From Structural Syntax to Constructive Adpositional Grammars

Federico Gobbo
Research Center "Informatica Interattiva"
University of Insubria, Varese, Italy
federico.gobbo@uninsubria.it

Marco Benini
DICOM
University of Insubria, Varese, Italy
marco.benini@uninsubria.it

Abstract

The importance of the research made by Tesnière (1959) for the concepts of dependency and valency cannot be underestimated. However, his Structural Syntax remains still uninvestigated in most part. In this paper, a formal grammar model that follows Tesnière's intuitions and concepts as much as possible is proposed. This model is called constructive adpositional grammar. This paper explains the linguistic and formal reasons behind such a research plan.

1 Introduction

Research in dependency linguistics acknowledges a lot from the work by Lucien Tesnière, the French linguist who introduced, in modern times, the key concepts of dependency and valency. Nonetheless, unlike valency, there is no agreement among scholars and specialists on how to treat precisely the concept of dependency. In fact, even if the theoretical assumption behind all dependency-based models is fairly the same, namely "the syntactic structure of sentences resides in binary asymmetrical relations holding between lexical elements" (Nivre, 2005, 6), in practice the use of this assumption is different among authors. For example, in Topological Dependency Grammar (TDG), proposed by Debusmann (2001), there are two different forms of dependencies, called "syntactic dependency tree (ID tree)" and "topological dependency tree (LP tree)", while Mel'čuk (1988) postulates three types of syntagmatic dependency relations: semantic dependency, syntactic dependency and morphological dependency. How did it all begin? In other words, how Tesnière *really* defined dependency? What can be saved – and adapted – for a dependency-based linguistic model that is formally feasible with modern mathematical and computational tools?

2 Governor, dependent, connection

Tesnière (1959) can be considered the *summa* of his work, being more than 600-pages long, where his language analysis system, called Structural Syntax (in French: *syntaxe structurale*), is explained in detail.[1] That work was published posthumously, and for this reason it is not always coherent in all parts; however, every paragraph is numbered referring to a Chapter that belongs to an internal Book (from A to F) belonging to a Part (from 1 to 3). In the sequel, references to that work will take the original form. For instance, paragraphs 1–8 of Chapter 21 belonging to Book A of Part 1 will be referred like this: (1, A, ch. 21, par. 1–8). Analogously, it was decided to retain the original numbers of Tesnière's examples (*stemma*) in order to help the reader in the comparison between Structural Syntax and the model presented in this paper, while new examples are numbered through capital letters.

Figure 1: How connection works for Tesnière

[1] Unfortunately Tesnière (1959) is still not available in English. All English translations are proposals written especially for this paper.

In (1, A, ch. 1, 4-12) the notion of connection (*connexion*) is presented. In Figure 1 the first examples of Tesnière (1959) are shown: in *Alfred parle* ('Alfred speaks'), the verb *parle* is the governor (*régissant*), the noun *Alfred* being the dependent (*élément subordonné*). Their relation, "indicated by nothing" (1, A, ch. 1, 4) is their connection. Connections can be recursive, as in example 2 *mon ami parle* ('my friend speaks'): governors are put above, dependents are put thereunder. It is interesting to note that Tesnière did not use the word 'dependency' (*dependance*) but 'connection'. This choice becomes clear when the dichotomy 'structural *vs*. semantic influence' is introduced (1, A, ch. 2, 3).

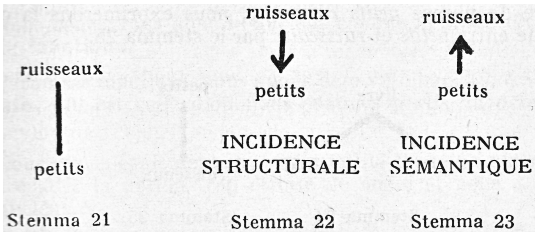

Figure 2: Structural and semantic influence

Figure 2 shows that two connections between the elements of *petits ruisseaux* ('little streams') are possible: either the governor *ruisseaux* structurally influences the dependent *petits*, or the dependent semantically influences the governor – i.e., by grammaticalization, e.g. in the proverb *Les petits ruisseaux font les grandes rivières*, 'tall oaks from little acorn grow', literally, "little streams form big rivers". Here, it seems that the French linguist wants to give the apart status of dependency only to semantically influenced connections. Unfortunately, this crucial point is never mentioned anymore throughout the book (more than 600 pages); in fact, only generic, underspecified connections are actually used.

In sum, Tesnièrian Structural Syntax shows a triple in order to describe the connection between two linguistic elements: governor, dependent, connection. Moreover, it is admittable that connections can be generic, structurally or semantically influenced. The depicting of this triple through unary trees – called *représentation stemmatique*, let it be 'structural syntactic trees' hereafter – made by the author seems not to be the best way to describe such a structure, under a formal point of view. For this reason, the model proposed here makes use of a special form of binary trees, called 'adpositional trees', in brief *adtrees*.

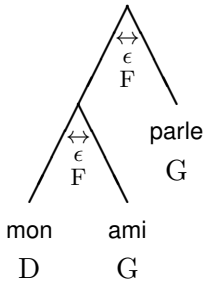

Figure 3: How connection is reinterpeted

Figure 3 shows the reinterpretation of example 2 in terms of adtrees instead of structural styntactic trees, where all structural elements become far more evident. Governors (indicated by G) are put on the right leaf of the binary trees, while dependents (indicated by D) are put on the left ones. The third element of the triple, is put as a "hook" under the root of the tree, (indicated by F, for 'final'). What Tesnière conceived as the connection, can be represented as *adposition*. In fact, in many languages of the world what gives the final character (F) to the connection is a preposition, a postposition or another form of adposition: this fact gives the same dignity to morphology and syntax, unlike Tesnière's tenet (see section 3 below). In the case of example 2, as the connections between *mon ami* and *parle* and *mon* and *ami* are morphologically unmarked, i.e., they are syntactic adpositions, epsilons (ϵ) are put accordingly (figure 3). The influences behind connections are left underspecified through left-right arrows ($\leftrightarrow$).

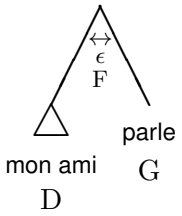

Figure 4: Information hiding explained

The main linguistic difference between the proposed structure, using adpositional trees, which are binary, and Tesnièrian structural syntactic trees, which are unary, is the following: from the point of view of the verbal governor *parle*, the dependent is the whole tree depicting *mon ami* (Figure 4). On the contrary, in Structural Syntax, apparently only *ami* is dependent of *parle* (figure 1 right). Furthermore, the small triangle ($\triangle$) indi-

cates that a binary tree was "packed" in order to increase human legibility: however, no information was lost, but only *hidden* – i.e., it is always possible to get that information explicit. No such possibility is present in Tesnière (1959). However, a single example of something similar to information hiding is provided, when dealing with grammaticalization (1, A, ch. 29, par. 21).

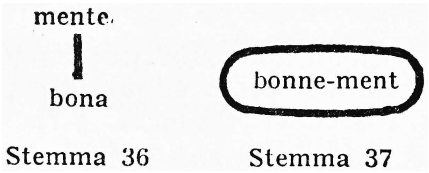

Figure 5: Grammaticalization example

Figure 5 shows how the Latin syntactic expression *bona mente* ('with a good mind') became in French *bonne-ment* ('quite simply') for grammaticalization.[2]

This example lead to the goal of providing a coherent treatment in terms of binary trees of all features explained in Structural Syntax – and consequently, in terms of the underlying formal model, as explained below. In fact, one of the great problems in Tesnière (1959) is that the examples (*stemma*) are illustrated in different ways throughout the work, where some information got lost during the way, and other introduced – for instance, connection influence, as presented above, got completely lost.

The explicitation of the triple 'governor, dependent, connection' let the structure to be illustrated with the recursively use of adtrees – partially hidden when needed – retrieving Tesnière's structural information whenever possible.

3 Word classes and syntactic functions

Tesnière (1959) quite early introduces a set of symbols which "express the deep nature [of structure] without keeping the accidental contingencies" (A, ch. 33, par. 1). For Tesnière, morphology is the "shallow facet" while syntax is the "essential facet" of structure, i.e., Humboltian *Innere Sprachform* – in modern terms, deep structure (1, A, ch. 12, note 1). This emphasis on syntax is a severe limit, perhaps a cultural heritage of the times when the French linguist lived, where so much importance was given to morphology, almost neglecting syntax. However, now times changed, and it is possible to express both syntactic and morphological information in the same binary tree structure (although, from a formal and computational point of view, syntax and morphology still may be kept separate for practical reasons). In other words, the model proposed here aims to extends Structural Syntax in order to comprehend morphologic phenomena. As it considers morphosyntax as a whole, and it has adpositions (interpreted as morphosyntactic connectors) as the central concept, it was called 'Constructive Adpositional Grammars' – the term 'constructive' will be explained in section 6 below, devoted to the formal model.

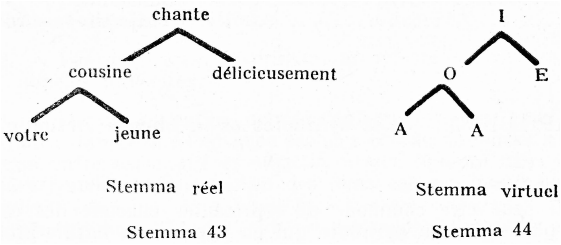

Figure 6: Tesnierian classes and functions at work

Figure 6 shows the instantiated example 43 and its abstract syntactic tree (example 44, i.e. *stemma virtuel*).

(43) votre jeune cousine chante délicieusement
'your young cousin sing lovely'

Tesnière individuates four 'word classes' (*classes de mots*) as generally and cross-linguistically valid. Their markers, which indicate their respective 'syntactic functions' (*fonctions syntactiques*), are the following:

I = verbants (presence of predication),
O = statives (expressions of reference),
E = circumstantials (modifiers of verbants),
A = adjunctives (modifiers of statives).

There is general agreement among linguists that the presence of expression of reference (i.e., "things") and the presence of predication (i.e., "events") are conceptual archetypes, i.e., always-valid universals of language (Langacker, 1987; Tomasello, 2003, for example).

Within the Standard Average European sprachbund, verbants (I) include verbs and interjections, while statives (O) include common and proper nouns, personal pronouns. Normally verbants and

[2] Grammaticalization is, roughly speaking, the 'paradox of change' (Coseriu) for which yesterday's morphology becomes today's syntax, and vice versa, paraphrasing Givón.

statives are the governors of their respective structural trees, while their modifiers play the role of dependents. Let adjuntives (A) be the modifiers of statives, including adjectives, determiners, possessive pronouns. Finally, let circumstantials (E) be the modifiers of verbants, e.g., in English, adverbs and adverbials. Figure 6 (right) shows that both modifiers (A and E) are dependents, respectively of the stative (O) *cousine* and the verbant (E) *chante*, and in fact they are put below in the Tesnièrian abstract syntactic tree.

Tesnière (1959) explains that the choice of the vowels is a borrowing from the planned language Esperanto, used as a "mnemonic tool" (1, A, ch. 33, par. 1). While the original Tesnièrian vowels are retained here for adherence with the original, in order to help the reader in the comparison of the two models, their original names, like "substantives" or "verbs", were not adopted in Constructive Adpositional Grammars, being too closely related to the grammar tradition belonging to the Standard Average European sprachbund (Haspelmath, 2001). However, it is worth noticing that Tesnière (1959) gives examples in many different languages through the book, e.g. French, German, Latin, Russian, Greek, but also Coptic, Chinese, Samoan, Turc, Tatar, Votiak, in order to show how Structural Syntax is valid across sprachbunds.

In a completely independent way, Whorf (1945) addressed the problem of grammar categorization out of Standard Average European, with results similar to Tesnière's. Since Whorf's names are valid across typologically distant sprachbunds, they were adopted here, with some adaptation.

The main difference between the two authors is the concept of selective and collocational lexemes introduced by Whorf (1945). He noticed that in every language some lexemes he calls selective have their proper grammar category carved inside, as in the English adjunctive (A) *honest*. No collocation can turn the grammar category of the selective adjunctive, but only morphology, e.g., *honest-y*, in order to obtain a stative (O), or *honest-ly*, in order to obtain a circumstantial (E).

By contrast, collocational lexemes are defined only if put into the syntagmatic axis: in isolation, we can have cues about their most probable function, but we cannot be certain. For instance, the English lexeme *walk* is probably a verbant (I), as in the phrase *I* ***walk*** *in the park*. Nonetheless, it can also be a stative (O), as in *Let's have a* ***walk*** or even an adjunctive (A), as in ***walk****ing distance*.

For this reasons, within Constructive Adpositional Grammars instead of 'word classes' it is preferred to say 'grammar characters', as the characters are can be applied or not to morphemes following the adtree where they are collocated, while selective lexemes are retained as a special case.

4 Adpositional trees and valency

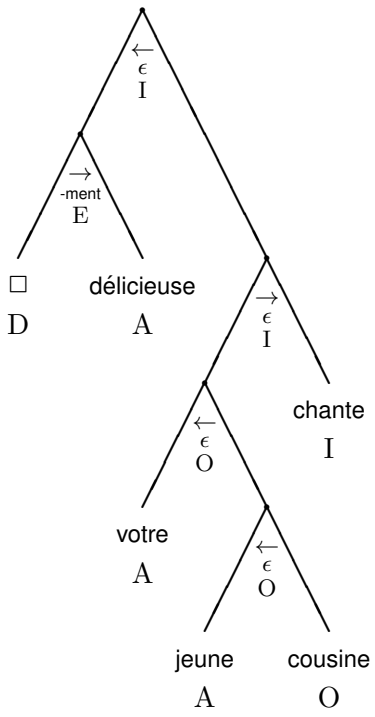

Figure 7: Reinterpretation of examples 43-44

Adtrees retain all the features of Tesnière's model in a single, unique representation, as shown in the adtree of examples 43-44 (figure 7).

First, both the concrete and abstract syntactic trees (i.e., *stemma réel* and *virtuel*) are represented together. Moreover, the verbant *chante* and the stative *cousine* are the governors of their respective adtrees, as expected from example 44. The reader is invited to note that the final grammar character of the stative group *votre jeune cousine* is indicated by the syntactic adpositions (ϵ); analogously, the stative-verbant connection is syntactic as well. On the contrary, the adverb *délicieusement* is obtained by the application of the suffix *-ment* which act as an adposition, imposing the circumstantial grammar character E to the adjective *délicieuse*. The dependent branch in this case is left underspecified (D), while structurally it is blocked (□), i.e., it cannot be furtherly expanded by the application of other morphemes. This adtree shows

how syntactic and morphological connections are treated in the same way. Finally, unlike structural syntactic trees, the trajectories of information prominence are rendered explicitly in the adtree of examples 43-44 (figure 7).

4.1 Trajectories of information prominence

The Tesnièrian dichotomy 'structural *vs.* semantic' influence was the source of one of the core features of Constructive Adpositional Grammars. Typological research on ergativeness has shown that a good grammar theory "would have to recognise that there are three basic syntactic-semantic primitives (A, S and O) rather than just two 'subject' and 'object'– however these are defined" (Dixon, 1994, 236). The arrows proposed by Tesnière (1959) are a cue for the solution of this problem. Within a stative-verbant connection, if the stative actively "does" the action, then the stative will be the most prominent element of the pair: in the terms proposed by Langacker (1987), the stative (O) will be the trajector (tm) while the verbant (I) will be the landmark (lm). Therefore, the trajectory of information prominence will be left-right ($\rightarrow$). In other words, the stative, being the dependent, is prominent (tr), and hence the connection will be a *dependency* ('semantic influence', according to Tesnière). Conversely, if the action marked by the verbant (I) "happens" to the stative (O), then the verbant will be trajector (tr) and the stative landmark (lm): the trajectory will be right-left ($\leftarrow$) accordingly. As the verbant is the governor, the connection will be a *government* ('structural influence', according to Tesnière). Therefore, the word 'dependency' assumes a very technical and precise meaning within the adpositional paradigm. It is important to note that what stated for the stative-verbant connection is valid for every grammar character connection, as exemplified in figure 7.

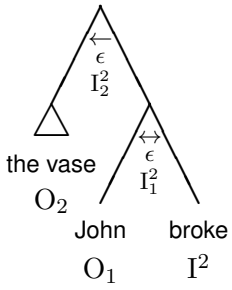

Figure 8: Adtree of *John broke the vase*

The adtree of *John broke the vase* is a good example of how trajectory of information prominence is treated within the adpositional paradigm. Let assume that our world model is standard, i.e., vases are inanimated objects, without will or beliefs, and John is a man. [3] While John can have broken the vase by accident (government, $\leftarrow$) or willingly (dependency, $\rightarrow$), the vase for sure happened to be broken, from its point of view, and hence its connection is a government ($\leftarrow$).

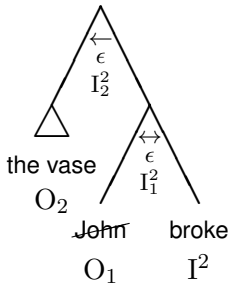

Figure 9: Adtree of *the vase broke*

Trajectory of information prominence explains why some "subjects" are statives in dependency – $\overrightarrow{O_1}$, 'A' for Dixon (1994) – while others are in government, i.e. $\overleftarrow{O_1}$, 'S' for Dixon (1994). In fact, the adtree of *the vase broke* (figure 8) can be considered a reduced or *transferred* adtree of *John broke the vase* (figure 9), where the subject (either in government or dependency, i.e., generically O_1) got lost. Before to deal with the concept of transference, which is derived from the Tesnièrian *translation* – explained in Part 3 of Tesnière (1959) – it is necessary to explain how valency is treated within the model proposed here.

4.2 The treatment of valency

The introduction of valency by Tesnière (1959) is one of the most successful part of his Structural Syntax, as it was adopted in most dependency-based frameworks in its fundamental traits:

> one could indeed compare the verb to a kind of **crossed atom**, which can attract a number more or less high of actants, in proportion to the number more or less high of hooks needed to maintain the respective dependencies (2, D, ch. 97, par. 3).

The concepts of valency and actants, i.e., how many statives are needed to saturate the valency

[3] Constructive Adpositional Grammars are agnostic in respect of world models.

value, are taken as such in Constructive Adpositional Grammars.

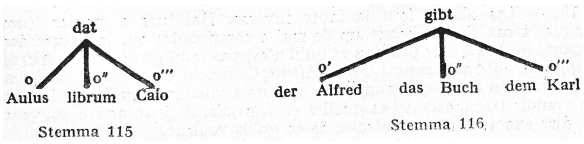

Figure 10: Examples of trivalent verbs

Figure 10 shows how Tesnière sometimes indicates the numbers of the actants saturating the valency value, in case of a trivalent verb. The examples are in Latin and in German, where an English equivalent can be *Alfred gave the book to Charles*.

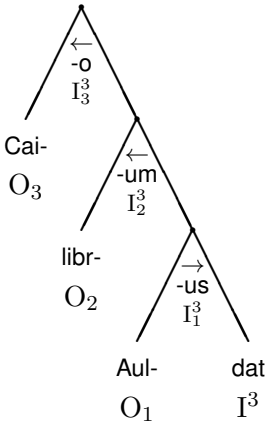

Figure 11: Adtree of example 115

Figure 11 shows the adtree of example 115 in Latin. In Constructive Adpositional Grammars, the verbant is the governor of the phrasal adtree (with 'phrase' meaning a tree governed by a uniquely identified verbant). If the verbant is a trivalent verb, as in example 115, three actants (i.e., O_1, O_2, O_3) are provided to saturate the valency value, along with their respective adtrees. The superscript number of the verbant indicates the absolute valency value – e.g., I^2 for a bivalent verb. The subscript number of the verbant indicates the degree of saturation in that point of the adtree, while the subscript of the actant indicates the actant number, following Tesnière's usage (figure 10). Example 115 shows that Latin substantive finals act as adpositions of the stative-verbant connection, with an indication of information prominence: *Aulus* ('Alfred') performes the giving (*dat*) and hence it is in dependency (→), while the giving happens both to *Caio* ('Carl'), being the beneficiary, and *librum* ('the book'), i.e., the actual object which was given, are both in government (←).

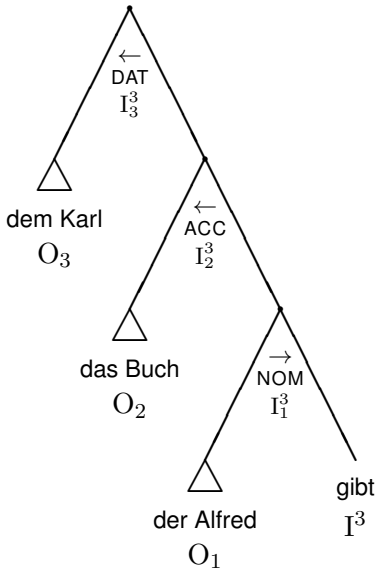

Figure 12: Adtree of example 116

Sometimes adpositions are marked through sememes, i.e., structural well-defined traits within a given language, even if the morph – the explicit morphemic signifier – is absent. For instance, example 116 shows that in German the case markers, like DATIVE, are not morphologically marked in the stative-verbant connection, but still well present in every German speaker's competence. In these cases, sememes can be written explicitly instead of epsilons, for clarity, if there is no possible ambiguity.

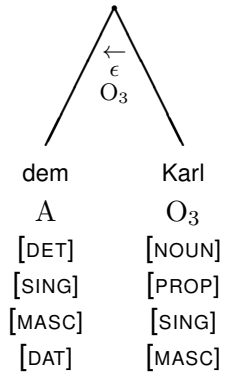

Figure 13: Detail of example 116

The detail of the adtree hidden under the third actant *dem Karl* (figure 13) shows that the sememe DATIVE is an additional trait under the adjunctive grammar character. Moreover, the adtree clarifies that there is a number agreement requirement, indicated by the sememe SINGULAR, between the stative *Karl* and the adjunctive *dem*, in order to

get everything work under a grammatical point of view. No such level of detail is present in Structural Syntax, most probably because Tesnière was not interested in such a direction. However, it is important that such level of detail is possible within the model here proposed if needed, e.g. for language learning purposes.

5 Transference

Every language – regardless of its sprachbund – has a class apart within its morphemes devoted to convey the most part of semantics, called lexemes. In fact, while the concept of 'word' is central only in the grammar traditions of the Standard Average European, the distinction of lexemes within a language's morpheme set is valid in general. For example, in Japanese non-lexemes are written in *kana* (syllabic script), while lexemes are written in *kanji* (Chinese logographic characters).

Lexemes are morphemes devoted to represent the relation between the language and the non-linguistic world, with a particular attention to reference. In structural syntactic trees, they are put above, being the governors, and similarly in the adpositional paradigm they are put in the rightmost leaves of their respective adtrees.

Tesnière (1959) noted that the most part of the non-lexical morphemes have the function of "turning" the grammar character of the lexemes they are applied to (3, A, ch. 161, par. 6).

(290) un exemple frapp-**ant** (I > A)
'a strik-**ing** example'
(292) liber Petr-**i** (O > A)
'Peter**'s** book'

The French suffix *-ant* (in 290) is applied to verbant lexemes in order to transfer their syntactic function to adjunctive, while the Latin suffix *-i* (in 292) is applied to stative lexemes in order to transfer their syntactic function to adjunctive as well. Of course, there are a lot of differences between the two adjunctives: in Constructive Adpositional Grammars, they would be expressed by different sememes and trajectories of information prominence. This kind of differences are not well formalized in Structural Syntax; however, the fundamental intuition that the morpheme set of a language can be divided in lexemes and non-lexemes on a functional syntactic basis is a remarkable part of Tesnière's heritage in the adpositional paradigm, since its definition in Gobbo (2009). This kind of morphemes (prepositions, pospositions, derivational suffixes and so on) were called by Tesnière (1959) *translatifs* (in the model proposed here, morphological, explicit adpositions) and the phenomenon as a whole was called *translation*, while in English "an equivalent may be *transference*, as the word *translation* has already the meaning of the French 'traduction"' (3, A, ch. 154, par. 2).

In the development of the formal model on which Constructive Adpositional Grammars are based, the role of transference is growing of importance. Tesnière (1959) devoted a lot of pages to transfer chains, from 'simple transference' (*translation simple*, e.g. I > O) to 'double transference' (*translation double*, e.g. O > A > O) until, at a limit, sevenfold transference (*translation septuple*). Complex transfer chains, i.e., double or more, can be explained in terms of recursive, nested adtrees, but this solution has two limits. First, from a linguistic point of view, there is no relation between an abstract adtree and the others belonging to the same language – 'abstract adtree' meaning what Tesnière called a *stemma virtuel*, i.e., an adtree without morphemic instantiation. Second, from a computational point of view, the constructive adpositional grammar of a given language, which should contain at least two sections – the first for morphemes, their grammar characters, basic transfers and the appropriate sememes, the second for the set of admittable abstract adtrees – will grow inconveniently. In fact, one of the goals of the adpositional paradigm is to give a convenient description of natural language grammars, both linguistically and computationally.

5.1 Abstract adtrees as constructions

The Tesnièrian concept of transference shows that most part of the dictionary is in reality the result of transfer chains: for this reason, a constructive dictionary can be built upon the lexemes and a set of transfer chain *patterns* in order to perform grammar character changes. In a cognitive linguistic perspective, these patterns of usage of form-meaning correspondences, that carry meaning beyond the meaning of the single componing morphemes, are called *constructions* (Croft, 2001; Goldberg, 2006). As a side note, the community of cognitive linguistics recognized Structural Syntax as a complementary, although dated, approach (Langacker, 1995).

After the study of Tesnièrian transference, it seemed more reasonable to see abstract adtrees as constructions instead of describing grammar only in terms of adtrees, so that the relations between constructions are formally represented in terms of adtree transformations, i.e., Tesnièrian transference rendered in formal terms. For example, the active-passive diathesis transference (2, D, ch. 101–102) can be expressed in terms of adtree transformations. Basically, the most primitive construction is the active diathesis, with all valency saturated by the possible actants, then a chain of adtree transformations permits to obtain the desired construction.

(A) $(\text{Carl})_{\text{O}}$ $(\text{slept in})_{\text{I}}$ $(\text{the beds})_{\text{O}}$.
(B) $(\text{the beds})_{\text{O}}$ $(\text{were slept in})_{\text{I}>\text{I}}$ $(\text{by Carl})_{\text{O}>\text{O}}$.
(C) $(\text{Carl's})_{\text{O}>\text{A}}$ $(\text{sleeping})_{\text{I}>\text{O}}$.

Examples A-B-C were annoted with the main grammar characters of the respective adtrees in order to help the reader in the knowledge of the use of transference within the model proposed here. In particular, example A shows an instantiation of the active diathesis construction of the English verbant *to sleep in*, while example B shows the correpondent passive construction. It is worth noticing that two transfers were performed in order to obtain the appropriate form of the verb ($\text{I} > \text{I}$) and of the SLEEPER actant ($\text{O} > \text{O}$). Moreover, example C is an example of nominalization: the SLEEPER actant was transferred into a saxon genitive construction ($\text{O} > \text{A}$) while the *ing*-construction transferred the verbant into a stative ($\text{I} > \text{O}$).

It is possible to write down classes of lexemes following the admittable patterns of adtree transformations. For example, it can be easily tested that the verbants *to sleep in* and *to melt in* belong to different classes of English verbants:

(D) the ice cube melted in the oven.
(E) *the oven was melted in by the ice cube.
(F) the melting of the ice cube.

Example D is structurally identical to example A; nevertheless, the passive construction obtained by the adtree transformation is ungrammatical (exampleE), while a different adjunctive construction, head by the adposition *of*, is to be preferred to saxon's genitive (example F). A full treatment of adtree transformation would need at least another paper devoted to it, so it is left as a further work.

English	**wh**-ere	**th**-ere	**wh**-en	**th**-en
French	où	là	**qu**-and	alors
Latin	u-bi	i-bi	**qu**-ando	**t**-um
German	**w**-er	**d**-a	**w**-ann	**d**-ann

Table 1: Tesnièrian analysis of correlatives

5.2 Second-order transference

Tesnière (1959) introduces the second-order transference (*translation du second degré*) in order to explain "what the traditional grammar had already implicitly regarded apartly with the name of 'subordination".' (3, D, ch. 239, par. 2). For example, the sentence *Alfred espère qu'il réussira* ('Alfred hopes that he will achieve') is a second-order transference from the verbant phrase *Alfred réussira* ('Alfred will achieve') to the stativized phrase *qu'il réussira* ('that he will achieve'; 3, D, ch. 241, par. 15). This kind of second-order transference is indicated with the symbol: $\gg$; e.g., a verbant-stative second-order transfer will be indicated as such: $\text{I} \gg \text{O}$.

Tesnière (1959) noticed that the *translatifs* – in the model proposed here, adpositions – devoted to second-order transference show a high degree of regularity in many different languages (3, D, ch. 240, par. 6, adapted in Table 1).

In Constructive Adpositional Grammars there is no need of a second-order level because of the expressive power of the mathematics underlying the formal model (see next section 6). What is retained from the Tesnièrian analysis is the observation that correlatives are *double* morphemes, made by a fixed part (e.g., *wh-* in English), that is appended to the governor phrase, and a flexible part (e.g., the English *-ere* for PLACE and *-en* for TIME) that is put in the adtree of the adtree of the subordinate phrase.

(H) I know where she goes.

Figure 14 shows the adtree of example H. The adtree of *where she goes* is intact in its inner construction: the relevant fact is that the correlative adposition *wh-ere* transfers the phrase from verbant to the second actant stative ($\text{I} > \text{O}_2$), from the point of view of the construction of *I know [where she goes]*. As the reader can see, adtrees can represent correlatives without any need of a second-order level of analysis.

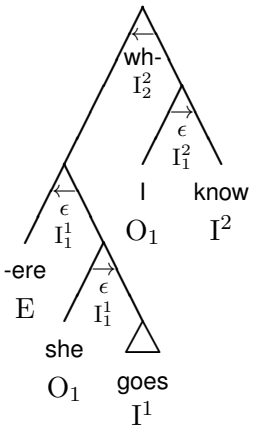

Figure 14: Adtree of example H

6 Sketches of the formal model

Tesnière (1959) asserts that "the use of symbols [grammar characters $\mathrm{O}, \mathrm{I}, \mathrm{E}, \mathrm{A}$, authors' note] in grammar is equivalent to the use of calculus in algebra" (1, A, ch. 33, par. 10). This statement implies that Structural Syntax can be formalised, at least theoretically.

In the fields of mathematical and computational linguistics there are many natural language grammar formalisms currently under investigation. In particular, the constructive adpositional formalism can be put into the realm of the so-called "categorial grammars"—i.e., representations of natural language grammars in terms of categories (Morril, 2007). At the present stage, the formal model is intended as a guiding reference for the development of linguistic concepts (GobboBenini, 2011). In fact, 'constructive' is intended linguistically as pertaining constructions (as already defined) and mathematically as pertaining constructive mathematics, i.e., any formal, mathematical construct used here have a constructive existence. In other words, it is possible to find an algorithm, non necessarily efficient, to construct any entity of the model.

In particular, adtrees and constructions together form a category, called **AdTree**, in the mathematical sense (MacLane, 1998; Borceux, 1994). A mathematical category is an algebraic structure composed by two classes, the *objects* and the *arrows*; arrows lie between two objects, the *source* or *domain*, and the *target* or *codomain*. Also, a category states that there are distinct arrows, the *identities*, one for every object A and such that the source and the target are A. Moreover, a category is equipped with a partial operation allowing to compose two arrows whenever one has the domain which is the target of the other one. Composition is required to be associative and identities act as one expects with respect to composition.

Intuitively, there is an arrow f from A to B whenever we can construct the B tree starting from the A tree applying the construction f. We do allow complex constructions obtained by sequentially composing simpler ones; if f and g are constructions such that $f(A) = B$ and $g(B) = C$, that is, if f maps A into B, and g constructs C from B, then $g \circ f$ is the construction which maps A into C by doing g after f.

It is possible to observe that, calling M the free monoid over the alphabet of morphemes of some natural language, i.e., the set of all possible (finite) sequences of morphemes obtained by juxtaposition, the functions mapping the trees in **AdTree** into the sequences of M comprehend the textual renderings of adpositional trees. If the attention is restricted to *contravariant functors*, i.e., the functions preserving the identical transformation and the reverse composition of adpositional trees, what is obtained is a class of functions which is called *presheaves over* M. Requiring that a presheaf maps the morphemes in the adtree into themselves in the monoid, what is obtained is exactly the lexicalizations of adtrees. In other words, there is a subclass of presheaves which directly corresponds to the texts the adtrees represent and which encodes the transformations that constitute the grammar. It is this space of presheaves which is generally understood as the subject of linguistics. Moreover, considering *endofunctors* on **AdTree**, i.e., functions mapping each adtree into another adtree, and each construction into another one such that they preserve identities and composition, it easily follows that each linguistic transformation, e.g., the mapping of active to passive diathesis, is an endofunctor. In turn, an endofunctor can be represented as an arrow between presheaves, thus showing that the mathematical model of the presheaves space is rich enough to represent and to reason about the foundational elements of Constructive Adpositional Grammars.

As a side effect of this intended model of interpretation, it follows that whatever construction over adtrees which is built by combinatorially composing the fundamental constructions, is an arrow. Lifting the structure of the **AdTree** category into the spaces of presheaves, which is a cat-

egory, it is possible to reason in a larger and richer environment, where the full power of mathematical methods can be applied: in fact, the presheaves space is a *Grothendieck topos* (MacLane, 1992; Johnstone, 2002), one of the richest mathematical structures available.

7 Conclusion

The impressive work by Tesnière (1959) is a constant source of inspiration for the definition of Constructive Adpositional Grammars. It is quite astonishing that nobody until now – as far as the authors know – has proposed a dependency-based model that makes use of the grammar characters proposed by the French linguist, i.e. $\mathrm{O}, \mathrm{I}, \mathrm{E}, \mathrm{A}$, which are the ground on which Structural Syntax is actually built. Such heritage could be the topic of another paper.

Directly formalising Structural Syntax, which was the first hypothesis considered, is simply not possible, essentially for two reasons. First, pragmatically Tesnière (1959) is a posthumous publication, and hence there are formal and linguistic incongruences which cannot be overcome; in particular, the unary tree (*répresentation stemmatique*) used by the author is ever-changing within the text, and not optimal to represent the triple 'governor, dependent, connection', for the reasons exposed in this paper. Second, Tesnière, working in the 1930-50 years, was a son of his time: he could take advantage of the fruits of the great tradition of linguistic structuralism that spread out in francophone Europe in the first half of the past century, but on the other hand he could not have the proper formal and mathematical instruments to be applied to his linguistic results – as category theory was introduced by Samuel Eilenberg and Saunders Mac Lane in the 1940s, and in those times it was not mature enough for linguistic applications.

Nonetheless, Constructive Adpositional Grammars, standing on the shoulders of Tesnière, can be considered a derivative work of Structural Syntax in many aspects, all of which were presented in this paper.

References

Francis Borceux. 1994. *Handbook of Categorical Algebra I*. Cambridge University Press: Cambridge.

William Croft. 2001. *Radical Constructions Grammar*. Oxford University Press: Oxford.

Ralph Debusmann. 2001. *A declarative grammar formalism for dependency grammar*, PhD thesis. Universität des Saarlandes.

Robert M. W. Dixon. 1994. *Ergativity*. Cambridge University Press, Cambridge.

Federico Gobbo and Marco Benini. 2011. *Constructive Adpositional Grammars: Foundations of Constructive Linguistics*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Federico Gobbo. 2009. *Adpositional Grammars: A Multilingual Grammar Formalism for NLP*, PhD thesis. University of Insubria.

Adele E. Goldberg. 2006. *Constructions at work*. Oxford University Press: Oxford.

Martin Haspelmath. 2001. The European linguistic area: Standard Average European, in: Martin Haspelmath, Wulf Oesterreicher and Wolfgang Raible. *Handbuch der Sprach- und Kommunikationswissenschaft*. Mouton de Gruyter, Berlin & New York. 1492–1510.

Peter Johnstone. 2002. *Sketches of an Elephant: A Topos Theory Compendium*, Volume 2. Cambridge University Press: Cambridge.

Ronald W. Langacker. 1995. Structural syntax: the view from cognitive grammar, in: Germain Dondelinger, Francoise Madray-Lesigne and Jeannine Richard-Zappella. *Lucien Tesnière aujourd'hui*. Peeters Publishers, Leuven. 13–37.

Ronald W. Langacker. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford University Press, Stanford.

Igor Mel'čuk. 1988. *Dependency Syntax: Theory and Practice*. The SUNY Press, Albany, N.Y.

Saunders Mac Lane. 1998. *Categories for the Working Mathematician*. Springer Verlag, Berlin.

Saunders Mac Lane and Ieke Moerdijk. 1992. *Sheaves in Geometry and Logic*. Springer Verlag, Berlin.

Glyn Morril. 2007. A Chronicle of Type Logical Grammar: 1935–1994. *Research on Language & Computation*, 5(3):359–386.

Joakim Nivre. 2005. *Dependency grammar and dependency parsing*, Technical report. Växjö University.

Lucien Tesnière. 1959. *Éléments de syntaxe structurale*. Klincksieck, Paris.

Michel Tomasello. 2003. *Constructing a language: A Usage-Based Theory of Language Acquisition*. Harvard University Press, Harvard.

Benjamin L. Whorf. 1945. Grammatical Categories. *Language*, 21(1):1–11.