Making Verb Argument Adjunct Distinctions in English

Jena D. Hwang
University of Colorado at Boulder

In natural language processing, identifying a verbs argument structure is important for many tasks including parsing, text simplification, and semantic role labeling. However, while most formal theories of grammar generally agree that there is a distinction between constituents that are arguments and those that are adjuncts, linguists do not yet agree on how to define what it means to be an argument or an adjunct. Nevertheless, these efforts have been fruitful in gaining a general understanding of some of the characteristics of arguments and adjuncts. This paper will explore the distinct approaches taken by the linguistics community, focusing on verb argument and adjunct distinctions. This paper explores the reasons behind why this distinction is such a difficult issue in both the semantic and syntactic community.

1 Introduction

Most formal theories of grammar generally agree that there is a distinction that can be made between constituents that are arguments and those that are adjuncts. Syntactically speaking, arguments are typically considered to be constituents that are syntactically licensed or required by the head verb of the phrase. As for adjuncts, no such restriction or requirement is necessary for them to be present in a phrase. Semantically speaking, arguments are necessary participants in the event or state created by the verb and they participate in the manner specified by the verb's subcategorization frame. Adjuncts, on the other hand, unlike arguments, do not rely on the relational information conveyed by the verb. Rather they comment on the general action or state of the predicating unit – the verb and its arguments.

For natural language processing (NLP), identifying the verb's argument structure is important. In NLP's statistical parsing task, the automatic system generates the most statistically plausible parses for any given sentence. The automatic system picks from this set the most likely parse. It has been shown that providing a verb's subcategorization information during syntactic parsing can improve performance (Collins, 1999), for example by helping to resolve such issues as PP-attachment ambiguity (Hindle and Rooth 1993; Merlo and Esteve Ferrer 2006).

In addition to the parsing task, verb argument structure information is used in distinguishing different senses of a word that are likely to be associated with a particular argument frame (Dligach and Palmer, 2008), in marking required or optional elements in a sentence during machine translation (DeNeefe and Knight, 2009), and discriminating crucial information in a given document from that which is non-critical or parenthetical in text summarization and text simplification. Thus, the key piece in defining verb argument structure in NLP is identifying which constituents in a given sentence should be included in or excluded from the head verb's subcategorization frame.

Consequently, drawing a distinction between what is an argument and what is not is an important task for both the syntactic resources used in the NLP communities and the lexical resources that provide human annotated data for the training and testing of the automatic systems described above. For grammar formalisms used in NLP such as Tree Adjoining Grammar (TAG) and Lexical Functional Grammar (LFG), a clear definition of the predicate's subcategorization frame is necessary as this is the basis for establishing the verb's syntactic definitions (e.g. identifying tree family membership in TAG and representing functional structure in LFG for a verb). Furthermore, lexical and semantic resources such as FrameNet (Ruppenhofer et al., 2010), VerbNet (Kipper et al., 2008), and PropBank (Kingsbury and Palmer, 2003; Palmer et al., 2005), as well as lexical resources such as COMLEX Syntax (Grishman et al., 1994) and Combinatorial Categorical Grammar (CCG; Steedman 2000) provide data for training and testing of automatic systems. Effective supervised processing techniques, whether sentence parsing, machine translation or text simplification, depend highly on the quality and consistency of the annotation based on available resources.

Thus, this paper explores the distinct approaches taken by the linguistics community in making verb argument and adjunct distinctions and why the distinction is such a difficult issue for both the semantic and syntactic communities. Additionally, this paper will briefly touch on what this difficulty means to the current the lexical resources that handle data for automatic systems.

2 Semantic Intuitions Concerning the Argument-Adjunct Distinction

The task of making the distinction between arguments and adjuncts of a verb is in some sense a way of capturing a basic intuition that if a world event or activity must be described, such an event will necessitate participants (e.g. *birthday boy* in a *birthday party* or *snow* in a *snowstorm*) or other relevant information that is salient to the setting. And as it is in any setting, some information will be more crucial to the described event and other information will be less important (though not necessarily irrelevant). Thus, linguistic intuition is that in an event described by a verb, there will be key participants without which the event would not be complete. Such event will also include other peripheral information that provides descriptors of the general condition or circumstance of the state or event, which are not as central to the meaning of the verb.

2.1 Thematic Relationships

When we invoke our intuitions of which are the "necessary" participants in a given state or event described by a verb, generally speaking, we are referencing the semantics side of the issue. Take as an example a *giving* event as seen in the following sentence:

DOI: https://doi.org/10.25810/g2ev-pv70

(1) His father gave him a computer last night for succeeding in his studies.

Here, we have five elements or concepts expressed in the sentence: *His father*, *him*, *a computer*, *last night*, and *for succeeding in his studies*. In the case of the first three elements, the relationships they have with the verb are often referred to as thematic relations. This concept of thematic relations has been well discussed by numerous studies in the linguistics literature (cf. Fillmore, 1968, Jackendoff, 1972, Dowty, 1989). Thematic relations describe the roles participants play in the event or state created by the verb. In example (1), the *father* is the AGENT or GIVER as he takes on the role of the giving entity, *a computer* would be the THEME or TRANSFERRED ITEM, and *him* is the RECIPIENT of the computer. Furthermore, these participating roles are considered required or obligatory in such a way that if they were removed from the sentence the semantics of the sentence would be incomplete, as seen in the following sentences.

(2) ? His father gave.
(3) ? His father gave him.

For such utterances as (2) and (3) to be meaningful, the missing participant(s) (i.e. RECIPIENT and THEME for example (2) and THEME for example (3)) would have to be cited elsewhere and recoverable in the context. Thus, these participants are considered to play a direct role in the relational information conveyed by the verb, and therefore, necessary components of the semantics of the verb. Those participants that are in a thematic relationship with the verb are considered to be semantic arguments of the verb.

In contrast to the arguments, the last two elements in example (1) would be considered semantic adjuncts. Unlike arguments, adjuncts do not rely on the relational information conveyed by the verb. Rather they comment on the general action or state of the predicating unit – the verb and its arguments. The adverbial *last night* and the prepositional phrase *for succeeding in his studies* are present because they comment on the event described by the verb and its arguments. They are there to set up the context in which the event happens regardless of the specific meaning of the verb: the adverbial sets the time in which the *giving* takes place and the prepositional phrase describes causal events leading up to the event, which serve as a motivation for event's occurrence.

Thus, in general, the elements in the sentence that are in a thematic relationship with the verb and play a central role in the event or state presented by the verb, are considered to be arguments. These arguments are licensed and required by the verb to realize its full meaning. Those elements in the sentence that do not hold a specific relationship to the verb and provide contextual information "typically information about time, location, purpose or a result of an event" (Saeed, 1997) are considered adjuncts. Unlike arguments, the adjuncts do not care about the specific meaning of the verb. Rather, they modify the entire predicating unit.

Consequently, since the adjuncts are not licensed by the verb, they are considered to be applicable to a wider range of events (c.f. Cowper (1992, p.65)). That is, the same adverbial *last night* in (1) would retain a similar descriptive value even when used to describe other events like "Scotty laughed hard at the jokes *last night*" or "Bethany read the poetry beautifully *last night*". The same could not be said about the arguments. The noun phrase *a computer* only serves as a TRANSFERRED ITEM in the *giving* event. It would take on an entirely different role specified by the verb in the context of other events (consider "The computer was damaged when it fell to the ground" where *computer* is a PATIENT). As Cowper (1992) and Gawron (1988) note, self-evident adverbials like *last night* make a reference to a time and therefore are related to the verb in a specific way. However, one cannot say what thematic relation a noun phrase like *the computer* should hold if it is not in a relationship with a verb.

2.2 Problem of Semantic Intuition

These intuitions about the nature of argumenthood at first glance seem to be fairly straightforward. If these intuitions are indeed sufficient, then it would seem that determining the semantic representation of arguments and adjuncts should be a simple enough task. Consider the following example:

(4) We ate our supper on our balcony.

Intuitively speaking, central to the *eating* event are two participants: the one who eats and the entity that is eaten. The prepositional phrase provides a general location or setting in which *eating* takes place. It would be very simple if we could extrapolate from such an example and say that prepositional phrases like *on our balcony* could always be considered adjuncts as in example (4). However, as we know, this is not always the case. Consider the following example:

(5) I put the book on the table.

The locative prepositional phrase in (4) is distinguished from the same prepositional phrase in example (5), which would generally be recognized as the argument of the verb *put* as it is the location in which *the book* is placed, an element without which the semantics of a *putting* event would not be complete. That is, unlike the adjunct-like prepositional phrase in example (4), *on the table* in (5) would have to be classified as an argument. Take into consideration a few more examples, with attention to the adverbials used in the sentences:

(6) Scotty laughed *hard* at the jokes.
(7) Johnny hit the nail *hard* to drive it into the wood.
(8) Bethany read the poetry *beautifully*. *beautifully*.

DOI: https://doi.org/10.25810/g2ev-pv70

If indeed adjuncts are characterized by the ability to be applicable to a wider range of events as the literature suggests, it would stand to reason that adverbials such as *hard*, *quickly* or *beautifully* should be judged adjunct-like as they maintain a similar meaning across a variety of situations. For example, *hard* in (6) describes the intensity by which Scotty laughed at the jokes, just as it describes the intensity of force used by Johnny to drive the nail into the wood in (7). However, the adverb *well* in (10) stands as a counterexample.

(9) This child reads well.
(10) This book reads well.
(11) This duck eats well.

Here, the adverb *well* seems to do more than simply comment on the *reading* event. It has a direct effect on the interpretation of the meaning of the verb. The sentence in (10) could be used to express how gripping the said book is, which cannot be said about the sentence lacking this adverb: *This book reads*. In fact, the sentence is rendered nonsensical. In this particular usage of *read*, then, a manner adverbial such as *well* is a necessary component for completing the intended meaning and thus is considered to behave more like an argument than an adjunct despite our initial intuitions. Moreover, in example (9), it is not clear if the adverb *well* is behaving like an argument as in (10) or if is an adjunct as in (8) that happens to comment on the manner in which this particular child reads. This is even more evident in example (11), where the sentence could be interpreted to mean that the animal has a good appetite (i.e. adjunct reading) or that the cooked duck is good for eating (i.e. argument reading). Such a decision would likely depend on the correct identification of the reading of the sentence intended by its speaker or writer, which hopefully would be available in the context.

Finally, in certain cases, the distinction seems to depend on the lexical items present in the sentence. Compare the following pairs of sentences in examples (12) and (13), in which each sentence carries a prepositional phrase with locative information:

(12) a. I cooked the chicken on the grill.
b. I cooked the chicken on the patio.
(13) a. She kissed her mother on the cheek.
b. She kissed her mother on the platform. (Quirk et al., 1985, p.511)

Fillmore (1994, p.159) would consider the phrases in the (a) sentences to carry "information that fills in details of the internal structure of an event", which he calls *frame internal* information, while those in the (b) sentences provide "incidental

attending circumstances of that event, the frame-external information"[1]. This seems to indicate that the reading of the *on*-phrase in the first sentences should lead to an argument-like reading. The second sentences should lead to an adjunct-like reading. The appropriate reading depends on whether the the prepositional phrase is viewed as a modifier of the location of the undergoer argument (an argument-like reading) or as a modifier of the location of the agent argument (an adjunct-like reading).

3 Dimensions of Distinction

The level of uncertainty discussed in the previous section is reflected in the linguistic literature on how researchers think these arguments and adjuncts should be characterized. While linguists agree that there is indeed a distinction to be made between arguments and adjuncts, the researchers have not yet converged on how to define what it means to be an argument or an adjunct, and how the boundary between the two should be characterized. As Przepiorkowski (1999) writes, "[a]lthough the [argument]/adjunct dichotomy is supposed to play a central role in the Chomskyan version of generative linguistics, there is no generally agreed upon classification of kinds of dependents, nor is there a generally accepted analysis of adjuncts" (Ibid., p.257). Nevertheless, those efforts have been fruitful for gaining a general understanding of some of the characteristics of arguments and adjuncts. Here are a few quotes that reflect varied definitions found in the literature:

> [A]rguments are something that lexical heads have; they are the central participants in the scene that the head presents, and thus in the situations the head is instantiated by. (Gawron, 1988, p.111)

> Complements tend to be (though not always) *obligatory*, whereas Adjuncts are *always optional* (Radford, 1988, p.263)

> Another classic observation involving the *do so* anaphora is that arguments in VP are closer to the verb than other adjuncts. (Culicover and Jackendoff, 2005, p.128)

Varied as they are, the above characterizations of arguments and adjuncts display clear themes. We have already seen the first general distinction expressed by Gawron (1988) earlier in this section. It speaks to our semantic understanding of arguments as central participants in the predicate. Secondly, another trend in argument and adjunct distinction seen in the literature, as expressed by Haegeman

[1] Quirk et al. (1985, p.510-511) makes a similar observation. However, he labels what Fillmore calls frame-internal elements "predication adjuncts" and frame-external elements "sentence adjuncts"

DOI: https://doi.org/10.25810/g2ev-pv70

(1991), Radford (1988) and Quirk et al. (1985), is the notion of obligatoriness and optionality, in which arguments are generally considered obligatory, while adjuncts are considered optional.

A final distinction that is often found in the linguistic literature is one based on the syntactic structure of a sentence. Culicover and Jackendoff (2005) have chosen to use the words *closer to the verb* to describe the observation that arguments, in English, sit next to the verb and it is rare that other constituents are allowed to intervene between the verb and its argument. Along with the notion of *closeness*, there is the observation that arguments form a constituent with the verb and they act within the sentence as a single unit. And as we will see in this paper, the free interchange between the terms complement and arguments, as seen in Haegeman (1991) and Radford (1988), also speaks to this point.

3.1 Structural Distinction

For a discussion of the distinctions proposed by the proponents of the classical or transformational grammar, we will begin with the *do so* test introduced by Lakoff and Ross (1976) for two reasons. The *do so* test is a widely accepted substitution test used to distinguish arguments from adjuncts in linguistic literature appearing in many studies where the topic of argumenthood is discussed. Secondly, it is a very clear and concise illustration of what role the hierarchical syntactic structure has on the distinction of argument and adjuncts. Through the course of this section, we will see that for Principles and Parameters (P&P), Minimalist Theory (MP), and other transformational theories of syntax the argument and adjunct distinction is highly dependent on the structural configuration of the verb phrase.

3.1.1 The *Do So* Test

The test is based on the observation that *do so* serves as an anaphoric substitute to the verb and its arguments (c.f. Culicover and Jackendoff (2005, p.124-127); Cowper (1992, p.31); Quirk et al. (1985, p.81-82)). The following examples illustrate the test:

(14) Sue cooked lunch yesterday, and Fred *did so* today. [*did so* = cooked lunch]
(15) *Sue cooked lunch, and Fred *did so* dinner. [*did so* = cooked]

The test is that if *do so* cannot refer to the verb without one of its constituents, then that constituent must be an argument. The concept behind this test is that a verb and its arguments form a V' and the anaphor *do so* must refer to the unit as a whole. For example, in (14), *do so* is perfectly happy to refer back to the verb and its object. However, when *do so* refers to the verb alone, as in (15), the sentence is rendered ungrammatical. Adjuncts, on the other hand, can be included as an antecedent of the *do so* or they can be left out (Culicover and Jackendoff (2005); Haegeman (1991)). This can be seen in examples (16) and (17).

(16) Mary will cook the potatoes for fifteen minutes in the morning, and Susan will *do so* **for twenty minutes in the evening**. [*do so* = cook the potatoes]
(17) Mary will cook the potatoes for fifteen minutes in the morning, and Susan will *do so* **in the evening**. [*do so* = cook the potatoes for fifteen minutes]
(18) Mary will cook the potatoes for fifteen minutes in the morning, and Susan will *do so* too. [*do so* = cook the potatoes for fifteen minutes in the morning]
(19) *Mary will cook the potatoes for fifteen minutes in the morning, and Susan will *do so* **the vegetables**. [*do so* = cook]

The *do so* test, thus, tells us the time adverbials *for X minutes* and *in the X* are adjuncts, while the object *the potatoes* must be an argument as the sentence in (19) fails. Here are other examples in which the judgement for argument and adjunct distinction from the *do so* test lines up with our general semantic intuitions we have seen earlier in this paper:

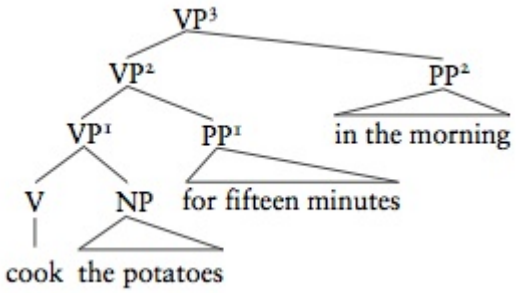
(20) a. *His father gave him a gift, and his mother *did so* **a card**.
b. *I put a book on the table, and she *did so* **on the counter**.
(21) a. We ate our supper on our balcony, and they *did so* **on their porch**.
b. Bethany read the poetry beautifully, but Carla *did so* **quite poorly**.

The test correctly predicts that the phrases in bold (20) are arguments and that the phrases in bold (21) are adjuncts. Before we turn to the studies that have disputed the validity *do so* test, we will quickly touch on the conclusions P&P generally derives from these and other constituency tests.

3.1.2 Importance of structural configuration

In traditional syntactic approaches such as P&P, the *do so* test has often been cited as evidence for an embedded V' or VP structure over an alternative flat structure. As Culicover and Jackendoff (2005) writes, syntactic behaviors based on *do so* substitutions seen in examples (16)-(19) lead "to the conclusion that the maximal VP consists of a nested structure of VP's, along the lines of [(22)]:"

(22)



Mary will cook the potatoes for fifteen minutes in the morning. (cf. (16)-(19))

This agrees with the structures proposed by Cowper (1992) and Haegeman (1991). Haegeman (1991) argues that a flat structure has "no internal hierarchy between constituents of V [in which] all VP-internal constituents are treated as being on equal footing" (ibid., p.79). Consequently, it can only account for examples like (18), where the the *do so* substitutes for the entire VP. However, she argues, such a tree would only "be expected to affect either the top-node VP, i.e. the entire VP [...], or each of the VP-internal constituents, that is to say V or NP or PP" (ibid., p79-80). Moreover, it does not provide a proper account for the fact that in examples like (16) and (17) *do so* only substitutes in for a portion of the VP. What would be necessary, Haegman suggests, is a hierarchical structure with V' nodes at V and at each of the PP levels, to account for substitution behaviors as seen in the *do so* test. In other words, the complement of the verb will be closer or local to the V (i.e. the argument will take up the complement position), the rest of the non-complements or adjuncts will join sequentially to a higher V' node, where there is one V' node for each of the adjuncts.

This, then, begs the question of how should syntax know which constituents should occupy the which positions in the trees. The *theta grid* of the verb holds the necessary mapping that assigns the arguments to the complement positions (c.f. Carnie, 2006, p.223-226; Cowper, 1992, p.65; Haegeman, 1991, p.296-297). It includes theta roles[2] that the verb selects for. Then through a "predictable" and deterministic association (c.f. Chomsky, 1995, p.30-33; Cowper, 1992, p.64-69), the theta roles are assigned to the complement positions on the tree. Unlike arguments, adjuncts are not contained in the theta grid. Though it is not explicitly expressed in any of the P&P literature examined in this paper, the assumption is that all other constituents are systematically attached to a V' node which dominates, or eventually dominates depending on the number of adjuncts present, the V' in which the verb is attached.

Thus, in P&P the prime distinction between the adjunct and argument is in the differences in the position in the tree occupied by the constituents, and ultimately this decision is up to the theta grid that assigns the positions (i.e. nodes sharing the same V' as the verb).

Furthermore, the syntactic test above and other diagnostic tests not covered in this section[3] are based on the assumption that the syntactic configuration has

[2]Note here that the *term theta* role should be distinguished from the term *thematic role* as used in semantics. Theta roles are thematic relations assigned by the verb to a particular position in the syntax (Cowper, 1992). Unlike semantics, in which an argument can hold more than one thematic role (e.g. the thematic role of *his father* in "His father gave him a gift" verb-specific label such as GIVER but also could be labeled with more of a general label AGENT). In syntax, the Theta Criterion indicates that there only can be one theta role assigned per constituent in a theta grid.

[3]Aside from the *do so* diagnostics, there are other tests that are often used to establish argumenthood. Amongst popular syntactic diagnoses are *ordering restriction* (Radford, 1988, p.235; Culicover and Jackendoff, 2005, p.130; Cowper, 1992, p.32), ellipsis test (Radford, 1988, p.236),

much to say about the distinction. And in display of circular reasoning, these very tests have been used as evidence for the said configurational differences between arguments and adjuncts. These assumptions, as we will see in the next section are problematic.

3.2 Where Structural Argumentation Fails

As a quick reminder, the *do so* diagnostic claims that V' dominating the verb is an all-or-nothing unit – the anaphor *do so* acts on the whole V' (e.g. (23)), and the ordering restriction states that the argument must be realized closer to the verb, before the adjunct (e.g. (24)). Moreover, since *do so* is a substitution test, it will replace the V' or the VP iteratively (e.g. (25)).

(23) Emily ate a burger on Thursday, and John *did so* on Friday.
*Emily ate a burger on Thursday, and John *did so* a pizza on Friday.
(24) Emily ate [a burger] [on Thursday].
*Emily ate [on Thursday] [a burger].
(25) Emily ate a burger slowly on Thursday,
... but Emily *did so* quickly on Friday. [*did so* = ate a burger]
... and Emily *did so* on Friday. [*did so* = ate a burger slowly]
... and John *did so* too. [*did so* = ate a burger slowly on Thursday]

3.2.1 Culicover and Jackendoff (2005)

As a counter to these claims, Culicover and Jackendoff (2005) point out that the antecedent of *do so* "is not necessarily a continuous portion of another sentence" (ibid, p125), where the *do so* can refer to a span of constituents that are contiguous as in (26) but also noncontinguous as in (27). In addition this, they also note that the anaphor can make access to portions of the VP that could not possibly constitute a single constituent within the P&P framework (Culicover and Jackendoff, 2005, p.126-127).

(26) Robin slept for twelve hours in the bunkbed, and Leslie *did so* **on the futon**. [*do so* = slept for twelve hours]
(27) Robin slept for twelve hours in the bunkbed, and Leslie *did so* **for eight hours**. [*do so* = slept ... on the bunkbed]
(28) Robin broke the window with a hammer, Mary *did [it/the same thing/etc]*[4] to the table top. [*do so/did it* = broke ... with a hammer]

and x-happen test (Culicover and Jackendoff, 2005, p.284-285).

[4]See Culicover and Jackendoff (2005, p.126-127) for a discussion on how anaphoras such as *do it* and *do the same thing* act like *do so* in standing in for a subportion of a VP.

DOI: https://doi.org/10.25810/g2ev-pv70

(29) Robin proved to George that Mary would win the race, and Bill *did [it/the same thing/etc]* regarding Susan. [*do so/did it* = proved to George that ... would win the race]

If indeed *do so* makes anaphoric reference to a complete V' at a time, the sentences (27), (28) and (29) should not be grammatically feasible. In the case of (28), it is not only the case that the *do so* is accessing noncontiguous portions of the VP, but also it is leaving out the object *the window*, and therefore, a complement from reference. In the case of (29), the *do so* is substituting for far more than a single verb and its arguments – it is making reference to both arguments and adjuncts of two different verbs. What's more, the constituent "left out" of the substitution, which by the *do so* test should be an adjunct, is the argument of the embedded predicate *win*. Thus, *do so* cannot consistently substitute for the verb plus its arguments in a previous sentence (Ibid, p.127), and it is not always the case that the adjuncts will be those that can be successfully "left out" from the substitution. In other words, the *do so* test fails to distinguish arguments from the adjuncts.

3.2.2 Przepiorkowski (1999)

A similar argument is made by Przepiorkowski (1999). Citing Miller (1992), he argues that *do so* does not have to refer to the meaning of the entire maximal V'/VP node that contains the verb and its arguments. In each of the cited examples below, Przepiorkowski points out that *did so* refers to the meaning of the verb and not the V'/VP.

(30) a. John spoke to Mary, and Peter *did so* **to Ann**. [*did so* = spoke]
b. John spoke to Mary, and Peter *did so* **with Ann**. [*did so* = spoke]
c. John kicked Mary, and Peter *did so* **to Ann**. [*did so* = kicked]

The semantic analysis tells us that the phrase *to Anne* in (a) is the GOAL of the verb *speak*, a verb that directly participates in the event of speaking. Consequently, dropping this argument (i.e. *John spoke*) would alter the meaning of the utterance. On the other hand, the *do so test* incorrectly assesses the phrase *to Anne* in example (30) (a) as an adjunct. A similar problem with the *do so* test is seen in examples (b) and (c) as well where *with Ann* is the GOAL and *to Ann* is the PATIENT, respectively.

Framing Miller (1992, p.96), Przepiorkowski argues that "acceptability of a PP complement *do so* [...] is not whether or not the corresponding complement of the antecedent verb is within the VP of the antecedent, but whether or not the PP complement is acceptable as a complement for the main verb *do* with a thematic role compatible with that which the corresponding complement of the antecedent verb has with respect to the antecedent verb" (Przepiorkowski, 1999, p.290). In other words, if the PP complement after *do so* is thematically the same as the PP complement in a sentence where the antecedent verb sits, then the sentence should

also be acceptable. Consequently, the following sentences in which the PP complement that follows *do so* has a different thematic role than the one accompanying the antecedent, are judged unacceptable as seen in example (31):

(31) a. *John spoke to [Mary]-GOAL, and Peter *did so* [**for Ann**]-BENEF.
b. *John kicked [Mary]-PATIENT, and Peter *did so* [**for Ann**]-BENEF.

3.3 Distinction is Semantic

The eventual conclusion that both Culicover and Jackendoff (2005) and Przepiorkowski (1999) arrive at is that the distinction between arguments and adjuncts belongs to the semantic, and not syntactic, level of analysis. Przepiorkowski (1999) suggests that in Head-Driven Phrase Structure Grammar[5] all adjuncts should be treated as arguments and therefore be included in the argument structure (ARG-ST feature). The ARG-ST feature in the lexical entry of the verb specifies the arguments of the verb. Through an adjunct-addition lexical rule, the lexical item would gain the adjuncts that appear in the sentence[6]. In a similar manner, Culicover and Jackendoff (2005) advocate for a flat structure for a verb, its arguments, and its adjuncts, which puts arguments and adjuncts on the same footing in their syntactic formalism (see Ibid, Chapter 5 and 6). It is at the Conceptual Structure, which contains semantic, aspectual, referential, and functional information that the distinction between arguments and modifiers of the verb is made. That is, both views argue against the structural distinction of argumenthood; rather they propose that the distinction should be made at the semantic layer of description. Thus in the following section, we will return to the semantics side of the issue and look at the distinctions made based on the concepts of obligatoriness and optionality.

3.4 Obligatoriness, Optionality, and Specificity

When considering arguments and adjuncts, the general tendency is to associate obligatoriness with arguments and optionality with adjuncts. This follows from the notion, discussed in section 2.1, that argument labels are given to the participants in a verb's event or state and adjunct labels are given to the 'extra' information that provide settings to the verb. However, the issue is complicated by the fact that it is not the case that all participants involved in an event or a state are always expressed in all possible sentences nor should it be that all participants involved should be expressed in all sentences.

[5]Head-Driven Phrase Structure Grammar (HPSG) is a lexically driven constraint-based generative grammar. See Pollard and Sag (1994) for an introduction to the formalism.

[6]For the full development and use of the adjunct-addition lexical rule and other related features that feed into the addition of adjuncts to the ARG-ST, see Chapter 9 of Przepiorkowski (1999).

DOI: https://doi.org/10.25810/g2ev-pv70

3.4.1 Distinction Amongst Arguments

Jackendoff (2002) makes a distinction between the *semantically obligatory/optional* and the *syntactically optional*:

> It is often said that *eat* "licenses an optional argument". However this conflates semantic and syntactic argument structure. The character being eaten is part of one's understanding, whether it is expressed or not. This becomes clearer by comparison with the verb *swallow*. Its syntactic behavior is identical to that of *eat* [...]. But although one cannot eat without eating *something*, one can *swallow* without swallowing anything. That is, *swallow* differs from *eat* in that its second *semantic* argument is optional. [...] More generally, we need at least to be able to say how many semantic arguments a verb licenses, and which of them are obligatorily expressed. (Ibid., p.134)

Here, Jackendoff draws a clear distinction between what is optional in the semantic sense from the syntactic sense of optionality. This discussion is further developed in Culicover and Jackendoff (2005, p.174-176). The distinction made is this: if a participant is implied in the semantics of the verb, then the participant is *semantically obligatory* for the verb. Otherwise, it is *semantically optional* for the verb. These definitions hold independently of the syntactic expression. Consider the following examples:

(32) a. He swallowed/ate (the food).
b. He swallowed, but he didn't swallow anything.
c. *He ate, but he didn't eat anything.

(33) a. He kicked/threw the pumpkin (down the stairs).
b. He kicked the pumpkin, but it didn't move at all.
c. *He threw the pumpkin, but it didn't move at all.

According to the above definition, in example (32), since INGESTED ITEM is implied in the semantics of the verb *eat* as tested in example (c), this participant is semantically obligatory for *eat*. In the same way, in example (33), the PATH OF MOTION is implied in the meaning of the verb *throw*; it is considered a semantically obligatory argument of *throw*. In contrast to *eat* and *throw*, the thing swallowed for *swallow* and the path of motion for *kick* are not implied by the verbs' meanings, and therefore these arguments are considered semantically optional for the respective verbs. Whether or not the argument is syntactically expressed does not change the labels. In fact, semantically obligatory arguments that are not explicitly expressed in the syntax, Culicover and Jackendoff (2005) term *implicit arguments* (Ibid., p.175).

3.4.2 Distinction Amongst Obliques

In Culicover and Jackendoff (2005), the terms obligatoriness and optionality refer to the distinctions made within arguments. For Koenig et al. (2003), as we will see, the terms refer to the different adjuncts. These are not incompatible arguments, as we will see in the following section. Rather, they represent differing cuts at the description, which are based on differing perspectives from which the issue is examined.

In section 2.1, it was discussed that if a constituent is entailed by the event or state described by the verb, then that constituent is considered an argument. By this definition, the non participants would be classified as adjuncts. At first blush, this definition sounds reasonable – it generally does a good job in describing the observations we have made concerning thematic relations. However, Koenig et al. (2003) brings up a problematic issue with this style of definition, namely, the classification of obliques. Here is a reformulation of the definition for argument by (Koenig et al., 2003, p.72):

> **Semantic Obligatoriness Criterion** (SOC): If *r* is an argument participant role of predicate *P*, then any situation that *P* felicitously describes includes the referent of the filler of *r*.

Thus by Koenig et al. (2003)'s account, the SOC is the criterion by which an argument is recognized from a set of constituents around the verb. However it is insufficient, as they note that if SOC holds, then the italicized obliques in the following sentence would have to also be considered as arguments:

(34) Marc knits *in his office during lunch*.

They write, that SOC is too inclusive of what constituents it "lets in" as an argument. This definition as it stands would include ones that should be classified as adjuncts.

> If you knit, you must knit somewhere; in other words, any situation described by the predicate corresponding to the English word *knit* includes a location in which the event occurred. [...] Again, if the SOC were the sole determinant of argumenthood, the denotations of time expressions such as *during lunch* would qualify as semantic arguments, a conclusion contradicting most linguists' intuitions. (Ibid., p.72)

The problem is that most conceivable events or states have to be located in a certain space and time (Koenig et al., 2003, p.73). Semantic components that describe location, time, and beneficiaries, which Fillmore (1994) calls *circumstantials*, are entailed in just about any situation. Thus, to stop the SOC from overgeneralizing, Koenig et al. (2003) suggest that a second definition criterion related to the specificity of the verb should be introduced:

DOI: https://doi.org/10.25810/g2ev-pv70

> **Semantic Specificity Criterion** (SSC): If *r* is an argument participant role of predicate *P* denoted by verb *V*, then *r* is specific to *V* and a restricted class of verb/events.

SSC forces the choice of argument to be specific to the verb. That is, SSC considers a participant role (as determined by the SOC) as an argument only if the role is asked to "bear additional properties aside from those which are characteristic of the role" (Ibid., p.73). Take the following sentences, as example, paying special attention to the the object or the THEME of the verb:

(35) Marc sang *a song* yesterday.
(36) Marc wrote *a song* yesterday.

By the SOC, we recognize that *a song* plays a role in both verbs (i.e. THEME in both cases). By the SSC we recognize that *a song* for the specific verb *sing* additionally takes on a unique property that is not present in *a song* participating in the *writing* event, namely a quality that requires vocal folds. In a similar manner, the *song* in (36) takes on a unique property that is not present in (35), namely the written quality of the song. To say it in another way, SSC capitalizes on the fact that verbs (or certain classes of verbs) semantically 'color' their arguments slightly differently, and these roles that take on a different 'color' of meaning would be considered by SSC to be an argument. As to the adverbial *yesterday*, the claim is that the meaning is held constant across two or more verbs (or classes of verbs), and therefore it must be an adjunct. That is, the adverbial passes the SOC, but fails the SSC.

What is not discussed in Koenig et al. (2003) are examples of cases in which it is either difficult to detect the "additional properities" the argument bears above and beyond the regular participant roles. Here is an example in which the extra semantic coloring of the argument is not as evident:

(37) Marc put the apple *on the porch*.
(38) Marc ate the apple *on the porch*[7].

Our semantic intuitions tells us that the oblique in (37) is an argument while the oblique in (38) is an adjunct. In order to correctly classify the locative PP in (37), according to the SSC, it would have to hold a meaning that is slightly different from that in (38). It would be up to each reader's judgement to decide if it passes the SSC test; however, the semantic distinction for such locatives is slight at most and difficult to make. This is true not only for locatives, but also for roles such as goals, directionals, and benefactives. The question we ask here is if there is a

[7] Intended reading: *Marc ate the apple while on the porch.*

reliable way of testing if a given constituent holds additional semantic meaning that would classify it as an argument. In the rest of the paper, which is not discussed here, Koenig et al. (2003) present linguistic judgement studies where the SSC is put to test. Given the studies, it seems reasonable that such cases would have been addressed. However, they are not explicitly discussed.

3.4.3 Comparing (Culicover and Jackendoff, 2005) and (Koenig et al., 2003)

In Culicover and Jackendoff (2005), obligatoriness and optionality were dimensions of distinctions amongst the constituents already classified as arguments. That is, obligatory arguments were distinguished from optional arguments, but they were both arguments nonetheless. Here in Koenig et al. (2003), they are seeking to establish a clearer definition for arguments in such a way that amongst obliques it allows as arguments only those obliques that coincide with our intuition of argumenthood.

As noted above, these distinctions are not incompatible. In the general linguistic literature, there is a systematic ambiguity as to how the terms *obligatory* and *optional* are used. The first approach to defining obligatoriness or optionality is to note that amongst arguments there are those that are obligatory for the completion of the predicate's event or state, and those that comment on the setting of said event or state. Here is a graphical illustration:

obligatory	*optional*	
arguments/complement		adjuncts
`semantically obligatory`	`semantically optional`	

This analysis is consistent with Culicover and Jackendoff's (2005) view. A distinction is made within subcategorized constituents into those that are semantically obligatory and those that are semantically optional. The last box in the third row is blank as all adjuncts are considered to be optional.

The second cut can be made through the syntactic core-oblique argument layer. When the cut is made through this layer, then semantic obligatoriness spans over both core and oblique arguments since either can be a potential semantic participant in the verb's event or state. Optionality spans only over the oblique arguments. Here is a graphical illustration of the overlap between what is semantically obligatory/optional and what is core/oblique.

obligatory	*optional*
core arguments	oblique arguments

The distinctions presented by Koenig et al. (2003) are better represented by this analysis. In their case, the focus was on correctly separating argument obliques from adjunct obliques so that the obliques that provide setting or circumstantial information about the predicate, such as those in (34) repeated in (39), would be

DOI: https://doi.org/10.25810/g2ev-pv70

classified as adjuncts.

(39) Marc knits *in his office during lunch*.

Thus, Koenig et al. (2003) and Culicover and Jackendoff (2005) differ, but not incompatibly so, in their use of the terminology of obligatoriness based on the differing approaches the two studies have taken. What remains the same in both cases is that they see the distinction not as a simple dichotomy between arguments and adjuncts coinciding with obligatoriness and optionality, but rather something more complex than that.

3.5 Dowty (2003)

So far we have operated under the assumption that, despite the disagreement on where the distinction should be made, there is a distinction to be made. More specifically, we have examined studies that try to identify semantic and syntactic criteria by which we can say that certain constituents are arguments and certain others are adjuncts. But what would happen if we posit there is no dividing line between arguments and adjuncts? That is, even if there are decidedly "argument-y" things and "adjunct-y" things in a sentence as our intuitions seem to suggest, clear argument and adjunct categories might not exist.

Dowty (2003) suggests that viewing the argument-adjunct issue as a case of clear dichotomy may not be the right analysis. He argues that most distinctions that syntax and semantics have tried to draw between arguments and adjuncts have failed precisely because there is no single clean line that can be drawn between what is the argument-like constituent and the adjunct-like constituent. His solution on the problem of argumenthood is that of a dual analysis, where any given constituent of the VP can be analyzed as both argument and adjunct. As a point of illustration, Dowty (2003) does a short case study on the preposition *to* (Ibid., p.8-11). Consider the following examples:

(40) Mary kicked the ball *to the fence*.
(41) Mary explained the memo *to John*.

What Dowty seeks to claim is that there is a semantic similarity between the two prepositional phrases headed by *to*: they both indicate a physical or abstract location at which something arrived as a result of the action of the predicate. However, the two phrases are also different. In example (40), the PP headed by *to* expresses the new location at which the object arrives at as a result of the action. This is distinguished from the *to* phrase in (41), whose meaning is less compositional and more argument-like than the dative phrase in (40). For example, (41) "does not mean that memo itself came to be at/near John, but only that the information contained in the memo came to be more fully understood by John, as a result of Mary's explanation" (Ibid., 9).

Dowty (2003) points out the usual analyses for these examples either focus on the semantic similarity or on the semantic difference between the two PPs. If analysis recognizes the similarity of the phrases, the two PPs are assigned with the thematic label of GOAL. If analysis wants to recognize that the two phrase are distinct in their semantic expressivity, however, both the argument-like or the adjunct-like uses of the preposition *to* is assigned a different semantic representation. The problem with these approaches, as he writes, is that the former thematic analysis fails to recognize that the two uses of *to* have different semantic values. The latter approach fails to recognize that the two usages are semantically related. Thus, the dual analysis view, which Dowty proposes, remediates both issues by positing, first, an adjunct analysis for both expressions to serve as a launching point for secondary, argument, analysis:

> The idea behind the dual analysis view can be thought of [...] as the claim that the locative adjunct analysis of *all* occurrences of *to*, *from* and other locative prepositions is a PRELIMINARY ANALYSIS which serve language-learners[8] as a semantic "hint" or "crutch" to figuring out the idiosyncratic correct meaning of the complement analysis for the non-locative instance: a preliminary adjunct analysis of the to-PP (as locative) [(40)] gives way to a complement analysis of to-PP structure as in [(41)]. (Ibid., 10)

Dowty (2003)'s claim is that "virtually all complements have a dual analysis as adjuncts" and any adjunct can potentially be reanalyzed as a complement (Ibid., 12). Thus, according to his analysis, the quality of arguments and adjuncts lies in their placement along a directed continuum; every VP constituent other than the verb goes through an adjunct analysis *before* it can get to the argument analysis. While it's unclear from the text when exactly it is that adjuncts would also be analyzed as arguments, at least in this way both argument and adjunct analyses should be available for any given sentence.

4 Implications for NLP and Conclusion

In essence, all of the observations made and views proposed by the linguistics literature on the argument and adjunct distinction show that there is some degree of

[8]Dowty (2003) has a small section devoted to the implications of the dual analysis view, and how it is cognitively a more feasible explanation for language learners who are picking up the argument and adjunct distinction. He notes that if learners first access the adjunct analysis and use it as a clue to learning the argument analysis, the learning burden would be softened. This is an interesting argument as a basis of his views, but there is only one small paragraph expanding on this claim, so there isn't much more that could be said about it.

DOI: https://doi.org/10.25810/g2ev-pv70

argument/adjunct distinction to be made, but it is seemingly without a clear solution. From a linguistic perspective, this is not problematic since it provides ample room for discussion and continuous evaluation of existing theories. However, the same cannot be said for NLP. NLP has to be able to say something definite about the argument and adjunct distinction, since machines have no intuitions to rely on.

Creators of resources for NLP clearly recognize the challenge. Take the English Penn Treebank (ETB; Taylor 1996), as an example. ETB is a syntactic resource that provides the NLP community with manually parsed corpora with phrase structure-like parses, including a wide variety of genres such as written news (e.g. Wall Street Journal), spoken news (e.g. CNN, NBC), and weblogs. The topic of argument and adjunct distinction, as they note, is not a trivial issue. Recognizing the difficulty of argument and adjunct distinction, the ETB creators have chosen not to distinguish arguments and adjuncts at the syntactic level and simply make all post verbal constituents sisters to the verb:

> Unfortunately, while it is easy to distinguish arguments and adjuncts in simple cases, it turns out to be very difficult to consistently distinguish these two categories for many verbs in actual contexts. [...] After many attempts to find a reliable test to distinguish between arguments and adjuncts, we abandoned structurally making this difference. Instead we decided to label a small set of clearly distinguishable roles, building upon syntactic distinction only when the semantic intuitions were clear-cut. However, getting annotators to consistently apply even the small set of distinctions discussed here was fairly difficult. (Taylor et al., 2003)

Their solution was to use the label *closely related* 'CLR', instead, which "marks constituents that occupy some middle ground between argument and adjunct of the verb phrase. These roughly correspond to 'predication adjuncts', prepositional ditransitives, and some 'phrasal verbs' (Bies et al., 1995). The definition as it stands is somewhat vague and there is no specific indication as to what *predication adjuncts*, *prepositional ditransitives* and *phrasal verbs* actually are. As they note, in practice, even the CLR distinction was difficult to make and it was not always the case that CLRs are used as consistently as perhaps ETB intended.

Ideally, the best solution for any NLP application would be to have a specific set of features or criteria that makes a clean distinction between constituents that are arguments and those that are adjuncts. However, it is clear that there is no one set of criteria that suffices for distinctions across all verbs. If Dowty is actually correct in that all elements in the sentence can be given both an argument and an adjunct analysis, it would suggest that the quest for cleanly identifying the distinction might be misguided.

References

Ann Bies, Mark Ferguson, Karen Katz, and Robert MacIntyre. *Bracketing Guidelines for Treebank II Style Penn Treebank Project*, 1995. URL `http://www.sfs.uni-tuebingen.de/~dm/07/autumn/795.10/ptb-annotation-guide/root.html`.

Andrew Carnie. *Syntax: A Generative Introduction*. Blackwell, 2006.

Noam Chomsky. *The Minimalist Program*. The MIT Press, 1995.

Michael Collins. *Head-Driven Statistical Models for Natural Language Processing*. PhD thesis, University of Pennsylvania, 1999.

Elizabeth A. Cowper. *A Concise Introduction to Syntactic Theory*. The University of Chicago Press, 1992.

Peter W. Culicover and Ray Jackendoff. *Simpler Syntax*. Oxford University Press, 2005.

Steve DeNeefe and Kevin Knight. Synchronous tree adjoining machine translation. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2009.

Dmitriy Dligach and Martha Palmer. Word sense disambiguation with automatically retrieved semantic knowledge. *International Journal of Semantic Computing (IJSC)*, 2(3): 365–380, 2008.

David Dowty. On the semantic content of the notion 'thematic role'. *Properties, Types and Meaning*, 2:69–130, 1989.

David Dowty. *The Dual Analysis of Adjuncts/Complements in Categorial Grammar*. Mouton de Gruyter, 2003.

Charles J. Fillmore. The case for case. In Emmon Bach and R. Harms, editors, *Universals in Linguistic Theory*. Holt, Rinehart, and Winston, New York, 1968.

Charles J. Fillmore. Under the circumstances. In *Proceedings of the 20th Annual Meeting of the Berkeley Linguistics Society*, Berkeley, California, 1994.

Jean Mark Gawron. *Lexical Representations and the Semantics of Complementation*. New York: Garland Publisher, 1988.

Ralph Grishman, Catherine Macleod, and Adam Meyers. Comlex syntax: Building a computational lexicon. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Kyoto, Japan, 1994.

Liliane Haegeman. *Introduction to Government and Binding Theory*. Blackwell, Oxford UK and Cambridge USA, 1 edition, 1991.

Donald Hindle and Mats Rooth. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1), March 1993.

Ray Jackendoff. *Semantic Interpretation in Generative Grammar.* The MIT Press, Cambridge, Massachusetts, 1972.

Ray Jackendoff. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press, 2002.

DOI: https://doi.org/10.25810/g2ev-pv70

Paul Kingsbury and Martha Palmer. Propbank: The next level of treebank. In *Proceedings of Treebanks and Lexical Theories*, Växjö, Sweden, 2003.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. A large-scale classification of english verbs. *Language Resources and Evaluation Journal*, 42(1):21–40, 2008.

Jean-Pierre Koenig, Gail Mauner, and Breton Brienvenue. Arguments for adjuncts. *Cognition*, 89:67–103, 2003.

George Lakoff and John R. Ross. Why you can't do so into the sink. In J. McCawley, editor, *Notes from the Linguistic Underground*, volume 7 of Syntax and Semantics. Academic press, New York, 1976.

Paola Merlo and Eva Esteve Ferrer. The notion of argument in prepositional phrase attachment. *Computational Linguistics*, 32(3):341–378, 2006.

Philip H. Miller. *Clitics and Constituents in Phrase Structure Grammar*. Garland, New York, 1992.

Martha Palmer, Dan Guildea, and Paul Kingsbury. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1):71–105, March 2005.

Carl Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar.* University of Chicago Press, Chicago, 1994.

Adam Przepiorkowski. *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. PhD thesis, University of Tübingen, November 1999.

Randolph Quirk, Sidney Greenbam, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. Longman, London and New York, 1985.

Andrew Radford. *Transformational Grammar: A First Course*. Cambridge University Press, 1988.

Josef Ruppenhofer, Michael Ellsworth, Miriam R. L. Petruck, Christopher R. Johnson, and Jan Scheffczyk. Framenet ii: Extended theory and practice. Technical report, International Computer Science Institute, 2010.

John I. Saeed. *Semantics*. Blackwell, 2nd edition, 1997.

Mark Steedman. *The Syntactic Process*. The MIT Press, 2000.

Ann Taylor. *English Penn Treebank Guidelines*, 1996. URL `http://www-users.york.ac.uk/~lang22/TB2a_Guidelines.htm`.

Ann Taylor, Mitchell Marcus, and Beatrice Santorini. The penn treebank: An overview. In Anne Abeillé, editor, *Building and Using Parsed Corpora*, volume 20 of *Text, Speech and Language Technology*. Springer, 2003.