Hudson on heads: About distributional criteria

Sylvain Kahane (Modyco, Université Paris Nanterre & CNRS)

1. Introduction

This text discusses the distributional criteria used in the definition of syntactic headedness. The notion of head is particularly important in dependency syntax, where the syntactic structure cannot be defined without assigning a head to every syntactic unit, but at least since the advent of X-bar syntax (Chomsky 1970, Jackendoff 1977), almost all syntactic theories acknowledge the notion of head.

In this paper, we argue in favor of the central role of distributional criteria in the definition of the syntactic head, and discuss the pitfalls in the statement of these criteria and the consequence of a rigorous application of these criteria to the dependency structure.

This paper is dedicated to Dick Hudson, who, with Igor Mel'čuk, is one of the two linguists who were prominent in the discussion about the criteria used to define dependency structure in the second half of the 20th century.[1] As we will see, Mel'čuk (1988) laid great stress on distributional criteria but in fact circumvented them in practice, while Hudson (1984, 1987) applied them quite systematically but never really stated them clearly.

We will show that there are three distributional criteria, which we call the Positive distributional criterion with removal (section 3), the Negative distributional criterion with removal (section 4), and the Distributional criterion without removal (section 5). We start by recalling the close relation between the notion of head and the dependency tree (in particular for readers not very familiar with these notions) (section 2) and we finish by showing how the distributional method can be initiated on the head of a sentence (section 6) and how the different criteria are related to each other (section 7).

2. Heads and dependency trees

The relation between dependency trees and heads will be exemplified by sentence (1).[2]

(1) *Bill has never met the girl who lives here.*

We propose a dependency tree for (1) in Figure 1.

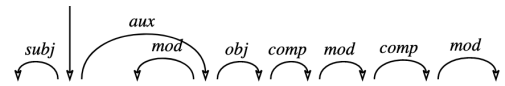

Figure 1. A dependency tree for (1)

It is important to understand that the shape of the dependency tree we associate with a given sentence depends on the criteria we favor for the definition of the head. For instance, in our dependency tree, the determiner *the* governs the noun *girl*, which means that we consider the determiner to be the head of the phrase *the girl who lives here*. Such a phrase will be called a

[1] Of course, many other linguists contributed to the development of dependency syntax during this period, especially Eva Hajičová, who created, with her team, the first dependency-based treebank for a natural language (Hajičová *et al.* 2001).

[2] Some linguists, such as Tesnière (1959) or Hudson (1984, 1987), advocate a more complex dependency structure. We will come back to this point latter, when we discuss *wh*-words.

substantive phrase.[3] This term does not presuppose which element, the noun or the determiner, is the head. The word *Bill* forms a substantive phrase on its own. Such words will be called substantives. Pronouns such as *she* are also substantives, but common nouns are generally not in languages such as French or English, where determiners are compulsory in some syntactic positions.[4] Note that we also consider that the auxiliary, which carries the finiteness feature, is the head of the sentence and that the relative pronoun *who* is the head of the relative clause.

A dependency tree defines numerous units. All the connected subparts of the dependency tree, what Osborne *et al.* (2012) call catenae, are potentially interesting units. Due to the tree-structure of the dependency tree, all these units are naturally headed. For instance, *has never met* is a catena and, as a subpart of the dependency tree, inherits its own tree-structure. According to the analysis given in Figure 1, this unit is headed by *has*.

Conversely, to define the syntactic structure of a sentence, we must first decide which sub-units of the sentence are relevant. If we have identified three units U, A, and B such that U = A B, we know that A and B combine together to form U and we can postulate a connection between A and B. To convert this connection into a dependency between A and B, a second step is needed: we must decide which of the two is the head of U. If A is the head of U, then A governs B, or in other words, B depends on A. For instance, U = *the girl*, A = *the*, and B = *girl* are three sub-units of (1), which means that *the* and *girl* are connected. Note that *the girl* is not a constituent of (1) from the point of view of phrase structure grammar (the projection of *the girl* is *the girl who lives here*), but it is nevertheless a linguistic sign that belongs to (1), because it is a linguistic sign that contributes to the meaning of the whole sentence. If we have criteria to decide that *the* is the head of *the girl*, we conclude that *girl* depends on *the*. Applying the same reasoning to every triplet of sub-units U, A, B of the sentence such that U = AB will give us a dependency tree (Gerdes & Kahane 2013, Kahane & Gerdes 2022). What we need are criteria to identify units and criteria to identify heads of units. We will focus on the latter point.

It is interesting to remark that Tesnière (1959), who is credited with being the first to develop a linguistic theory based on dependency, did not state any criterion. Tesnière was a mentalist, who held that: “Each word in a sentence is not isolated as it is in the dictionary. The mind perceives **connections** between a word and its neighbors. The totality of these connections forms the scaffold of the sentence.” (*ibid*: ch. 1). As noted by Garde (1994), Tesnière used the notion of connection (his term for *dependency*) systematically and applied it to all the constructions he knew and it is a fact that his system works, which justifies his choices of syntactic representation. “It is finally this *pragmatic* utility which justifies *a posteriori*, in each connection, the attribution to one of the terms the role of governor and to the other the role of subordinate” (*ibid*: 96). Tesnière merely remarked (1959: ch. 3) that “every subordinate follows the fate of its governor”. As pointed out by Garde (1994: 96), “This remark, which does not have the status of a definition in his eyes, nevertheless seems to us to be implicitly the key to his conception. From this point on, we can define the superior term *as the one that ensures the linkage of the whole connection with the context in which it is inserted* (On this definition, inspired in part by Mel'čuk 1964, see Garde 1977). Thus the verb is the main term of the clause

[3] Originally, the term *substantive* has been used for nouns, while *noun* (from Latin *nōmen*) included both nouns and adjectives. Since the term *substantive* is no longer used today, we take the liberty of giving it this new meaning, especially since a substantive phrase generally refers to a substance (while a noun alone can generally not refer).

[4] In English, plural nouns and mass nouns can be used without a determiner (while the particular interpretation of such constructions could allow us to postulate a zero determiner). In French, nouns can only be used alone in very special constructions where determiners are banned (see section 4).

because it is the verbal form that determines the independent or subordinate status of the clause." We will present Garde's (1977) very insightful work in section 5.

3. The positive distributional criterion with removal

When deciding which part of a combination U = AB is the head, the simplest case is when A and B can stand alone. Note that it is generally not the case that A and B have this property. For instance, in the combination *the girl*, neither *the* nor *girl* can be easily used alone: clearly *the* can never be used without a noun and there are very few contexts where a singular noun like *girl* can be used without a determiner.[5]

If A and B can stand alone, it is possible to study the distributions of A and B independently. What we call the *distribution* of a unit X is the set of syntactic positions that X can occupy, which can also be described as the set of possible governors for X. Here arises an apparent paradox: to define the head of unit U we need to know what the possible governors of U are. This paradox can be lifted by deciding first what the head in non-controversial cases is and then applying distributional criteria for the remaining cases. An even neater solution is developed in section 6.

We can now state what we call the positive distributional criterion with removal.

> **Positive distributional criterion with removal.** If U = AB, A can stand alone (i.e., B can be removed), and U and A have the same distribution, then A is a head of U.[6]

We do not exclude the possibility for both A and B to be heads of U. If B can also stand alone and the distribution of B is different from the distribution of U we can state that A is *the* head of U.

This criterion was probably first stated by Bloomfield (1933: §12.10), who used the term *form-class* for the distributional classes:

> "Every syntactic construction shows us two (or sometimes more) free forms combined in a phrase, which we may call the *resultant* phrase. […] In subordinative endocentric constructions, the resultant phrase belongs to the same form-class as one of the constituents, which we call the *head*: thus, *poor John* belongs to the same form-class as *John*, which we accordingly call the head; the other member, in our example *poor*, is the attribute. The attribute may in turn be a subordinate phrase: in *very fresh milk* the immediate constituents are the head *milk*, and the attribute *very fresh*, and this phrase, in turn, consists of *ranks* of subordinative position: in *very fresh milk* there are three: (1) *milk*, (2) *fresh*, (3) *very*."

The form-class is defined in the following terms:

> "Certain English words and phrases can appear in the actor position, certain others in the action position. The positions in which a form can appear are its *functions* or, collectively, its *function*. All the forms which can fill a given position thereby constitute a *form-class*." (*ibid*: §12.2)

---

[5] Even for a plural noun, such as *girls*, which can stand alone, we can say that the determiner is compulsory, because, in case of absence of a determiner, *girls* must obligatorily be interpreted as an indefinite phrase.

[6] I should say *roughly* the same distribution. As Hockett (1958: 184) put it: "Some constructions are such that the form-class of the constitutes is similar to the form-class of at least one of the immediate constituents. Here "similar" means that the two ranges of privileges of occurrence largely overlap. The grammarian would prefer to speak of identity rather than similarity, but languages are not completely tight-knit, and greater precision would be spurious."

It is interesting to see how Harris (1951: chapter 16) developed Bloomfield's notion of distributional classes, trying to define all the classes of English. He started by the distributional classes of morphemes and then combined them to obtain new classes. Sometimes, the combination does not give a new class: For instance, the combination of an adjective and a noun is "substitutable" with a noun (A N = N, section 16.21 on *Non-repeatable Substitutions*). And sometimes it gives new classes: For instance, from N^1 for bare nouns (*boy*), he defines N^2 for nominal forms ($N^2 = N^1$-*s*) and then N^3 for adjective-noun combinations ($N^3 = A\ N^3$, the combination is iteratable) and finally N^4 for combinations of N^3 with a determiner ($N^4 = T\ N^3$). Harris did not introduce the notion of head and considered that each N^{i+1} is an extension of N^i and not really a new class (*ibid*: Note 14, p. 266). If T N is not equivalent to N it is because "we cannot substitute T N for N in A $N^3 = N^3$ for we would derive a non-extant A T $N^3 = N^3$ (*Swiss some cheeses*)."

Hudson (1984: 90) used the Positive distributional criterion with removal to state that the determiner is the head of the substantive phrase:[7]

> "Many of these words (e.g. *all, three, some, this, which*) can occur on their own, with an understood noun, in positions otherwise available only to the nouns (e.g. as inverted subject), so they must themselves classified as nouns […]"

Hudson (1987: 121) used this criterion more extensively and more clearly:

> "We now turn to Comp+S, where of course we have to take account of examples like *since he left* and *what he did* as well as of those like *that we control those penguins* which Zwicky cites. Once again the questions about which part is the distributional equivalent of the whole and about which part is obligatory are closely interconnected. […] Take a complementizer like *since*, which is much more typical. Here it is possible to omit the clause, by anaphoric ellipsis similar to the process responsible for VP ellipsis (compare *I haven't seen him since we had that argument* and *I haven't seen him since*); but it is not possible to omit the complementizer (compare **I haven't seen him we had that argument*)."

---

[7] The link between this argument and the list of criteria stated by Hudson is not self-evident. Hudson (1984: 78) proposed the following list of criteria for the head: "I shall briefly mention the main types of property to which the asymmetry applies:

a) temporal order (already illustrated)
b) possibility of occurrence – whether the modifier [= dependent] may, or must occur at all rests with the head, and not vice versa.
c) inflectional selection (i.e. traditional 'government') – the head decides which inflectional form of the modifier occurs;
d) lexical selection (alias 'colocation') – the head selects a particular, formally specified, word (e.g. *decide* selects the preposition *on*);
e) semantic structure – the head provides a structure into which the modifier fits (e.g. *brown* defines the colour of the eyes, rather than *eyes* defining the application range of *brown*)."

The first criterion (temporal order) was discussed a page before: "What is the basis of the asymmetry between a head and its modifier? More concretely, how do we know which member of an established companion pair is the head? The most general answer is that it is the head that provides the link between the modifier and the rest of the sentence, rather than vice versa. For example, the position of *brown* in *She has brown eyes* is fixed in relation to *eyes*, and not vice versa (*brown* must precede *eyes, eyes*, as object, must follow *has*; its position after *brown* is just the converse of the latter's position before it, and it is this position which is specified in the grammar.)" I do not think that this criterion is operational headedness. We can state that "the subject goes before the verb" or that "the verb goes after the subject"; both ways are statable. It is because the verb is the governor of the subject that we prefer the first formulation rather than the other way around. Only the second criterion (possibility of occurrence) considers a distributional criterion and we can assume that it was this criterion that Hudson used for the determiner.

The limits of the Positive distributional criterion with removal were pointed out by Bloomfield (1933: sect 12.20) himself who considered that some phrases do not have any head:

> "The resultant phrase may belong to a form-class other that of any constituent. For instance, *John ran* is neither a nominative expression (like *John*) nor a finite verb expression (like *ran*). Therefore we say that the English actor-action construction is *exocentric*: the resultant phrase belongs to the form-class of no immediate constituent."

But there is one problem in Bloomfield's reasoning: We do not know how the distributional class of *ran* can be compared with the distributional class of *John ran*, due to the fact that *ran* can never stand without a subject. This leads us to the next point.

4. Negative distributional criterion with removal

Mel'čuk (1988: 132) defined the positive distributional criterion with removal as follows, calling the distribution the *Surface-Syntactic valency* (*SS-valency*):[8]

> "**Criterion B.I** (imposition of passive SS-valencies). The SS-head of w1 — w2 [the combination of the words w1 and w2] is the wordform that determines the passive SS-valency of the phrase to a greater degree than the other wordform."

He completes this criterion by saying:

> "In several cases, one of the two wordforms, for example w1, can never be used alone (that is, without w2) as a dependent element in syntactic roles where the whole phrase w1 — w2 appears. In such a case, the passive SS-valency of w1 is that of the phrase w1 — w2, on the condition, however, that the latter is different from the passive SS-valency of w2." (*ibid*: 133)

He illustrates the second point with U = *to Paris*. As w2 = *Paris* has a distribution that is very different from U and w1 = *to* cannot stand alone,[9] Mel'čuk considers that w1 and U have the same distribution. According to criterion B.I, then w1 is the head of U.

It would be preferable to express matters more directly: As w2 = *Paris* has a distribution that is very different from U, the combination of w2 with w1 = *to* substantially changes the distribution of w2 and then w1 is the head of U. We do not need to attribute a distribution to w1, which can never stand alone. We obtain the following criterion, which I call the negative distributional criterion with removal.

> **Negative distributional criterion with removal.** If U = AB, B can stand alone (i.e., A can be removed), and U and B do not have the same distribution, then A is a head of U.

We remove the condition that A cannot stand alone, because even if A can stand alone, the fact that the combination of A with B changes the distribution proves that B is not a true dependent. The criterion can moreover be reformulated as follows:[10]

---

[8] More precisely, Mel'čuk calls *valency* the set of syntactic connections a unit has with other units. He divides the valency into the *active valency*, which is the set of dependent positions, and the *passive valency*, which is the unique connection with a governor. I do not find this terminology appropriate, especially for modifiers that "actively" select their governor.

[9] As pointed out by the reviewers, *to* can occur without an overt complement in some constructions (as with *come to* 'regain consciousness' and *pull the door to*). But if the example were changed to U = *from Paris*, then there could be no discussion.

[10] The equivalence of the two variants can be done in 2 steps: first "if p and q, then r" can be replaced by "if *p*, then [if *q*, then *r*]" and then by "if *p*, then [if no *r*, then no *q*]", which gives us "If U = AB and B can stand alone,

> **Variant of the negative distributional criterion with removal.** If A can stand alone and B is a dependent of A, the combination of A with B cannot change the distribution of A.

The Negative distributional criterion with removal can be applied to Bloomfield's example of U = *John ran*. As B = *John* does not have the same distribution as U, then A = *ran* is the head of U.

Although Mel'čuk stated an equivalent of the Negative distributional criterion with removal, he used it erroneously. He did not really discuss the question of the head of the substantive phrase, but only said, about Fr. *une fille* 'a girl', that "*une* is syntactically dependent on *fille*, since it is the noun that imposes its passive valency on the noun phrase" (*ibid*: 114). This reasoning is fallacious, because it is not true that the noun imposes its distribution on the substantive phrase in French. Indeed, the noun can rarely be used without a determiner (2) but, where it can, it cannot commute with a substantive phrase (3)-(4).

(2) a. *J'enseigne la syntaxe.* 'I teach [the] syntax.'
b. **J'enseigne syntaxe.*

(3) a. *On parle syntaxe.* 'We speak (about) syntax.'
b. **On parle la syntaxe.*

(4) a. *un livre de syntaxe* 'a book of syntax'
b. **un livre de la syntaxe.*

The Positive and Negative distributional criteria with removal are generally used together, as in the argument by Hudson (1987: 121) for Comp as the head of a complementizer phrase given in section 3. Let us see again the end of the citation, where the Negative Negative distributional criterion with removal is used:

> "Take a complementizer like *since*, which is much more typical. Here it is possible to omit the clause, by anaphoric ellipsis similar to the process responsible for VP ellipsis (compare *I haven't seen him since we had that argument* and *I haven't seen him since*); but it is not possible to omit the complementizer (compare **I haven't seen him we had that argument*)."

Here the use of the Negative criterion is justified, because B = *we had that argument* can stand alone without any problem, so we can conclude that A= *since* is a head of U = AB from the fact that B and U do not have the same distribution.

We can complete Hudson's argument by saying that the Negative criterion can also be used for A = *that*. Even if B = *we control those penguins* and U = AB can commute in some contexts (after *I think* for instance), they cannot commute in many contexts (after *since* for instance, or as the root of a sentence). Therefore B and U do not have the same distribution and A is a head.

5. Distributional criterion without removal

There is third distributional criterion that does not refer to the distribution of A and B and does not need to hypothesize that A or B can stand alone. To my knowledge, this criterion was first stated in a little-known work by Paul Garde published in 1977. Garde is a French Slavist who only marginally worked on dependency syntax, but he was interested in Tesnière's work (who was also a Slavist) and he also read Mel'čuk's early work in Russian (as shown in the passage quoted in Section 2).

---

then [if A is not a head of U (i.e. A is a true dependent of B), then U and B must have the same distribution]". It remains to exchange the role of A and B.

Here are the three criteria proposed by Garde (1977: 8) to state that A is the head of AB:[11]

"1) in a given context, B can be removed, not A:

X (AB) → X (A) and not *X (B)

2) in a given context, B can be replaced by something else, whereas A cannot:

X (AB) → X (AB′) and not *X (A′B)

3) a change in the context leads to a change of A, not of B:

X (AB) → Y (A′B) and not *Y (AB′)."

Garde's criterion 1 aims at the Negative distributional criterion with removal. As it is stated, it is not appropriate: For instance, in the case of B = *that* and A = *we control those penguins*, we will conclude that A is the head because in the context X = *I know*, B can be removed and not A. To be correct, the criterion must specify that it is applicable only if B can stand alone, because if it cannot we will always have *X(B). Moreover it is not necessary to require A to be able to stand alone. If we want to have a criterion based on X(A), it will be the Positive distributional criterion with removal, which states that A and AB have the same distribution, that is that X(AB) → X(A) for *every* context X (and not just a given one).

We can reformulate the Negative distributional criterion with removal in Garde's style:

1′) although B can stand alone, in a given context, A cannot be removed:

X (AB) and not *X (B)

What particularly interests us here are Garde's second and third criteria. They state that A is the head of AB if modifications on the governor X of AB are correlated with modifications on A, not on B. This is stated in two steps: commutations on B are always possible, while in some contexts a commutation on A is not possible (criterion 2); commutations on the context X can force a commutation on A, but have no impact on B (criterion 3).

Again, Garde's formulation is not completely satisfying; as with criterion 1, it is not necessary to require that what must be verified by A must not be verified for B. We propose the following reformulation.

Distributional criterion without removal: If U = AB, A can commute with an A′, and U′ = A′B does not have the same distribution as U, then A is a head of U.

In other words, if B depends on A, then B must not modify the distribution of A and a commutation on B does not change the distribution of the unit it forms with A.

Note that every commutation assumes a mutual exclusion: in other words, if A commutes with A′ in the context X, then *X(AA′).

We can reformulate the Distributional criterion without removal in Garde's style as follows:

2′) in a given context, A cannot be replaced by something else:

X (AB) and not *X (A′B)

3′) a change in the context leads to a change of A:

X (AB) and *Y (AB), but Y (A′B)

[11] My translation. The roles of A and B have been interchanged.

Note that criteria 2′ and 3′ are very similar: 2′ states that, for some A′, the distribution of A′B is not included in the distribution of AB, while 3′ states that, for some A′, the distribution of AB is not included in the distribution of A′B.

A particular case of the Distribution criterion without removal is what Zwicky (1985) called the Morphosyntactic locus criterion: the head of a unit U is "the constituent on which inflectional marks will be located. […] One way in which a constituent can 'characterize' a construct is that it can be the bearer of the morphosyntactic marks of syntactic relations between the construct and other syntactic units." This argument is used to choose the verb as the head of the clause (rather than the subject) and the auxiliary as the head of a complex verbal form (rather than the lexical verb), because the verb, or its auxiliary in the case of complex forms, bears the mode imposed by the governor of the clause. And a change in the mode changes the distribution of the clause. For instance, it is the auxiliary *be* that bears the subjunctive imposed by the construction in (5a) and this form could not be used in the context of (5b).

(5) a. […] *a World Health Organisation recommendation that all salt be iodised*. (GUM)

b. *All salt is iodised following the World Health Organisation recommendation.*

Zwicky (1985:7) himself extended the Morphosyntactic locus in cases without inflection, which fall under the Distributional criterion without removal:

> "English P + NP has no actual inflection locus: the NP does bear the marks of person and number, but person and number play no role in the distribution of P + NP constructs; and P in English is uninflected. However, there is a phenomenon long recognized as analogous to case inflection here, namely the choice of one preposition over another in constructions like *inform Sandy of the news* and *tell the news to Sandy*, where the prepositions *of* and *to* mark particular syntactic arguments of the verbs. The analogy suggests that P is the morphosyntactic locus in P + NP."

In other words, the choice of the verb (*inform* or *tell*), which is the context of the construct, is correlated with the choice of the preposition.

Before that, Hudson (1984: 84) stated that interrogative pronouns, such as *what* in *I don't know what he said*, have a double syntactic position by more or less using the Distributional criterion without removal (see the second part of his argument):[12]

> "The word *what* depends on *know*, but it also depends on *said*. The latter connection is obvious – *what* is object of *said* – but the former is less obvious. One argument is that *what* can occur on its own, without the rest of the interrogative clause, as in *he said something, but I don't know what*. In this sentence, *what* must depend on *know*, because there is nothing else for it to depend on, so we may assume that the same is true even

[12] The double position of *wh*-words was also defended by Tesnière (1959: ch. 246): "Take for example the pronoun in the French sentence: *l'homme qui écrit* 'the man who writes'. The pronoun *qui* is a translative of the transferred clause the verb of which is *écrit*, because it transfers this clause to an adjective subordinate to *l'homme*. It is also the first actant of the same verb *écrit*, since as an anaphor it represents the word *homme* and is therefore the first actant of *écrit*. […] The relative pronoun is hence a word that has a **double nature**."

On the contrary, Mel'čuk (1988: 26) argues that "the relative pronoun depends syntactically only on the main verb of the relative clause. True, it depends on its own antecedent as well, but solely morphologically." Although he stated criteria in great detail, Mel'čuk never uses them to justify an analysis. In a certain sense, he is satisfied if his syntactic representations allow him to model everything that he wishes to model (Mel'čuk's goal is to describe the correspondence between meanings and texts and dependency trees are just go-betweens in this correspondence). And if he ever agrees that one of his analyses contradicts his criteria, he will change the criteria, not the analyses, for sure.

> when there is something else to depend on. Another argument is that if we take *what* as an instance of the more general category 'interrogative word' (also containing *whether*), then we have a simple way of saying that verbs like *wonder* must take an interrogative clause as their complement."

In other words, an interrogative clause does not have the same distribution as an ordinary clause (the former can be the object of *wonder* not the latter) and if it is an interrogative clause it is because it contains an interrogative word. Consequently, the interrogative word is the head of the interrogative clause.

Hudson (1987: 121) used the Distributional criterion without removal even more clearly in his answer to Zwicky and he presents it as a natural extension of the Morphosyntactic locus:

> "For example, the verb *doubt* allows either *whether* or *if* as complementizer in its complement (e.g. *I doubt whether/if it's true*). This fact cannot be stated in semantic terms, as allowing an interrogative structure as complement, because most kinds of interrogative structure are not in fact allowed after *doubt* – compare **I doubt why he did it* – and in any case, *whether* and *if* are here interchangeable with *that* (e.g. *I doubt that it's true*) without change of meaning. Similarly verbs seem to differ as to whether they accept either *if* or *whether* – so *discuss*, for example, seems to prefer *whether* strongly to *if* (e.g. *We were discussing whether/?if we had enough money*). In other words, the subcategorization for verbs must be able to distinguish between clausal complements which are introduced by different complementizers just as it must be able to be sensitive to the prepositions in prepositional complements. In short, Comp has at least as good a claim to be taken as the morphosyntactic locus in Comp + S as S does."

The Distributional criterion without removal gives one argument for the noun as a head of the substantive phrase, because a commutation on the noun can change the distribution. For instance, *three men* has a quite different distribution from *three tables*, because many verbs accept only +human phrases as subject (**three tables think that's a good idea*). This can be considered as a semantic constraint rather than a syntactic one, but there are cases where the commutation entails a completely different distribution: some nouns, when combined with a determiner, have an adverbial distribution and can modify a verb: *I stayed three days; I came three times*. It is of course absolutely impossible for ordinary nouns: **I stayed three men; *I came three tables*.

It is not our goal to discuss the question of the head of the substantive phrase in detail (see Hudson 2004, forthcoming, or Osborne 2021), but just to see how the distributional criteria work and are used.

6. The non-circularity of distributional criteria for headedness

The distributional criteria we consider are based only on the syntactic context. We are not interested in the linear context. And by the syntactic context, we only mean the set of potential governors. We are not interested in the dependents of the unit: The head of a unit is the contact point with the governor of this unit.

As already noted by Mel'čuk (1988:115), the need to know where the governor of the unit under investigation is opens up an apparent circularity in the definition of the head: The notion of head is based on the notion of governor, which is equivalent to the notion of head (A is the governor of B if and only if A is the head of AB). In fact, there is no paradox in this definition because we can start by a particular case which does not involve a governor strictly speaking. This is the case of the head of the sentence, the root of the dependency structure, which is not governed.

A sentence is an illocutionary unit. This gives us the following criterion.[13]

> **Distributional criterion for the head of a sentence.** The head of a sentence is the element that bears the illocutionary force and will be affected if the illocutionary force is modified.

Let us see how this criterion applies to English: English has special constructions for a question or a negation involving a particular position of an auxiliary or modal verb. In (6), *came* is replaced by *did ... come* when the assertion becomes a question. According to the previous criterion, this shows that the verb bears the illocutionary force and is affected when the illocutionary force changes. Moreover, the negation is carried by the auxiliary or modal verb, indicating that it is the head.

(6) a. *Peter came.*
b. ***Did*** *Peter come?*
c. *Peter* ***didn't*** *come.*

French also has a particular marking for questions and negations that clearly identifies the auxiliary as the head of a complex verbal form. Questions are marked by a pronominal enclitic on the head, while negations have a double marking by two adverbs, *ne* and *pas*, that surround the head.

(7) a. *Pierre a dormi.* 'Peter has slept.'
b. *Pierre a****-t-il*** *dormi ?* 'Has Peter slept ?'
c. *Pierre* ***n'a pas*** *dormi.* 'Peter hasn't slept.'

For some languages, it is even clearer. For instance, in Korean, there is a special morpheme *da* that marks an assertion (8a), and that can commute with three other values for question (8b), order (8c), and suggestion (8d).[14]

(8) a. *Yongi-ka cha-leul bo-n-****da*** 'Yongi sees the car.'
Yongi-NOM car-ACC see-PRES-**DEC**

b. *ney-ka cha-leul bo-****ni*** 'Do you see the car?'
you-NOM car-ACC see-**INT**

c. *ney-ka cha-leul bo-a-****la*** 'You, look at the car!'
you-NOM car-ACC see-INF-**IMP**

d. *ney-ka cha-leul bo-****ca*** 'You should look at the car.'
you-NOM car-ACC see-**PROP**

7. Link between the different distributional criteria

The Distributional criterion *without* removal is, in my opinion, the most important and powerful criterion for headedness, but it has, to my knowledge, never been clearly distinguished from the Negative distributional criterion with removal. For instance, Zwicky (1993: 297) gives an interesting formulation of headedness that evokes the Distributional criterion without removal,

---

[13] Mel'čuk (1988 :115) does not give any criterion, considering the question obvious: "this circularity is eliminated because the concept of passive SS-valency includes the capacity of the unit in question to be the absolute head of an utterance. This capacity is postulated in this book for the finite verb. The dominant syntactic role of the finite (i.e., tensed) verb is intuitively evident. Taking it for granted, one proceeds by induction, the finite verb constituting the induction basis. The other cases of syntactic dependency are reduced, sometimes in several steps, to dependency on the finite verb, that is, on the grammatical predicate of the sentence."

[14] NOM = nominative, ACC = accusative, PRES = present, DEC = declarative, INT = interrogative, INF = infinitive, IMP = imperative, PROP = propositive.

but he concludes by identifying it with other criteria and the examples he gives do not allow us to distinguish them:

> "With specific reference to its external syntax, the Head is the determinant in a somewhat different sense: the distribution of the construct as a whole is predictable from properties of the Head, the properties of a Dependent being irrelevant or 'transparent' in the matter, so that the Head determines what is in effect the lexical subcategory of the construct as a whole. […] For external purposes, *demonstrate that the earth is flat* has a distribution predictable from the properties of *demonstrate*, with the Argument *that the earth is flat* transparent for these purposes, and *very red tomatoes* has a distribution predictable from the properties of *tomatoes*, with the Modifier *very red* transparent for these purposes. This is the 'distributional equivalence' criterion for headship of Zwicky (1985), which was based on Bloomfield's (1933: 194) formulation in terms of 'same form class'."

The Negative distributional criterion *with* removal can be seen as a limit subcase of the Distributional criterion *without* removal. Indeed, when B can stand alone and A is removable, it means that A has the "empty unit" in its paradigm of commutation. The Distributional criterion without removal says that A is the head if at least one commutation on A changes the distribution. If the removal of A is seen as a commutation with emptiness, the fact that B has not the same distribution as AB means that at least one commutation on A (the commutation with emptiness) changes the distribution.

Conversely, if A can stand alone and B is removable, we can add the removal of B as one of the possible commutations on B and see the comparison of the distributions of A and AB as a particular case of the initial condition of the Distributional criterion without removal ("B can commute with a B′ and U′ = AB′ does not have the same distribution as U = AB"). The Negative distributional criterion with removal appears then as a particularly strong criterion which states that only the particular case of the comparison of the distribution of A and AB is sufficient to conclude that A is the head.

8. Conclusion

We conclude this paper by saying that, to our knowledge, the distributional criteria, which are essential for the characterization of the head of a syntactic unit, have never been completely stated properly. I hope we have convinced the reader that they cannot simply be stated in terms of the distribution of the constituents of the unit under study because the distribution of a unit can only be considered if this unit can stand alone. That is why it is very difficult to prove the head status of word that can generally not stand alone, such as determiners, prepositions, or auxiliaries. For such words, we need to use a distributional criterion that does not involve the removal of one of the constituents of the unit under study. Although he never stated such a criterion in black and white, Dick Hudson was the first one to really try to justify his choices of analysis in dependency syntax by using distributional criteria and to use both distributional criteria with and without removal.

Acknowledgements

I would like to thank Bas Aarts and And Rosta for valuable remarks on this paper.

References

Bloomfield, Leonard (1933) *Language*. New York: Holt, Rinehart and Winston.

Chomsky, Noam (1970). Remarks on nominalization. Jacobs, Roderick A., Peter S. Rosenbaum, eds. *Studies on Semantics in Generative Grammar*, 11-61.

Garde, Paul (1977) Ordre linéaire et dépendance syntaxique : Contribution à une typologie, *Bulletin de la Société de Linguistique de Paris*, 72(1), 1-26.

Garde Paul (1994). Syntaxe et sémantique chez Tesnière. *Linguistica*, *34*(1), 95-99.

Gerdes, Kim, Sylvain Kahane (2013). Defining dependencies (and constituents). *Frontiers in Artificial Intelligence and Applications*, *258*, 1-25.

Hajičová, E., Hajič, J., Hladká, B., Holub, M., Pajas, P., Řezníčková, V., Sgall, P. (2001) The current status of the Prague Dependency Treebank. *Proceedings of the International Conference on Text, Speech and Dialogue*, Springer, Berlin/Heidelberg, 11-20

Harris, Zellig S. (1951) *Methods in Structural Linguistics*, The University of Chicago Press.

Hockett, Charles F. (1958) *A Course in Modern Linguistics*, New York: MacMillan Company.

Hudson, Richard (1984) *Word Grammar.* Oxford: Blackwell.

Hudson, Richard (1987) Zwicky on heads. *Journal of linguistics*, *23*(1), 109-132.

Hudson, Richard (2004) Are determiners heads? *Functions of Language*, 11, 7–43.

Hudson, Richard (forthcoming) Arguments against the universality of D and determiners. In Solveiga Armoskaite, Martina Wiltschko (eds.) *The Oxford Handbook of Determiners*. Oxford University Press.

Jackendoff, Ray (1977) *X syntax: A study of phrase structure.* Linguistic Inquiry Monographs Cambridge, Mass.

Kahane, Sylvain, Kim Gerdes (2022) *Syntaxe théorique et formelle : Modélisation, Unités, Structures*, Berlin: Language Sciences Press.

Mel'čuk, Igor (1964) *Avtomatičeskij sintaksičeskij analiz, Tom I* [Automatic syntactic analysis, Vol. 1], Novosibirsk: Izdatel'stvo SO AN SSSR.

Mel'čuk, Igor (1988) *Dependency Syntax: Theory and Practice*. State University of New York Press.

Osborne, Timothy, Michael Putnam, M., Thomas Gross (2012) Catenae: Introducing a novel unit of syntactic analysis. *Syntax*, *15*(4), 354-396.

Osborne, Timothy (2021) NPs, not DPs: The NP vs. DP debate in the context of dependency grammar. *Acta Linguistica Academica*.

Tesnière, Lucien (1959) *Éléments de syntaxe structurale*, Klincksieck [transl. by T. Osborne, S. Kahane, 2015, *Elements of structural syntax*, John Benjamins].

Zwicky, Arnold (1985) Heads. *Journal of linguistics*, *21*(1), 1-29.

Zwicky, Arnold (1993) Heads, bases and functors. In *Heads in grammatical theory*, 292-317.