Negative Stripping in English: A Corpus-based Approach

Geonhee Lee[1], Jong-Bok Kim[2†]
Kyung Hee University

Geonhee Lee, Jong-Bok Kim. 2021. Negative Stripping in English: A Corpus-based Approach. *Language and Information*, 25.3, 23-42. Stripping, also called bare argument ellipsis, is taken to be an elliptical construction that unexpresses everything from a clause except one constituent (e.g., *Susan works at night, and Bill too.*). It can also accompany the negator *not*, which we call Not-Stripping (e.g., *Alan likes to play volleyball, but not Sandy.*). This paper reviews key grammatical properties of Not-Stripping in English and investigates its authentic uses with the online corpus COCA (Corpus of Contemporary American English). The paper then performs some quantitative and qualitative research of the extracted data and discusses theoretical implications. The result from corpus investigation shows that Not-Stripping preferably used as PP remnant. Its register distribution also shows that Not-Stripping is most frequently used in fiction register followed by spoken, and magazine registers which may indicate that Not-Stripping is mainly used in colloquial context. In terms of category match/mismatch types, the construction is overtly used as mismatching relation between the remnant and the correlate. In addition, in case of merger/sprouting types, Not-Stripping seems to favor merger type, but sprouting type can be regarded as frequently used type. The grammatical function of both remnant and correlate are preferably used as modifier. The data for Not-Stripping are dealt with move-and-delete operations to postulate putative clausal source. Observing such attested data challenging the derivation of negative stripping from syntactic movement operations, the paper challenges a direct licensing construction-based analysis of the construction.

Key words: stripping, ellipsis, corpus, negation, move-and-delete, direct licensing

[1] First author, Kyung Hee University, Department of English Language and Literature, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, E-mail: geonhee.lee@khu.ac.kr

[2†] Corresponding author, Kyung Hee University, Department of English Language and Literature, 26 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, E-mail: jongbok@khu.ac.kr

1. Introduction

Stripping is a type of ellipsis that elides everything in a clause except for one constituent, as illustrated in the following examples from Wikipedia (Hankamer & Sag 1976, Merchant 2003, Wurmbrand 2017, Kim & Abeillé 2019):

(1) a. Susan works at night, and Bill too.
     b. Why did Sam call, and Bill too?
     c. Should I do it, or you?
     d. Chris said yesterday that he knew it, and today too.
     e. She asked the kids to stay, and the adults too.

As seen from these, the second conjunct has only one constituent, and the elided part in (1a) which is *works at night* seems to be under identity with corresponding parts of the preceding clause. Also, stripping in (1c) only contain a single constituent, whereas in other examples, the expression *too* is followed by the second conjunct, indicating that Stripping can be used in various ways. Note that Stripping can also accompany the negator *not*, as seen in the following attested data:

(2) a. One or two others chuckled, but not Mary. (COCA 1996 FIC)
     b. He came back with two other officers. They were angry, but not with me. (COCA 2017 MAG)
     c. That's a daunting challenge, but not impossible. (COCA 2017 NEWS)
     d. That is a bit of a rhetorical question, but not completely. (COCA 2012 BLOG)

Such Not-Stripping (a.k.a negative stripping) examples, a subtype of Stripping, typically correct an assumption in the preceding context. This paper focuses on Not-Stripping in English.

There are two possible directions to account for such (Not)-Stripping: sentential and direct licensing approaches. Sentential approaches, dominant in generative grammar, take stripping to arise from complete sentential syntax together with move-and-delete operations (Merchant 2004, Weir 2014), whereas direct licensing approaches introduce no further syntax beyond what appears at the surface (Ginzburg & Sag 2000, Culicover & Jackendoff 2005, Jacobson 2016, Kim 2020). This paper performs a corpus investigation of Not-Stripping using the corpus COCA (Corpus of Contemporary American English). It suggests that the yielded data (total 1,136 tokens of negative NP fragments we extracted) raise nontrivial issues in postulating clausal sources for Not-Stripping. For instance, similar to (2), there are a significant number of

Not-Stripping examples that their putative clause of stripped clause would be unacceptable:

(3)   a. I have moved on to new ones, but not he. (COCA 2012 WEB)
        b. Most artists at that level have sponsors, but not me. (COCA 2016 MOV)

This paper first reviews key grammatical properties of Not-Stripping in English and then investigates the uses of Not-Stripping in real-life with COCA. The paper in particular performs basic quantitative and qualitative research of the extracted corpus data and discusses their possible theoretical implications.

2. Some Key Properties

As noted in the beginning, Not-Stripping is a type of bare argument ellipsis that leaves out a single constituent from the putative clausal source, and typically accompanies a focus-sensitive negation or particle like *not* or *too* (Hankamer & Sag 1976, Merchant 2003). The syntactic category of the stripped remnant with the negator *not* has a wide range as it can be NP, AP, PP, VP, or even CP:

(4)   a. Mary ate an apple, but not [NP an orange].
        b. He's already found that may be a nice idea, but not [AP easy]. (COCA 1995 SPOK)
        c. John likes to sing in his room, but not [PP in the park].
        d. The pancakes take center stage at this restaurant, but not [AdvP today]. (COCA 2012 SPOK)
        e. I was ready to declare victory, but not [VP relax my vigilance]. (COCA 1999 NEWS)
        f. She did tell him that her father had been an officer, but not [CP that he had been a lieutenant general].

The remnant of Not-Stripping can bear a variety of grammatical functions:

(5)   a. John likes swimming, but not Paul. [Subject]
        b. John likes swimming, but not surfing. [Direct object]
        c. T/T is just a payment method, but not payment term. (COCA 2012 WEB) [Predicative complement]
        d. They divvy up liberty into thousands of bits; dozens, maybe, but not thousands. (COCA 2012 WEB) [Prepositional object]

e. The secret ingredients for US fracking are being shared with China, but not with the EPA. (COCA 2012 BLOG) [Oblique complement]

f. That's a daunting challenge, but not impossible. (COCA 2017 NEWS) [Modifier]

As observed above, the remnant is introduced by the coordination marker *and* or *but*, implying that Stripping is applied in coordinating structures:

(6)   a. Everybody gets to eat what they want, and not me. (COCA 2006 MOV)

       b. Dad might take it, but not me. (COCA 2017 FIC)

Such examples indicate that the second conjunct is not an NP, but has some clausal properties. Note that Not-Stripping cannot occur in the subordinating clause (Wurmbrand 2017):

(7)   a. *Bill went to the store, because not Susan.

       b. *John left early, after not Mary.

(Not-)Stripping is taken to be surface anaphora that requires a coherent syntactic antecedent, as noted from a typical example in the following (Hankamer & Sag 1976):

(8)   Alan likes to play volleyball, but not Sandy △.

       △ = likes to play volleyball

The construction does not license deep anaphora, as seen from examples in the following example with a discourse antecedent (Hankamer & Sag 1976, Merchant 2003):

(9)   [Context: Sag plays William Tell Overture on recorder]

       Hankamer: #Yeah, but not very well.

The speaker makes her response based on the context, without referring to a linguistic antecedent. This shows that Not-Stripping is rather a syntactically controlled anaphoric phenomenon.

3. Move-and-delete Analyses

Within the Minimalist Program, (Not-)Stripping has been taken to involve clausal ellipsis

with move-and-delete operations (Merchant 2003). The move-delete approaches introduce an unpronounced 'clause' source with the operation of moving the focused remnant to a functional projection and deleting the remaining parts (Merchant 2001, van Craenenbroeck 2010, Weir 2014, Wurmbrand 2017). This is illustrated in the following:

(10)   John met Mary yesterday, but not$_i$ Sally$_j$ [J̶o̶h̶n̶ ̶d̶i̶d̶ t_i m̶e̶e̶t̶ t_j].

In (10), the negator *not* and the direct object *Sally* in the unpronounced clause are moved outside the clause as the focused remnant and the rest of the unpronounced clause is deleted. The supporting arguments for such move-and-delete approaches mainly come from connectivity effects (Merchant 2016). For instance, the preposition of the remnant must be identical to that of the correlate as in (11a) and this could be explained by assuming that it is derived from a clausal source in (11b):

(11)   a. John relies on Mary, but not on/*of Susan.
        b. John relies on Mary, but [John does not rely on/*of Susan].

P-stranding in Not-Stripping also seems to support the clausal analysis. English, as a preposition stranding language, allows the optional presence of a preposition in an elliptical construction like Sluicing. Not-Stripping allows the optionality of a preposition:

(12)   a. I spoke with Kim yesterday, but not (with) Lee.
        b. I rely on Kim, but not (on) Lee.

Setting aside the treatment of *not*, we may account for the preposition optionality in (12a) by assuming that the derivation involves a clausal source like the following:

(13)   a. With Lee, I̶ ̶d̶i̶d̶ ̶n̶o̶t̶ ̶s̶p̶e̶a̶k̶ ̶y̶e̶s̶t̶e̶r̶d̶a̶y̶.
        b. Lee, I̶ ̶d̶i̶d̶ ̶n̶o̶t̶ ̶s̶p̶e̶a̶k̶ ̶w̶i̶t̶h̶ ̶y̶e̶s̶t̶e̶r̶d̶a̶y̶.

In (13a), the preposition is pied-piped with the focused NP *Lee*, while in (13b), it is stranded.
    Connectivity effects can also be observed when the remnant is a reflexive, as in (14a). The postulation of its clause source as in (14b) then would easily predict its licensing.

(14)   a. John$_i$ relied on Mary, but not himself$_i$.
        b. John relied on Mary, but John$_i$ did not rely on himself$_i$.

The clausal property could be found from the meaning of the remnant:

(15)   Abby speaks passable Dutch, but not Ben.
        (not Ben = Ben does not speak passable Dutch.)

The proper meaning of the remnant *not Ben* is a clausal one as given in the parentheses.

   The clausal property of the second conjunct is also supported from the presence of a sentential adverb as in (16):

(16)   a. They shop at places like Banana Republic or Bloomingdales, but probably not at Barneys.
        b. They will also offer such systems to their customers, but possibly nat the 20kWh scale.
        c. The bevels are ridiculous, but fortunately not too noticeable.

As seen in (16), sentential adverbs such as *probably*, *possibly* and *fortunately* can appear before XP that holds sentential meaning.

   The possibility of having a sloppy reading indicates the clausal property of Not-Stripping (Merchant 2003):

(17)   You can keep Rosa in her room the whole day, but not Zelda.

The sentence in (17) could have not only a strict reading as in (18a) but also allows a sloppy reading as in (18b):

(18)   a. You can keep Rosa$_i$ in her$_i$ room the whole day, but you cannot keep Zelda in her$_i$ room the whole day
        b. You can keep Rosa$_i$ in her$_i$ room the whole day, but you cannot keep Zelda$_j$ in her$_j$ room the whole day.

   In addition to the clausal properties of Not-Stripping, there seem to be some supporting arguments for moving the focused expression to the clause initial position. One often cited argument concerns the fact that Not-Stripping is sensitive to Island constraints (Lappin 1996, Drubig 2003, Merchant 2004):

(19)   a. *We have interrogated [the burglar who stole the car already], but not the diamonds. [Complex NP Constraint]

b. *[The request that John submit articles to the journal] pleased John, but not book reviewers. [Complex NP Constraint]

c. *[That Mary was late again] annoyed Peter, but not John. [Sentential Subject Constraint]

As discussed so far, there seem to be arguments that support the postulation of clausal sources and movement operations for Not-Stripping. In what follows, we will see if the attested data can support this view.

4. Corpus Data and Methodology

4.1. Data Collection

In collecting the authentic data of Not-Stripping, we used the corpus COCA, which is composed of more than one billion words in 485,202 texts from evenly distributed registers, including 24-25 million words each year from 1990 to 2019. To extract relevant data, we used simple search strings including *but not* and some regular expressions. With the additional restriction of registers excluding TV, BLOG, and MOV, we initially obtained a total of 54,264 tokens of Not-Stripping. From these, we randomly extracted 700 tokens, excluding irrelevant examples like the following:

(20)   Yes, today's Black woman wants a man, but not just any man will do. (COCA 1992 MAG) [finite clause]

After manually excluding such examples, we had a total of 205 tokens and performed a quantitative and qualitative analysis of them.

4.2 Methodology

In analyzing the tokens, we introduced 6 variables: 1) registers, 2) remnant category types, 3) correlate category types, 4) category match/mismatch, and 5) merger/sprouting types[1] and

---

1) According to Chung et al. (1995), the merger refers to a type that the remnant has a concrete correlate

6) grammatical function types. Assigning registers as a variable, we intend to see how the actual sample data from COCA is distributed in terms of register: whether it is used in the formal or informal context. For the remnant and correlate category types, we try to identify the mapping relation between those two: whether they have the exact same identity or not, and if not, then how its correlate categories are distributed in each remnant type. For the merger and sprouting types, we try to figure out whether the remnant requires a linguistic antecedent or not. With the variable for grammatical functions, we expect to identify if there are any preferences for the remnant of Not-Stripping.

We also performed a statistical analysis to understand the central tendency and its dispersion (Gries 2013, Lee 2016). Since the variables used for the token analyses are categorical variables, the central tendency is observed as the mode value. In addition, dispersion of the data is measured by Relative Entropy represented as H_{rel}, which is a test for measuring dispersion for the data of categorical variables. If H_{rel} value is close to 1, it represents that the distribution of data is equally frequent among different levels within the variable; if the value is close to 0, it indicates that it is in a single level of distribution.

4.3 Findings

4.3.1 Overall distribution

The following table shows the frequency of the category of the remnants, hinting general preference of each remnant type:

<Table 1> Frequencies of the remnant types by category

AP	AdvP	NP	PP	VP
34	41	44	66	20

As seen in Table 1, PP is the most frequently used remnant type in the 205 instances. The PP frequency is followed by NP, AdvP, and so on. The H_{rel} value for the frequencies is 0.9581534, implying that its data dispersion is moderately equally dispersed in all levels of the variable.[2]

4.3.2 Register Distribution

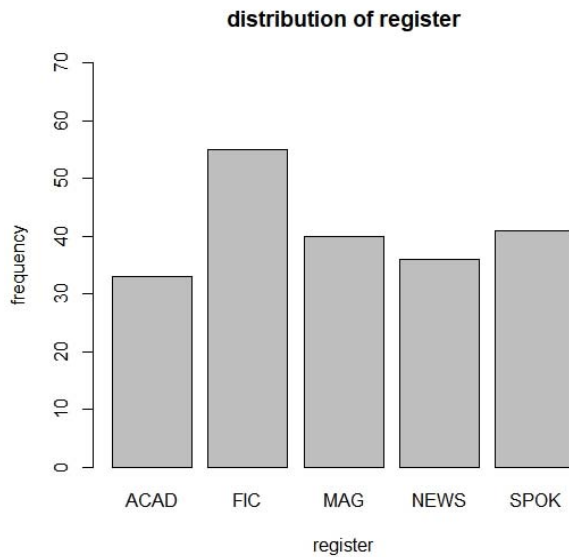The instances of Not-Stripping with respect to their registers show us in what context

---

in the linguistic antecedent, whereas the sprouting refers to the type that the remnant has no overt correlate.
2) A caveat is that only looking at this macro-level of the limited data is not enough to draw any generalizations.

Not-Stripping is preferably used. As mentioned earlier, the registers of the corpus consist of five types: academic, fiction, magazine, news, and spoken. The following table illustrates the overall register distribution of Not-Stripping among five registers:

distribution of register

<Figure 1> Distribution of Not-Stripping from COCA by registers

As in Figure 1, the frequency of the fiction register is the highest, followed by spoken and magazine registers. This implies that Not-Stripping has a preference in colloquial context rather than in formal contexts like news and academic registers. However, except fiction register, it is notable to see fairly evenly distributed frequencies from other registers such as academic register. Academic register consists of formally written texts where informal usage of language are rather inappropriate. The following sentence is the instance used in academic register:

(21)   He can take preventive measures against terrorism but not against Iranian missiles. (COCA 2015 ACAD)

The figure also indicates that except for the fiction register, the general distribution of the construction is equally dispersed. A statistical analysis can tell us if such distributions are indeed relatively equally distributed. Statistically, 55 instances in the fiction register is the

mode value that shows the central tendency regarding register distribution, which is a categorical variable. The dispersion can be calculated with relative entropy, and the H_{rel} value is 0.9899249. Since the value is close to 1, its dispersion is relatively equal meaning that there is no significant register difference in the uses of Not-Stripping.

4.3.3 Category Match and Mismatch types

In typical Not-Stripping cases, the remnant and its correlate share the same grammatical category:

(22)   a. *Alan* likes to play volleyball, but not *Sandy*.
        b. Sports are *tricky*, too, but not *impossible*.

As seen from the examples, the category of the remnant in the second conjunct matches that of its correlate in the first conjunct. However, there are also a variety of category mismatching cases between the two, as seen from the following table:

<Table 2> Distribution of Match and Mismatch type

remnant	MATCH	MISMATCH	H_{rel}
AP	15	18	0.9940302
VP	16	4	0.7219281
NP	38	7	0.6235702
PP	11	55	0.6500224
AdvP	6	35	0.6006086
Total	86	119	0.981226

As the Table 2 shows, the frequencies of mismatching outnumber those of the matching (119 vs. 86). This implies that Not-Stripping allows a number of category mismatching cases between the remnant and its correlate. However, as the H_{rel} value is 0.981226, the category mismatching is equally dispersed in all levels of the variable.

In what follows, we look into the category matching and mismatching cases in each remnant type. First, let us see the AP remnant, which has the strongest H_{rel} value. The following table shows the correlate type with the AP remnant:

<Table 3> Distribution of correlates with the AP remnant

NP	VP	AP	PP	AdvP	COVERT
3	2	15	1	0	12

The AP correlate has the highest frequency number, which implies that the AP remnant prefers to have the matching category. However, note that there is a covert correlate, which is not overtly realized but could be provided in the context. This covert correlate type is the second highest frequency, hinting that Not-Stripping can be context-dependent. Also, considering two types of adjectives: predicative and attributive, the predicative usage of AP as a remnant is more frequently used in Not-Stripping, as illustrated in the following examples, respectively:

(23) a. The views are [AP lovely] from the lighthouse, but not [AP worth the lengthy drive to see them]. (COCA 1998 NEWS) [Match-Merger]
      b. They are men, [AP different] from us in ways, but not [AP very different]. (COCA 2001 FIC) [Match-Merger]
      c. The lenses should be [AP dark enough] to shield your eyes from bright lights and glares but not [AP so dark] that they compromise your vision in a low-light environment. (COCA 2003 MAG) [Mismatch-Sprouting]

(24) That will be a [AP hard] task, but not [AP impossible]. (COCA 2012 WEB)

Statistically, its mode value is 15 from the AP remnant, and its dispersion value, H_{rel} is 0.9940302. This means that it is equally distributed among the levels within the variable.

The following table shows the distribution of frequencies regarding correlates of the VP remnant:

<Table 4> Distribution of correlates with the VP remnant

NP	VP	AP	PP	AdvP	COVERT
1	16	2	0	0	1

In the case of VP remnant, VP correlate type is most frequently used. This means that similar to AP remnants, category matching between the remnant and the correlate is strongly favored. Mismatching cases are quite rare. The following examples show matching and mismatching tokens:

(25) a. If it can, then Labour might [VP fray] at the edges but not [VP shatter]. (COCA

2015 MAG) [Match-Merger]

     b. ...told by a doctor who was taught [VP to clench] his teeth but not [VP to cry]. (COCA 2011 FIC) [Match-Merger]

     c. ...especially when they were [PP both in Rome] but not [VP living together]. (COCA 1990 MAG) [Mismatch-Merger]

The mode value of matching tokens is thus 16, and its dispersion H_{rel} value is 0.7219281.

    As the following table shows, when the remnant is NP, the correlate favors matching NP:

<Table 5> Distribution of correlates with the NP remnant

NP	VP	AP	PP	AdvP	COVERT
37	1	0	2	0	3

As other remnant types, NP remnant has a preference of having an overt correlate and favors to have the matching correlate. Some illustrative examples are given in the following:

    (26)   a. People will snap up [NP dogwoods and azaleas] that bloom once, but not [NP roses]. (COCA 2006 NEWS) [Match-Merger]

          b. Some NRA officials said [NP the climate] was shifting slightly, but not [NP the tide]. (COCA 1990 SPOK) [Match-Merger]

          c. Good science is [VP being done], but not [NP the kind of science we need]. (COCA 2013 ACAD) [Mismatch-Merger]

As shown in these examples, the NP remnant can have either a category matching correlate or a mismatching correlate. However, it dominantly has a matching correlate. Statistically, its H_{rel} value is 0.6235702, implying that its dispersion is relatively in a single level rather than evenly distributed among other correlate types.

    The PP remnant somewhat behaves differently from the other remnant types. In particular, the remnant prefers to have a covert correlate, as seen from its frequency in the following table:

<Table 6> Distribution of correlates with the PP remnant

NP	VP	AP	PP	AdvP	COVERT
0	0	0	11	3	52

    The following examples show Not-Stripping with a PP remnant:

(27) a. There may be much more control over crime, but not [PP in a good way]. (COCA 2014 NEWS) [Mismatch-Sprouting]

   b. As an aside, Antarctica did host dinosaurs, but not [PP until the Jurassic Period]. (COCA 2019 MAG) [Mismatch-Sprouting]

   c. ...creating a fetal-tissue bank that could use tissue [PP from miscarriages], for instance, but not [PP from abortions]. (COCA 2001 ACAD) [Match-Merger]

As noted from the examples (27a) and (27b), the PP remnant can have a covert correlate. In terms of dispersion, the H_{rel} value is 0.6500224, meaning that it is not equally but rather moderately dispersed in a certain level compared to the other remnant types.

The remnant can be an AdvP as well, and the distributional frequencies of its correlate types are given in the following table:

<Table 7> Distribution of correlates with the AdvP remnant

NP	VP	AP	PP	AdvP	COVERT
1	1	0	4	6	29

As shown in the table, the AdvP remnant prefers to have an overt correlate rather than a covert one. The following includes illustrative examples:

(28) a. "True love, " she'd [AdvP sometimes] say, but not [AdvP very often]. (COCA 2008 FIC) [Match-Merger]

   b. It was better than a stark dismissal, but not [AdvP much]. (COCA 2008 FIC) [Mismatch-Sprouting]

   c. Children become desensitized to pain, mistrustful and full of rage, but not [AdvP always]. (COCA 1990 SPOK) [Mismatch-Sprouting]

Statistically, the relative entropy H_{rel} value is 0.6006086. This implies that this remnant type is not preferably used and further is not equally dispersed.

4.3.4 Distribution of Merger and Sprouting cases

We have seen the overall distributive pattern of Not-Stripping and analyzed the extracted data with variables including the grammatical category of the remnant. Results show that each remnant type behave differently. The statistical analysis of relative entropy with H_{rel} values also indicates that the inferences drawn from the analysis are statistically significant.

In addition to this, we compared tokens with an overt correlate (merger cases) and those with a covert correlate (sprouting) in order to understand the anaphoric nature of the remnant. The following table shows the frequencies of merger and sprouting with respect to each remnant type:

<Table 8> Distribution of merger and sprouting with respect to the remnant category

remnant	MERGER	SPROUTING	H_{rel}
AP	22	11	0.9182958
AdvP	12	29	0.8721618
PP	15	51	0.7732267
NP	42	3	0.3533593
VP	19	1	0.286397
Total	110	95	0.9961345

As shown from the frequencies, the distribution of merger and sprouting types is not significantly different. Considering that Stripping is surface anaphora, we expect to have a strong preference for merger types, but Not-Stripping often allows context-dependent sprouting instances. Statistically speaking, the mode value for this variable is 110 from merger type, and its relative entropy H_{rel} is 0.9961345 meaning that its dispersion is equally distributed since the value is close to 1.

Let us see each remnant type and which type is preferably used in Not-Stripping. As for the AP remnant, the merger type outnumbers the sprouting type. Interestingly, there are some patterns found in terms of grammatical functions in both types. In the case of merger type, the adjective is used as a predicative complement, while in the case of sprouting type, it is preferably used as a modifier. Statistically, the value of relative entropy H_{rel} is 0.9182958, indicating that it is equally distributed among various levels of this variable.

<Table 9> Frequencies of merger and sprouting types in the AP remnant

MERGER	SPROUTING
22	11

(29) a. Alas, he is armed but not dangerous. (COCA NEWS 1998)

     b. ...a bottle of fentanyl went, which set off gossip among the staff, but not much
       more. (COCA 2015 MAG)

Adverbial remnants have a different tendency from AP remnants: the sprouting outnumbers the merger type, as shown in the following table. Syntactically, they are all modifiers and the mode value for this remnant type is 29, and its H_{rel} is 0.8721618.

<Table 10> Frequencies of merger and sprouting types in the AdvP remnant

MERGER	SPROUTING
12	29

(30) a. She'd sometimes say, but not very often. (COCA 2008 FIC)
      b. She used to be flat and sort of ugly and now she's just flat but not totally. (COCA 1992 FIC)

As for the PP remnant, the frequency of sprouting is much higher than that of merger type. Most of the PP remnants are used as oblique complements in merger cases, but they are used as modifiers in sprouting cases. Statistically speaking, the mode value for this is 53, and the relative entropy value H_{rel} is 0.7732267.

<Table 11> Frequencies of merger and sprouting types in the PP remnant

MERGER	SPROUTING
15	51

(31) a. It's in Intel's interest for several customers to have about 10% market share but not for any of them to be much beyond. (COCA 1994 MAG)
      b. Old Gavril came out, but not before shaving himself with his best razor. (COCA 2004 FIC)

As for the NP remnant, merger is more favored than sprouting. For the merger type, the most frequently used grammatical function is direct object, followed by subject and predicative complement. On the other hand, for the sprouting cases, the modifier is the most used one. The mode value is 40, and its relative entropy value H_{rel} is 0.3533593, which is the second lowest value among other remnant types.

<Table 12> Frequencies of merger and sprouting types in the NP remnant

MERGER	SPROUTING
42	3

(32) a. They can open their own documents but not your desktop check register. (COCA 2005 MAG)

b. Church hierarchy, of course, has a major role in facilitating that growth, but not the only role. (COCA 2007 MAG)

For the VP remnant, the merger instances outnumber the sprouting instances. Unlike other remnant types seen above, the VP remnant has rather a strong preference as the frequency number of merger type is 19 out of 20. For the merger type, the remnant is largely used as predicate. Thus, the mode value for this is 19, and its relative entropy H_{rel} is 0.286397, which is the lowest value among the five remnant types.

<Table 13> Frequencies of merger and sprouting types in the VP remnant

MERGER	SPROUTING
19	1

(33) He would hurt me – maybe, possibly, try to beat me up, but not physically try to kill me. (COCA 1996 SPOK)

4.3.5 Distribution of grammatical functions

To identify the grammatical function of the remnant in Not-Stripping can also help us understand its grammatical properties as well as relations with its correlate. This helps to see if there is any difference between syntactic categories and grammatical functions in Not-Stripping. The following table shows the frequencies of grammatical functions for the remnants and the correlates:

<Table 14> Distribution of grammatical functions of the remnants

Function	Frequency
Modifier	114
Pred-comp	26
Direct-obj	24
Oblique-comp	17
Predicative	16
Subject	8
Total	205

<Table 15> Distribution of grammatical functions of the correlates

Function	Frequency
Modifier	20
Pred-comp	25
Direct-obj	23
Oblique-comp	18
Predicative	18
Subject	9
N/A(covert)	92
Total	205

As seen from the tables, the remnants are used as modifiers in most cases, possibly due to the flexibility of modifiers used in a variety of syntactic environments. As for the correlates, covert correlates whose grammatical functions are not clearly identified are favorably used. Except for the covert uses, the predicative uses are preferred. One interesting property we have observed is that, as given in Table 16, there is no significant difference in the instances of matching and mismatching between the grammatical functions of the remnant and its correlate:

<Table 16> Frequency number of match/mismatch type of grammatical functions

MATCH	MISMATCH
102	103

4.4 Discussion

The extracted corpus data of Not-Stripping have shown us that the remnant is necessary to have matching correlate. However, some exceptions are plausible when there are no overt correlate which appear in the form of sprouting. In terms of negation, sentential reading is dominant in this construction. However, it does not fully explain the phenomenon of Not-Stripping as there are still some room to accept for the other.

Most of the instances in Not-Stripping seem to have close correlation to the antecedent clause as there were only few cases of remnants that have mismatching or no correlate. Also, it could be regarded that the negation mostly takes large scope rather than narrow scope, indicating that Not-Stripping has clause as its putative source. However, it is still questionable to assume that it is possible to introduce a clausal source for all remnants in Not-Stripping, as a simple introduction of clause source is unacceptable as in (34b):

(34)   a. Abby speaks passable Dutch, but not Ben.
       b. *Abby speaks passable Dutch, but not Ben speaks passable Dutch.

Even if (34b) is still acceptable, the scope of negation in (34b) is not sentential but constituent. A more plausible source would be the following:

(35)   Abby speaks passable Dutch, but Ben does not speak passable Dutch.

This would indicate that there is another movement process for the negator *not*. Direct generation approach also takes similar challenges: how to get the sentential scope of negation and how to account for the connectivity effects and island sensitivity we observed.

   Attested examples like (36) also challenge the postulation of a clausal source for the remnant in Not-Stripping: (36) would violate the subject condition (a condition on extraction domain that states the impossibility of extraction from subjects, Chomsky 1986):

(36)   I'd known that someone would steal my laptop for this test, but not who. (COCA 2008 MAG)

Further issues arise from the interpretation of the negator:

(37)   a. Man can live without a lot of things, but not water. (COCA 2012 WEB)
       b. Cellutech has developed a hydrophobic product that is designed to soak up oil, but not water. (COCA 1996 ACAD)

In examples like (37a), the negator is sentential negation while in those like (37b), it is constituent. In other words, the stripped clause would be operated to have sentential reading in (37a) whereas the stripped clause in (37b) would have non-sentential reading, which also questions an uniform clausal source in Not-Stripping.

   Considering such issues in the postulation of clausal sources and movement operations for Not-Stripping, a feasible alternative seems to pursue direct licensing approaches suggested by Ginzburg & Sag (2000), Culicover & Jackendoff (2005), Kim (2020). The scope of this paper leaves this direction for future research.

5. Conclusion

As discussed in this paper, Not-Stripping is an elliptical construction whose semantic resolution seems to refer to a clausal source. To understand its uses in real-life, we have investigated extracted corpus data with several variables. Each variable has shown us some interesting facts. The register variable shows us that Not-Stripping is preferably used in informal contexts like fiction and spoken data. The category matching variable between the remnant and its correlate also shows us an unexpected property of the construction: the category mismatch outnumbers that of the matching. Related to this, the data also show us that covert correlates (sprouting) are also often used in Not-Stripping even if the construction has been taken to be surface anaphora. We have also seen that most of the attested data for Not-Stripping could be accounted for by move-and-delete operations with the postulation of clausal sources for each instance, but there are also non-trivial examples that challenge such derivational approaches. The data also seem to challenge direct licensing approaches in which the stripped remnant functions as a non-sentential utterance.

<References>

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. Santa Barbra, California: Greenwood Publishing Group.

Chung, Sandra, William Ladusaw & James McCloskey. 1995. Sluicing and Logical Form. *Natural Language Semantics* 3(3). 239–282.

van Craenenbroeck, Jeroen. 2010. *The Syntax of Ellipsis: Evidence from Dutch dialects*. New York, NY: Oxford University Press.

Culicover, Peter & Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Drubig, H Bernhard. 2003. *Toward a Typology of Focus and Focus Constructions*. Berlin: Walter de Gruyter.

Ginzburg, Jonathan & Ivan Sag. 2000. *Interrogative Investigations*. Stanford, California: CSLI Publications.

Gries, Stefan. 2013. *Statistics for Linguistics with R: A Practical Introduction*. De Gruyter Mouton.

Hankamer, Jorge & Ivan Sag. 1976. Deep and Surface Anaphora. *Linguistic Inquiry* 7(3). 391–428.

Jacobson, Pauline. 2016. The Short Answer: Implications for Direct Compositionality (and vice versa). *Language* 92. 331–375.

Kim, Jong-Bok. 2020. Negated Fragments: A Direct Interpretation Approach. *Korean Journal of*

English Language and Linguistics 20(3). 427–449.

Kim, Jong-Bok & Anne Abeillé. 2019. Why-Stripping in English. *Linguistic Research* 36. 365–387.

Lappin, Shalom. 1996. *The Handbook of Contemporary Semantic Theory* chap. The Interpretation of Ellipsis. 145–175. Oxford: Blackwell.

Lee, Yong-hun. 2016. *Corpus Linguistics and Statistics using R*. Seoul: Hankook Publishing Company.

Merchant, Jason. 2001. *The Syntax of Silence: Sluicing, Islands and the Theory of Ellipsis*. Oxford: Oxford University Press.

Merchant, Jason. 2003. Remarks on Stripping.

Merchant, Jason. 2004. Fragments and Ellipsis. *Linguistics and Philosophy* 27(6). 661–738.

Merchant, Jason. 2016. Ellipsis: A Survey of Analytical Approaches. Ms., University of Chicago, Chicago, IL.

Weir, Andrew. 2014. *Fragments and Clausal Ellipsis.* University of Massachusetts, Amherst, Mass. dissertation.

Wurmbrand, Susi. 2017. Stripping and Topless Complements. *Linguistic Inquiry* 48(2). 341–366.