

Implementing Argumenthood Diagnostics for Prepositional Phrases: Comments on Merlo and Esteve Ferrer

Carson T. Schütze

UCLA

2012

1. Introduction

Merlo and Esteve Ferrer (2006; hereafter M&EF) propose a set of heuristics for automatically determining from a corpus whether prepositional phrases (PPs) in the English sentence frame [Verb NP ___] (e.g., *saw the cop with binoculars*) are arguments versus adjuncts of the preceding verb or noun. Because the diagnostics used by linguists often are not suitable for computer implementation, M&EF devise gradient approximations of properties to which the diagnostics are sensitive. These are weighted and combined into an estimate vector such that “the feature values in the vector are corpus-based numerical equivalents of the grammaticality diagnostics used by linguists to decide whether a PP is an argument or an adjunct” (page 344). M&EF show that this measure is quite successful at determining the argument/adjunct nature of attachments (as diagnosed by human judgments) in a test sample after “training” on a distinct sample from the same corpus.

My main goal in this paper is to programmatically suggest ways in which the performance of this and similar algorithms might be improved, by showing how it is possible to more closely approximate the underlying argument/adjunct properties using the kinds of automated calculations employed by M&EF. A secondary goal is to clarify the linguistic nature of the diagnostics themselves. More detailed discussion and caveats concerning most of the points covered here can be found in Schütze (1995).

2. Diagnostics implemented by Merlo and Esteve Ferrer

2.1 Head Dependence¹

The degree to which a PP depends on a head (noun or verb) for its interpretation is often seen as the fundamental semantic indication of the extent to which it is an argument. For instance, one could determine an appropriate meaning for the PPs in ex-

¹ In M&EF’s paragraph “Head Dependence,” beginning on page 345, there is apparently a negation missing: “...because they are [not] necessary for the correct interpretation of the semantics of the [head].” Also, M&EF do not note that their examples (10), (11), (13), (14), and (15) are taken directly from Schütze (1995).

ample (1) without knowing what the verbs in those sentences were, indicating that those PPs are adjuncts, while this is clearly not true for examples (2a–b), where *invested* and *jumped* provide vital information on how to interpret *in*.

- (1) a. My friend arrived in a car.
- b. The escapee foraged in the woods.
- (2) a. My friend invested in a car.
- b. The escapee jumped in a car.

M&EF estimate degree of head independence by “the dispersion of the distribution of the different heads that co-occur with a given PP” (page 346). However, what linguistic theory would predict is that the degree of head independence (adjuncthood) should be proportional to the number of different heads (types) that co-occur with PPs headed by the same preposition used in the same sense. This quantity differs in two relevant ways from that of M&EF: The NP complement of the preposition does not matter, but the meaning of the preposition does. I therefore propose making two corresponding adjustments to the implementation. By way of illustration, the fact that the PP *in a car* occurs in two environments with two different verbs in example (2) is not evidence about how likely that PP is to be an adjunct, because *in* is clearly being used in two different senses. What would instead constitute such evidence would be a data set like example (1).

In the applications M&EF have in mind, the sense of the preposition might not be established before the site and nature of attachment are determined, in which case their estimates could not take this into account. Nevertheless, in one of their implementations they apply a heuristic that could go part-way towards this, by grouping contexts according to the semantic class of the noun heading the NP complement of the preposition (and that of the direct object of the verb).² This will separate out certain senses; for example, *in a {minute/hour/day/...}* cannot get either of the interpretations shown in example (2) because these nouns refer to units of time. Still, the sentences in example (2) do show that a particular NP complement of a particular preposition can be compatible with multiple senses, and it suggests that considering the semantic class of the verb would further improve the algorithm. (Evidently, the limitations of WordNet prevented this latter improvement from being realized.) I therefore propose a separate in-

² This was not the stated purpose of the heuristic, however. M&EF motivate it by noting that the data would be too sparse if they based their counts on particular combinations of nouns and verbs.

vestigation to determine the most effective preposition-sense disambiguation formula in the *V NP PP* configuration, to be applied here and in other heuristics discussed below.

2.2 Optionality

M&EF assert that “in most cases, PP-arguments are obligatory elements of a given sentence whose absence leads to ungrammaticality” (page 346), as in example (3). Surveying the examples of PP arguments of verbs in §2 of Schütze (1995), this seems not to be true: Example (4) shows that optional ones are plentiful.³

- (3) a. John put the book *(in the room).
- b. Lou handed a book *(to the kids).

- (4) a. John informed his friend (of the danger).
- b. Chris rented/described/offered the movie (to Mary).
- c. They asked the committee (for some information).
- d. They talked/complained (to the landlord) (about the tenants).

Fortunately, M&EF implement a gradient version of this diagnostic: Argumenthood is more likely the greater the conditional probability of a PP given a particular verb. (Again, the sense of the preposition should be relevant, and the approach suggested in §2.1 should be employed.) This strategy will be led astray in situations that the authors acknowledge: “Some adjuncts frequently co-occur with certain heads” (page 372).

M&EF do not apply this diagnostic for arguments to nouns because, with circumscribed exceptions (Grimshaw 1990), arguments to nouns are never obligatory. But as I have shown, PP arguments to verbs are often not obligatory either, and we would thus expect the gradient version of the diagnostic to work similarly in both cases. That is, that fact that *of physics* is an argument of *student* makes it more likely to follow *student* than, say, the adjunct *with red hair*—the fact that an *of*-PP is not obligatory following *student* does not undermine the logic that a phrase that completes the meaning

³ While M&EF acknowledge the existence of optional PP arguments, their only example is instrumental *with*-PPs. They assert that the contrapositive is exceptionless: “All adjuncts are indeed optional” (footnote 1, page 346). This is controversial: *The Space Needle towers *(over Seattle)*, *We treated the antiques *(with great care)*, and so on appear to be counterexamples. See Levin (1993), Jackendoff (1990), and Dowty (2003) for discussion of so-called subcategorized adjuncts.

of the head is more likely than one that adds incidental modification.⁴ The optionality diagnostic should therefore be implemented for nouns as well as verbs.

2.3 Iterativity and Ordering

M&EF assert that “in a sequence of several PPs only the first one can be an argument, whereas the others must be adjuncts” (pages 346–7). This is false: Example (4d) illustrate sequences of two argument PPs. M&E’s example that is supposed to show this, example (5), actually demonstrates that arguments of the same type cannot be iterated, whereas adjuncts of the same type can.

- (5) a. *Chris rented the gazebo to yuppies, to libertarians.
 b. Kim met Sandy in Baltimore, in the hotel lobby, in a corner.

There *is* a generalization about ordering of arguments versus adjuncts, which M&EF note in passing: Arguments must generally appear closer to the head they modify than adjuncts, as illustrated in examples (6)–(7).

- (6) a. a member {of Parliament with gray hair/*with gray hair of Parliament}
 b. a politician {from Paris with gray hair/with gray hair from Paris}
- (7) a. They complained {to the LANDLORD after the flood/?*after the FLOOD to the landlord}.⁵
 b. They complained {to the LANDLORD about the tenant/about the TENANT to the landlord}.

⁴ Using all occurrences of the head (verb or noun) as the baseline of the conditional probability is not appropriate, in that it penalizes PPs that happen to occur with heads that are more often used without any following PP. For example, suppose *student* occurs 60% of the time with no following PP, 35% with an *of*-PP, and 5% with a *with*-PP. The huge gap between the rate of *of*-PPs and the others ought to count as evidence that the former are arguments. But suppose *member* occurs with no PP only 20% of the time, with an *of*-PP 70%, and with other PPs 10%. The ratio of *of*-PPs to others is the same as for *student*, yet if we use the proposed baselines we will conclude that *of*-PPs are twice as likely to be arguments of *member* as of *student*. Intuitively they may be *more* likely, but not that much more likely.

⁵ Small caps indicate prosodic emphasis, placed on the nonfinal PP to prevent derived word orders from sounding good (Pollard and Sag 1987). The ordering requirement discussed in the main text applies to underlying orders only.

The consequences of this in English are the following: For a given head, two argument PPs can appear in any relative order, as in example (7b), two (or more) adjunct PPs can appear in any relative order, as in example (6b), but when both types of PPs are co-present, arguments must precede adjuncts, as in examples (6a) and (7a).

M&EF’s implementation is stated as “The probability of a PP being able to iterate, and consequently being an adjunct, can be approximated as the probability of its occurrence in second position in a sequence of PPs” (page 347). Obviously, given the above, there is no reason in principle why such a contingency should hold. Occurring second in a sequence of PPs might in practice correlate to some degree with being an adjunct if most head tokens appear with only zero or one PP arguments. But I assume it would be preferable, and should ultimately be more successful, to try to base algorithms on properties that are not merely coincidental.

The ordering diagnostic illustrated in examples (6) and (7) is typically used by linguists when they already know the status of one PP with respect to a head; a second PP’s status can then be determined by its ordering options. This situation will not arise in the batch learning scenarios discussed by M&EF, though their approach does not preclude incremental applications where it could. Abstracting away from these details, once certain PPs have been determined with high confidence to be arguments (of a particular head or semantic class of heads), any PPs that precede them (in the same environments) can be taken to also be arguments according to this test; similarly, once certain PPs have been determined with high confidence to be adjuncts, any PPs that follow them can be taken to also be adjuncts.

2.4 Copular Paraphrase

M&EF propose that the probability that a PP attached to a noun can be paraphrased by a copular relative clause, which is a property of adjuncts, “can be approximated by the probability of its occurrence following a construction headed by a copular verb, *be*, *become*, *appear*, *seem*, *remain*” (page 347), relative to its occurrences following all verbs (examples below). An initial problem with extending the class of traditional copulars to *appear* and *seem* is that these often take a PP headed by *to* as an argument (expressing an experiencer); the heuristic will wrongly assess such a PP as an adjunct, and will also miss the opportunity to gather evidence for a subsequent PP being an adjunct, because now it will not directly follow the “copular,” as in *Pat seems to Chris under control*, where *to Chris* will be seen as an adjunct and *under control* will be ignored altogether. Since *to*-PPs are almost never adjuncts to nouns, the algorithm could be tweaked to ignore them when they follow *appear* and *seem*.

M&EF consider it a shortcoming that their implementation does not restrict the copular verbs that it looks for to those inside relative clauses, but this per se is not relevant for the diagnostic, as noted by Grimshaw (1990). Parallel contrasts in the two environments are shown in examples (8) and (9).

- (8) a. the destruction of the city
 b. *the destruction that was of the city
 c. *The destruction was of the city.
- (9) a. the albums on the shelf
 b. the albums that were on the shelf
 c. The albums were on the shelf.

If a PP can follow a copular verb then it *can be* an adjunct. Thus, not restricting the search to relative clauses in fact benefits the algorithm, since it increases the sample size.

The purpose of the linguistic diagnostic, however, is to determine whether the PP actually *is* an adjunct in combination with a particular noun. Borrowing an example from Abney (1989), the fact that we can say *the man (who) was [in a Volvo]* is consistent with the fact that the same PP attached to a different noun is an argument: *his interest [in a Volvo]*, as shown by the ungrammaticality of **His interest was in a Volvo*. Therefore, important information is being lost by M&EF's procedure by not including the identity of the subject of the copula (or at least the semantic class of its head noun) as part of the calculation, because we expect the same (kinds of) nouns to take a given PP (actually, PPs introduced by a particular preposition in a particular sense) as an adjunct within their NP and as a predicate across a copula. Likewise, (kinds of) nouns that never take a given PP as a predicate across a copula should never take it as an adjunct within their NP either.⁶ These are the co-occurrences that should be tracked when employing this heuristic, making use once again of preposition sense disambiguation as discussed in §2.1.

2.5 Deverbal Nouns

M&EF take any PP immediately following a deverbal noun to be an argument. But deverbal nouns can perfectly well be modified by adjunct PPs, just like their counterpart verbs can, as in example (10). M&EF's explanation of the basis for the diagnos-

⁶ With one systematic exception: *the woman with grey hair* does not correspond to **The woman was with grey hair*; rather, the verb *have* must be used.

tic misses a crucial step: They say, “PPs following a deverbal noun are likely to be arguments, as the noun shares the argument structure of the verb” (page 347).⁷ But it is not that a PP in this position *in general* is likely to be an argument. Rather, this is true only if it has the same form and same sense as a PP following the verb from which the noun is derived, *and* that PP is itself an argument, as in example (11).

- (10) a. to meet in the afternoon
b. a meeting in the afternoon

- (11) a. to participate in the riot
b. a participant in the riot

Clearly it would require considerable effort to test for the correct set of conditions.

Furthermore, there are other ways that verbal arguments can line up with derived nominal arguments. An direct object argument of a verb may show up inside a PP as argument to a noun.

- (12) a. to desire a solution
b. the desire for a solution

Additionally, for one of the nine nominalizing suffixes used by M&EF to identify deverbal nouns, *-ee*, it is the subject of the verb that corresponds to a PP argument of the noun.

- (13) a. The president appointed someone.
b. an appointee of the president

As a result, there would be three sub-cases involved in finding a verbal argument to match up with a noun-following PP (which might turn out not to be attached to that noun at all). And because many deverbal nouns pattern like example (12), very often the verb will have a direct object that *might* correspond to the NP inside the PP that follows the noun, with no way to be sure whether it actually does. My conclusion is that this test should be dropped.

⁷ Given the use of “are likely,” it is surprising that this alone among the diagnostics was implemented as a binary variable. It turns out to involve the greatest uncertainty.

3. Diagnostics not implemented by Merlo and Esteve Ferrer

3.1 Iterativity

As discussed in §2.4 and illustrated in example (5), an argument of a particular kind cannot be iterated while an adjunct can. This could be approximated by saying that the probability of a PP being an argument declines to the extent that it co-occurs in a sequence of PPs with another one headed by the same preposition used in the same sense. The same sense restriction is to avoid misdiagnosing the two argument PPs as adjuncts in a sequence like *The colonel spoke with the commander with his walkie-talkie*.

3.2 Pro-form replacement

This set of diagnostics relies on the fact that when a pro-form such as nominal *one* or verbal *do so* replaces a contentful noun or verb, respectively, it must also subsume all internal arguments that co-occurred with that head, as illustrated in examples (14a) and (15a). Such pro-forms are free to co-occur with adjuncts, however, as in examples (14b) and (15b).

- (14) a. *I climbed to the top of the hill, but not to the one of the mountain.
b. She likes the car in the garage, not the one in the driveway.
- (15) a. *John put a book on the table, and Sue did so on the shelf.
b. John filled out the form during the lecture, and Mary did so after the lecture.

Although a corpus does not contain ungrammatical examples marked as such, we can take advantage of the following consequence of this pattern: Any PP that co-occurs with a pro-form and modifies the head that has been replaced by that pro-form must be an adjunct. From this we can create an easily-computable gradient approximation: A PP is more likely to be an adjunct the more frequently it occurs following a form of *do so* or nominal *one*. As with M&EF’s diagnostics, these counts will include a certain amount of noise, e.g., because some PPs following *one* will actually be VP-attached rather than NP-attached, if the attachment site ambiguity has not yet been resolved.

3.3 Extraction

PP arguments of an object noun are more easily extracted (via pied piping) and extracted from (with preposition stranding) than PP adjuncts. PP arguments of a verb are more easily long-distance extracted and extracted from than PP adjuncts when that verb heads a clause that forms a so-called weak island. (Details are discussed in Appen-

dix A of Schütze and Gibson [1999]; for a theoretical survey see Szabolcsi [2006]). Relevant contrasts are exemplified in (16) and (17).

- (16) a. Which problem did the President suggest [a solution to *t*]?
- b. *What kind of collar did you buy [a shirt with *t*]?

- (17) a. ?To which friend do you wonder [whether John gave the book *t*]?
- b. *On which day last week do you wonder [whether John bought the book *t*]?

M&EF sensibly used this diagnostic only in hand coding the test sentences for argument/adjuncthood, since such sentences would be far too sparse in a corpus.

In their example of how this test was applied to elicit human judgments (their Figure 3), they give the following corpus sentence, where the PP whose status is at issue is bracketed (construed as modifying the verb): *Americans will learn more about making products [for the Soviets]*. The sentence they used for the extraction test, *Who do you wonder whether Americans will learn more about making products for?*, was unnecessarily complex: The only elements relevant for the test are *make*, *products*, and *for the Soviets*; all the preceding material, and the fact that this VP occurs in a gerund, is not relevant to the status of the PP. Preceding material can, however, create additional constituents out of which extraction needs to occur, degrading the test sentence for irrelevant reasons. M&EF acknowledge that corpus sentences “had to be simplified so that the verb and prepositional phrases were in a main clause” (footnote 6, page 356), but PRO-*ing* gerunds such as *making products* have many clause-like properties, so the authors did not achieve this desideratum. They should have tested the sentence *Who do you wonder whether Americans will make products for?*, which is easier to perceive as (essentially) grammatical.

Another problem with M&EF’s use of extraction judgments is that they treated extraction with preposition stranding, as in example (16), and extraction with pied piping, as in example (17), as separate tests for purposes of tallying how many tests pointed to argument versus adjunct status. But in fact the two kinds of extraction pattern exactly the same in this regard, so this counting technique gave undue weight to the extraction test. (Both versions are needed because some prepositions resist stranding, while those that allow it often sound stilted with pied piping, and there is interspeaker variation. The tally should ideally be based on whichever variant each subject rated as more grammatical.)

4. General comments

Concerning the status of the argument/adjunct distinction as construed in theoretical linguistics, M&EF assert that “native speakers’ judgments on the argument and adjunct status of PPs are very unstable” (page 373). No evidence for this assertion is provided; it may be based on a misunderstanding of the notion of stability—see Cowart (1997) for relevant discussion and experiments. They continue, “No explanation is usually proposed of the fact that the tests of argumenthood are often difficult to judge or even contradict each other” (page 373). In fact, Schütze (1995) suggested explanations for both of those properties in numerous cases.

In conclusion, by taking the linguistic theory behind argumenthood diagnostics more seriously we can hope to improve upon the already impressive results demonstrated by M&EF in automatically determining the nature of PP attachments. Only actual implementation will tell us for certain, of course. I have also tried to provide some guidance to researchers who want to employ these diagnostics in eliciting human judgments, for whatever purpose that might prove useful.

References

- Abney, Stephen P. 1989. A computational model of human parsing. *Journal of Psycholinguistic Research*, 18:129–144.
- Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Sage, Thousand Oaks.
- Dowty, David. 2003. The dual analysis of adjuncts and complements in Categorical Grammar. In Ewald Lang, Claudia Maienborn, and Cathrine Fabricius-Hansen, editors, *Modifying Adjuncts*. Mouton de Gruyter, Berlin.
- Grimshaw, Jane. 1990. *Argument Structure*. MIT Press, Cambridge, MA.
- Jackendoff, Ray. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.
- Levin, Beth. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.
- Manning, Christopher. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay, and Stephanie Jannedy, editors, *Probabilistic Linguistics*. MIT Press, pages 289–314.
- Merlo, Paola and Eva Esteve Ferrer. 2006. The notion of argument in prepositional phrase attachment. *Computational Linguistics*, 32:341–377.
- Pollard, Carl and Ivan A. Sag. 1987. *An Information-Based Syntax and Semantics. Volume 1: Fundamentals. CSLI Lecture Notes Number 13*. Center for the Study of Language and Information, Leland Stanford Junior University.

- Schütze, Carson T. 1995. PP attachment and argumenthood. In Carson T. Schütze, Jennifer B. Ganger, and Kevin Broihier, editors, *Papers on Language Processing and Acquisition, MIT Working Papers in Linguistics 26*. Cambridge, MA, pages 95–151.
- Schütze, Carson T. and Edward Gibson. 1999. Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, 40:409–431.
- Szabolcsi, Anna. 2006. Strong vs. weak islands. In Martin Everaert and Henk van Riemsdijk, editors, *The Blackwell Companion to Syntax, Volume IV*. Blackwell, Malden, MA, pages 479–531.