



**HAL**  
open science

# Les trois dimensions d'une modélisation formelle de la langue : syntagmatique, paradigmatic et sémiotique

Sylvain Kahane

► **To cite this version:**

Sylvain Kahane. Les trois dimensions d'une modélisation formelle de la langue : syntagmatique, paradigmatic et sémiotique. *Revue TAL : traitement automatique des langues*, 2015, 56 (1), pp.39-63. halshs-01740584

**HAL Id: halshs-01740584**

**<https://shs.hal.science/halshs-01740584v1>**

Submitted on 3 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Les trois dimensions d'une modélisation formelle de la langue : syntagmatique, paradigmatic et sémiotique

Sylvain Kahane

Modyco, Université Paris Ouest & CNRS

92001 Nanterre cedex

sylvain@kahane.fr

---

*RÉSUMÉ.* Nous montrons que tout ensemble de productions d'une langue peut être appréhendé comme un espace à trois dimensions : l'axe syntagmatique est celui des combinaisons de signes au sein du texte, l'axe paradigmatic est celui des commutations possibles en chaque point du texte et l'axe sémiotique est celui interne aux signes, qui lie le signifié au signifiant, le sens au texte. Toute modélisation formelle d'une langue se doit d'effectuer un recouvrement de cet espace tridimensionnel. De plus, pour capter les régularités de la langue, un modèle cherchera à couvrir les portions les plus importantes possibles de cet espace avec chaque règle, ce qui est complexe en raison de l'interdépendance des trois dimensions. Pour mieux comprendre les enjeux, nous comparerons deux stratégies illustrées par un modèle bien connu, LTAG, et un autre similaire, GUST, qui s'en inspire tout en privilégiant un découpage selon les axes syntagmatique et sémiotique.

*ABSTRACT.* We show that every set of productions of every natural language can be described as a three-dimensional space: the syntagmatic axis is the axis of combinations of signs within the text, the paradigmatic axis is the axis of commutations in each position in the text, and the semiotic axis is the axis internal to the signs, that links the signified to the signifier, the meaning to the text. Every formal modeling of a natural language has to cover this 3-dimensional space. Moreover, in order to capture the regularities, a model seeks after covering the largest possible parts of this space with each rule, which is difficult due to the mutual dependency of the three dimensions. In order to better understand the issues, we will compare two strategies illustrated by a well-known model, LTAG, and another similar one, MTUG, which is inspired by the former but favors a cutting along the syntagmatic and semiotic axes.

*MOTS-CLÉS :* signe linguistique, grammaire formelle, LTAG, grammaire d'unification polarisée, grammaire d'unification sens-texte, interface syntaxe-sémantique, formalisme basé sur des règles, grammaire d'arbre, structure de dérivation.

*KEYWORDS:* linguistic sign, formal grammar, LTAG, Polarized Unification Grammar, Meaning-Text Unification Grammar, syntax-semantics interface, rule-based formalism, tree grammar, derivation structure.

---

## 1. Introduction

Dans cet article, nous souhaitons montrer que construire une grammaire formelle revient à recouvrir un espace formel à trois dimensions, dimensions que nous appellerons *syntagmatique*, *paradigmatique* et *sémiotique*. Si les deux premières dimensions sont bien connues (Saussure, 1916 : 170-75 ; Hjelmslev, 1943 [1971] : 55), la troisième est rarement nommée à notre connaissance bien que présente dans tout modèle linguistique. L'axe *sémiotique* est la dimension qui lie le signifié au signifiant (Saussure, 1916 : 97-99), l'expression au contenu (Hjelmslev, 1943 [1971] : 65-79)<sup>1</sup>, le sens au texte (Mel'čuk, 1988 : 43-50).

Toute modélisation linguistique se doit d'effectuer un recouvrement de cet espace à trois dimensions, mais il y a plusieurs stratégies pour cela. Les modèles recherchent en général une certaine économie, ils cherchent à saisir les régularités des langues et à généraliser ce qui peut l'être. Du point de vue du recouvrement, cela signifie que l'on essaiera de grouper les structures similaires et donc d'éviter de les traiter par des règles séparées. Nous comparerons deux stratégies de recouvrement que nous illustrerons par un modèle linguistique<sup>2</sup> bien connu, LTAG (Joshi, 1987 ; Abeillé, 1991 ; Candito, 1996), et par un autre moins connu, GUST (Kahane, 2001 ; Kahane et Lareau, 2005 ; Lareau, 2008 ; Kahane, 2013). Pour ce dernier, nous montrerons comment le fait de séparer davantage selon les dimensions sémiotique et syntagmatique permet de regrouper selon la dimension paradigmatique, à l'inverse des choix de LTAG.

Les deux modèles que nous considérons sont des modèles symboliques qui construisent les différentes structures associées à un énoncé en combinant des structures élémentaires. Ces structures élémentaires sont des portions de l'espace à trois dimensions que nous allons présenter. Ce type de modèles se prête bien à la démonstration que nous souhaitons faire. Nous pensons cependant que les questions abordées ici concernent directement toute modélisation linguistique, qu'il s'agisse de modèles déclaratifs ou procéduraux, de modèles à base de règles ou de contraintes, de modèles stochastiques acquis sur corpus ou de grammaires écrites à la main : dans tous les cas, il faudra décider quelles structures sont associées aux énoncés et comment l'ensemble des énoncés et de leurs structures est couvert par le modèle.

La section 2 présentera les trois dimensions de la langue et montrera leur interdépendance. La section 3 contrastera deux stratégies de recouvrement de l'espace linguistique, celle de LTAG et celle de GUST.

---

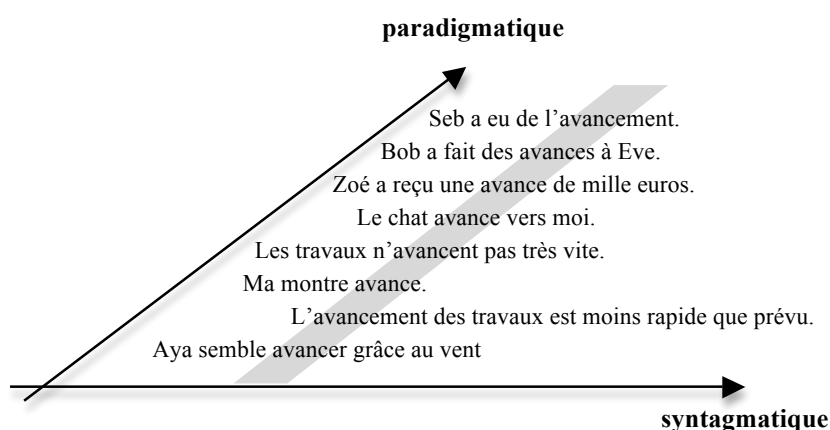
1. « Nous cesserons pour le moment de parler des signes car, ne sachant pas ce qu'ils sont, nous cherchons à les définir, pour parler de ce dont nous avons constaté l'existence, c'est-à-dire de la fonction sémiotique posée entre deux grandeurs : *expression* et *contenu*. » (Hjelmslev, 1943 [1971] : 66).

2. Nous distinguons le modèle LTAG du formalisme TAG. Un modèle repose sur des choix théoriques, qui sont en partie indépendants du formalisme utilisé pour les implanter.

## 2. Les trois dimensions

### 2.1. Syntagmatique, paradigmatic, sémiotique

La première dimension, sur l'axe *syntagmatique*, est celle des signes en présence (Saussure, 1916 : 171), celle de la combinaison des signes. Si on se place au niveau des signifiants des signes, la dimension syntagmatique se déroule le long de l'axe temporel, où les unités se succèdent (bien qu'au découpage « segmental » en phonèmes et morphèmes se rajoutent des informations dites « suprasegmentales », comme la prosodie ou les gestes et mimiques liés à la communication). Au niveau des signifiés, les combinaisons sont bien différentes et des éléments non contigus peuvent être en relation. Construire un modèle linguistique suppose de décider comment découper les énoncés selon la dimension syntagmatique. Comme nous allons le voir, on ne peut répondre à cette question sans considérer les autres dimensions.



**Figure 1.** Coupe de la « langue » selon les axes syntagmatique et paradigmatic (paradigme de *avanc-* /avãs/)

La deuxième dimension, sur l'axe *paradigmatique*, est celle des signes en absence, celle du paradigme des *environnements*<sup>3</sup> où une unité de première dimension peut se trouver. Partant d'un corpus représentatif (d'un fragment) de la langue, nous pourrions répertorier toutes les occurrences d'une forme donnée, par exemple le signifiant *avanc-* /avãs/ (figure 1). Construire un modèle suppose alors

---

3. Lorsqu'on évoque la dimension paradigmatic, on pense généralement au paradigme des éléments qui commutent dans un environnement donné et qui déterminent une position syntaxique. Nous adoptons ici la position *duale* qui consiste à fixer un élément et à regarder le paradigme de ses environnements.

de décider quelles occurrences sont des occurrences de la même chose et lesquelles sont des occurrences d'objets différents. L'ensemble des occurrences de *avanc-* (matérialisé en gris dans la figure 1) devra être recouvert par les règles du modèle. Par exemple, si on se place au plan syntaxique, on notera que toutes les occurrences considérées appartiennent à (et sont donc recouvertes par) trois unités différentes : les noms AVANCE et AVANCEMENT et le verbe AVANCER. Si on se place au plan sémantique, on pourra distinguer l'acception d'AVANCER signifiant 'se déplacer vers l'avant' (*Le chat avance vers moi*) d'autres acceptions où il n'y a pas de déplacement physique (*Ma montre avance ; Les travaux n'avancent pas*) et on aura donc un recouvrement plus granulaire.

La troisième dimension, sur l'axe *sémiotique*, est celle interne aux signes, celle qui lie signifié et signifiant, celle qui va du sens au texte (synthèse) et du texte au sens (analyse). Les signes linguistiques ne s'organisent pas de la même façon selon qu'on les considère du point de vue de leur signifiant ou du point de vue de leur signifié et nous en déduisons qu'il existe plusieurs niveaux d'organisation des unités (Sgall, 1967 ; Jackendoff, 2002 : 6). Nous considérons ainsi un niveau de représentation sémantique, celui des sens ou signifiés, et un niveau de représentation morphologique, celui des formes ou signifiants. Nous ne confondons pas le niveau morphologique, où les objets sont les signifiants eux-mêmes, avec les niveaux phonologique et graphique, dont les objets sont les phonèmes et les graphèmes.

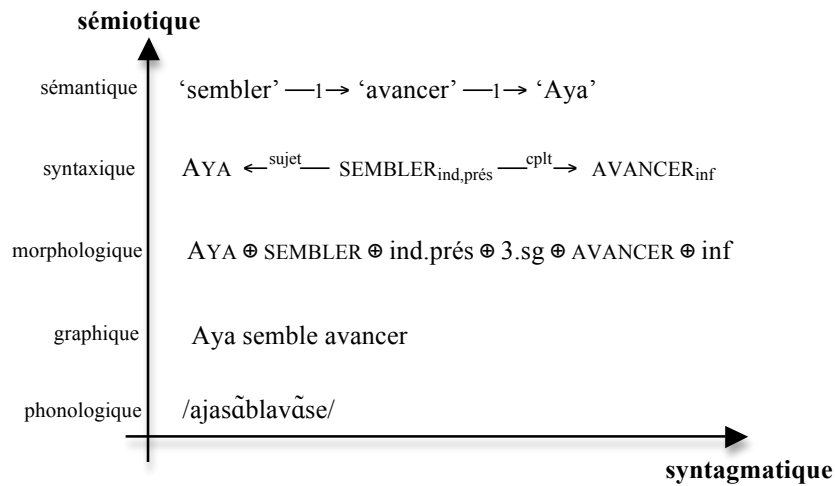
La plupart des modèles considèrent, en plus des plans morphologique et sémantique, un niveau d'organisation hiérarchique, la structure syntaxique, indépendant et différent de l'organisation des signifiants comme de celle des signifiés. D'autres niveaux d'organisation peuvent être encore considérés, comme la structure communicative (*information packaging*), la structure topologique (Gerdes et Kahane, 2001) ou encore la structuration opérée par la prosodie. Certains modèles dits stratifiés, comme la théorie sens-texte (Mel'čuk, 1988 : chap. 2), présentent ces différents niveaux d'organisation ordonnés<sup>4</sup> du plus profond (la représentation sémantique) au plus proche de la surface (la représentation phonologique). Qu'ils soient ou non ordonnés, leur diversité constitue une dimension orthogonale aux deux précédentes. Un modèle linguistique devra être capable d'appréhender ces différents niveaux d'organisation et beaucoup de modèles ne lieront pas directement la représentation sémantique à la représentation phonologique, mais utiliseront des niveaux intermédiaires en reliant par exemple la représentation sémantique à la représentation syntaxique (interface syntaxe-sémantique)<sup>5</sup>. Nous donnons figure 2

---

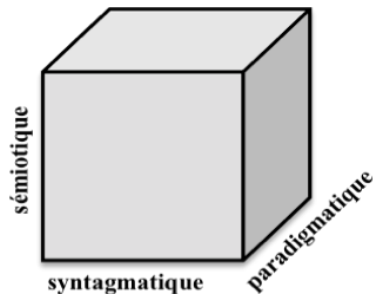
4. En fait, même un modèle très stratifié comme la TST avec ses sept niveaux de représentation ne peut pas ordonner tous les « niveaux » d'organisation. Ainsi en TST, la structure communicative fait partie de la représentation sémantique en parallèle avec la structure prédicative (le graphe de relations prédicat-argument).

5. Le fait que la plupart des éléments sémantiques soient interfacés avec des éléments syntaxiques n'exclut pas que certains puissent être interfacés directement avec le niveau phonologique. Par exemple, une interrogation pourra être réalisée par un élément lexical (*Est-ce que Marie vient ?*) ou un enclitique (*Marie vient-elle ?*) et donc déclencher une règle

différentes représentations de la phrase *Aya semble avancer*. Pour une vue des trois dimensions de l'espace, on combinera les figures 1 et 2, comme le montre la figure 3.



**Figure 2.** Coupe de la « langue » selon les axes syntagmatique et sémiotique (pour *Aya semble avancer*)



**Figure 3.** Les trois dimensions du modèle

Si l'on revient à l'axe syntagmatique, on voit que les unités se combinent différemment selon les niveaux. Du point de vue de leur signifiant, les signes se succèdent dans un ordre linéaire (à l'exception de quelques cas comme les apophonies et les amalgames). Du point de vue de leur signifié, les signes

---

d'interface sémantique-syntaxe, mais dans tous les cas, elle sera réalisée par un contour prosodique particulier et déclenchera une règle d'interface sémantique-phonologie. Un modèle linguistique doit donc être capable de gérer des règles de différents emplans selon l'axe sémiotique.

fonctionnent comme des prédicats qui se lient aux signes qui remplissent les positions argumentales que ces prédicats ouvrent (Frege, 1879 [1967] : 23 ; Jespersen, 1937 [1971] : 15 ; Žolkovskij et Mel'čuk, 1967 ; Kamp, 1981 ; Mel'čuk 1988 : 55 ; 2012). Les relations syntagmatiques sont donc des relations de succession au niveau des signifiants (ou plan morphologique), alors qu'elles forment un graphe de relations prédicat-argument au niveau des signifiés (ou plan sémantique). On aura noté que nous considérons comme relation syntagmatique toute relation de combinaison de deux signes, qu'elle soit de niveau morphologique, sémantique ou autre. Le fait que les combinaisons diffèrent selon les niveaux (par exemple, pour reprendre l'exemple de la figure 2, AYA se combine avec AVANCER au niveau sémantique, mais pas aux autres niveaux) justifie de faire de la dimension sémiotique une dimension orthogonale à la dimension syntagmatique. L'orthogonalité de la dimension sémiotique avec la dimension paradigmatique, plus évidente encore, sera développée à la section suivante (voir figure 5 notamment).

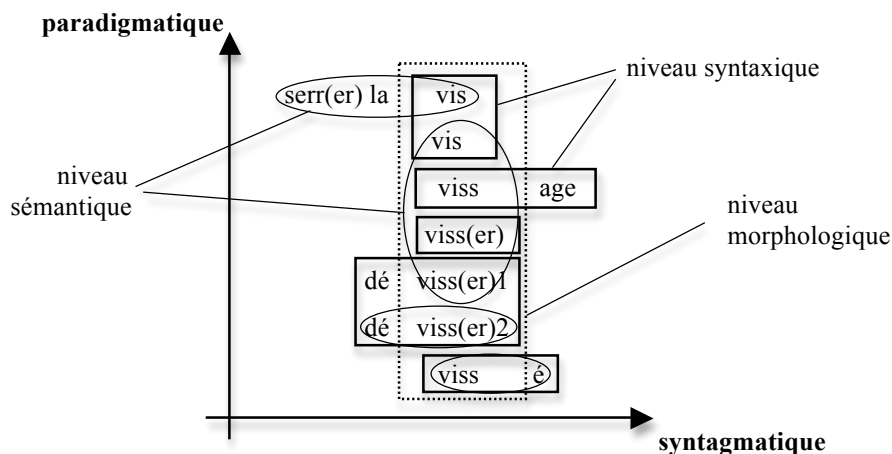
Tout acte linguistique suppose de parcourir l'espace linguistique dans ses trois dimensions. Considérons un locuteur ou un système de génération de textes qui, partant d'un sens, souhaite produire un texte. Il devra parcourir l'axe *sémiotique* en allant du sens au texte. Pour chaque élément de sens, il devra sélectionner un signe linguistique et donc parcourir l'axe *paradigmatique* des choix qui s'offrent à lui. Enfin, il devra consommer la totalité du sens qu'il veut exprimer et donc parcourir l'axe *syntagmatique* en combinant les signes linguistiques choisis, produisant ainsi des mots les uns à la suite des autres, c'est-à-dire un texte.

## 2.2. L'interdépendance entre les dimensions

Nous allons maintenant étudier la façon dont les signes linguistiques recouvrent l'espace linguistique. La façon dont ils « découpent » l'espace linguistique selon une dimension dépend fortement de la position du « plan de coupe » selon les autres dimensions. En particulier, le découpage selon les dimensions syntagmatique et paradigmatique dépend du niveau sémiotique considéré : il est bien connu en effet que les signes ne se découpent pas de la même façon selon leur signifiant ou leur signifié. Ceci est vrai pour la dimension syntagmatique, comme pour la dimension paradigmatique.

Commençons par la dimension syntagmatique. De nombreuses unités indécomposables dans leur signifié sont formées de plusieurs unités syntaxiques : ce sont les locutions, comme <sup>1</sup>SE RETOURNER DANS SA TOMB ou <sup>1</sup>TIRÉ PAR LES CHEVEUX<sup>1</sup>. À l'inverse, beaucoup de lexèmes, qui forment des unités indécomposables au niveau syntaxique, comme AVANCEMENT, BLESSURE ou HABILEMENT, peuvent être décomposés au niveau morphologique, voire sémantique (en raison de la régularité de la composition des signifiés des morphèmes lexicaux et dérivationnels en jeu). De même, certaines unités indécomposables au niveau morphologique composent plusieurs unités syntaxiques ou sémantiques. On parle alors d'amalgame

(au /o/ = À ⊕ LE), de fusion (le *-ons* /ɔ̃/ de *chantons* = 1<sup>re</sup> personne ⊕ pluriel) ou encore d'apophonie (*chevaux* = CHEVAL ⊕ pluriel ; *feet* = FOOT ⊕ pluriel).

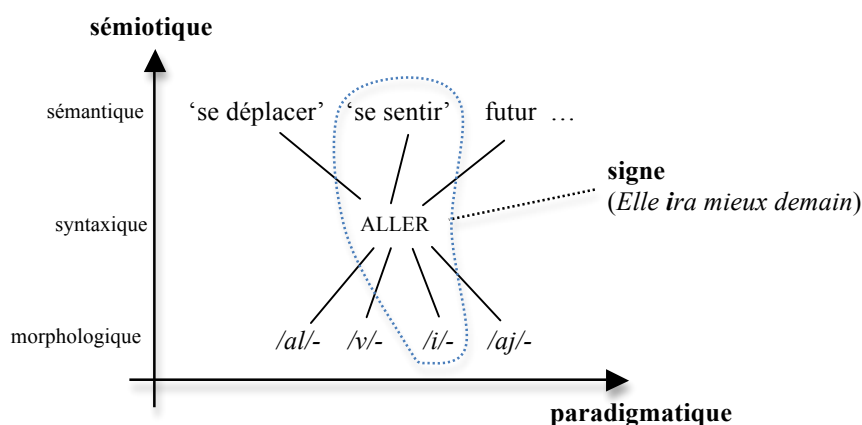


**Figure 4.** Coupe de la « langue » selon les axes syntagmatique et paradigmatique et projection du regroupement des signes à différents niveaux sémiotiques

Considérons maintenant la dimension paradigmatique en plus et prenons un morphème comme *vis* ‘tige filetée en hélice’ (figure 4) : il s’utilise tel quel comme nom (VIS), mais aussi comme radical du verbe VISSER ‘mettre une vis’, lequel possède une acception métaphorique VISSÉ (*Il est **vissé** sur sa chaise toute la journée*), qui doit être distinguée au niveau sémantique. Le verbe VISSER se combine compositionnellement avec le suffixe *-age* pour donner le nom VISSAGE ‘fait de visser’ et avec le préfixe *dé-* pour donner DÉVISSER1 ‘enlever une vis’, lequel possède également une acception métaphorique DÉVISSER2 (*L’alpiniste a **dévissé***). Nous prenons un acception très large du terme *morphème* : il désigne tout ensemble de signes ou quasi-signes qui possèdent le même signifiant et des signifiés sémantiquement liés. Par quasi-signes, nous entendons des éléments comme le *vis* de DÉVISSER2 ou de ‘SERRER LA VIS’ ‘réprimer certaines tendances au relâchement’, qui n’ont pas de signifié propre, mais dont la contribution au signifié de la combinaison dont ils font partie peut être rapprochée des signifiés des autres occurrences de *vis*. Nous rendons ainsi compte du fait que DÉVISSER est un dérivé de VIS en déclarant que l’un et l’autre contiennent le même morphème *vis* et du fait que la dérivation de DÉVISSER1 est régulière en déclarant qu’il s’agit de la même acception du morphème *vis* dans VIS et DÉVISSER1 (mais pas dans DÉVISSER2). Alors que DÉVISSER1 et VISSAGE, sont décomposables au niveau sémantique (*déviss*er1 est à *vis*ser ce que *dém*onter est à *mon*ter) mais pas syntaxique, une locution comme ‘SERRER LA VIS’ est décomposable au niveau syntaxique, mais pas sémantique. La figure 4 montre la projection sur le plan formé des axes syntagmatique et paradigmatique des regroupements de signes aux différents niveaux sémiotiques.



Voyons maintenant une propriété remarquable des signes qui est de s'organiser en faisceaux. Une entité syntaxique comme le verbe ALLER possède à la fois plusieurs acceptions (signifiés) et plusieurs allomorphes (signifiants) (figure 5). Chaque association entre une acception et un allomorphe de ALLER forme un signe distinct. Nous dirons que ALLER est à la fois polysémique (et se découpe en plusieurs entités sémantiques) et polymorphique (et se découpe donc en plusieurs entités morphologiques). Il s'agit dans les deux cas, pour les signifiés comme pour les signifiants, d'un découpage des occurrences de ALLER selon l'axe paradigmatique, mais à des niveaux différents (sémantique vs morphologique). Les deux « faces » du faisceau, les signifiés et les signifiants, sont indépendantes : chaque signifié peut correspondre à n'importe quel signifiant (à l'exception de l'acception futur, comme dans *Elle allait partir*, qui possède un paradigme de conjugaison plus restreint) et chaque signifiant peut correspondre à n'importe quel signifié. Les signes de ALLER forment ainsi un *faisceau*. L'indépendance entre le choix d'un signifié et le choix d'un signifiant permet de séparer complètement l'interface syntaxe-sémantique de l'interface syntaxe-morphologie<sup>6</sup>.



**Figure 5.** Dimensions paradigmatique et sémiotique (pour le faisceau de signes ALLER)

Le découpage selon l'axe paradigmatique fait au niveau syntaxique est un troisième découpage en partie indépendant des découpages aux niveaux sémantique et morphologique. Deux unités syntaxiques différentes peuvent avoir le même signifiant, comme VIS et VISSER (dont les signifiants sont /vis/). Certaines unités grammaticales partagent également une partie de leurs formes : ainsi l'impératif en français emprunte-t-il à la fois des formes du présent de l'indicatif (*prends !, finis !*) et du subjonctif (*aie !, sois !, sache !*). De la même façon, le marqueur du réfléchi,

6. Cette organisation est reflétée dans les approches traditionnelles qui séparent nettement les lexiques décrivant les sens (les dictionnaires) des lexiques décrivant les formes (les tables de conjugaison).

SE, possède à la 1<sup>re</sup> et à la 2<sup>e</sup> personne les mêmes formes que les pronoms personnels (*je me rase* vs *il me rase*).

De même que deux unités syntaxiques peuvent avoir le même signifiant, deux unités syntaxiques peuvent avoir le même signifié. Si l'on considère que des phrases comme *Bob a fait la sieste pendant une heure* et *La sieste de Bob a duré une heure* sont des paraphrases, on doit admettre que DURER et PENDANT sont deux lexicalisations d'un même sens, l'une verbale et l'autre adverbiale, et que donc les deux unités lexicales ont le même signifié.

Nous avons jusque-là présenté les niveaux sémantique et morphologique comme les niveaux des signifiés et des signifiants, mais la situation est plus complexe : certains signes ne couvrent pas tout l'axe sémiotique. Certains signes n'ont pas de contribution sémantique : ainsi le lexème ALLER de la locution <sup>1</sup>ALLER SE FAIRE CUIRE UN ŒUF<sup>1</sup> est la même entité syntaxique ALLER avec la même allomorphie que celle de la figure 5, mais il n'a pas ici de contribution sémantique autonome. C'est en quelque sorte un « demi-signé ». De nombreux marqueurs syntaxiques, comme la conjonction de subordination QUE dans *Marie annonce qu'elle part* ou la préposition DE dans *l'annonce de son départ*, n'ont pas non plus de contribution sémantique (on parle communément de signes « vides ») et sont donc des signes dont le signifié est de niveau syntaxique. De même, l'accord en genre des adjectifs sert uniquement à marquer le lien entre l'adjectif et le nom avec lequel il se combine : son signifié est une signification grammaticale (de niveau syntaxique) et non un sens. À l'inverse, certains signes, comme le présent en français, ont une contribution sémantique mais pas de réalisation phonologique. On parle alors de signifiant « zéro », mais on pourrait tout aussi bien considérer que leur signifiant n'est pas au niveau morphologique. Une fois cette porte ouverte, on peut tout à fait accepter qu'une construction, comme celle qui lie le verbe à son sujet (cf. la construction *actor-action* de Bloomfield, 1933 : ch. 12.1), est un signe qui ne contribue pas au niveau sémantique et qui au niveau morphologique donne une spécification d'ordre des mots (voir l'usage que nous en faisons à la section 3.1). C'est la position des grammaires de construction (CxG), qui utilisent le terme *construction* là où nous préférons conserver le terme *signe*.

Les conséquences de l'interdépendance des trois dimensions sont importantes pour qui souhaite développer un modèle linguistique. Le fait que les regroupements syntagmatiques et paradigmatiques ne soient pas les mêmes aux différents niveaux sémiotiques suppose qu'on ne peut espérer avoir un recouvrement simple de l'espace tridimensionnel formé par les énoncés et leurs différentes représentations aux différents niveaux sémiotiques. Nous allons explorer cette question maintenant.

### 3. Recouvrement de l'espace linguistique

Un modèle linguistique se doit de recouvrir la totalité de l'espace linguistique que nous venons de décrire, puisqu'il doit pouvoir associer à toute forme tous les sens

qui lui correspondent (analyse) et à tout sens toutes les formes qui lui correspondent (synthèse). Nous allons présenter deux stratégies, celle de LTAG, qui a l'avantage de la simplicité, et celle de GUST, plus granulaire, mais au prix d'une complexité plus grande du formalisme.

### 3.1. Le recouvrement de LTAG

Le modèle LTAG (*Lexicalized Tree-Adjoining Grammar*) s'est imposé dans les années 1990 (Abeillé, 1991 ; XTAG, 1995) comme un modèle à la fois simple et puissant permettant de développer assez facilement des grammaires à large couverture exploitables en TAL (de La Clergerie, 2005). Nous ne ferons pas une présentation exhaustive de ce modèle bien connu et bien documenté ; nous nous contenterons d'en souligner les caractéristiques pertinentes pour notre discussion.

Le principe de LTAG est de construire la structure syntaxique d'un énoncé en combinant des morceaux de cette structure, appelés *arbres élémentaires*. Il s'agit d'un découpage de la structure syntaxique selon la dimension syntagmatique. L'originalité (et la force) de LTAG est que ce découpage de la structure syntaxique est en fait effectué selon des critères sémantiques. Autrement dit, les arbres élémentaires TAG correspondent à des unités sémantiques. Il résulte de ce choix que la *structure de dérivation*, c'est-à-dire la structure qui décrit comment les arbres élémentaires ont été combinés pour dériver la structure de l'énoncé (Vijay-Shanker, 1987), peut être interprétée comme une représentation sémantique de l'énoncé (Candito et Kahane, 1998 ; Joshi et Vijay-Shanker, 2001).

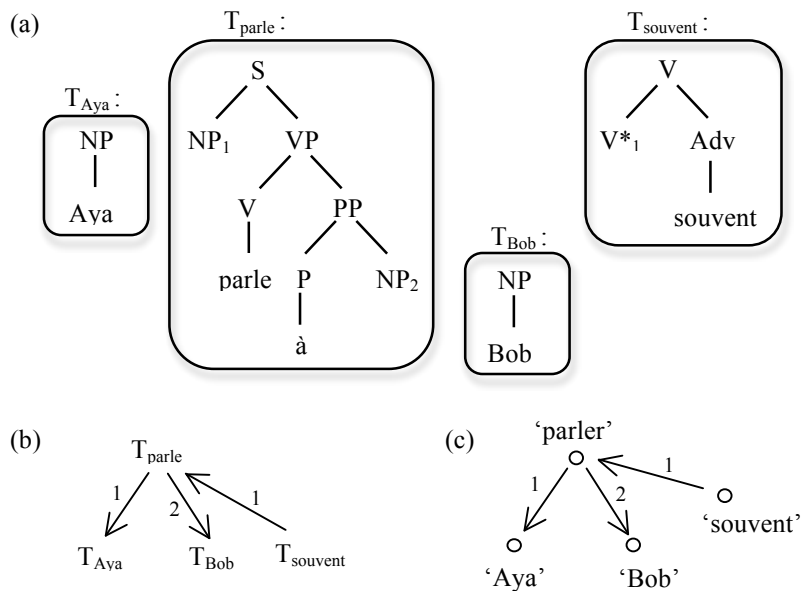
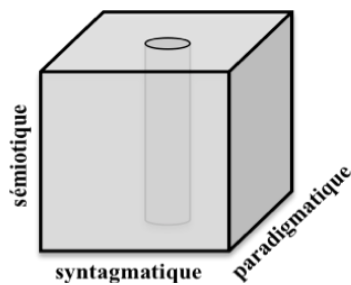


Figure 6. Dérivation de *Aya parle souvent à Bob*

Nous donnons figure 6(a) un ensemble d'arbres élémentaires permettant de dériver la phrase *Aya parle souvent à Bob*. En (b), nous donnons la structure de dérivation où une flèche vers le bas représente une substitution et une flèche vers le haut une adjonction (selon les conventions proposées par Candito et Kahane, 1998) et en (c) le graphe de relations prédicat-argument correspondant (Mel'čuk, 2012).

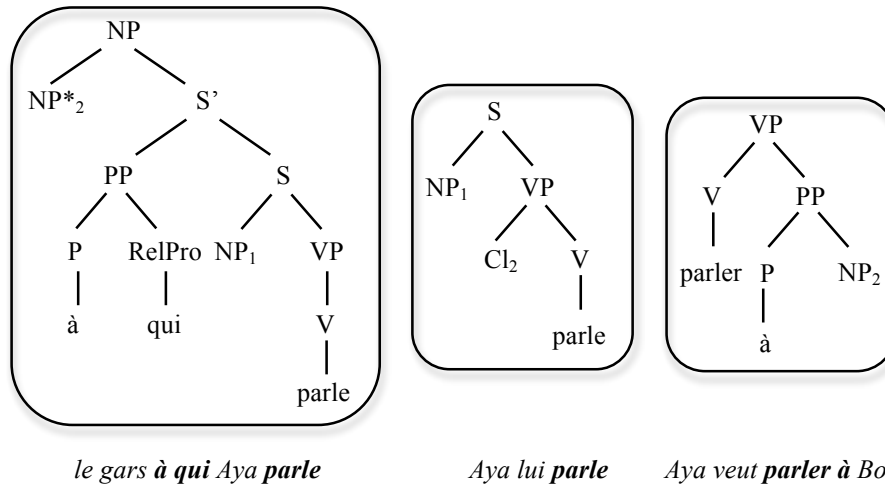
Les arbres élémentaires de LTAG, qui sont en quelque sorte les règles de cette grammaire, ont un large empan sémiotique puisqu'ils décrivent la totalité du signe : la règle correspond à une unité sémantique et elle décrit entièrement sa combinatoire syntaxique, y compris l'ordre linéaire (figure 7). Le lexique et la morphologie flexionnelle constituent néanmoins des modules séparés : un arbre comme  $T_{\text{parle}}$  résulte en fait de la combinaison d'un arbre du même type, mais avec des ancres lexicales non spécifiées, d'une entrée lexicale pour PARLER indiquant que ce verbe se construit avec la préposition à et d'une règle de flexion indiquant que  $\text{PARLER}_{\text{ind,prés,3,sg}} = \textit{parle}$ .



**Figure 7.** Empan d'un arbre LTAG

Le fait qu'un arbre LTAG a un large empan syntagmatique et sémiotique a un coût : une telle règle possède un faible empan paradigmatique. Autrement dit, les autres emplois de l'unité lexicale PARLER, pour la même acception avec la même sous-catégorisation, devront être réglés avec d'autres arbres élémentaires (figure 8).

Cette particularité de LTAG entraîne une explosion du nombre de règles associées à une même unité lexicale, même si certaines versions de LTAG (de La Clergerie, 2005) permettent de factoriser les arbres et de réduire nettement leur nombre. Il existe bien sûr des moyens de gérer ce grand nombre de structures alternatives en « découpant » les informations qu'elles contiennent dans une « métagrammaire » qui va générer la grammaire LTAG proprement dite (Candito, 1996 ; Crabbé, 2005). Néanmoins ces métagrammaires sont écrites dans des formalismes différents de la grammaire TAG proprement dite et les « règles » de la métagrammaire sont utilisées dans une phase préliminaire pour précompiler la grammaire TAG. Nous allons présenter un autre modèle, qui se rapproche des métagrammaires LTAG en couvrant de grands emfans paradigmatiques, tout en pouvant être utilisé sans précompilation. Nous reviendrons sur LTAG ensuite.



**Figure 8.** Quelques autres arbres élémentaires pour PARLER

### 3.2. Le recouvrement de GUST

La grammaire d'unification sens-texte (GUST) (Kahane et Lareau, 2005) est une grammaire formelle dont l'architecture s'inspire de la théorie sens-texte (TST), mais dont les mécanismes formels s'apparentent à TAG, puisque les structures associées à un énoncé sont obtenues par la combinaison de structures élémentaires. GUST est écrite dans le formalisme des grammaires d'unification polarisées (GUP) (Kahane, 2004), où le contrôle de la saturation des structures dérivées par l'unification des structures élémentaires est assuré par la polarisation des objets des structures.

Sur le plan théorique, GUST repose sur un niveau syntaxique intermédiaire où la structure de l'énoncé est représentée par un arbre de dépendance (non ordonné). Deux modules séparés s'interfaçent avec le niveau syntaxique : d'une part, un module d'interface syntaxe-sémantique qui met en relation la structure syntaxique avec une structure sémantique sous la forme d'un graphe de relations prédicat-argument, comme dans la TST (Mel'čuk, 1988 ; 2012) ; d'autre part, un module dit topologique composé de règles de linéarisation qui met en relation l'arbre de dépendance avec un ordre linéaire sur les nœuds de l'arbre (et qui construit en même temps une structure de constituants topologiques) (Gerdes et Kahane, 2001). Nous présenterons ici uniquement l'interface syntaxe-sémantique et ses derniers développements (Kahane, 2013). Nous allons montrer que le découpage en deux modules selon la dimension sémiotique, ainsi qu'un découpage syntagmatique très fin, permettent d'avoir des règles avec un grand empan selon la dimension paradigmatique, à la différence de LTAG.

### 3.2.1. Le formalisme de base

Le formalisme de base de l'interface syntaxe-sémantique de GUST est celui de Nasr (1995), avec quelques améliorations proposées par Kahane et Lareau (2005). La figure 9 représente en (a) cinq structures élémentaires (les flèches discontinues entre ces structures indiquent quels objets s'unifient quand on les combine), en (b) la structure dérivée, qui est un arbre de dépendance, et en (c) la structure de dérivation, interprétable comme un graphe sémantique. Nous allons préciser tout cela.

Une structure en GUP est fondée sur des *objets*. Les structures de l'interface syntaxe-sémantique de GUST présentée ici contiennent trois types d'objets : 1) des nœuds, correspondant aux lexèmes, 2) des dépendances syntaxiques, représentées par des flèches, 3) des *grammèmes* ou morphèmes flexionnels (Mel'čuk, 1988 ; Lareau, 2011), représentés par des losanges (cf. par exemple  $R_{\text{indicatif}}$  à la figure 9). Les objets sont liés à trois sortes d'éléments : 1) d'autres objets (par exemple une dépendance est liée à un nœud source et un nœud cible), 2) des *valeurs atomiques* (étiquettes ou valeurs de traits), 3) des *polarités*.

Les polarités se distinguent des valeurs atomiques par la façon dont elles se combinent. Lorsque deux structures (élémentaires) se combinent, au moins un objet d'une des structures doit être identifié avec un objet de l'autre structure. Quand deux objets sont identifiés, les éléments auxquels ils sont liés doivent eux-mêmes se combiner : les objets et les valeurs atomiques sont identifiés (c'est l'unification), tandis que les polarités se combinent selon une opération particulière appelée le *produit sur les polarités*. Nous considérons deux polarités dans cet article :

- = insaturé = « blanc » ;
- = saturé = « noir ».

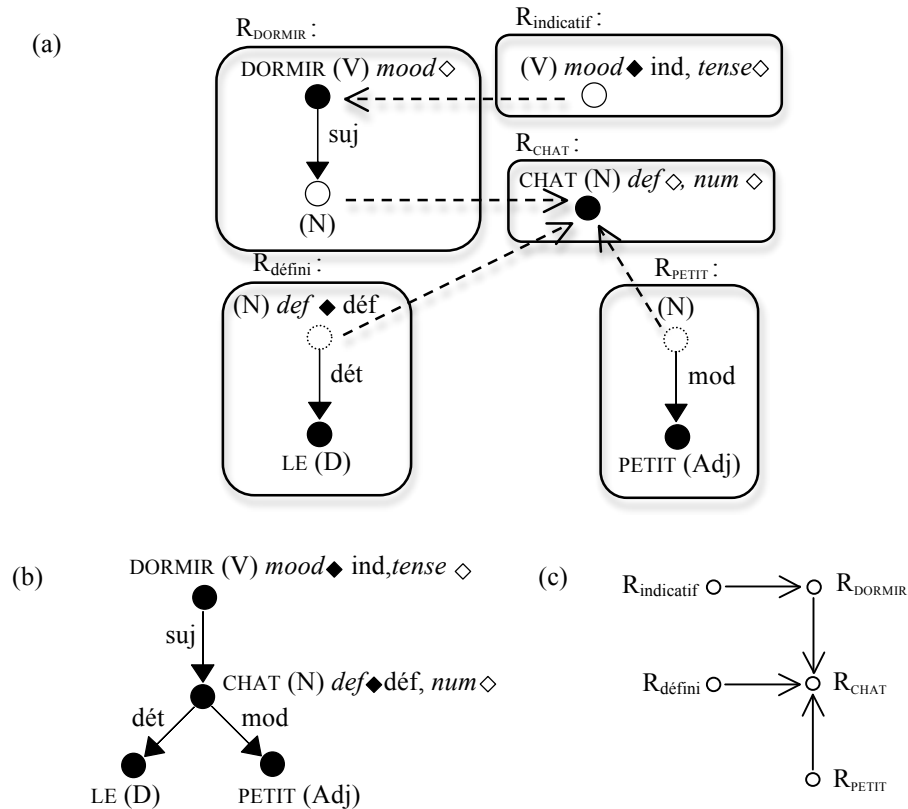
La polarité blanche est la seule polarité avec laquelle les autres polarités peuvent se combiner (■.■ = échec) et cette polarité est en plus l'identité du produit (□.□ = □ et □.■ = ■). Les polarités blanches sont insaturées, c'est-à-dire qu'elles doivent obligatoirement se combiner avec une polarité noire.

Nous associons à chaque objet un triplet de polarité. Quatre sortes d'objets sont ainsi considérés dans les structures élémentaires :

- = [■,□,□] = ressource = « noir » ;
- = [□,■,□] = besoin = « blanc » ;
- ⦿ = [□,□,□] = contexte = « pointillé » ;
- = [■,□,■] = à supprimer = « gris ».

Les objets noirs sont ceux construits par la règle et les objets blancs ceux à construire (voir aussi les polarités positive et négative de Kahane, 2004). L'unification des deux (noir + blanc) donne un objet [■,■,□] (dit « visible »), que nous représentons également en noir pour ne pas multiplier les conventions de représentation. Les objets gris servent à supprimer des besoins : l'unification blanc + gris donne un objet [■,■,■] (dit « invisible »). Les objets « invisibles », qui ne

peuvent plus être manipulés par aucune règle, ne seront pas représentés dans nos dérivations.



**Figure 9.** Dérivation de *Le petit chat dort*

Les trois polarités de nos triplets jouent des rôles différents. La première polarité est celle de la grammaire syntaxique proprement dite : elle indique quels sont les éléments qui sont construits au niveau syntaxique. La deuxième polarité n'est pas essentielle : elle sert à contrôler qu'une ressource corresponde toujours à un et un seul besoin (cf., par ex., les commentaires de la figure 11 sur l'unicité du sujet). La troisième polarité est elle essentielle : elle contrôle l'interface avec la topologie, c'est-à-dire avec le module qui assure la linéarisation (voir Kahane et Lareau (2005) pour l'introduction des *polarités d'interface*). Autrement dit, un objet  $[\blacksquare, \blacksquare, \square]$  doit être manipulé par la grammaire topologique, tandis qu'un objet  $[\blacksquare, \blacksquare, \blacksquare]$  est invisible pour ce module.

D'autres polarités peuvent encore être ajoutées, que nous n'introduisons pas pour des raisons de lisibilité. Une polarité permet de contrôler que la structure visible pour l'interface avec la topologie est bien un arbre (Kahane, 2004). Une autre

polarité indique quels sont les objets qui sont visibles pour l'interface avec la sémantique et permet ainsi de mieux contrôler l'interprétation sémantique de nos structures de dérivation (cf., par exemple, les commentaires de la figure 13 concernant les prépositions régimes).

La figure 10 explicite nos conventions de représentation : un objet noir est, comme nous l'avons dit, un objet dont la polarité est  $[\blacksquare, \square, \square]$ , cette polarité étant la valeur de la fonction *pol* ; le nom du lexème et sa partie du discours sont les valeurs atomiques des fonction *lex* et *pos*. Les grammèmes de mode et de temps sont des objets liés au nœud par des fonctions (*mood* et *tense*) et ont eux-mêmes des polarités et des valeurs associées<sup>7</sup>.

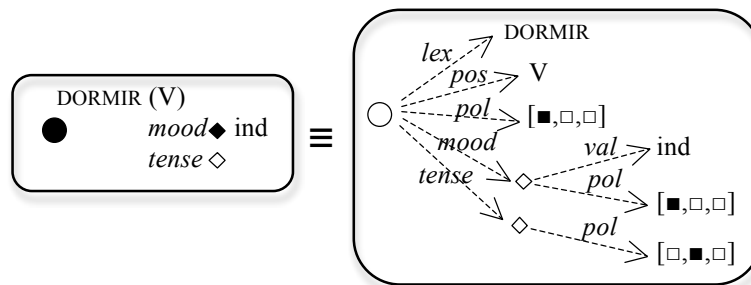


Figure 10. Nos conventions

Revenons à la figure 9. La règle  $R_{\text{DORMIR}}$  indique que le lexème DORMIR requiert un sujet nominal et un mode (ici la catégorie du mode inclut la finitude et peut prendre la valeur *infinitif*). La dépendance *sujet* et le lexème DORMIR (polarisés en noir) sont construits par la règle, tandis que le nom sujet et le mode (polarisés en blanc) sont des requêtes. La règle  $R_{\text{indicatif}}$  attribue le mode indicatif et requiert un temps en conséquence. La règle  $R_{\text{CHAT}}$  introduit le nom CHAT et requiert la définitude et le nombre pour lui. La règle  $R_{\text{défini}}$  adjoint le déterminant défini LE à un nom et sature son grammème de définitude. La règle  $R_{\text{PETIT}}$  adjoint PETIT à un nom. Les cinq règles se combinent entre elles pour donner la *structure dérivée* (b) qui est un arbre de dépendance. On notera que cette structure n'est pas saturée et requiert encore un temps pour DORMIR et un nombre pour CHAT (nous avons laissé de côté tous les grammèmes d'accord qui ne jouent pas de rôle dans l'interface syntaxe-sémantique). Le graphe (c) indique comment les règles se sont combinées et constituent la *structure de dérivation*. Le sens des flèches va de la requête à l'élément qui la remplit, c'est-à-dire du prédicat à son argument, et la structure de dérivation peut donc être interprétée comme une représentation sémantique (Mel'čuk, 2012).

7. Nous avons légèrement simplifié la figure 11. Une fonction *host* lie chaque grammème à son nœud hôte afin que l'identification de deux grammèmes entraîne bien l'identification des nœuds hôtes (Kahane et Lareau, 2005).



### 3.2.2. Des règles pour les dépendances

Il est possible de séparer les règles qui construisent les nœuds lexicaux des règles qui construisent les dépendances (cf. la distinction entre règles *nodales* et *sagittales* faite par Kahane et Mel'čuk (1999)). Nous modifions en conséquence nos précédentes règles en introduisant des règles séparées pour la dépendance. La figure 11 montre deux règles dont la combinaison ( $L_{\text{DORMIR}} \oplus D_{\text{sujet}}$ ) donne la règle  $R_{\text{DORMIR}}$ <sup>8</sup>. Séparer règles nodales et sagittales revient à segmenter davantage selon la dimension syntagmatique. L'idée qu'une dépendance est un objet en soi revient à Tesnière<sup>9</sup>, mais est aussi sous-jacente dans les constructions de Bloomfield (1933 : ch. 12.1) et l'idée a été reprise par les grammaires de construction (CxG), pour lesquelles les constructions sont des éléments du « lexique » au même titre que les lexèmes. Dans cet article, nous ne traitons pas directement la question de l'ordre linéaire. On notera néanmoins que c'est par les dépendances qu'est géré l'ordre des mots : à chaque dépendance est associée une spécification d'ordre qui place le dépendant par rapport au gouverneur (Gerdes et Kahane, 2001). Cette spécification d'ordre constitue en quelque sorte le signifiant de la dépendance, son signifié étant la relation syntaxique dans l'arbre de dépendance.

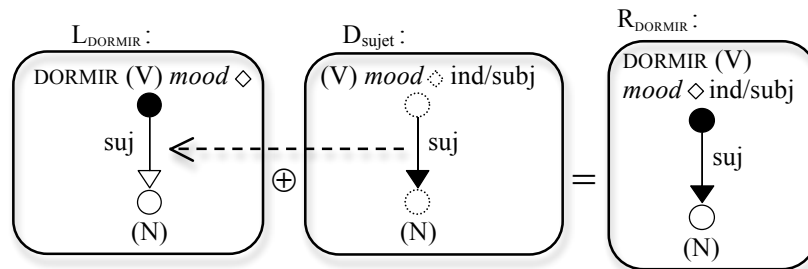


Figure 11. Règles nodale et sagittale.

La règle  $D_{\text{sujet}}$  permet de saturer la dépendance *sujet* si le verbe est au mode indicatif ou subjonctif. Autrement dit, bien que le sujet soit un argument du verbe, la réalisation syntaxique du sujet est, quant à elle, contrôlée par le mode (cf. la position du sujet sous InflP en syntaxe X-barre). La dépendance *sujet* de  $D_{\text{sujet}}$  est un objet noir qui doit obligatoirement se combiner avec un objet blanc, ce qui empêche qu'il y ait deux sujets sur un verbe.

8. Nous utiliserons les conventions suivantes pour nommer nos règles : L pour une règle nodale (construisant un lexème), D pour une règle sagittale (construisant une dépendance), G pour une règle grammaticale (construisant un grammème) et C pour une construction autre. R est utilisé pour les règles combinées de la sous-section 3.2.1.

9. « [...] une phrase du type *Alfred parle* n'est pas composée de deux éléments 1° *Alfred*, 2° *parle*, mais bien de **trois** éléments, 1° *Alfred*, 2° *parle* et 3° la connexion qui les unit et sans laquelle il n'y aurait pas de phrase. Dire qu'une phrase du type *Alfred parle* ne comporte que deux éléments, c'est l'analyser d'une façon superficielle, purement morphologique, et en négliger l'essentiel, qui est le lien syntaxique. » (Tesnière, 1959 : ch. 1)

Quand le verbe est à un autre mode, infinitif ou participe, le sujet n'est pas réalisé par une dépendance sur le verbe même. Par exemple, au passé composé (*a dormi*), le sujet « monte » sur l'auxiliaire AVOIR. La règle  $G_{\text{passé\_composé}}$  donnée figure 12 réalise cela grâce à la polarité grise : la requête d'un sujet de DORMIR va s'unifier avec une dépendance grise et devenir invisible pour les autres modules<sup>10</sup>. Elle sera remplacée par une dépendance *sujet* sur AVOIR. Le résultat de l'unification de la dépendance *sujet* blanche de  $L_{\text{DORMIR}}$  et de la dépendance *sujet* grise de  $G_{\text{passé\_composé}}$  donne un objet entièrement saturé, qui n'est pas représenté, car il ne jouera plus aucun rôle ensuite.

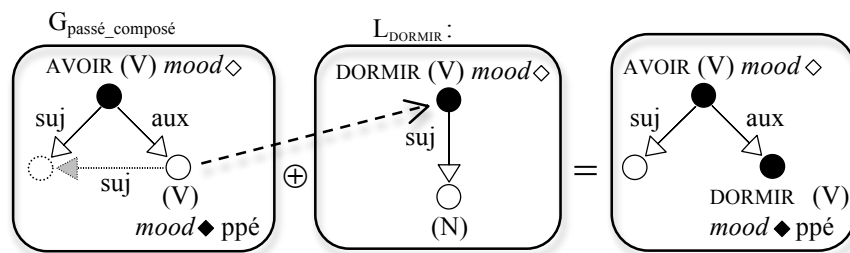


Figure 12. Dépendance invisible.

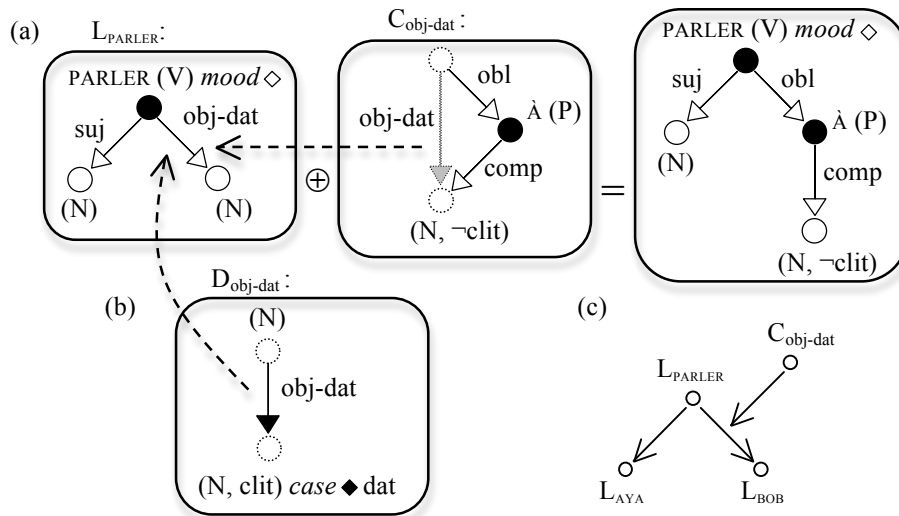
Le fait de permettre de supprimer une dépendance et de la remplacer par une autre est similaire à ce que propose la Relational Grammar (Perlmutter, 1980). On peut d'ailleurs reprendre ici l'idée des fonctions profonde et de surface de Fillmore (1968) (également exploitée par Candito (1996) dans l'écriture de sa métagrammaire LTAG). Une *fonction de surface* peut apparaître dans une structure syntaxique finale : elle doit pouvoir être instanciée par une règle de saturation comme  $D_{\text{sujet}}$  pour la fonction *sujet* (figure 11). Une *fonction profonde* est la fonction portée par une dépendance blanche, qui peut être instanciée (la fonction profonde est alors aussi une fonction de surface) ou qui peut être rendue invisible et remplacée par une autre dépendance avec éventuellement une autre fonction<sup>11</sup>.

Les fonctions profondes (et la possibilité de renommer ou déplacer une dépendance) permettent d'élargir l'empan paradigmatique de nos règles lexicales. Par exemple, le complément d'objet indirect du verbe PARLER peut soit être introduit par une préposition (*Aya parle à Bob*), soit être réalisé par un clitique (*Aya lui parle*). Nous aurons pour ces deux cas une même règle lexicale  $L_{\text{PARLER}}$  où l'objet a la

10. Une règle similaire a été proposée par Kahane (2001), où une dépendance grise est appelée une quasi-dépendance. L'idée de considérer une quasi-dépendance comme une dépendance avec une polarité d'interface avec la topologie saturée est apparue après (Kahane et Lareau, 2005) et est expliquée dans (Lareau, 2008).

11. Notons au passage qu'une dépendance n'est jamais « effacée » purement et simplement. Elle est toujours remplacée par une autre dépendance, ce qui revient à la renommer ou à la déplacer. Il est possible que cette précaution réduise la complexité du formalisme.

fonction profonde *obj-dat* (figure 13). La dépendance *obj-dat* de PARLER peut être remplacée par une construction avec la préposition À ( $C_{obj-dat}$ , cas (a) dans la figure 13) ou être instanciée lorsque le dépendant est un clitique ( $D_{obj-dat}$ , cas (b)). La structure de dérivation (c) indique que  $C_{obj-dat}$  se combine avec le lien qui unit  $L_{PARLER}$  avec son complément. Même si l'objet principal avec lequel  $C_{obj-dat}$  se combine a bien été introduit par  $L_{PARLER}$ , nous utilisons une convention de représentation pour la structure de dérivation qui explicite le fait que  $C_{obj-dat}$  a également  $L_{BOB}$  dans son contexte.



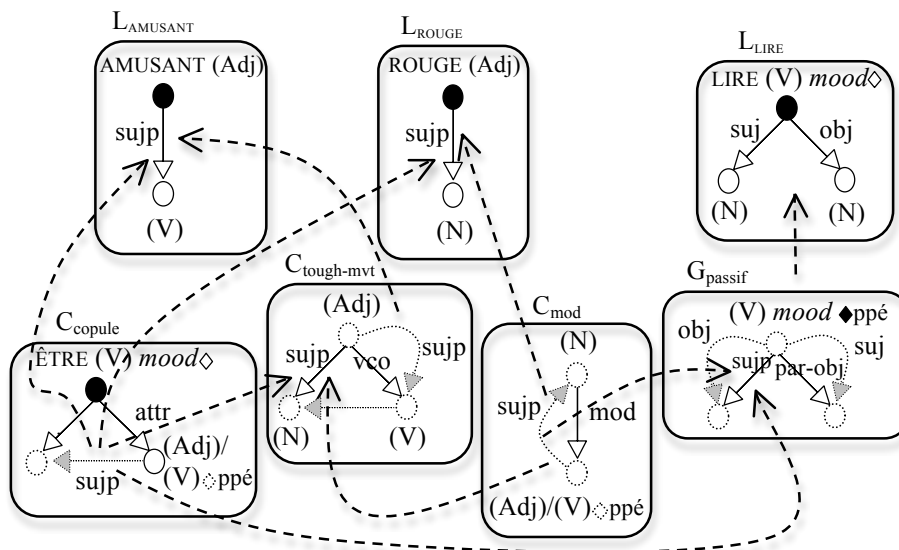
**Figure 13.** Fonctions profondes et de surface

La règle  $C_{obj-dat}$ , qui introduit À, n'est pas considérée comme une règle lexicale, mais comme une réévaluation de la connexion entre deux règles lexicales. Cette règle sera ignorée lorsque la structure de dérivation sera interprétée au niveau sémantique. Le découpage en règles nodales et sagittales nous a donc permis d'introduire une seule règle  $L_{PARLER}$  pour les deux constructions de PARLER (avec préposition ou avec clitique)<sup>12</sup>. Cette règle pour PARLER a un empan paradigmatique nettement supérieur à une règle LTAG (ainsi qu'aux règles des précédentes versions de GUST, comme Kahane, 2001).

Donnons un deuxième exemple montrant comment obtenir des règles lexicales avec un large empan paradigmatique. Les adjectifs apparaissent dans deux constructions de base : comme épithètes, ils modifient un nom (*le livre rouge*) et comme attributs, ils forment un complexe verbal avec la copule (*le livre est rouge*).

12. Outre le fait que la préposition est absente quand l'argument est pronominal, une raison d'introduire la préposition séparément est qu'elle peut être répétée en cas de coordination : *Aya parle à Bob et à Eve*.

Dans les deux cas, ‘rouge’ est un prédicat sémantique et ‘livre’ est son argument sémantique. Mais certains adjectifs peuvent aussi prendre un événement comme argument et réaliser cet argument par un verbe (*lire cet article est amusant*). De tels adjectifs ne peuvent pas modifier leur argument puisque c’est un verbe et que la construction épithétique est donc impossible. Mais, en revanche, le complément d’objet direct de l’argument peut être promu (ce qu’on appelle le *tough-movement*) et devenir ainsi le gouverneur syntaxique de l’adjectif (*un article amusant à lire*). En conséquence, on ne peut pas considérer que la construction épithétique est la construction de base de tout adjectif, puisqu’elle est impossible quand l’argument est verbal. Nous proposons alors que la construction de base d’un adjectif comporte une fonction profonde *suip* (*sujet profond*) (figure 14). Cette fonction n’existera pas comme fonction de surface, mais les trois constructions dont nous avons parlé pourront s’appliquer sur elle : la construction épithétique ( $C_{mod}$ ), la construction attributive ( $C_{copule}$ ) et la construction du *tough-movement* ( $C_{tough-mvt}$ ). On notera que la construction épithétique peut seulement s’appliquer si l’argument est réalisé par un nom ou après le *tough-movement*, lorsqu’un nom a été promu *suip*.

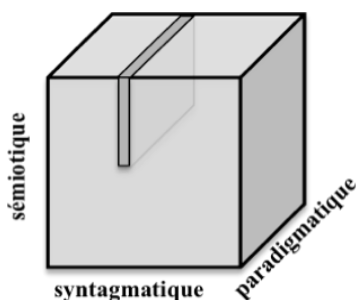


**Figure 14.** Fragment de grammaire pour les constructions adjectivales et passives

La même fonction profonde *suip* sera également utilisée pour le participe passé passif qui peut aussi entrer dans une construction attributive (*le livre a été volé*) ou épithétique (*le livre volé par Bob*). La construction passive du verbe marquée par le grammème de participe passé est ainsi obtenue par combinaison avec la construction de base du verbe (c’est-à-dire l’actif) en utilisant à nouveau la polarisation grise ( $G_{passif}$ , figure 14).

Nous avons ainsi décrit avec seulement trois règles ( $C_{\text{copule}}$ ,  $C_{\text{tough-mvt}}$ ,  $C_{\text{mod}}$ ) quatre constructions différentes (puisque  $C_{\text{copule}}$  et  $C_{\text{mod}}$  peuvent être utilisées directement ou être combinées avec  $C_{\text{tough-mvt}}$ ), qui vont elles-mêmes se combiner avec tous les adjectifs. Il faudra ajouter encore quelques règles pour avoir l'ensemble des constructions adjectivales (notamment l'impersonnel *il est amusant de lire ce livre*), mais on est évidemment très loin des 2 560 schémas d'arbres élémentaires pour les adjectifs proposés pour la grammaire LTAG par Barrier (2002).

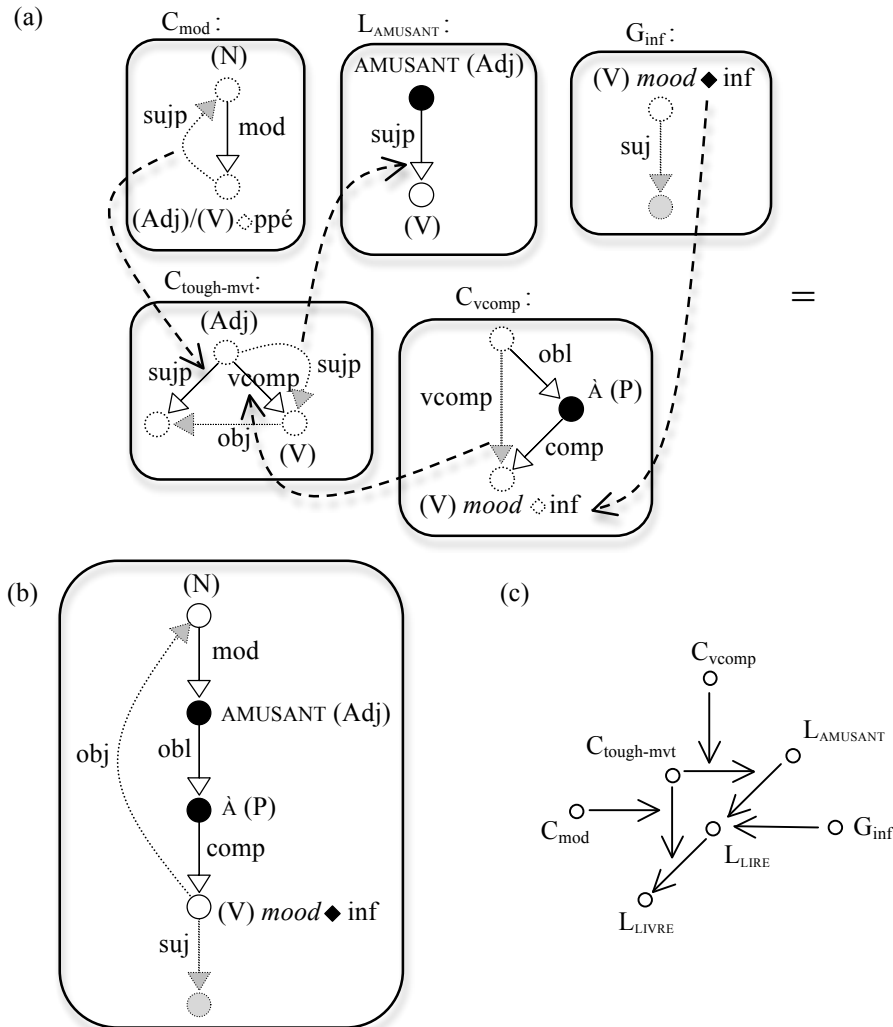
En séparant, comme nous l'avons fait, la description d'une unité lexicale de la description des constructions avec lesquelles elle est compatible, nous obtenons une grammaire avec des règles aussi bien lexicales que grammaticales ayant un empan maximal dans la dimension paradigmatique, c'est-à-dire des règles qui pourront être utilisées dans une grande variété d'énoncés. Ces règles ont en contrepartie un faible empan sémiotique (ce sont des « demi-signes ») et un empan syntagmatique minimal (figure 15).



**Figure 15.** Empan d'une règle GUST

### 3.3. Comparaison

Nous avons vu que les règles de LTAG et celles de GUST recouvraient l'espace linguistique de manière différente (cf. la comparaison des figure 7 et 15). Nous allons voir sur un exemple que les règles GUST peuvent être précompilées pour obtenir une grammaire similaire à LTAG. La figure 16 montre en (a) une combinaison de règles de GUST qui donne la structure (b) équivalente à l'arbre LTAG  $T_{\text{amusant}}$  de la figure 17. Cette combinaison est partielle, puisqu'il faudra encore combiner la structure dérivée de la figure 16 avec des règles sagittales pour chaque dépendance, lesquelles règles devront d'ailleurs spécifier l'ordre linéaire pour que la structure obtenue soit vraiment équivalente à l'arbre  $T_{\text{amusant}}$ .

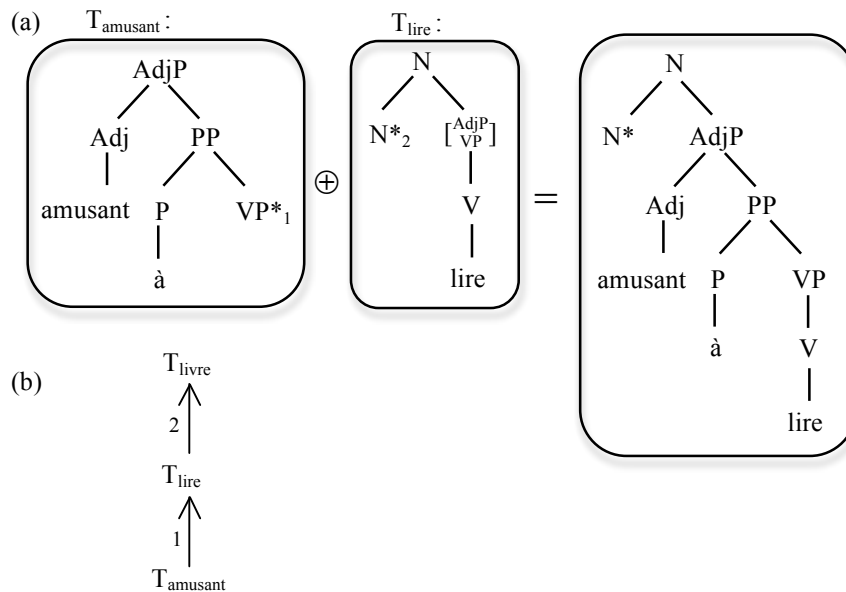


**Figure 16.** *Tough-movement* en GUST

La dérivation présentée dans la figure 16 étant assez complexe, nous allons donner quelques explications. On peut voir en (b) la structure dérivée et en (c) la structure de dérivation. Cette dernière nous indique que  $L_{AMUSANT}$  se combine avec  $L_{LIVRE}$ , qui se combine à son tour avec  $L_{LIVRE}$ , nous donnant ainsi les relations sémantiques souhaitées. La règle  $G_{inf}$ , qui réalise le grammème infinitif, indique que le sujet de LIVRE ne sera pas réalisé au niveau syntaxique (et prendra une valeur indéfinie similaire à ‘on’). La règle  $C_{tough-mvt}$  manipule les liens syntaxiques qui lient AMUSANT à LIVRE et LIVRE à LIVRE : le premier est une relation *subj* qui devient *vcomp* et le deuxième une relation *obj* qui monte sur l’adjectif et devient *subj*. Ces deux

relations sont à nouveau transformées : la première donne une préposition *à* par  $C_{vcomp}$  et la deuxième donne une relation *mod* par  $C_{mod}$ . La structure dérivée pourra se combiner avec la structure canonique du verbe LIRE (voir la règle  $L_{LIRE}$  de la figure 14).

La modélisation du *tough-movement* en LTAG est non triviale et n'a jamais été proposée à notre connaissance. Comme précédemment, on veut que la structure de dérivation contienne les relations prédicat-argument (c'est-à-dire sémantique) et donc que *amusant* s'adjoigne sur *lire*, qui s'adjoit lui-même sur *livre* (figure 17 (b)). Cela est délicat en TAG standard et il est préférable d'adopter, comme dans Rogers (1999) ou Kahane *et al.* (2000), une version de TAG où les catégories sont dans les traits *top* et *bottom*. L'arbre  $T_{lire}$  a ainsi un nœud  $[t : AdjP, b : VP]$  où l'adjonction est obligatoire (puisque la structure dérivée est bien formée seulement si les traits *top* et *bottom* de tous les nœuds peuvent s'unifier). Contrairement à ce que nous avons vu en GUST, il sera nécessaire en LTAG d'introduire une règle particulière pour LIRE lorsqu'il est utilisé dans un *tough-movement* (cf. l'arbre  $T_{lire}$  de la figure 17). Ceci est d'autant plus regrettable que le *tough-movement* s'applique à n'importe quel verbe transitif et n'est donc en rien une particularité lexicale de LIRE.



**Figure 17.** *Tough-movement* en LTAG

Comme on le voit, la grammaire GUST est beaucoup plus granulaire, mais il est toujours possible de la précompiler pour obtenir des règles équivalentes à celles d'une LTAG. En d'autres termes, la grammaire GUST est plus ou moins équivalente

à une métaگرامmaire TAG (Candito, 1996 ; Crabbé, 2005), c'est-à-dire une grammaire capable de générer les arbres de la grammaire TAG proprement dite. L'un des avantages de GUP est d'offrir un formalisme où peuvent être écrites des grammaires équivalentes aussi bien à la grammaire qu'à la métaگرامmaire TAG et donc de pouvoir utiliser les deux conjointement. Avec GUP, il est possible de ne précompiler que certaines règles (les combinaisons les plus courantes par exemple ou les dernières utilisées) et/ou de maintenir des combinaisons de règles précompilées en même temps que les règles qui ont été utilisées dans les combinaisons.

#### **4. Conclusion**

Nous espérons avoir convaincu notre lecteur que tout ensemble de productions d'une langue peut être envisagé comme un espace à trois dimensions. Des trois dimensions, la dimension syntagmatique est la mieux connue et le découpage syntagmatique a été largement exploré depuis Saussure et Bloomfield. La nature de la dimension sémiotique est encore imparfaitement connue et la question du nombre et de la nature des différents niveaux d'organisation d'un énoncé reste, selon nous, une question ouverte cruciale. La dimension paradigmatique est de loin la plus complexe. En effet, le paradigme des environnements possibles de toute forme est potentiellement illimité et il n'est même pas certain que le recouvrement de l'ensemble des occurrences d'une forme puisse être discrétisé. Il n'y a qu'à voir les nombreuses discussions des lexicographes sur le nombre d'acceptions de chaque entrée de dictionnaire et sur la possibilité même de découper une entrée en acceptions.

La modélisation linguistique suppose d'être capable de recouvrir l'espace linguistique et il existe différentes stratégies pour cela. Nous pensons qu'un modèle linguistique ne peut pas être composé uniquement de descriptions de signes linguistiques couvrant tout l'empan sémiotique, du sens à la forme de surface. Les signes linguistiques s'organisent en faisceaux et il est préférable de considérer des « demi-signes » qui ne couvrent qu'une partie de l'empan sémiotique, puisqu'il y a une forme d'indépendance entre la contribution sémantique du signe et sa forme (cf. le cas d'un faisceau de signes comme le lexème ALLER qui est à la fois polysémique et polymorphique). En formalisant des « signes » d'empan sémiotique restreint, qu'il s'agisse de signes lexicaux ou de constructions grammaticales (y compris des dépendances syntaxiques), on peut obtenir un modèle très modulaire, où les règles couvrent un empan syntagmatique minimal (et représentent donc des signes indécomposables) tout en couvrant un empan paradigmatique maximal (et pouvant donc être utilisées dans un très grand nombre d'énoncés).

#### **Remerciements**

Je remercie François Lareau, ainsi que deux relecteurs de TALN et les trois relecteurs de TAL, pour leurs très nombreuses remarques sur les précédentes



versions de cet article, lesquelles m'ont permis, je l'espère, de rendre mon propos nettement plus clair.

## 5. Bibliographie

- Abeillé A., Une grammaire lexicalisée d'arbres adjoints pour le français, Thèse de Doctorat, Univ. Paris 7, 1991.
- Barrier N., « Une MétaGrammaire pour les adjectifs du français », *Actes TALN*, 2002, Nancy, p. 351-357.
- Bloomfield L., *Language*, New York, Holt, Rinehart and Winston, 1933.
- Candito M.-H., « A principle-based hierarchical representation of LTAGs », *Proc. COLING*, 1996, Copenhagen.
- Candito M.-H., Kahane S., « Can the derivation tree represent a semantic graph? An answer in the light of Meaning-Text Theory », *Proc. TAG+4*, 1998, Philadelphia, p. 21-24.
- Crabbé B., Représentation informatique de grammaires d'arbres fortement lexicalisées : le cas de la grammaire d'arbres adjoints. Thèse de doctorat, Université Nancy 2, 2005.
- de La Clergerie E. Villemonte, « From metagrammars to factorized TAG/TIG parsers », *Proc. IWPT*, 2005, Vancouver.
- Fillmore Ch. J., « The Case for Case », in Bach & Harms (éds.): *Universals in Linguistic Theory*, New York, Holt, Rinehart, and Winston, 1968, p. 1-88.
- Frege G., « Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought », 1879, [trad. in van Heijenoort J., *From Frege to Gödel - A source book in mathematical logic, 1879-1931*, Cambridge, Harvard University Press, 1967, p. 1-82].
- Gerdes K., Kahane S. « Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy », *Proc. ACL*, 2001, Toulouse.
- Hjelmslev L. *Prolégomènes à une théorie du langage*, 1943, [trad. du danois par U. Canger, Paris, Éditions de Minuit, 1971].
- Jackendoff R. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, 2002.
- Jespersen O. *Analytic Syntax*, 1937, [trad. de A.-M. Léonard (1971), *La syntaxe analytique*, Éditions de Minuit].
- Joshi A., « Introduction to Tree Adjoining Grammar », in Manaster Ramer (éd.), *The Mathematics of Language*, Amsterdam, Benjamins, 1987, p. 87-114.
- Joshi A. K., Vijay-Shanker K., « Compositional semantics with lexicalized tree-adjoining grammar (LTAG): How much underspecification is necessary? », *Computing Meaning*, Netherlands, Springer Verlag, 2001, p. 147-163.
- Kahane S., « A fully lexicalized grammar for French based on Meaning-Text theory », *Proc. CILing*, 2001, Mexico, Springer Verlag, p. 18-31.
- Kahane S., « Grammaires d'unification polarisées », *Actes TALN*, 2004, Fès.

- Kahane S., « Predicative Adjunction in a Modular Dependency Grammar », *Proc. DepLing*, 2013, Prague.
- Kahane S., Candito M.-H., De Kercadio Y., « An alternative descriptions of extractions in TAG », *Proc. TAG+5*, 2000, Paris, p. 115-122.
- Kahane S., Lareau F., « Meaning-Text Unification Grammar: Modularity and polarization », *Proc. MTT*, 2005, Moscou.
- Kahane S., Mel'čuk I., « La synthèse sémantique ou la correspondance entre graphes sémantiques et arbres syntaxiques – Le cas des phrases à extraction en français contemporain », *TAL*, 40:2, 1999, p. 25-85.
- Kamp H., « A theory of truth and semantic representation », 1981, [republié in Portner P., Partee B. H., *Formal semantics - The essential readings*, Blackwell, 2002, p. 189-222].
- Lareau F., Vers une grammaire d'unification Sens-Texte du français : le temps verbal dans l'interface sémantique-syntaxe, Thèse de doctorat, Univ. de Montréal/Univ. Paris 7, 2008.
- Lareau F., « Grammemes », *Proc. MTT*, 2011, Barcelone, p. 145-154.
- Mel'čuk I., *Dependency Syntax: Theory and Practice*, Albany, SUNY, 1988.
- Mel'čuk I. *Semantics: From Meaning to Text*, Amsterdam, Benjamins, 2012.
- Nasr A., « A Formalism and a Parser for Lexicalised Dependency Grammars », *Proc. IWPT*, 1995, Prague, p. 186-195.
- Perlmutter D., « Relational grammar », in E. Moravcsik & J. Wirth (éds.), *Syntax and semantics: Current approaches to syntax* (Vol. 13), New York, Academic Press, 1980, p. 195-229.
- Rambow O., Joshi A., « A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena », in L. Wanner (éd.), *Current Issue in Meaning-Text Theory*, Pinter, London, 1994.
- Sgall P., *Generativní popis jazyka a česká deklinace* [Generative description of language and Czech declension], Academia, Prague, 1967.
- Shieber S., Schabes Y., « An alternative conception of tree-adjoining derivation », *Computational Linguistics*, 20:1, 1994, p. 91-124.
- Rogers J., « Generalized Tree-Adjoining Grammar », *Proc. MOL*, 1999, Orlando, p. 189-202.
- Saussure F. de, *Cours de linguistique générale*, Paris, 1916.
- Tesnière L., *Éléments de syntaxe structurale*, Paris, Klincksieck, 1959.
- Vijay-Shanker K., A Study of Tree Adjoining Grammars, PhD thesis, Univ. of Pennsylvania, 1987.
- XTAG-Group, A Lexicalized Tree Adjoining Grammar for English, Technical Report IRCS 95-03, Univ. of Pennsylvania, 1995.
- Žolkovskij A., Mel'čuk I. « O semantičeskomoj sinteze » [Sur la synthèse sémantique (de textes)], *Problemy Kybernetiki* [Problèmes de Cybernétique], 19, 1967, p. 177-238, [trad. fr. in *T.A. Information*, 2, 1970, p. 1-85].