

PRÉSENTATION

Sylvain KAHANE¹

La phrase est un **ensemble organisé** dont les éléments constituants sont les **mots**. Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des **connexions**, dont l'ensemble forme la charpente de la phrase. [...] Les connexions structurales établissent entre les mots des rapports de **dépendance**.

Tesnière 1959 (1967: 11, 13)

On appelle grammaire de dépendance toute grammaire formelle qui manipule comme représentations syntaxiques des structures de dépendance. Autrement dit, les grammaires de dépendance ne renvoient pas à un cadre formel particulier. Ainsi trouve-t-on parmi les grammaires de dépendance aussi bien des grammaires de réécriture que des grammaires non génératives. De plus, si nombre de modèles linguistiques s'affirment comme des grammaires de dépendances, il est aussi de nombreux modèles qui, sans le revendiquer clairement, s'apparentent fortement à des grammaires de dépendances, comme LFG, TAG, les grammaires catégorielles ou même HPSG. Il faut en effet rappeler que tout arbre syntagmatique pour lequel chaque constituant possède un sous-constituant tête induit naturellement un arbre de dépendance. Inversement, un arbre de dépendance induit naturellement un arbre syntagmatique en prenant comme constituant les projections des sous-arbres de l'arbre de dépendance. De ce fait, grammaires syntagmatiques et grammaires de dépendance ne s'opposent guère que par le fait de mettre en avant une structure plutôt que l'autre.

We call dependency grammar every grammar which handles dependency structures as syntactic representations. In other words, dependency grammars do not refer to a particular formal framework: dependency grammars include rewriting grammars as well as non-generative grammars. And while a number of linguistic models are presenting themselves as dependency grammars, there are also many models which, without claiming it, have clear similarities with dependency grammars, such as LFG, TAG, categorial grammars or even HPSG. Recall that every phrase structure tree for which each constituent has a head sub-constituent naturally induces a dependency tree. Conversely, a dependency tree naturally induces a phrase structure tree by taking projections of sub-trees of the dependency tree as constituents. As a result, constituency and dependency differ above all in which of the two kinds of structure they put first.

¹ LaTTiCe/TALaNa (Univ. Paris 7) & Univ. Paris 10 - Nanterre (sk@ccr.jussieu.fr)

S'il est difficile de dater l'apparition de la notion de dépendance syntaxique en linguistique, la première théorie linguistique basée sur la dépendance est incontestablement celle de Tesnière (1934, 1959), sans minimiser des travaux précurseurs, comme la syntaxe du russe de Peřkovskij (1934). Au même moment, Hays (1960) développait la première grammaire de dépendance, tandis que Gaifman (1965) établissait les liens entre les grammaires de dépendance de Hays, les grammaires catégorielles de Bar-Hillel et les grammaires de réécriture de Chomsky. À l'exception de la grammaire de Robinson (1970), les grammaires de dépendance se sont ensuite surtout développées en Europe, notamment autour de Sgall et Haji ová à Prague (Sgall *et al.* 1986) et de Mel' uk à Moscou (Mel' uk 1988), ainsi qu'en Allemagne (cf., par ex., la classique grammaire de l'allemand de Engel 1992) et au Royaume Uni autour de Anderson (1971) et Hudson, la France restant curieusement à l'écart.

Le retour des grammaires de dépendance au premier plan est dû à la concomitance de deux évolutions de la recherche en linguistique théorique et en linguistique computationnelle : la nécessaire prise en compte des informations lexicales et la réintégration de la composante sémantique dans les descriptions linguistiques et les systèmes de TAL. Pour le lexique, la chose est claire : en mettant la lexie au centre de la structure syntaxique, il est immédiat d'exprimer les relations lexicales comme la sous-catégorisation et le régime, et les grammaires de dépendances n'ont jamais cessé de prendre en considération de telles informations. À l'extrême, les grammaires complètement lexicalisées (comme TAG ou les grammaires catégorielles), qui construisent des phrases en combi-

It is difficult to date the appearance of the notion of syntactic dependency in linguistics, the first linguistic theory based on dependency is uncontestedly that of Tesnière (1934, 1959), without minimizing his precursors such as the Russian grammar of Peřkovskij (1934). At the same time, Hays (1960) was developing the first dependency grammar, while Gaifman (1965) was establishing the links between Hays' dependency grammar, Bar-Hillel's categorial grammar and Chomsky's rewriting grammars. Excepting the grammar of Robinson (1970), dependency grammars have been essentially developed in Europe, notably around Sgall and Haji ová at Prague (Sgall *et al.* 1986) or Mel' uk at Moscow (Mel' uk 1988), as well as in Germany (e.g. the classical German grammar of Engel 1992) and in United Kingdom around Anderson (1971) and Hudson, France staying oddly on the sidelines.

The return of dependency grammars to the forefront of debate is due to two parallel developments in the research in theoretical and computational linguistics: the necessary consideration of lexical information and the reinstatement of the semantic component in linguistic descriptions and in NLP systems. For the lexicon, things are clear: by putting the lexical unit at the center of the syntactic structure, the linguist takes upon himself the obligation to express lexical relations such as sub-categorization and government and dependency grammars have never stopped taking such information into account. In the extreme, fully lexicalized grammars (such as TAG or categorial grammars), which build sentences by combining elementary structures associated to words, natu-

nant des structures élémentaires associées aux mots, produisent naturellement des liens entre les mots et donc des structures de dépendance. Pour la sémantique, les structures de dépendance, en permettant de dissocier l'ordre des mots et la structure syntaxique proprement dite, se rapprochent davantage d'une représentation sémantique que ne le fait une structure syntagmatique. Mieux encore, les relations sémantiques prédicat-argument, parfois appelées dépendances sémantiques, bien que devant être distinguées des dépendances syntaxiques, coïncident en partie avec celles-ci (cf. Kahane & Mel'uk 1999 dans le précédent numéro de la revue TAL).

Ce numéro consacré aux grammaires de dépendance regroupent 9 articles qui couvrent les principaux courants dans le domaine.

L'article de Richard Hudson est une présentation de la composante syntaxique de son modèle linguistique, Word Grammar, l'un des modèles de référence en grammaire de dépendance. La présentation est orientée vers l'analyse des structures non projectives, c'est-à-dire, en termes de constituants, des discontinuités. Cet article, fort bien écrit, contient une présentation de la notion de dépendance, qui devrait satisfaire le lecteur non-initié.

L'article d'Eva Haji ová, à travers la question de la constitution d'un corpus syntaxiquement et sémantiquement annoté pour le tchèque, le Prague Dependency Treebank, donne un aperçu du cadre théorique qu'elle a développé avec Sgall.

L'article d'Alexander Dikovskiy et Larissa Modina présente les principaux travaux mathématiques sur les grammaires de dépendance dans les années 60 et 70 en URSS. L'article renvoie à de nombreuses publica-

tionnellement produisent des liens entre les mots et ainsi des structures de dépendance. Concernant la sémantique, les structures de dépendance, en permettant de dissocier l'ordre des mots et la structure syntaxique proprement dite, se rapprochent davantage d'une représentation sémantique que ne le fait une structure syntagmatique. Mieux encore, les relations sémantiques prédicat-argument, parfois appelées dépendances sémantiques, bien que devant être distinguées des dépendances syntaxiques, coïncident en partie avec celles-ci (cf. Kahane & Mel'uk 1999 dans le précédent numéro de la revue TAL).

This issue, dedicated to dependency grammars, contains 9 papers covering the main streams in the domain.

The paper of Richard Hudson is a presentation of the syntactic component of his linguistic model, Word Grammar, one of the reference models in dependency grammar. The presentation is geared to the analysis of non-projective structures, that is, in phrase-structure terms, discontinuities. This paper, very well written, contains a presentation of the notion of dependency, which should satisfy the non-initiated reader.

The paper of Eva Haji ová, which discusses the question of the constitution of a syntactically and semantically annotated Czech corpus, the Prague Dependency Treebank, gives an overview of the theoretical framework which she has developed with Sgall.

The paper of Alexander Dikovskiy and Larissa Modina presents the main mathematical works on dependency grammars in the 60's and 70's in USSR. The paper refers to a lot of publications which have re-

tions restées inconnus ici, qui touchent à des questions centrales et qu'il n'est pas trop tard pour découvrir.

L'article de Ingo Schröder, Wolfgang Menzel, Kilian Foth et Michael Schulz propose une grammaire de dépendance dont les règles se présentent comme des contraintes pondérées à satisfaire. Un langage de contraintes facilitant l'implémentation est décrit et illustré par des règles pour l'allemand.

L'article de Diego Mollá, Gerold Schneider, Rolf Schwitter et Michael Hess présente un système d'extraction de réponses, qui, par l'intermédiaire d'arbres de dépendance, construit des formes logiques minimales. Cet article, illustré par de nombreux exemples simples, montre bien comment une structure de dépendance peut être exploitée pour obtenir une représentation sémantique.

Les quatre articles qui suivent s'intéressent au traitement de constructions non projectives, notamment les extractions, et présentent des liens évidents entre eux, ainsi qu'avec l'article de Hudson.

L'article de Vincenzo Lombardo et Leonardo Lesmo présente une grammaire de dépendance lexicalisée à la Gaifman traitant les constructions non projectives par l'introduction de traits Slash et de nœuds vides. En plus des extractions, divers phénomènes de coordination de l'anglais sont décrits.

L'article de Sylvain Kahane présente une grammaire de dépendance lexicalisée traitant les constructions non projectives par un mécanisme particulier permettant de regrouper des chaînes de nœuds en des bulles. En conséquence, cette grammaire, à la différence des autres, ne manie pas des structures de dépendance pures.

mained unknown here, which deal with important questions and which it is not too late to discover.

The paper of Ingo Schröder, Wolfgang Menzel, Kilian Foth and Michael Schulz proposes a dependency grammar whose rules are weighted constraints to be satisfied. A constraint language making the implementation easier is described and illustrated by rules for German.

The paper of Diego Mollá, Gerold Schneider, Rolf Schwitter et Michael Hess presents an answer extraction system, which, through the intermediary of dependency trees, builds minimal logic forms. This paper, illustrated by numerous simple examples, shows well how a dependency structure can be exploited in order to obtain a semantic representation.

The four following papers look into the treatment of non-projective constructions, notably extractions, and present some evident links between them, as well as with Hudson's paper.

The paper of Vincenzo Lombardo and Leonardo Lesmo presents a lexicalized dependency grammar à la Gaifman dealing with non-projective constructions by using Slash features and empty nodes. Besides extractions, various English coordination phenomena are described.

The paper of Sylvain Kahane presents a lexicalized dependency grammar dealing with non-projective constructions by a particular mechanism allowing for grouping of some strings of nodes into bubbles. Consequently, this grammar, unlike the others, does not handle true dependency structure. A description of

Une description des relatives en français est proposée.

L'article de Norbert Bröker compare différents traitements des constructions non projectives, dont celui de Lombardo et Lesmo. L'accent est mis sur une grammaire de dépendance traitant l'introduction des constructions non projectives par des règles indépendantes permettant de remonter une dépendance dans l'arbre. Un algorithme d'analyse polynomial est décrit en détail.

L'article de Tomáš Holan, Vladislav Kuboš, Karel Oliva et Martin Plátek propose une mesure du degré de discontinuité d'un syntagme (lié à l'enchâssement des constructions non projectives) et montre qu'en tchèque, à la différence de l'anglais, ce degré est non borné. Des liens avec la complexité des algorithmes d'analyse, à comparé avec ceux de l'article de Bröker, sont également établis.

En conclusion, ce numéro spécial présente un panorama assez complet des développements actuels des modèles linguistiques basés sur les dépendances, même si on peut regretter que soient bien peu représentés les travaux portant sur des systèmes plus appliqués. Citons, puisque nous sommes dans une revue française, deux systèmes d'envergure développés en France : le générateur de texte développé à LexiQuest par José Coch (1996) et intégré au système MultiMétéo (Coch 1998) de génération de bulletins météo multilingues et l'analyseur en flux de Jacques Vergne (2000) qui a remporté l'action Grace, portant sur l'évaluation des étiqueteurs pour le français. Pour d'autres travaux, on pourra également se reporter aux actes du récent atelier sur le traitement automatique par des grammaires basées sur la dépendance (Kahane & Polguère 1998) et au portail

French relatives is proposed.

The paper of Norbert Bröker compares various treatments of non-projective constructions, including Lombardo and Lesmo's one. It concentrates on a dependency grammar dealing with non-projective constructions by independent rules enabling us to lift a dependency in the tree. A polynomial parsing algorithm is described in detail.

The paper of Tomáš Holan, Vladislav Kuboš, Karel Oliva and Martin Plátek proposes a measure of the degree of discontinuity of a phrase (linked to the embedding of non-projective constructions) and shows that, in Czech, contrary to English, this degree is non bounded. Some links with the complexity of parsing algorithms, to be compared with these of Bröker's paper, are also established.

To conclude, this special issue presents a rather complete panorama of recent developments of linguistic models based on dependency, even if we regret the lack of more applied works. As we are in a French journal, let us mention two ambitious systems developed in France : the text generator developed at Lexiquest by José Coch (1996) integrated into the MultiMeteo system (Coch 1998) of generation of multilingual weather forecasts and the flux parser of Jacques Vergne (2000), which was the winner of the *action Grace* evaluation of the French taggers. For other works, it is possible to consult the proceedings of the recent workshop on the processing of dependency-based grammars (Kahane & Polguère 1998) and the official web site of dependency grammar <http://ufal.mff.cuni.cz/dg.html>.

officiel des grammaires de dépendance <http://ufal.mff.cuni.cz/dg.html>

En plus des membres du comité de rédaction, nous remercions pour leurs relectures Norbert Bröker, Damien Genthial (TRILAN / Grenoble), Geert-Jan Kruijff (Univ. Charles, Prague), Igor Mel'uk (OLST, Univ. de Montréal), Alexis Nasr (Univ. d'Aix-Marseille), Alain Polguère (OLST, Univ. de Montréal) et Leo Wanner (Univ. de Stuttgart).

Besides the members of the editorial committee, we thank for their reviews Norbert Bröker, Damien Genthial (TRILAN / Grenoble), Geert-Jan Kruijff (Charles Univ., Prague), Igor Mel'uk (OLST, Univ. of Montreal), Alexis Nasr (Univ. of Aix-Marseille), Alain Polguère (OLST, Univ. of Montreal) and Leo Wanner (Univ. of Stuttgart).

REFERENCES

- ANDERSON John (1971) : *The Grammar of Case: Towards a Localist Theory*, Cambridge, Cambridge University Press.
- COCH José (1996) : "Overview of AlethGen", *Proc. 8th Int. Workshop on Natural Language Generation (INLG'96)*, Herstmonceux, Vol.2, pp. 25-28.
- COCH José (1998) : "Interactive generation and knowledge administration in MultiMeteo", *Proc. 9th Int. Workshop on Natural Language Generation (INLG'98)*, Niagara-on-the-Lake, pp. 300-303.
- ENGEL Ulrich (1992) : *Deutsche Grammatik*.
- GAIFMAN Haïm (1965) : "Dependency systems and phrase-structure systems", *Information and Control*, 18, pp. 304-337.
- KAHANE Sylvain & MEL'UK Igor (1999) : "La synthèse des phrases à extraction en français contemporain. Du réseau sémantique à l'arbre syntaxique", *T.A.L.*, 40, n° 2, pp. 25-85.
- KAHANE Sylvain & POLGUÈRE Alain (eds) (1998) : *Processing of Dependency-Based Grammars, COLING-ACL'98 Workshop*, Montréal.
- MEL'UK Igor (1988) : *Dependency Syntax : Theory and Practice*, New-York, State University of New-York Press.
- PEÆKOVSKIJ Aleksandr (1934) : *Russkij sintaksis v nau nom osveänii* [Russian syntax: a scientific approach], Moscow, U pedgiz.
- ROBINSON Jane, "Dependency structures and transformational rules", *Language*, 46, pp. 259-285.
- SGALL Petr, HAJICOVÁ Eva & PANENOVÁ Jarmila (1986) : *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Dordrecht, Reidel.
- TESNIÈRE Lucien (1934) : "Comment construire une syntaxe", *Bulletin de la Faculté des Lettres de Strasbourg*, 7, 12^e année, pp. 219-229.
- TESNIÈRE Lucien (1959) : *Éléments de syntaxe structurale*, Kincksieck, Paris.
- VERGNE Jacques (2000) : *Étude et modélisation de la syntaxe des langues à l'aide de l'ordinateur - Analyse syntaxique automatique non combinatoire*, Thèse d'HDR, Université de Caen.