analyses for noncomparative and comparative PG, treating the former as a more restricted variant of stripping. The latter idea is apparently quite compatible with the results reported here, specifically the contrastive focus status of the remnants. Thoms' syntactic analysis predicts that only noncomparative PGs should be subject to island constraints (though the details of how comparative PGs escape them are not fully worked out), something which appears to be true, but which I would suggest should be accounted for on discourse grounds. The nonconnectivity problems, which appear both in comparative and noncomparative PG, might be addressed by postulating null prepositions (Thoms, p.c.). However, cases without appropriate syntactic antecedents clearly remain a problem.

[15]Following the line or research summed up in Frazier 2013, an alternative approach to all the data discussed in this section would be to consider that the cases which are apparently problematic for the remnant raising analysis are in fact ungrammatical, but repaired ('recycled', to use the terminology of Frazier and her colleagues). It is beyond the scope of this paper to explore the potential advantages and disadvantages of such an approach.

constructions of VPE are not sufficiently known to allow well-supported conclusions. Bos and Spenader 2011 provide some information, but as their study is based on the *Wall Street Journal*, it is not representative of registers with very different properties, such as the spoken register.

PG occurs mainly in comparative environments As shown in Table 1, this is clearly true. In fact, as mentioned, the larger corpus studied here shows an even stronger bias in favor of comparatives than noted in previous studies. And, contrary to VPE, noncomparative PG is almost exclusively a spoken register phenomenon. But these differences in distribution may not be as relevant as one might initially think, for several reasons. First, VPE itself very frequently occurs in comparative structures (Bos and Spenader 2011 find 31% of cases, far less than for PG, but still far more than the percentage of comparative clauses overall). Second, as is the case with PG, the properties of VPE are significantly different in comparative and noncomparative structures. For instance, it was noted by Levin 1986:3 that noncontrastive adjuncts are dispreferred after VPE (an observation corroborated by psycholinguistic experiments reported in Miller and Hemforth 2014). However, it appears that this is not true in comparative VPE.

PG prefers to have the same subject as its antecedent, as opposed to VPE Both Levin and Hoeksema argue that PG exhibits a preference for having the same subject as its antecedent and, in this, differs from VPE. The COCA corpus data confirms the preponderance of same subject PG: 79% of the cases are same subject cases. However, Miller 2011 found that 83% of his sample of 122 occurrences of VPE from the COCA had the same subject as their antecedent. If this result is representative of the COCA in general, the alleged difference disappears.

PG does not allow voice mismatches whereas VPE does Merchant 2008 claims that contrary to VPE, PG does not allow voice mismatches and attributes this to a difference in the target of deletion in the two cases, namely, in VPE the target is a node lower than [voi(ce)], whereas with PG the target contains [voi(ce)]. It is important to note that he qualifies this in a footnote, suggesting that there might be variation in the target of deletion for PG, an idea which is again mentioned in Merchant 2013, citing other studies that have claimed that voice-mismatch is possible in PG. The data from the corpus suggest that the situation is more complex. Indeed, it contains 10 occurrences (out of 1415) of PG with voice mismatches, among which those in (20):

(20) a. A whole poached wild striped bass should be taken to the table as you would a Thanksgiving turkey or a crown roast of pork, with a twinkle of extravagance. (News)

b. I mean for her to be dressed—and addressed—as we would Becky Sharp, or Ophelia, or Elizabeth Bennet, or Mrs. Ramsay, or Mrs. Wilcox, or even Hester Prynne. (Mag)

c. These savory waffles are ideal for brunch, served with a salad as you would a quiche. (Mag)

This should be compared to the data of Bos and Spenader (p.c.), who find no examples of voice mismatch with VPE in the 487 occurrences of their sample of the *Wall Street Journal*. It thus seems at first sight that voice mismatches are actually more frequent with PG than with VPE. This conclusion requires significant qualification.

To begin with, contrary to VPE, mismatches are only found in comparative PG. With noncomparative PG they are systematically very degraded. In the light of the discourse conditions on felicitous voice mismatch with VPE proposed by Kertz 2010, 2013, the absence of

voice mismatches with noncomparative PG can be easily explained, given the discourse conditions on noncomparative PG proposed above. Indeed, Kertz argues that voice mismatches are only acceptable in cases of what I have called Aux-Choice VPE (specifically without additional Subj-Choice). Otherwise, they lead to violations of an information-structural constraint on contrastive topics. Since, as discussed above, noncomparative PG can never be simply Aux-Choice, the unacceptability of mismatches can be expected to follow from a similar constraint. The question then arises of why Aux-choice comparative VPE structures do not seem to allow voice mismatches (as noted by Kertz). If we consider (20a) above, turning it into VPE with a contrasting subject clearly strongly reduces its acceptability: *??A poached bass should be taken to the table as an excellent butler would.* It thus appears that the presence of the remnant plays a crucial role in making these acceptable: it provides a correspondent to the subject of the antecedent which is absent in the VPE variant. The topic clearly warrants further research.

PG does not occur cataphorically, contrary to VPE Though cataphoric uses of PG are very hard to construct, there is one example in the COCA, which does not seem to be a speech error:

(21) Behind them, disguising her desire, one catches a poignant glimpse of the youthful, shaved-headed Cather. As it did me, work rescued Willa Cather. (Fic)

Bos and Spenader 2011 found four cases of cataphoric VPE in the 487 examples of their corpus, which suggests a significant difference in frequency. A preliminary investigation of cataphoric VPE in the COCA suggests a possible account for this discrepancy: cataphoric VPE appears to be almost always Aux-choice. As PG is never Aux-choice, this might explain the difference.

PG involves a single auxiliary Levin 1986 and Hoeksema 2006 mention that though sequences of auxiliaries are possible with VPE, they are hard to attest with PG (though Levin did collect one example: *I processed everybody's [check] but I must not've yours.* Levin (1986), p.18, ex. (34)). This may again be due to the impossibility of simple Aux-choice PG. Indeed a simple search on the COCA suggests that VPE with multiple auxiliaries is almost never Subj-choice. Namely, searching for the sequence "could|would have too." (the most frequent context for Subj-choice VPE) in the COCA provides only two examples of Subj-choice *would have too* and none of *could have too*, whereas searching for "could|would too." gives 47 examples of Subj-choice *would too* and 19 of *could too.* On the other hand, Aux-choice uses of VPE with *could have* and *would have* are much more frequent (several hundred of each).

PG does not occur with infinitival to *contrary to VPE* It is a striking fact that VPE is possible after infinitival *to*, but that PG is not. This seems to be a robust property: to my knowledge not a single attested example of PG with *to* has been found in corpora, and constructed examples are hopeless (e.g. **She may not visit you but she has to me.*) However, it is possible that this is once again linked to a conspiracy of discourse factors. A quick look at the COCA suggests that almost all cases of VPE with *to* are Aux-choice. Beyond this, it may also be the case that VPE after *to* in comparatives is dispreferred more generally. In the following examples from the COCA, *to* is followed by *do* (this is one of the rare contexts where nonfinite auxiliary *do* is possible in American English, as opposed to British English, where it has a broader variety of uses). The variants where *to* is removed appear to be degraded:

(22) a. Here Santayana, like Dickens, delights in the existence of ordinary humanity as

Emerson never quite manages to do. [Compare: ?as E. never quite manages to.]

b. Rather, they should respond by taking the opinion's reasons seriously and exploring those reasons' implications—as this Article hopes to do. [Compare: ?as this Article hopes to.]

c. maybe he would have overcome public censure and gained acceptance for the possibility of being both Mohawk and Christian, as later converts to Protestantism apparently were able to do. [Compare: ?were able to.]

6 Conclusion

This study presents preliminary conclusions based on a first analysis of the corpus data. Some questions have been left almost completely unanswered, such as why noncomparative PG is apparently less acceptable than comparative PG. For other problems, only a sketch of an analysis has been provided. This is the case for the statement of the discourse conditions on PG, as well as for the details of the way in which the overgeneration of the suggested auxiliary-as-proform analysis can be reigned in by semantic constraints and processing preferences. On all these questions, it will be necessary to further investigate the corpus materials gathered and to conduct psycholinguistic experiments to test hypotheses in full detail.

It seems to me, however, that the central merit of the present study is that it shows that the complexity of the data on PG has been vastly underestimated. More generally, a study like this one makes apparent the limitations of the standard generative methodology of data collection by introspection. Even as professional linguists, our introspective abilities are simply unable to come up with the relevant range of data. And, in a case like the one under study, this clearly has important consequences for the mainstream theoretical proposals.

Beyond this, a corpus study makes apparent some of the Usage Preferences (UPs, see Miller 2013) governing the use of PG. Because of the cumulative effect of multiple UP violations, it is crucial to base one's analyses on examples that conform to UPs, which is not possible until these have been established. For instance, Lasnik 1999 cites a single example of a comparative PG at the beginning of his paper, but invents all of his other data, which are noncomparative with full NP subjects. The present study makes apparent that such cases are dispreferred, even when nothing else is wrong with them, so that any other problems they might exhibit (e.g. lack of parallelism, less accessible antecedents) might lead to a feeling of strong unacceptability. It is thus crucial to understand the usage factors that make examples more or less acceptable if one is to disentangle what are truly grammatical constraints from other factors affecting acceptability.

References

Ariel, Mira. 1990. *Accessing noun phrase antecedents.* London: Routledge.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman grammar of spoken and written English.* Harlow: Longman.

Bos, Johan, and Jennifer Spenader. 2011. An annotated corpus for the analysis of VP ellipsis. *Language Resources and Evaluation* 45:463–494.

Cornish, Francis. 1999. *Anaphora, discourse, and understanding: Evidence from French and English.* Oxford: Oxford University Press.

Culicover, Peter, and Ray Jackendoff. 2005. *Simpler syntax.* Oxford: Oxford University Press.

Davies, Mark. 2008-. The Corpus of Contemporary American English: 450 million words, 1990-present. http://corpus.byu.edu/coca/.

Dubey, Amit, Patrick Sturt, and Frank Keller. 2005. Parallelism in coordination as an instance of syntactic

priming: Evidence from corpus-based modeling. In *HLT '05 Proceedings of the Conference on Human Langugage Technology and Empirical Methods in Natural Language Processing*, 827–834. Stroudsburg, PA: Association for Computational Linguistics.

Francis, Hartwell S., Michelle L. Gregory, and Laura A. Michaelis. 1999. Are lexical subjects deviant. In *Papers from the 35th Regional Meeting of the Chicago Linguistic Society*, 85–97. Chicago, IL.

Frazier, Lyn. 2013. A recycling approach to processing ellipsis. In *Diagnosis in syntax*, ed. Lisa Lai-Shen Cheng and Norbert Corver, 485–501. Oxford: Oxford University Press.

Frazier, Lyn, and Charles Clifton, Jr. 2001. Parsing coordinates and ellipsis: Copy α. *Syntax* 4:1–22.

Gengel, Kirsten. 2013. *Pseudogapping and ellipsis.* Oxford: Oxford University Press.

Hardt, Daniel. 1993. Verb phrase ellipsis: Form, meaning, and processing. Doctoral Dissertation, University of Pennsylvania. Distributed as IRCS Report 93-23.

Hoeksema, Jack. 2006. Pseudogapping: its syntactic analysis and cumulative effects on acceptability. *Research on Language and Computation* 4:335–352.

Jayaseelan, K. A. 1990. Incomplete VP deletion and gapping. *Linguistic Analysis* 20:136–155.

Kertz, Laura. 2010. Ellipsis reconsidered. Doctoral Dissertation, University of California, San Diego, La Jolla, CA.

Kertz, Laura. 2013. Verb phrase ellipsis: The view from information structure. *Language* 89:390–428.

Kubota, Yusuke, and Robert Levine. 2014. Pseudogapping as pseudo-VP ellipsis. In *LACL 2014, LNCS 8535*, ed. Nicholas Asher and Sergei Soloviev, 122–137. Berlin: Springer-Verlag.

Kuno, Susumo. 1981. The syntax of comparative clauses. In *Papers from the Seventeenth Regional Meeting of the Chicago Linguistic Society*, 136–155. Chicago, IL.

Lasnik, Howard. 1999. Pseudogapping puzzles. In *Fragments: Studies in ellipsis and gapping*, ed. Shalom Lappin and Elabbas Benmamoun, 68–97. Oxford: Oxford University Press.

Levin, Nancy. 1986. *Main verb ellipsis in spoken English.* Outstanding Dissertations in Linguistics. New York: Garland.

Merchant, Jason. 2008. An asymmetry in voice mismatches in VP ellipsis and pseudogapping. *Linguistic Inquiry* 39:169–179.

Merchant, Jason. 2013. Voice and ellipsis. *Linguistic Inquiry* 44:77–108.

Miller, Philip. 1990. Pseudogapping and *Do so* substitution. In *Papers from the 26th Regional Meeting of the Chicago Linguistic Society*, 293–305. Chicago, IL.

Miller, Philip. 2011. The choice between verbal anaphors in discourse. In *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium, DAARC 2011*, ed. Iris Hendrickx et al., volume 7099 of *Lecture Notes in Artificial Intelligence*, 82–95. Berlin: Springer.

Miller, Philip. 2013. Usage preferences: The case of the English verbal anaphor *do so*. In *Proceedings of the 20th International Conference on Head-Driven Phrase Structure Grammar*, ed. Stefan Müller, HPSG Online Proceedings, 121–139. Stanford: CSLI Publications.

Miller, Philip, and Barbara Hemforth. 2014. Verb phrase ellipsis with nominal antecedents. Ms. Université Paris Diderot.

Miller, Philip, and Geoffrey K. Pullum. 2014. Exophoric VP ellipsis. In *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag*, ed. Philip Hofmeister and Elisabeth Norcliffe, 5–32. Stanford, CA: CSLI.

Schachter, Paul. 1978. English propredicates. *Linguistic Analysis* 4:187–224.

Seoane Posse, Elenea. 1999. The consolidation of the indirect and prepositional passive in Early Modern English: evidence from the Helsinki Corpus. *Estudios Ingleses de la Universidad Complutense* 7:119–139.

Sharifzadeh, Saghie. 2012. Recherches sur *do*, lexique et grammaire. Doctoral Dissertation, Université de Paris-Sorbonne, Paris.

Thoms, Gary. To appear. Pseudogapping, parallelism and the syntax of focus. *Syntax*.

Université Paris Diderot, Sorbonne Paris Cité, EA 3967 CLILLAC-ARP
philip.miller@univ-paris-diderot.fr