

Pour citer cette notice :

Mazziotta (N.), 2023, « Récurtivité », in *Encyclopédie Grammaticale du Français*, en ligne : <http://encyclogram.fr>

1. INTRODUCTION

La récurtivité (parfois nommée *réursion* par influence de l'anglais) est un concept aux contours un peu flous (Lowenthal 2014 : v-vi) présent dans de nombreux domaines du savoir et dans de nombreuses œuvres artistiques. Dans une optique interdisciplinaire revendiquée comme telle, Hauser et al. (2002) militent pour que soit posée (et testée) l'hypothèse selon laquelle la récurtivité est la propriété du langage humain qui distingue ce dernier de toutes les formes de communication animale. Même s'il n'est pas certain que cette hypothèse soit irréfutable – D. Everett a défrayé la chronique en présentant le pirahã comme une langue humaine non réursive (2005) –, la notion est importante dans l'épistémologie de la linguistique. Elle a principalement été exploitée dans les approches les plus formalisées en syntaxe (champ sur lequel nous nous focaliserons), en morphologie (v. p.ex. ; Booiij 2007 : 76 pour une explication simple sur les composés ou Corbin 1987 : en partic. 496-500) en sémantique formelle (voir p.ex. Roussarie 2017 : 77-79 sur la grammaire de Montague). Il semblerait par contre que le concept soit peu adapté aux questions phonologiques (Neeleman et van de Koot 2006). Dès les années 1950, poursuivant l'objectif de construire un modèle capable de générer tous les énoncés d'une langue, Chomsky a montré par des moyens formels que la récurtivité permet de rendre compte du potentiel combinatoire illimité des signes linguistiques. De nombreux linguistes ont reformulé ce postulat après lui :

La réursion est un fait fondamental des langues naturelles : c'est la réursion qui permet de construire, à partir du stock nécessairement fini de signes linguistiques (dont dispose une langue donnée), un nombre infini d'expressions. (Mel'čuk et Milićević 2014 : vol. 2, p. 18)

Sans nous prononcer sur l'irréfutabilité de cette propriété du langage humain, nous expliquerons en quoi consiste la récurtivité en syntaxe et nous présenterons une partie des questions théoriques et pratiques qu'elle pose. Nous commencerons par introduire la notion de manière intuitive, avec des exemples qui ne sont pas des exemples linguistiques (§ 2). La section suivante (§ 3) sera consacrée à la manière dont la récurtivité se manifeste en syntaxe. L'exposé se focalisera sur deux logiques archétypales d'analyse des combinaisons de mots : l'approche « en constituants » et l'approche dépendancielle – qui, dans la pratique des modèles disponibles, sont souvent combinées. Nous montrerons ensuite (§ 4) ce que la nature formelle de la récurtivité implique au niveau de la manière dont on peut encoder l'analyse des constructions – ce terme sera entendu ici de manière très générale au sens de “signe décomposable d'un point de vue syntaxique”. Cette section sera l'occasion de distinguer les notions de *récurtivité* et d'*itérativité* en

examinant notamment le cas de la coordination. La dernière section examinera comment se manifeste la récursivité dans le corpus *French Treebank*, principalement pour déterminer les limites du potentiel récursif des constructions (§ 5).

2. PREMIÈRES APPROCHES

Avant de nous occuper de linguistique, nous proposons d'introduire la notion au moyen d'exemples intuitifs (§ 2.1) et formels simples (§ 2.2).

2.1. Approche intuitive

Il est plus aisé d'appréhender la récursivité en se focalisant tout d'abord sur des phénomènes qui ne sont pas linguistiques. Un des cas de récursivité les plus connus est celui des matriochkas (poupées russes). Ce jouet est constitué d'un ensemble de poupées imbriquées les unes dans les autres de manière à ce que chaque poupée s'ouvre en deux et contienne une poupée plus petite, qui elle-même contient une poupée plus petite, et ainsi de suite, jusqu'à atteindre la poupée la plus petite, qui ne s'ouvre pas et ne contient pas d'autre poupée. La Fig. 1 représente des matriochkas.



Figure 1. – Photographie de [matriochkas](https://commons.wikimedia.org/wiki/File:Russian-Matroschka_no_bg.jpg)
(Gnomz007~commonswiki, 17 Juillet 2005
[\[https://commons.wikimedia.org/wiki/File:Russian-Matroschka_no_bg.jpg\]](https://commons.wikimedia.org/wiki/File:Russian-Matroschka_no_bg.jpg)).

La plupart du temps, il serait possible de confectionner une poupée un peu plus petite que cette poupée finale et de continuer selon le même principe. Toutefois, il est nécessaire de s'arrêter à un moment car les poupées deviendraient trop petites pour être manipulées. D'autre part, même en poussant la démarche dans ses limites les plus extrêmes, on atteindrait finalement le niveau de la molécule, puis de l'atome, niveaux auxquels il est absurde d'encore parler de poupée.

Un autre exemple de récursivité est observable dans l'iconographie publicitaire du fromage *La vache qui rit* : les boîtes de fromage sont illustrées par une tête de vache qui porte des boucles d'oreilles en forme de boîtes de fromage illustrées de la même manière. On peut ainsi voir les motifs de la tête de la vache, les boucles d'oreille et les boîtes de fromage se répéter (Fig. 2). La répétition serait infinie si, atteignant une taille très petite, le motif ne devenait impossible à distinguer (ce qui arrive assez rapidement).



Figure 2. – Photographie de boîtes de fromage « La vache qui rit »
(Flavio Ensiki, 6 Avril 2006
[https://commons.wikimedia.org/wiki/File:La_vache_who_ris.jpg]).

Ces deux exemples permettent de comprendre ce qu'est une structure récursive (on parle aussi de « mise en abyme »). Les poupées contiennent des poupées, les représentations de boîtes de fromage comprennent des représentations de boîte de fromage. Dans les deux cas, il y a une relation d'inclusion plus ou moins directe qui unit deux occurrences réalisant un unique type d'unité. On parle d'*auto-référence*. On ne peut décrire complètement la poupée la plus grande sans faire appel au type « poupée ». Il y a là une forme inévitable de circularité, qui forme une boucle potentiellement infinie.

Toutefois, cette circularité n'est pas « grave » (pour reprendre les mots de Douglas Hofstadter 2000 : 149), notamment parce que la boucle est limitée. Cette limitation peut être complètement interne au système. Ainsi, qu'il soit question de poupées ou de dessins de boîtes de fromage, l'objet décrit atteint une sorte de point de rupture de la boucle : la poupée la plus petite ne contient pas d'autre poupée. Dans ce genre de cas, le système fixe ses propres limites et s'arrêtant après un certain nombre de boucles. Toutefois, la limitation peut également dépendre de contraintes extérieures au système : en l'occurrence, il peut arriver un moment où les poupées deviennent trop petites et où aucun humain ne serait capable de les manipuler pour les ouvrir, ou que le dessin de boîte de fromage soit trop petit pour permettre à un humain de distinguer un dessin similaire de moindres dimensions. Les limitations biomécaniques engendrent alors en pratique l'échec de la relation d'auto-référence.

Le même principe d'auto-référence et le même genre de limitation de la boucle potentiellement infinie permet, comme nous allons le voir, de modéliser certaines structures syntaxiques.

2.2. *Approches mathématique et informatique « naïves »*

D'un point de vue mathématique et procédural, la récursivité est un élément essentiel à la démonstration de la calculabilité des algorithmes. Dans cette optique, une fonction est dite *récursive* si elle se fait référence à elle-même (on parle d'*auto-référence*). Un exemple classique de fonction récursive est celle qui permet de calculer naïvement la suite de Fibonacci¹. Il s'agit d'une suite infinie d'entiers dont chaque terme est la somme des deux entiers qui le précèdent dans la suite (Knuth 2013 : 79 sv.) ; le début de la suite est donc :

1 Du surnom du mathématicien italien Leonardo Pisano qui l'a discutée en 1202.

0, 1, 1, 2, 3, 5, 8, 13, 21, etc.

Une fonction F qui permettrait de calculer l'entier à la position n dans la suite serait la suivante : si $n = 0$, $F_n = 0$; si $n = 1$, $F_n = 1$; et si $n > 1$, avec $F_0 = 0$ et $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$. L'entier dans la position n a ainsi la valeur de la somme de l'entier dans la position $n - 1$ et de l'entier dans la position $n - 2$. Pour calculer F avec $n > 1$, on mobilise la fonction elle-même, de manière similaire aux matriochkas imbriquées. Il y a donc *auto-référence* et, par conséquent, récursivité. Les valeurs de la suite de Fibonacci peuvent ainsi être calculées sans limite supérieure.

D'un point de vue pratique, notamment dans le domaine de l'informatique, il est généralement nécessaire de définir une limite aux algorithmes récursifs : de la même manière que la plus petite matriochka ne contient pas d'autre poupée pour des raisons concrètes, les algorithmes récursifs sont limités par les contraintes logicielles et matérielles de l'environnement où on les exécute. Ainsi, dans le célèbre langage de script *python*, on peut implémenter la fonction (au sens informatique) f qui calcule les valeurs des entiers de la suite de Fibonacci selon leur position n comme suit (le code, simple, est glosé en italiques à droite) :

```
def f(n) :                               # définir la fonction « f », avec l'argument « n »
    if n == 0 : return 0                  # si n vaut 0, renvoyer la valeur 0 et arrêter
    if n == 1 : return 1                  # si n vaut 1, renvoyer la valeur 1 et arrêter
    return f(n - 1) + f(n - 2)           # renvoyer la somme de f(n - 1) et f(n - 2) et arrêter
```

L'interpréteur traitant un appel à cette fonction opérera sans problème pour les petites valeurs de n . Par contre, pour les valeurs élevées, la complexité impliquée par l'auto-référence entraînera une expansion exponentielle et déraisonnable tant du temps de traitement que de la mémoire mobilisée. La limite de récursivité doit donc être définie de manière interne à l'algorithme, par exemple à l'aide d'une condition d'arrêt, de manière à ce que le système ne tombe pas à court de ressources et le message d'erreur *maximum recursion depth exceeded* "profondeur de récursivité maximale dépassée" est défini par défaut dans l'interpréteur². De ce fait, un appel à la fonction f provoquera une erreur si n est trop élevé, alors qu'aucun obstacle purement logico-mathématique ne s'oppose à son exécution. La cause de ce genre d'arrêt est alors externe au système, comme l'est l'impossibilité de manipuler des poupées devenues trop petites dans l'exemple précédent.

3. RÉCURSIVITÉ ET SYNTAXE

Noam Chomsky est souvent cité comme une référence importante sur la question de la récursivité dans le langage (voir notamment Chomsky 1964 : 86-92), car il fait usage de grammaires formelles pour modéliser la syntaxe dans une optique générative. La grammaire d'une langue serait censée permettre la génération de toutes les phrases possibles. Selon l'approche syntaxique classique, la récursivité est ce qui rend justement possible la création d'une infinité d'énoncés différents dans toute langue. On peut illustrer cela à l'aide de la subordination des propositions : les phrases peuvent en contenir d'autres, qui elles-mêmes peuvent en contenir d'autres, et ainsi de suite.

(1) (a) Nous espérons que *le nouveau pouvoir reconnaîtra qu'une implication (...) offre les meilleures chances (...)* (SUD-FR, id. fr-ud-train_01126)

2 <<https://docs.python.org/3/library/exceptions.html>>

- (b) Le nouveau pouvoir reconnaîtra qu'*une implication offre les meilleures chances*.
- (c) Une implication offre les meilleures chances.
- (d) Je sais que *nous espérons que le nouveau pouvoir reconnaîtra qu'une implication offre les meilleures chances*.

La phrase (1a) illustre ce genre de phénomène. La phrase (1b), qu'elle contient (en italiques), pourrait être employée de manière autonome. Elle contient elle-même la phrase (1c) (en italiques). On pourrait aller plus loin encore et intégrer (1a) à une phrase plus large, comme (1d). Chaque phrase contient une phrase de moindres dimensions, sauf (1c) qui est la plus petite, exactement comme les matriochkas de notre exemple introductif.

Les opérations spécifiques à la méthodologie linguistique permettent de mettre en évidence que la récursivité est présente dans le langage, indépendamment de la formalisation choisie pour le décrire (nous revenons sur ce point sous § 4). Ainsi, l'inclusion de la phrase *Une implication offre les meilleures chances* dans (1b) est démontrée en (2) par les tests de pronominalisation, de pseudo-clivage (EGF 3.1.4.2, « Constructions pseudo-clivées » : en partic. § 3.1.2) et de fragment-réponse.

- (2) (a) Le pouvoir le reconnaîtra/Le pouvoir reconnaîtra ça.
- (b) Ce que le pouvoir reconnaîtra, c'est qu'une implication offre de meilleures chances.
- (c) Qu'est-ce que le pouvoir reconnaîtra ? Qu'une implication offre de meilleures chances.

Pour comprendre comment les syntacticiens rendent compte de ce phénomène, nous allons voir comment la récursivité se manifeste, de manière assez similaire, au travers des structures arborescentes qui encodent les relations « partie-tout » (méréologiques) de l'analyse en constituants immédiats (§ 3.1) et les dépendances de la syntaxe dépendancielle (§ 3.2). La présentation est ici volontairement focalisée sur ces deux archétypes de description syntaxique, en commençant par l'analyse en constituants immédiats car, historiquement, c'est pour ces dernières que la question de la récursivité a d'abord été explorée de manière approfondie. La plupart des approches syntaxiques concurrentes dérivent en effet une partie de leurs principes de ces deux manières de concevoir les relations entre les termes en présence : soit par conceptualisation de relations entre les mots (approche dépendancielle), soit par identification de relations « partie-tout » (approche en constituants). Il est vrai que certains modèles prennent l'approche en constituants comme base explicite (p.ex. Grammaire de construction « radicale » de Croft 2001 : ch. 6), alors que d'autres revendiquent un ancrage dépendanciel – p.ex. Théorie Sens-Texte (Mel'čuk et Milićević 2014), Word Grammar (Hudson 2010). Toutefois, de nombreux modèles combinent volontiers les deux logiques, comme la *Lexical Functional Grammar* (Bresnan et al. 2016), la *Head-Driven Phrase-Structure Grammar* (Pollard et Sag 1994 ; voir notamment Kahane 2009), et autres « grammaires d'unification ». Du point de vue de la distinction entre les deux grandes options descriptives, il n'est pas impératif que la théorie soit homogène. C'est essentiellement dans le but d'exposer de manière simple les principes de la récursivité que nous mettons de côté les approches mixtes.

3.1. Analyse en constituants immédiats

Voyons comment la récursivité peut être modélisée à l'aide de l'analyse en constituants immédiats³ (ACI ; v. notamment Ruwet 1968 : 117-118 ; Touratier 2005). Pour commencer, nous rappellerons très brièvement les principes de l'ACI. Nous décrirons ensuite les cas intuitivement perçus comme présentant des signes de récursivité. Nous verrons que ces cas ne sont pas les plus simples du point de vue strictement formel, car ils présentent un type de récursivité dite « indirecte ».

L'ACI repose sur l'identification et la classification de groupes de mots en fonction de leurs propriétés distributionnelles⁴. Ces propriétés sont mobilisées pour construire des relations « partie-tout » (relations méreologiques) entre des constituants abstraits et leurs constituants immédiats. Dans l'approche classique⁵, la phrase P est décrite comme une construction qui contient deux constituants immédiats⁶ : un syntagme nominal SN et un syntagme verbal SV. Cela peut être modélisé par la règle notée (3), où l'opérateur « → » se lit « est réécrit en ». Cela traduit en un algorithme génératif une relation « partie-tout ».

$$(3) \quad P \rightarrow \text{SN SV}$$

Dans (1b), *le pouvoir* correspond au SN de cette règle, alors que l'ensemble *reconnaîtra qu'une implication offre les meilleures chances* correspond au SV (on peut d'ailleurs faire commuter ce dernier avec un simple verbe, comme *choisira* ou *décidera*).

3.1.1. Principe général de récursivité

La construction *une implication offre les meilleures chances*, qu'on peut définir comme une phrase P (elle-même constituée d'un SN et d'un SV), est contenue dans le SV *reconnaîtra qu'une implication offre les meilleures chances*, qui lui-même fait partie d'une autre phrase plus large. On peut inscrire cela de manière simplifiée dans un diagramme ensembliste, le mieux à même de rendre compte des relations d'inclusion entre les deux P (Fig. 3).

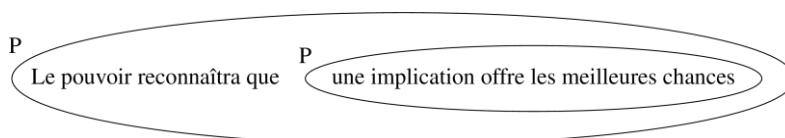


Figure 3. – Représentation ensembliste simplifiée de la subordination d'une proposition.

3 Nous concentrerons notre exposé sur l'examen d'analyses ne requérant pas une présentation théorique approfondie et tenues pour « classiques » en grammaire du français. Les écoles théoriques fondées sur l'ACI ont depuis longtemps complexifié l'appareil représenté ici (voir notamment § 4.2).

4 Il n'est pas nécessairement évident que les tests décrits dans la littérature scientifique depuis des dizaines d'années montrent vraiment une structure de constituants. Voir le volume 5 de *Language under discussion* (2018), qui est consacré à un débat sur la question à la suite de Osborne 2018.

5 Touratier 2005 ; cette approche mobilise les modélisations de type X-barre qui sont ensuite devenues dominantes.

6 Équivalent francophone du « S » de Chomsky 1965, précédemment noté en toutes lettres « Sentence ».

On voit dans le diagramme qu'une structure représentée par une bulle, étiquetée « P » et contenant *une implication offre de meilleures chances*, se trouve à l'intérieur de la bulle la plus large, également étiquetée « P ». Il est donc impossible de décrire exhaustivement le contenu de cette dernière sans mentionner une bulle plus petite étiquetée de la même manière. La description de P fait référence à P ; il y a donc *auto-référence*. Techniquement, on pourrait dire que P *fait partie de la dérivation de P*, ou que P se *domine* lui-même (Ruwet 1968 : 130). Le principe est identique à celui observé pour les poupées russes, où, pour décrire une poupée, il faut dire qu'elle contient une poupée (§ 2). La Fig. 4 (le symbole P figure en gras) correspond à une inscription classique de la structure de constituants sous forme d'arbre dit « enraciné » (D = déterminant ; GN = SN sans article, cf. Touratier 2005 : 117).

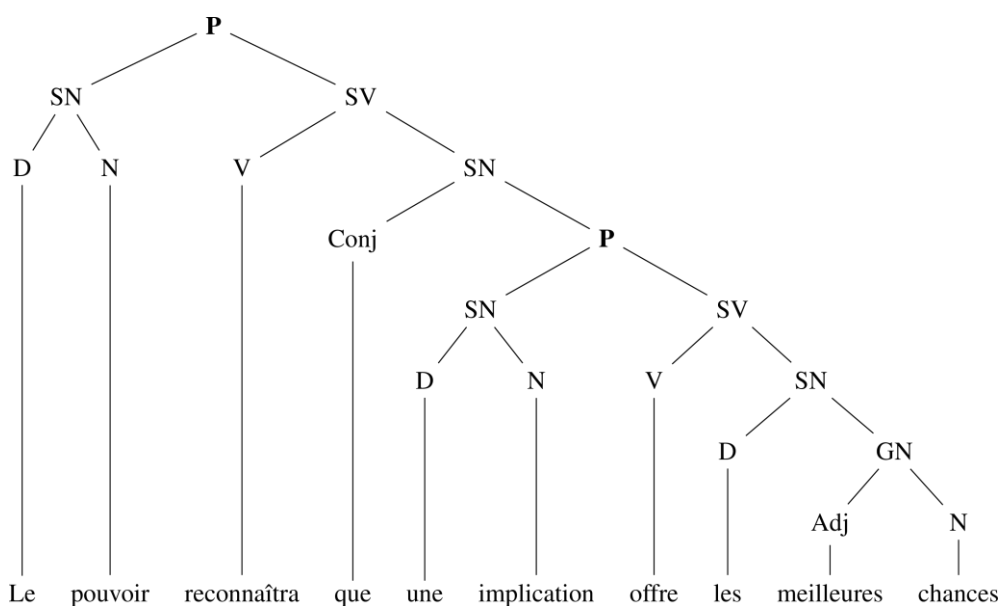


Figure 4. – Arbre ACI comportant une proposition subordonnée.

Pour bien comprendre les implications formelles de la récursivité en syntaxe, nous avons besoin du concept d'*arbre enraciné*⁷, dont nous résumons les caractéristiques principales en cinq points :

- Un arbre enraciné est une structure formelle constituée de *nœuds* et d'*arcs* (arêtes orientées) reliant ces nœuds deux à deux pour former un couple de sommets ordonnés. Cet ordre implique que la relation est asymétrique.
- L'arbre est une *structure connexe* car il existe un chemin continu suivant les arêtes entre un nœud et n'importe quel autre.
- L'arbre est une *structure acyclique* car il n'existe qu'un seul chemin d'un nœud à un autre.

7 Voir l'article « Tree (data structure) » sur *Wikipédia* <[https://en.wikipedia.org/wiki/Tree_\(data_structure\)](https://en.wikipedia.org/wiki/Tree_(data_structure))>. Voir également Mel'čuk et Milićević 2014: vol. 3, annexe I concernant les concepts formels utiles en linguistique.

– Dans un arbre enraciné, un (et un seul) des nœuds, nommé *racine*, n'est le second élément d'aucun arc.

– S'il n'est pas la racine de l'arbre pris dans son ensemble, tout nœud d'un arbre enraciné forme la racine d'un *sous-arbre complet* similaire à arbre complet à ceci près que sa racine dépend d'un autre nœud (Mel'čuk et Milićević 2014: vol. 3, p. 296).

Dans la modélisation dominante de l'ACI, l'analyse est encodée dans un arbre enraciné – nous mettons ici de côté la question de l'ordre des mots (voir toutefois § 3.1.3). Dans l'approche classique, nous avons vu que le symbole de départ était P. Ce dernier est ainsi placé en position de racine dans la Fig. 4. Les traits correspondent aux arcs orientés : l'extrémité supérieure correspond au premier sommet du couple, l'extrémité inférieure au second. Ces arcs expriment toujours un rapport d'inclusion et peuvent être lus de haut en bas « X contient Y » ou de bas en haut « Y fait partie de X ». La relation ainsi encodée est *transitive* au sens formel (si X contient Y et que Y contient Z, alors X contient Z). Pour le cas qui nous occupe, on peut suivre ces arcs de haut en bas vers la droite. On voit que la racine est connectée indirectement à un sous-arbre dont la racine est également un nœud P, ce qui permet de visualiser que la structure utilise une règle récursive.

En inscription algébrique (règles de réécriture), la même structure est représentée par (le symbole P figure en gras) :

- (4) **P** → SN SV
SN → D N
D → *le*
N → *pouvoir*
SV → V SN
V → *reconnâtra*
SN → Conj **P**
Conj → *que*
P → SN SV
SN → D GN
D → *une*
N → *implication*
SV → V SN
V → *offre*
SN → D GN
D → *les*
Adj → *meilleures*
N → *chances*

C'est l'existence d'une règle de réécriture $SN \rightarrow Conj P$ dans la grammaire qui permet la récursivité dont nous parlons. Mobiliser cette même règle un nombre infini de fois permet en théorie de créer une boucle sans fin. Toutefois, il n'y a jamais véritablement d'imbrication infinie de structures par récursion (ce point sera davantage développé sous § 5). Cela est modélisé par l'existence de règles de réécriture alternatives qui arrêtent la boucle récursive (tel un point d'arrêt dans un algorithme). Ainsi la règle (5) dit qu'un SV peut consister en un simple verbe, ce qui empêche la présence d'une complétive P. C'est grâce à l'application de règles de ce type que l'on peut achever la boucle, de la même manière que la plus petite des poupées d'un jeu de matriochkas ne s'ouvre pas (§ 2).

- (5) SV → V

En guise d'exemple supplémentaire, le locuteur natif du français aura également l'intuition de phénomènes récurrents dans l'exemple suivant.

- (6) le lancement d'une formation des enseignants du supérieur (FTB, texte 1116, phrase 137)

L'enchâssement des compléments déterminatifs en *de* dans le SN (6) est rendu possible par l'existence des règles récursives suivantes (SP correspond au syntagme prépositionnel):

- (7) SN → Art GN
GN → N SP
SP → Prép SN

Puisque la règle qui décrit la structure de SP contient un SN et qu'un SN peut contenir un GN qui peut contenir un SP, le système est récursif. De la même manière que pour les subordinées, c'est l'existence de règles alternatives qui permet de modéliser un SN qui s'achève. La règle (8) correspond à la combinaison d'un article et d'un nom, ce qui empêche la présence d'un complément déterminatif.

- (8) SN → Art N

Envisagée de la sorte, ce n'est pas la phrase (ni une quelconque structure) qui est récursive, mais les règles de la grammaire qui peuvent être mobilisées pour la construire : dès lors que, pour décrire une structure d'un certain type, il est possible de mentionner une structure du même type, il y a *auto-référence* et on est en face de règles grammaticales récursives.

3.1.2. Récursivité directe vs récursivité indirecte

D'autre part, on peut distinguer la récursivité *directe* de la récursivité *indirecte*. Dans les règles permettant la récursivité directe, le même symbole est employé à la fois à gauche et à droite de l'opérateur « → ». Dans les cas de récursivité indirecte, il faut continuer à développer plusieurs niveaux de dérivation pour retrouver le même symbole que le symbole de départ d'une règle. Il n'est en effet pas nécessaire qu'un symbole se domine directement lui-même pour qu'on parle de récursivité (Ruwet 1968 : 130). Les cas de la complétive et du complément déterminatif ci-dessus sont fréquemment modélisés à l'aide de dérivations présentant une récursivité indirecte. Par contre, une apposition comme dans (9) peut être modélisée à l'aide de la récursivité directe, par une règle de type (10) – ce genre de règle postule que l'apposition est intégrée à un SN et ne constitue pas un « détachement » lié à un actant (Neveu 2013 : 229).

- (9) La sculpture, *art de l'éternité*, défie le temps en recherchant des matériaux indestructibles (Tournier *apud* Benetti et Berrendonner 2012 : 287)

- (10) SN → SN GN

La règle (10) permet de décrire la combinaison de *la sculpture*, SN complet et pouvant assumer la position de constituant immédiat de P, avec le GN *art de l'éternité*, qui ne pourrait occuper cette place en raison de l'absence d'article. L'ensemble forme un SN complet qui contient donc directement un SN complet de dimensions moindres. Le même symbole peut éventuellement être employé plusieurs fois à droite de la flèche, comme on

le voit dans la première ligne de l'analyse (12) du SN (11) – repris de Monneret (1999 : 242-244).

(11) Pierre, un homme de confiance

(12) SN → SN SN
 SN → *Pierre*
 SN → Art GN
 etc.

3.1.3. Récursivité et ordre des constituants

Question fondamentale de la syntaxe dès les origines, la description de l'ordre des mots est une préoccupation qui apparaît très tôt dans les travaux portant sur l'ACI (Bloomfield 1933 : 197-201, Wells 1947 : § 55). La conséquence en termes de modélisation est que les arbres enracinés dans lesquels les ACI sont encodées sont généralement ordonnés. En toute rigueur, le formalisme tel que nous l'avons décrit jusqu'ici est insuffisant : les nœuds terminaux de l'arbre doivent être ordonnés dans la structure, ce que l'on représente diagrammatiquement par l'alignement des mots sur un axe similaire à celui de l'écriture. Si l'ordre des mots est encodé, il est pertinent de s'interroger sur les relations entre cet ordre des mots et la récursivité : cette dernière se manifeste-t-elle à la périphérie gauche ou droite, ou bien au milieu du constituant ? On parle alors respectivement de *récursivité à gauche*, de *récursivité à droite* et d'*emboîtement* (angl. *center-embedding*). Ruwet (1968 : 130-134) illustre la récursivité à droite à l'aide de compléments déterminatifs récursifs, comme dans (6). Il souligne que le français favorise la récursivité à droite – bien que la récursivité à gauche puisse s'y manifester, notamment en cas de cumul d'épithètes antéposées (13) –, tandis que l'anglais favorise la récursivité à gauche (14).

(13) un bon vieux petit bistrot

soit un_{GN}[bon_{GN}[vieux_{GN}[petit N[bistrot]]]]
 (les limites des constituants sont ici représentées par des crochets étiquetés du nom de leur classe syntagmatique)

(14) flue pipe support strap [littéralement 'sangle de support du conduit de fumée']

soit :_{GN}[flue_{GN}[pipe_{GN}[support N[strap]]]]

L'emboîtement, problématique, est décrit comme suit par Ruwet :

Il y a emboîtement chaque fois qu'une suite B « est entièrement incluse dans une suite A, avec quelque élément non-nul à sa gauche et quelque élément non-nul à sa droite » (Chomsky 1965 : 12) (c'est le cas où A = XBY, X et Y étant non-nuls). L'emboîtement (*nesting*) est donc un cas général dont l'auto-enchâssement n'est qu'un cas particulier. (1968 : 133)

Observons (15) : la hiérarchie de l'emboîtement est représentée par des crochets dans (15b).

(15) (a) Le rat que le chat que le chien a chassé a tué a mangé le fromage (Ruwet 1968 : 132)
 (b) Le rat [que le chat [que le chien a chassé] a tué] a mangé le fromage

Nicolas Ruwet souligne les limitations du cerveau humain associées à l'emboîtement :

J'ai indiqué que, dans la mesure où un modèle de performance repose, notamment, sur une mémoire finie, il y aura un grand nombre de phrases parfaitement grammaticales qui, au niveau de la performance, seront inacceptables ou incompréhensibles (comme c'est déjà largement le cas de la phrase [(15)] ci-dessus). Or, il semble bien qu'un organisme doté de mémoire finie est parfaitement capable de s'accommoder de la récursion de gauche et/ou de droite, c'est-à-dire de produire et/ou de comprendre un nombre indéfini de phrases comportant des éléments récursifs à gauche et/ou à droite en nombre indéfini. En revanche, on sait qu'un organisme doté d'une mémoire finie (comme les machines à calculer et comme l'organisme humain) est incapable de comprendre ou d'émettre des phrases comportant un trop grand degré d'*emboîtement* (cf. Miller et Chomsky 1963 : 468) [...] (Ruwet 1968 : 133)

En d'autres termes, non seulement les langues peuvent être caractérisées en fonction de la position des constituants récursifs (en lien évident avec la position relative du déterminant et du déterminé), mais il semblerait que, quelle que soit la langue, les structures emboîtées soient nettement plus rares, en raison de limitations cognitives comparables aux limitations de mémoire que nous avons mentionnées pour les programmes informatiques (§ 2.2).

3.2. Analyse dépendancielle

Dans les formalisations dépendancielles, la récursivité apparaît au niveau des relations de dépendance. L'existence de ces dernières repose sur l'hypothèse, non incompatible avec celle de l'ACI bien que souvent opposée dans les débats épistémologiques, que les mots sont liés entre eux par des rapports grammaticaux abstraits. On identifie et on classe ces derniers en fonction du comportement morphosyntaxique des unités : leur rapport à la pronominalisation, leur comportement par rapport au clivage, leurs propriétés distributionnelles, etc. (voir p.ex. Mel'čuk 2009). Dans les pratiques dominantes de ce type d'analyse, les arbres représentant le réseau que forment les dépendances sont également des arbres enracinés (§ 3.1), mais la manière dont le formalisme est employé pour encoder l'analyse diffère de l'ACI. Ainsi, (i) tout nœud, y compris la racine, correspond à un mot ; (ii) les arcs relient des couples de mots correspondant (dans cet ordre) au gouverneur et à un de ses dépendants ; (iii) les arcs sont étiquetés du nom de la relation syntaxique (« fonction ») qu'ils servent à encoder. Dans cette optique, la plupart des grammaires de dépendance placent le verbe en position de racine de l'arbre. Les théories dépendancielles abordent la plupart du temps l'ordre des mots dans un module séparé de celui de la description des relations syntaxiques, parfois selon des termes assez proches de ceux des analyses en constituants (Mel'čuk et Milićević 2014: vol. 2, pp. 215-227 ; Gerdes et Kahane 2006). Nous ne reviendrons donc pas sur ce point dans cette sous-section.

Nous allons montrer que la récursivité se manifeste au niveau des dépendances (§ 3.2.1) et au niveau des classes de mots hiérarchisés (§ 3.2.2).

3.2.1. Récursivité et sous-arbres de dépendances

Le diagramme à gauche dans la Fig. 5 représente une analyse simplifiée de (1b), telle que la Théorie Sens-Texte pourrait proposer (Mel'čuk et Milićević 2014 ; Kahane 2001) – les étiquettes ont la valeur suivante : *subj* = relation subjectale ; *obj-dir* = relation objectale directe ; *cj-sub* = relation conjonctionnelle de subordination ; *dét* = relation déterminative ; *mod* = relation modificative.

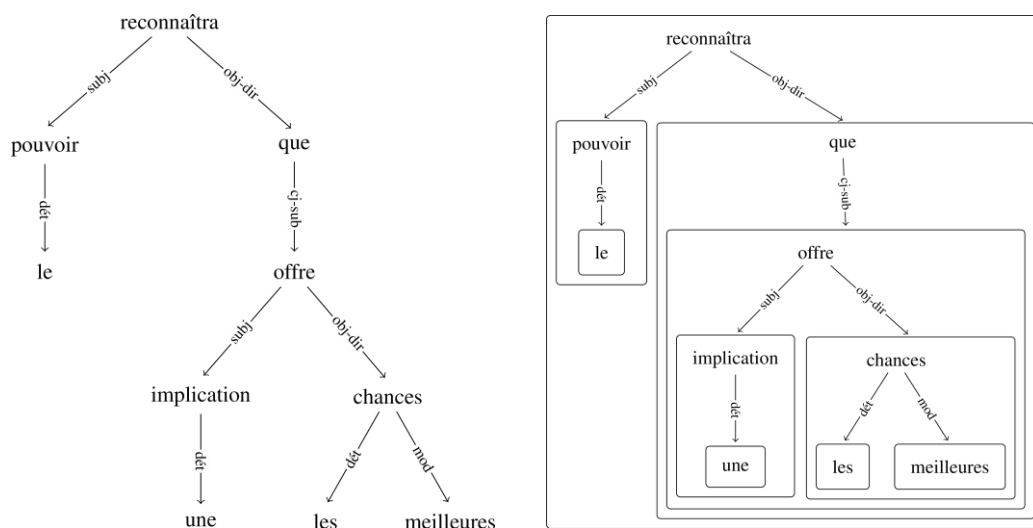


Figure 5. – Arbres dépendanciels comportant une proposition subordonnée (premièrement, selon la Théorie Sens-Texte ; deuxièmement, sous la forme d’un arbre de Beuzée-Glaadkij)

On peut en outre considérer que chaque nœud dépendant constitue le pivot par lequel l’ensemble du sous-arbre qu’il domine dépend de son gouverneur. Cela peut être encodé dans une structure que S. Kahane nomme « arbre de Beuzée-Glaadkij »⁸. Le second diagramme de la Fig. 5 représente l’analyse dépendancielle de (1b) inscrite dans un arbre de ce type. L’arbre entier, ainsi que chaque sous-arbre complet (même réduit à un seul nœud), est entouré d’un rectangle aux angles arrondis. Comme on le voit, les arcs sont orientés d’un mot (gouverneur) à un sous-arbre complet. Comme la racine de ce dernier est un mot unique, il est formellement équivalent de dire que le dépendant est le sous-arbre ou sa racine. La représentation de Beuzée-Glaadkij met en évidence que des rapports d’inclusion assez similaires à ceux observés pour les ACI apparaissent dans les arbres de dépendance⁹. Il en résulte que la description des sous-arbres de dépendance complets peut être rapprochée de celle des constituants (la relation « être un sous-arbre de » est également transitive au sens formel). La description exhaustive d’un sous-arbre pourra également faire intervenir une relation dépendancielle du même type que celle dans laquelle la racine du sous-arbre est observée. Dans la Fig. 6, les deux dépendances indiquées en gras sont du même type.

8 Des noms respectifs du grammairien qui en a eu l’intuition et du linguiste qui l’a dessinée pour la première fois (Kahane 2020 : 93).

9 Les inclusions sont cependant organisées de manière différente, étant donné qu’une bulle dépendra toujours d’un nœud constitué d’un nœud unique, alors qu’en ACI, une construction n’a pas nécessairement un constituant terminal.

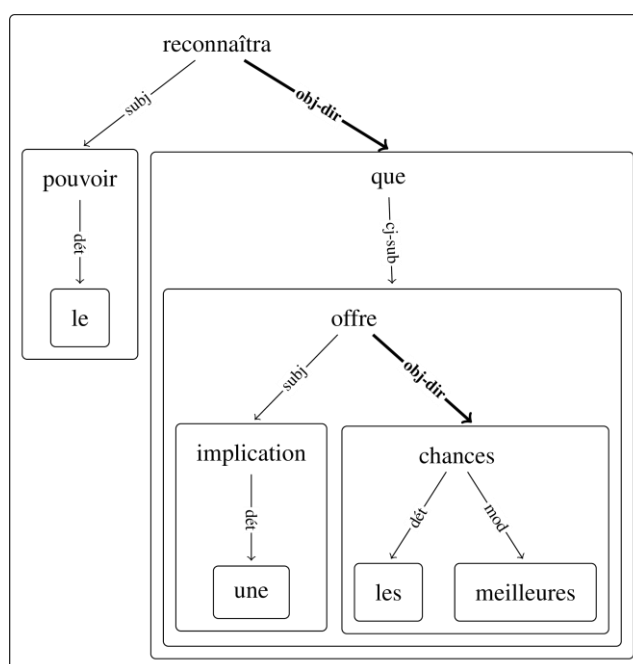


Figure 6. – Arbre de Beauzée-Glaadkij comportant une structure récursive observée au niveau des dépendances.

Pour décrire complètement le sous-arbre sous la dépendance *obj-dir* supérieure, il est nécessaire de faire référence au même type de relation. Il y a auto-référence, et donc récursivité, au niveau des relations de dépendance (arêtes de l'arbre), comme il y a récursivité au niveau des constituants (nœuds de l'arbre) en ACI¹⁰. Les mêmes distinctions entre la récursivité directe et indirecte sont aisément transposables.

3.2.2. Récursivité et classes de mots

La récursivité se manifeste au niveau des dépendances, mais également au niveau de la classe des mots impliqués dans les relations (Matthews 1981 : 81). Cela est particulièrement clair dans les règles de combinaison des mots avec leurs dépendants (ou *valence*). Dans les grammaires dépendanciennes, chaque mot appartient à une classe et est associé à un certain régime : quelles autres classes de mots peuvent dépendre de lui et quelle relation de dépendance les unit (v. p.ex. Mel'čuk et Milićević 2014 : vol. 1, pp. 281-283). La phrase est ainsi définie comme une structure dont le terme principal appartient à la classe des verbes et les règles grammaticales stipulent : (i) qu'un verbe peut avoir un mot appartenant à la classe des conjonctions qui dépend de lui ; (ii) qu'un mot appartenant à la classe des conjonctions de subordination a nécessairement un mot appartenant à la classe des verbes qui dépend de lui. Il en résulte une chaîne de dépendance comme (16) – les flèches représentent ici des dépendances. Dans ce genre de cas, on voit qu'un sous-arbre dont la racine est un verbe peut contenir un autre verbe.

(16) verbe → conjonction de subordination → verbe

¹⁰ Il est possible de montrer que cette distinction entre nœuds et arêtes est moins importante qu'il n'y paraît (Kahane et Mazziotta 2015).

La récursivité est également une propriété impliquée dans le positionnement de groupes syntaxiques plus ou moins complexes (Mel'čuk et Milićević 2014: vol. 2, pp. 17-18, divers procédés de mise en évidence ici rendus par l'italique) :

Lors de la synthèse d'une phrase, les syntagmes sont construits récursivement par l'emploi itéré de règles syntaxiques, et deviennent des *groupes syntaxiques*, séquences de mots-formes manipulées comme un tout par les règles d'ordre des mots. Par exemple, les syntagmes simples qui correspondent aux sous-arbres LA ← FILLE, PETITE ← FILLE et À → FILLE sont linéarisés ensemble pour construire le groupe prépositionnel à la *petite fille*, qui doit être positionné comme un tout par rapport au verbe gouverneur. De la même façon, on réunit des groupes syntaxiques en propositions et des propositions en phrases, toujours par *récursion*.

Ici encore, il y a lieu de parler de récursivité d'une manière similaire à celle observée pour l'ACI : c'est le positionnement de constructions dans des ensembles plus larges qui est considéré.

4. EFFET DES FORMALISMES

Dans cette section, nous commencerons par souligner les choix qui nous paraissent affecter la notion de récursivité dans la manière dont les syntacticiens s'approprient les formalismes (§ 4.1). Nous présenterons différents cas où la récursivité est clairement générée par les contraintes associées à la structure de la modélisation choisie (§ 4.2). Nous montrerons ensuite que la complexité de la récursivité (directe ou indirecte), voire son existence pour certaines classes, est liée à la hiérarchisation *a priori* de ces dernières (§ 4.3). Dans un quatrième temps, nous parlerons de la distinction entre la récursivité et l'*itérativité*, considérée comme une propriété formelle « concurrente » qui peut être exploitée pour distinguer des constructions de natures différentes (§ 4.4)

4.1. Du formalisme pur à la description linguistique

La théorie linguistique façonne ses modèles formels en fonction de considérations théoriques qui donnent du sens aux objets formels, qui sont par nature dépourvus de valeur sémiotique. La principale question qui découle de cette utilisation d'objets formels comme outils de description est celle des limites de leur isomorphisme, au sens où les modélisations formelles sont iconiques des relations qu'elles encodent et en reproduisent l'organisation. Il est délicat de déterminer si la récursivité observable au niveau de la structure construite par analyse correspond effectivement à une récursivité linguistique :

[T]here are multiple levels of analysis at which recursion can be measured, at least including algorithm, structure and mental representation. Analysing one of these levels does not necessarily allow inferences about the other two. For example, both recursive and non-recursive algorithms can be used to generate recursive structures. Therefore we cannot conclude that systems able to generate recursive structures necessarily use a recursive algorithm. (Martins et Fitch 2014 : 16)

Les syntacticiens s'entendent généralement sur leur objectif : modéliser de manière économique les règles de combinaison de mots ou de groupes de mots. Ce faisant, ils donnent un sens particulier aux structures arborescentes et aux graphes. Pour expliquer ce que l'on entend par *récursivité* en syntaxe, nous avons eu besoin de mobiliser le formalisme de l'*arbre enraciné*, qui sert à encoder les analyses. Or, de par sa définition mathématique, le formalisme arborescent est compatible avec la notion de *récursivité* et il est apte à l'encoder dans le cadre de la description des langues. Il devient possible de

parler de cette notion dès lors que l'objet formel qui encode la description le permet. On pourrait même aller plus loin en disant qu'on n'« observe » pas la récursivité comme un fait car elle est une part indissociable de la description et de la modélisation.

La sémiotique des arbres est fixée en fonction de l'option théorique choisie. Cette dernière est propre à l'épistémologie de la linguistique et contraint le formalisme. La récursivité se manifeste ainsi selon les termes d'une interprétation particulière de structures formelles non sémiotisées *a priori*. Ainsi, le principe de base de l'ACI est que tout constituant non terminal est décomposable en constituants plus petits : un constituant contient un constituant qui contient un constituant, etc.¹¹ De même, tout terme gouverneur d'un arbre de dépendance gouverne un autre terme qui peut également être gouverneur : un terme gouverne un terme, qui gouverne un terme, etc. (Askedal 2003 : 81). Les niveaux d'inclusion de l'ACI sont théoriquement illimités, de même que les niveaux formés par les associations successives de termes par des dépendances. D'autre part, excepté le niveau initial, tout niveau ultérieur doit être décrit en prenant pour référence le niveau précédent.

À ce niveau théorique de base viennent se greffer des contraintes supplémentaires dont le niveau de finesse et la portée dépendent encore des options théoriques choisies à l'intérieur de ces grandes approches : en grammaire dépendancielle, un Tesnière placera le verbe au-dessus de la hiérarchie de la phrase (1959), en ACI, une Haegeman plaidera pour que la phrase ne soit décomposée qu'en utilisant des dichotomies (1994 : 138-144). De même, la richesse de la classification des catégories et des relations aura évidemment un impact sur l'existence de récursivité. En effet, si les relations de dépendance ne sont pas typées, toutes sont potentiellement récursives¹². La théorie peut décréter que certaines unités ne peuvent avoir de dépendants ou de constituants immédiats, interdisant toute récursivité à un niveau inférieur. C'est le cas par exemple des mots-outils dits « subsidiaires » chez Tesnière (1959 : ch. 29).

On le comprend, si les structures de base permettent toujours la récursivité, les contraintes spécifiques à la théorie linguistique « enrichissent » les formalismes en les limitant. Se poser la question de l'adéquation de ces choix pour rendre compte des données relève évidemment de la responsabilité du linguiste. Chaque construction syntaxique demande qu'on y consacre une réflexion spécifique (et certaines génèrent des débats qui n'ont pas abouti à ce jour à l'ombre d'un consensus). Notre objectif n'est pas d'entrer dans une telle démarche évaluative, mais de comprendre comment certains choix théoriques mènent à des contraintes formelles qui affectent la présence de la récursivité.

11 Dans sa version minimaliste, le programme générativiste définit d'ailleurs la fonction de base MERGE, selon laquelle deux constituants fusionnent en un constituant plus large lui-même susceptible de se voir appliquer la même fonction (voir notamment Carnie 2011 : unit 12).

12 Ainsi les dépendances sous-spécifiées étiquetées *dep* dans le projet *Orféo* impliquent davantage de récursivité (Kahane et al. 2017 : § 3.5). Inversement, le projet *French Treebank* distingue la phrase principale (étiquetée *S*) des subordonnées (étiquetées *S-sub*, *S-rel* et *S-int* suivant les cas ; v. § 5), ce qui implique moins de récursivité si l'on ne tient pas compte de la hiérarchie des classes.

4.2. Artéfacts et stratification

Dans cette sous-section, nous nous occupons d'un problème spécifique à l'ACI : la présence de structures englobantes qui étalent les compléments de statut apparemment similaire sur plusieurs couches. Nous parlerons de la *stratification* de l'arbre de constituants (Kahane 1997 ; Kahane et Mazziotta 2015).

Observons (17) et (18). Pour que l'on puisse dire que ces phrases sont construites grâce à des règles récursives, il faut que les compléments « circonstanciels » soient considérés comme une hiérarchie de compléments.

- (17) Il étudia à Paris, dans un grand immeuble, au dernier étage, dans une chambre de bonne.
 (18) Il étudia à Paris pendant sept ans pour devenir médecin.

La perception intuitive (plutôt sémantique) est que ces compléments sont hiérarchisés dans le cas de (17). Ces compléments désignent des lieux progressivement plus précis, qu'on peut considérer comme imbriqués les uns dans les autres. Ainsi, la chambre est au dernier étage d'un grand immeuble qui se trouve lui-même à Paris. En revanche, dans (18), les compléments se trouveraient à un niveau hiérarchique identique selon le même raisonnement. Cependant, dès lors qu'il est question de modélisation syntaxique, l'intuition n'a idéalement plus sa place : le statut hiérarchique des compléments doit être déterminé sur la base de critères évaluables. En l'occurrence, tous les compléments circonstanciels de (17) et (18) sont effaçables, déplaçables, et clivables indépendamment les uns des autres. Si l'on s'en tient simplement à ces tests, leur statut est similaire – il serait éventuellement possible d'affiner la description en précisant que compléments exprimant le lieu n'ont pas la même distribution que les autres. Toutefois, pour des raisons d'économie (voir § 4.4.1), les formalisations ACI qui posent la réécriture $P \rightarrow SN\ SV\ SP$ (Monneret 1999 ; Soutet 1995) mobilisent des règles du type de (19) pour modéliser les compléments circonstanciels multiples (ce n'est évidemment pas la seule option descriptive possible).

- (19) $SP \rightarrow SP + SP$

L'application de cette règle autant de fois que nécessaire pour encoder tous les compléments SP limite le nombre de règles dont la grammaire formelle a besoin pour décrire les phrases contenant un ou plusieurs SP. Appliquer trois fois la règle (19) pour effectuer l'analyse de (17) mène à la décomposition représentée dans la Fig. 7. Chaque application de la règle crée une couche supplémentaire dans l'arbre. L'analyse résultante est ainsi *stratifiée* : des syntagmes de même statut d'un point de vue linguistique sont encodés à des niveaux différents dans l'arbre.

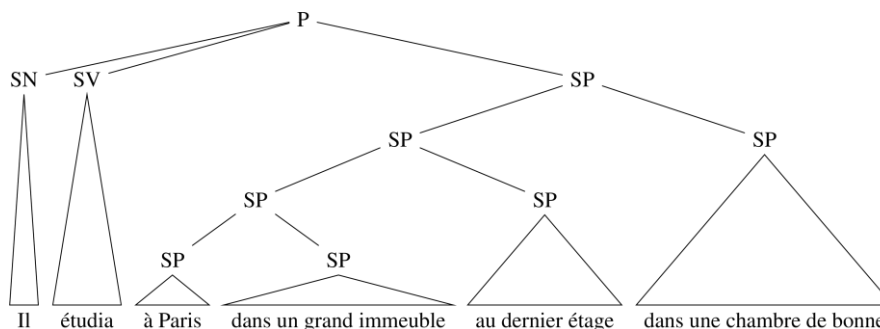


Figure 7. – Arbre ACI comportant une analyse stratifiée de SP compléments « circonstanciels ».

Le choix de ne pas distinguer entre différents types de SP aboutit à un formalisme qui mobilise une règle récursive pour encoder une concaténation d'éléments de même statut, seule solution compatible avec les contraintes formelles sélectionnées pour encoder les analyses. S'il est posé au niveau de l'axiomatique du modèle que les constructions contiennent un nombre de constituants immédiats limité, l'analyse ne pourra s'affranchir de cette limite. Les constituants « surnuméraires » ne pourront être intégrés qu'en ajoutant des strates supplémentaires. La plupart des ACI modernes ont posé la binarité de la décomposition en constituants immédiats comme une contrainte générale (Haegeman 1994 : 138-144). Dans la version très répandue de l'ACI, nommée « syntaxe X-barre » (Haegeman 1994 ; voir Coulomb 1991 pour un exposé synthétique), la structure de base du syntagme SX consiste en un niveau qui contient la « tête » X éventuellement suivie d'un syntagme complément et un niveau qui contient le constituant « barre » X', éventuellement précédé d'un syntagme spécifieur. La théorie pose le *principe d'adjonction*, selon lequel les syntagmes SX, ainsi que leurs projections intermédiaires X', peuvent être dédoublés autant que nécessaire pour rendre compte d'éléments dépendant de la même tête X. À titre illustratif, on peut représenter l'analyse de (20) à l'aide de l'arbre de la Fig. 8 (*I* = (in)flexion, *V* = lexème verbal, *D* = déterminant, *P* = préposition).

(20) Jean donne un livre à Marie.

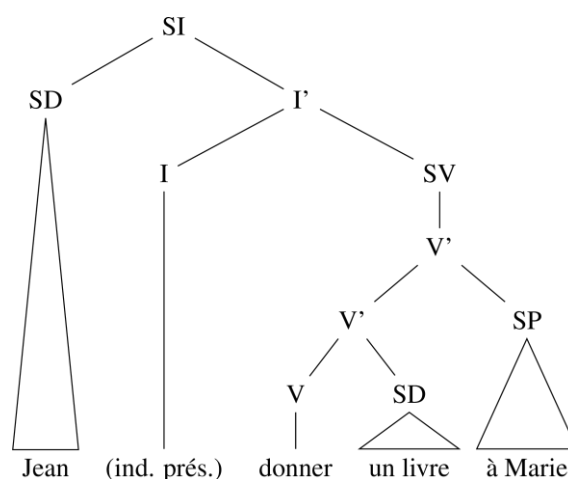


Figure 8. – Arbre X-barre comportant une analyse stratifiée des compléments régis du verbe.

On voit que le niveau V' est dédoublé de manière à accueillir les deux compléments du verbe *donner* (analyse du type de celle proposée par Laenzlinger 2003 : 41-42). La stratification instaure ainsi une hiérarchie *a priori* entre l'objet direct et l'objet indirect. Cela peut être considéré comme un moyen de représenter formellement la différence de statut qu'on leur reconnaît.

A contrario les arbres de dépendance classique font coïncider la non-stratification avec l'utilisation des nœuds dans le but exclusif d'encoder des mots (Kahane et Mazziotta 2015). Comme on l'a vu en observant les arbres de Beauzée-Glaadkij (§ 3.2), les sous-arbres correspondent toujours à un nœud encodant un mot. Aucune stratification causée par les contraintes formelles ne peut se manifester dans la structure – il s'agit d'un argument fréquemment mobilisé en faveur des analyses dépendancielles (p.ex. Maxwell

Gerdes 2019). Ainsi, le projet *Universal Dependencies*¹³ propose d'analyser les constructions prépositionnelles en plaçant la préposition en position de dépendant de son « régime », comme une marque subordonnée semblable à une flexion casuelle, mais réalisée dans un mot¹⁴. Le diagramme à droite de la Fig. 10 représente une telle analyse (*det* = déterminant, *nmod* = dépendant nominal, *case* = marqueur de cas). Dans ce diagramme, d'une part le mot *formation* est impliqué dans deux relations étiquetées *nmod* : une première fois en position de dépendant, une seconde fois en position de gouverneur ; d'autre part, les occurrences de *de* n'ont aucun dépendant. Le choix de modélisation consistant à placer la préposition en position de dépendant de l'élément qu'elle introduit implique donc que la récursivité est directe et que la dépendance *case* n'est pas récursive.

On déduit de ces comparaisons que la complexité de la récursivité, voire son existence, est au moins en partie associée aux conclusions de la linguistique théorique et aux choix formels et non à la nature linguistique intrinsèque des constructions.

4.4. Récursivité vs itérativité

Nous traiterons tout d'abord la distinction entre les concepts de *récursivité* et d'*itérativité* (§ 4.4.1). Nous montrerons ensuite succinctement comment le contraste entre ces alternatives peut être exploité pour modéliser la coordination (§ 4.4.2) et l'organisation au niveau énonciatif (§ 4.4.3).

4.4.1. Distinction et alternatives

La distinction entre les compléments multiples de même statut grammatical et les compléments ne pouvant apparaître qu'une seule fois est récurrente dans les théories syntaxiques. Elle constitue un paramètre permettant d'établir une taxinomie des configurations de constituants (Laenzlinger 2003 : 40 sv. ; Haegeman 1994 : 87-95) et des dépendances (Tesnière 1959 : ch. 56, § 4 ; Iordanskaja et Mel'čuk 2009 : 166-169), en particulier en ce qui concerne les compléments du verbe. La théorie syntaxique a donc les moyens d'évaluer la différence entre ces types de compléments indépendamment du formalisme, mais le choix de ce dernier va déterminer si les arguments répétables sont encodés par récursivité ou par *itérativité*. Cette dernière est définie en syntaxe comme une manière de combiner les constructions par une suite d'opérations indépendantes sans auto-référence.

Nous avons indiqué ci-dessus (§ 4.2) que la combinaison de la contrainte de la binarité et de la stratification des ACI impliquait la récursivité. Si on ne limite pas le nombre de constituants immédiats de chaque structure, il est possible d'analyser (17) sans stratification, de la manière inscrite dans le diagramme suivant (Fig. 11).

13 <<https://universaldependencies.org/>>

14 L'objectif est de favoriser le traitement homogène de nombreuses langues différentes et non d'encoder les structures les plus linguistiquement pertinentes pour chaque langue en particulier.

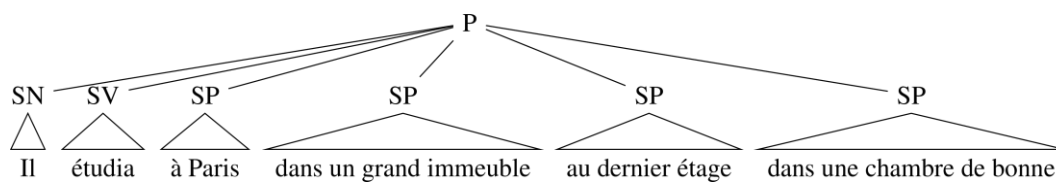


Figure 11. – Arbre ACI comportant une analyse non stratifiée de SP compléments « circonstanciels ».

Ce type de modélisation forcerait : (i) soit à poser l'existence d'une règle supplémentaire pour modéliser la combinaison d'un SN et d'un SV avec un SP, d'une autre pour modéliser la présence de deux SP, d'une autre encore pour trois SP, etc. ; (ii) soit à ajouter un autre type de règle permettant la présence d'un nombre non limité de SP comme constituants immédiats de P. La notation (23) par exemple, indique cette propriété par l'astérisque.

$$(23) \quad P \rightarrow SN \, SV \, SP^*$$

Il s'agit bien d'une autre conception de la structure des compléments « circonstanciels » et non d'une simple variation notationnelle : SP peut être actualisé un nombre indéterminé de fois. Pareille règle itérative n'est pas vraiment compatible avec le formalisme classique des règles de réécriture, mais elle exprime que l'ajout de chaque SP se fait indépendamment des précédents. Cette opération est réitérée autant que nécessaire. L'indépendance des opérations impliquées correspond nécessairement à une absence d'auto-référence. Par rapport à notre exemple des matriochkas, tout se passe comme si les poupées avaient la même taille et étaient posées les unes à côté des autres.

Les analyses dépendanciennes classiques ne permettent pas la stratification (§ 4.2). C'est là une de leurs limitations définitoires (Mazziotta 2020 : 139-140, Kahane et Gerdes 2022 : en particulier 440-443). En conséquence les compléments de statut grammatical identique sont ajoutés au même niveau sans manifester de récursivité. La Fig. 12 représente ainsi l'arbre de (17) – avec *circ* = dépendance circonstancielle.

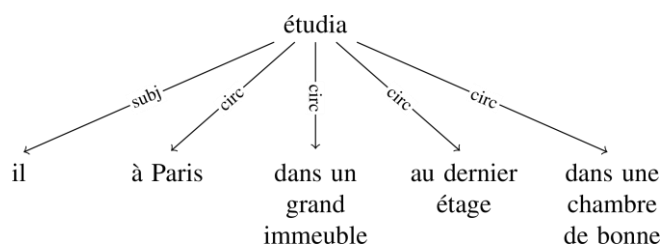


Figure 12. – Arbre dépendanciel comportant plusieurs compléments « circonstanciels ».

Comme on le voit, et de manière similaire à ce que nous venons d'observer pour les SP dans une ACI « non stratifiée », le cumul de quatre compléments circonstanciels ne présente aucune récursivité. Les compléments circonstanciels ont été ajoutés de manière itérative comme dépendants du verbe en utilisant toujours la même opération d'ajout d'une dépendance.

Dans la pratique, il n'est pas complètement évident de distinguer l'itérativité de la récursivité (MacWhinney 2009 : 407-408 ; Martins et Fitch 2014). Cela est principalement lié aux doutes soulignés à l'entame de § 4, quant à l'isomorphisme entre les algorithmes, les structures et ce qu'elles sont censées modéliser. Au vu des

alternatives possibles pour analyser les mêmes chaînes de signes linguistiques, cette question relève peut-être davantage de l'évaluation formelle que de l'analyse proprement linguistique. Certains n'hésitent d'ailleurs pas à réévaluer certaines structures originellement modélisées comme récursives et à proposer une interprétation itérative (Bickerton 2009). Dans certains cas, donc, la récursivité est une propriété du formalisme employé pour modéliser les structures linguistiques. Or, en algorithmique, il a été démontré que les formalismes des fonctions calculables par un automate étaient équivalents d'un point de vue expressif (thèse de Church-Turing), ce qui a comme principal corollaire que toute structure récursive peut être convertie en une structure itérative. Toutefois cette transformation se fait au prix d'une complexification de la description et d'un affaiblissement considérable des outils de visualisation employés.

4.4.2. Coordination et entassements

Le traitement de la coordination est en quelque sorte un cas d'école en ce qui concerne la distinction entre itérativité et récursivité. On peut résumer la description traditionnelle en ces termes :

Au sens traditionnel du terme, il y a coordination lorsque deux unités de *même niveau* et assurant la même fonction syntaxique sont reliées par une conjonction de coordination [...]. C'est l'absence de dépendance syntaxique entre les éléments reliés qui distingue la coordination de la subordination [...]. Il s'agit en somme d'un procédé *non pas hiérarchisant, mais séquentiel*, qui permet de démultiplier une catégorie de départ (proposition, syntagme ou mot) en une *chaîne de catégories identiques*. (Riegel et al. 2004 : 521, nous soulignons)

Cette conception correspond à une modélisation itérative de la coordination. Elle est loin d'être consensuelle, comme en témoignent les différentes options de formalisations qui ont pu être exploitées au sein même des grandes familles de modèles d'analyse en constituants (Mouret 2007) et de grammaire de dépendance (Mazziotta 2014 : 142-150). On peut retenir essentiellement que les options descriptives les plus contraignantes encodent les coordinations à l'aide des mêmes outils conceptuels que ceux employés pour encoder les subordinations, sur un mode qui privilégie la représentation de hiérarchies imbriquées. Ainsi, une règle de réécriture récursive non binaire comme (24) force à encoder l'analyse de l'exemple (25) en hiérarchisant les SN (Fig. 13).

- (24) $SN \rightarrow SN \text{ et } SN$
 (25) John et Paul et Georges et Ringo aiment les sous-marins jaunes (*apud* Ruwet 1968 : 160)

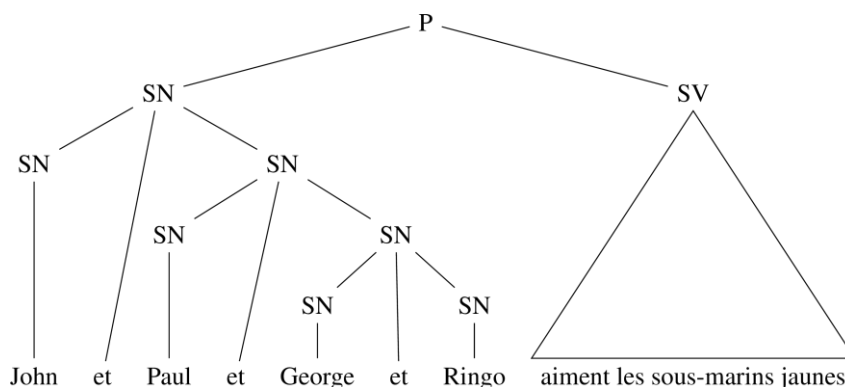


Figure 13. – Hiérarchisation des syntagmes coordonnés en ACI.

Chomsky remarque lui-même que la modélisation de la coordination en constituants contient « trop de structure » (Chomsky 1964, *apud* Ruwet 1968 : 161-162), car la stratification inhérente au formalisme force à positionner à des niveaux différents des constituants qui sont de même niveau grammatical.

Un effet de hiérarchisation similaire (mais sans stratification) peut être observé dans la modélisation de la coordination proposée par la Théorie Sens-Texte (Mel'čuk 2009 : 50-51), qui aboutit à l'analyse de la Fig. 14 (*coord* = relation coordinative ; *conj-coord* = relation conjonctionnelle coordinative).

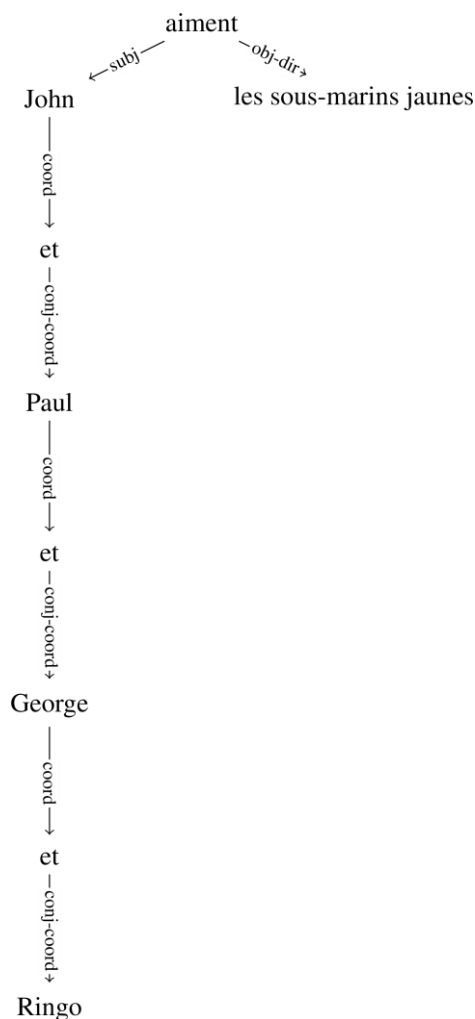


Figure 14. – Hiérarchisation des coordinations dans la Théorie Sens-Texte.

En d'autres termes, de nombreuses théories encodent de manière récursive des éléments que la tradition considère comme se combinant sur un mode itératif. Plus proches de cette conception, certaines modélisations contrastent avec ces options descriptives en proposant un traitement spécifique de la coordination. Ainsi, Lucien Tesnière (1959 : 2^e partie) propose de considérer la coordination (nommée « jonction » dans son œuvre) comme une relation distincte (non dépendancielle), qui est représentée par un trait horizontal dans ses diagrammes. Cette horizontalité contraste avec l'orientation de haut en bas des traits représentant les relations de subordination (Kahane 2012). Elle indique

que les éléments coordonnés « dédoublent » une position syntaxique et se situent au même niveau. Il en résulte une conception plate de la coordination, telle que représentée dans la Fig. 15.

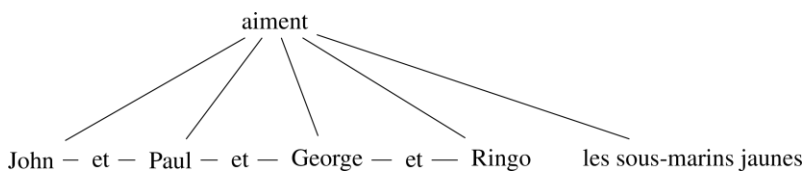


Figure 15. – Coordination « plate » d'après le modèle de Tesnière.

Les termes coordonnés sont ainsi ajoutés les uns derrière les autres, selon une logique itérative similaire à celle que nous avons décrite plus haut (§ 4.4.1). À première vue donc, l'itérativité et la récursivité correspondent respectivement à la coordination et à la subordination, ce qui paraît une synthèse élégante.

Les groupements de mots étant réalisés par des relations et ces dernières ne pouvant pas être elles-mêmes groupées (§ 3.2), il n'est pas possible de hiérarchiser les termes coordonnés. L'analyse du syntagme souligné dans (26) ne peut être faite de manière satisfaisante.

- (26) La question des limites du territoire [...] peut être abordée sous *les angles géographique, ethnique ou politique, mais aussi historique et culturel* [...]. (SUD-FR, id. fr-ud-train_00332)

Là où les ACI imposent l'utilisation de structures stratifiées, les modélisations dépendanciennes classiques sont incapables d'encoder exhaustivement la construction. Une solution à ce problème a été proposée à la suite des études menées sur le français parlé par le Groupe aixois de recherche en syntaxe (GARS, dans le sillage de Claire Blanche-Benveniste). Les travaux du GARS ont abouti au développement de modélisations capables de rendre compte des répétitions, des reformulations, des dysfluences, courantes en français oral spontané. Ces constructions particulières partagent des points communs avec la coordination, ce qui justifie qu'on traite cette dernière de la même manière (Bilger 1999). Dans l'optique dépendancielle du GARS, il s'agit de décrire de façon satisfaisante des structures où la même position syntaxique est saturée par plusieurs mots ou constructions qui font partie du même paradigme (ils peuvent se substituer les uns aux autres dans cette même position). Par exemple, (27a) présente plusieurs compléments de la construction impersonnelle d'existence *il y a* (« sujets réels »).

- (27) (a) Il y a des congrès des sardines des crabes des fleurs des algues toutes sortes de poissons (*apud* Bilger 1999 : 257)
- (b) Il y a des congrès
des sardines
des crabes
des fleurs
des algues
toutes sortes de poissons
- (c) Il y a { des congrès | des sardines | des crabes | des fleurs | des algues | toutes sortes de poissons }

Pour rendre compte de cette équivalence hiérarchique, les chercheurs ont utilisé différentes notations, comme (27b) (les éléments équivalents sont énumérés verticalement) ou (27c) (les éléments équivalents sont rassemblés entre accolades et séparés par des barres verticales), capables de représenter ce qu'ils appellent des « listes » ou « piles » paradigmatiques ou encore, terme que nous retiendrons ici, des « entassements ». Ces entassements permettent de modéliser les appositions et les coordinations, puisque les termes qui entretiennent ces relations sont d'une certaine manière interchangeables par rapport à l'élément dont ils dépendent.

Les entassements se combinent à une analyse dépendancielle en mobilisant une autre dimension pour énumérer (selon une logique itérative) les mots de même fonction (Kahane 2012). À ce niveau élémentaire de combinaison entre l'arbre de dépendance et l'entassement, il n'y a pas de stratification et pas de récursivité. Toutefois, certaines structures modélisées de la sorte comportent des imbrications d'éléments entassés, comme (28a), tiré de Gerdes et Kahane 2009.

- (28) (a) deux petites phrases deux vraies options qui dessinent votre route une route qui témoigne d'une certaine d'une bonne d'une très bonne conduite
- (b) { deux petites phrases | deux vraies options } qui dessinent { votre route | une route qui témoigne { d'une certaine | d'une bonne | d'une très bonne } conduite } (nous ajoutons l'accolade finale, manquante)
- (c) { } { { } }

Nous pouvons représenter l'essentiel de l'analyse de la hiérarchie des entassements en (28b) en la réduisant aux accolades qui les délimitent en (28c). On voit parfaitement qu'une paire d'accolades se situe à l'intérieur d'une autre, ce qui correspond à une imbrication du même type que celles qu'on observe en analyse en constituants.

Le cas des coordinations effectivement hiérarchisées est encore plus manifeste, comme observé dans (26), repris dans (29) :

- (29) (a) les angles géographique, ethnique ou politique, mais aussi historique et culturel [...].
 (b) les angles { { géographique, | ethnique | ou politique, } | mais aussi { historique | et culturel } }

L'entassement de *géographique, ethnique ou politique* avec *mais aussi historique et culturel* constitue un premier niveau, dans lequel chacun des éléments entassés est lui-même un entassement plus petit. La modélisation permet donc la récursivité tout en maintenant la possibilité d'encoder les termes de même niveau sur un mode itératif, sans introduire d'artefact.

4.4.3. Niveaux d'analyse

La distinction entre itérativité et récursivité permet en outre à certains chercheurs de délimiter le domaine des relations intrapropositionnelles du niveau supérieur de la construction du discours¹⁵.

15 Cette distinction entre les relations intrapropositionnelles et les relations entre énoncés n'est pas sans rappeler la distinction entre hypotaxe et parataxe, à laquelle se raccrochent certaines conceptions de l'évolution des langues. Selon une conception ancienne, certains posent que les

La question est justement de savoir si ce mode de récursivité par inclusion, qui caractérise *certaines* catégories micro-syntaxiques¹⁶ (et pas toutes...), peut être généralisé à tous les ordres de combinatoire linguistique, avec pour conséquence de les indifférencier, et d'indifférencier leurs catégories fonctionnelles constitutives. Nous sommes d'avis que non. (Groupe de Fribourg 2012 : 63, n. 15)

Utiliser les mêmes outils formels pour décrire tous les niveaux n'a en effet que peu de sens (voir également Deulofeu 2016 au niveau de la construction de l'énoncé). Le Groupe de Fribourg propose de parler d'une récursivité par « rebond », qui se distingue de l'auto-inclusion (Berrendonner 2012a : 202-204). Cette récursivité par rebond est considérée comme un caractère fondamental de l'enchaînement des unités minimales ayant une fonction communicative, des *énonciations* (Berrendonner 2012b : 335). Non sans générer une petite confusion terminologique, le Groupe de Fribourg parle de « mode de récursivité par itération » et utilise le symbole « ▸ » pour représenter les relations entre les énonciations de (30).

(30) les Égyptiens ▸ leurs dieux ▸ c'est les chats (Berrendonner 2012a : 202)

Ce symbole connecte le membre de gauche, qui prépare l'action de communication, au membre de droite, qui constitue l'action. On forme ainsi une sorte de chaîne où se succèdent préparations et actions. Ainsi, de droite à gauche, *c'est les chats* constitue l'action préparée par *leurs dieux*, qui, lui-même, constitue l'action préparée par *les Égyptiens*. L'enchaînement montre assez clairement qu'il n'y a pas lieu de parler d'auto-inclusion : chaque terme de l'opération représentée par « ▸ » peut être décrit par les relations auxquelles il participe indépendamment des autres relations.

5. APPROCHE EMPIRIQUE : LIMITATIONS DE LA RÉCURSIVITÉ

La récursivité est un concept théorique et abstrait qui a permis à l'école générativiste de proposer une modélisation générale de la syntaxe basée sur un nombre limité de classes et de règles. Nous avons évoqué précédemment certaines limites cognitives qui pèsent sur les cas d'emboîtement, dont la moindre réalisation implique un effort d'interprétation (§ 3.1.3). Cette observation, limitée à des structures intrinsèquement complexes, laisse cependant ouverte la question des limitations qui affectent la récursivité de manière générale : le nombre de fois où une règle peut effectivement être réemployée et les différences entre les catégories grammaticales.

langues sont originellement paratactiques et évoluent vers l'hypotaxe (v. p.ex. Antoine 1899). Nous nous garderons de suivre cette direction, qui n'est pas dépourvue d'une tendance à hiérarchiser les langues.

16 Les recherches sur le français parlé ont mené à la remise en cause de la notion de *phrase* (EGF : 2.2, « La notion de phrase ») et à la distinction entre deux types de syntaxe (Blanche-Benveniste et al. 1990, Berrendonner 1990, Benzitoun et al. 2011), qu'on désigne fréquemment des termes de *microsyntaxe* et de *macrosyntaxe*. Nous nous contenterons ici de rappeler à quoi réfèrent ces deux ordres d'organisation de l'énoncé. La *microsyntaxe* concerne les contraintes relationnelles entre les unités (rapports de subordination et de complémentation) ainsi que les rapports de coordination entre les unités régies. Selon l'approche de l'école d'Aix et de ses continuateurs, ces relations sont formalisables sous la forme d'un arbre de dépendance, nommé *unité relationnelle*. La *macrosyntaxe* concerne l'organisation linéaire de constructions par rapport à un noyau, qui permet à l'énoncé d'exister en tant qu'acte illocutoire. Ce noyau porte la *force illocutoire* de l'énoncé, c'est-à-dire qu'il détermine si l'énoncé est une assertion, une question, un ordre ou une exclamation.

En effet, il est évident que l'enchâssement récursif des propositions s'achève naturellement en raison de contraintes extérieures au système. Il y a plusieurs raisons à cela. Premièrement, la phrase réalise un énoncé, c'est-à-dire qu'elle implique l'existence d'un acte d'énonciation qui, pour aboutir, doit s'achever. Deuxièmement, de la même manière que nos facultés à manipuler des poupées minuscules sont limitées (§ 2.1), et de la même manière que les programmes informatiques dépendent des ressources qu'on leur alloue (§ 2.2), notre mémoire est mise à rude épreuve quand il est question de comprendre une phrase rendue trop longue par des enchâssements successifs. Ainsi, le message doit être transmis sous une forme qui ne soit pas obscure, selon la maxime de manière de Grice (1975 : 46-47). L'exemple (31), parfaitement bien formé du point de vue de sa structure, est peu acceptable s'il est envisagé du point de vue de son efficacité communicationnelle.

- (31) LE POMPIER – Mon beau-frère avait, du côté paternel, un cousin germain dont un oncle maternel avait un beau-frère dont le grand-père paternel avait épousé en secondes noces une jeune indigène dont le frère avait rencontré, dans un de ses voyages, une fille dont il s'était épris et avec laquelle il eut un fils qui se maria avec une pharmacienne intrépide qui n'était autre que la nièce d'un quartier-maître inconnu de la marine britannique et dont le père adoptif avait une tante parlant couramment l'espagnol et qui était, peut-être, une des petites-filles d'un ingénieur, mort jeune, petit-fils lui-même d'un propriétaire de vignes dont on tirait un vin médiocre, mais qui avait un petit-cousin, casanier, adjudant, dont le fils avait épousé une bien jolie jeune femme, divorcée, dont le premier mari était le fils d'un sincère patriote qui avait su élever dans le désir de faire fortune une de ses filles qui put se marier avec un chasseur qui avait connu Rothschild et dont le frère, après avoir changé plusieurs fois de métier, se maria et eut une fille dont le bisaïeul, chétif, portait des lunettes que lui avait données un sien cousin, beau-frère d'un Portugais, fils naturel d'un meunier, pas trop pauvre, dont le frère de lait avait pris pour femme la fille d'un ancien médecin de campagne, lui-même frère de lait du fils d'un laitier, lui-même fils naturel d'un autre médecin de campagne, marié trois fois de suite, dont la troisième femme... (E. Ionesco, *La cantatrice chauve*, VIII)

En guise d'autre exemple, on cite fréquemment celui du professeur allemand dont le cours ne serait qu'une longue phrase unique truffée de subordonnées (Hofstadter 2000 : 147). Or, dans cette langue, les verbes des subordonnées sont rejetés en fin de structure. En conséquence, les étudiants devraient attendre la fin du cours pour entendre, dans un ordre dont ils auront oublié la pertinence, tous les verbes en question. Cette restriction cognitive est notamment liée au fait que nous devons non seulement garder les unités linguistiques dans notre mémoire de travail, mais également les liens qu'elles entretiennent (Halford et al. 2005). Cette contrainte a un impact sur les structures syntaxiques effectivement employées. Il faut donc distinguer les structures observées du potentiel récursif. L'allégation selon laquelle « there is no longest sentence (any candidate sentence can be trumped by, for example, embedding it in “Mary thinks that...”), and there is no nonarbitrary upper bound to sentence length » (Hauser et al. 2002 : 1571) est valable au niveau du système grammatical envisagé comme une machine formelle. Cependant, ce potentiel est extrêmement contraint par les limitations cognitives et situationnelles. Il est donc légitime de se demander si, à partir d'un certain niveau d'auto-enchâssement, la récursivité n'est pas une propriété mobilisée par jeu littéraire, principalement par des locuteurs qui ont reçu une formation grammaticale, une « grammaire seconde » qui a développé leur sentiment linguistique (Blanche-Benveniste 1990).

Pour évaluer le comportement des structures récursives de manière empirique et à titre illustratif (l'examen approfondi dépasserait le cadre de cette notice), nous avons choisi d'extraire les structures du corpus *French Treebank* (FTB ; Abeillé et al. 2003) et de nous focaliser sur l'ACI pour en fournir une analyse sommaire. FTB comporte 21 550 phrases (ca 664 500 tokens). On gardera à l'esprit que les sections précédentes ont montré que la récursivité est un phénomène structurel dont la présence et la profondeur (le nombre de niveaux d'inclusion) dépendent de l'orientation théorique de la description et des contraintes sur le formalisme sélectionné pour encoder l'analyse qui en découlent. En conséquence, d'une part aucun résultat ne pourra être atteint si nous n'acceptons pas les principes d'encodage du corpus comme une donnée « axiomatique » ; d'autre part, les conclusions de toute procédure d'évaluation empirique des structures récursives dans un corpus ne pourront être généralisées qu'avec précaution. Cela posé, nous pouvons présenter les matériaux (§ 5.1) et examiner les cas de manière globale (§ 5.2), puis nous focaliser sur la récursivité directe (§ 5.3) et la récursivité indirecte (§ 5.4).

5.1. Présentation des matériaux

Pour commencer, nous avons donc besoin de présenter l'inventaire des classes syntagmatiques du corpus et la manière dont les annotations sont structurées. En outre, un certain nombre de pré-traitements ont été appliqués par nos soins et nécessitent un bref commentaire.

Les classes syntagmatiques sont les suivantes (Abeillé et al. 2015 : 7 ; nous les mentionnons sous la forme francisée que nous utiliserons par la suite) :

- SAdj (syntagme adjectival) ;
- SAdv (syntagme adverbial) ;
- SN (syntagme nominal) ;
- SP (syntagme prépositionnel) ;
- NV (noyau verbal, contenant le verbe et les clitiques) ;
- SV-inf (syntagme verbal infinitif) et SV-part (syntagme verbal participial) ;
- P (phrase maximale), ainsi que P-int, P-rel, P-sub (propositions subordonnées, relatives, ou internes : incises, parenthétiques, coordonnées) ;
- COORD (syntagme coordonné).

En ce qui concerne la structure des analyses, l'annotation est « plate » et il n'y a pas de contrainte de binarité (§ 4.2). La stratification est donc minimale. Cela implique notamment que : 1/ les constructions constituées d'un unique mot ne sont généralement pas considérées comme des syntagmes (cela est particulièrement notable en ce qui concerne les adjectifs et les adverbes) ; 2/ il n'y a pas de SV au niveau de la réécriture de P (la phrase transitive avec sujet et objet nominaux est analysée selon la règle (32) ; 3/ les compléments « circonstanciels » sont intégrés à P de façon itérative, comme dans la Fig. 11.

$$(32) \quad P \rightarrow SN \text{ NV } SN$$

Nous avons appliqué un certain nombre de pré-traitements au corpus avant l'extraction – d'autres choix auraient mené à d'autres résultats. Premièrement, les distinctions entre les sous-types de P (en particulier la distinction entre P et P-sub) occulteraient les phénomènes d'emboîtement de propositions. Nous avons donc réduit tous les sous-types de P à P, de manière à rendre compte de la même manière de toutes les subordonnées. À

l'inverse, la distinction entre SV-part et SV-inf est maintenue étant donné la grande différence de distribution de ces sous-types et l'absence de SV au niveau de la réécriture de P. Deuxièmement, la coordination posant des problèmes de modélisation (notamment le problème mathématique de la dimension sur laquelle elle se trouve dans le graphe d'annotation), les catégories dont la récursivité implique COORD, ainsi que COORD elle-même, ont été retirées de l'analyse qui suit¹⁷.

5.2. Généralités

Nous avons extrait les hiérarchies de constituants qui contiennent de la récursivité et nous les avons regroupés sous un patron d'imbrications qui exprime le chemin le plus direct correspondant à l'auto-emboîtement des symboles. Il en ressort 165059 structures récursives dont seuls 14 représentent au moins 1 % de ce total. Ils représentent ensemble près de 70 % des cas de récursivité du corpus. Par exemple, l'enchâssement direct de la complétive objet soulignée dans (33a) est résumé par le patron (33b). Le patron (34b), le plus fréquent, encode quant à lui la récursivité indirecte observée dans (34a).

(33) (a) Devant le Parlement réuni en session extraordinaire, Mr Hans Modrow a annoncé [...] *que la création d'une nouvelle police secrète "n'interviendrait pas avant le 6 mai"*. (FTB, texte 1106, phrase 15)

(b) P[P] (fq. abs. 6438, 3.9 %)

(34) (a) au bénéfice *de l'Europe tout entière* (FTB, texte 1105, phrase 3)

(b) SP[SN[SP]] (fq. abs. 41803, 25.33 %)

Il est très difficile d'évaluer les proportions relatives des patrons car il faudrait en toute rigueur tenir compte des cas qui sont imbriqués dans des cas plus complexes pour effectuer une évaluation statistique fiable. On se bornera ici et par la suite à des observations de surface : 8 des 14 patrons retenus impliquent une hiérarchie de SN et de SP, c'est-à-dire une imbrication plus ou moins profonde de compléments déterminatifs comme (35a), dont le patron est (35b).

(35) (a) son vernis de chef d'état à l'occidentale (FTB, texte 1112, phrase 69)

(b) SN[SP[SN[SP[SN]]]] (fq. abs. 8194, 4.96 %)

Les autres cas sont principalement représentés par les termes nominaux à l'intérieur des relatives (SN[P[SN]] ; fq. abs. 9176, 5.56 %), les relatives dans un terme nominal constituant immédiat de la phrase (P[SN[P]] ; fq. abs. 2367, 1.43 %), les appositions (SN[SN] ; fq. abs. 7146, 4.33), les propositions complétives sujet ou objet et les circonstanciellées (P[P] ; fq. abs. 6438, 3.9 %), ainsi que les SP compléments d'un participe employé comme épithète (SN[SV-part[SP[SN]]] et SP[SN[SV-part[SP]]] ; respectivement fq. abs. 4277, 2.59 % et 2031, 1.23 %). Il est frappant que ces cas ne présentent généralement pas une très grande profondeur d'imbrication et que la diversité des structures concernées est très limitée.

17 La coordination est une structure difficile à encoder. Elle force à poser des choix qui sont rarement consensuels. Pour le corpus FTB, un certain nombre de cas de dysfluences ou de constructions que nous considérerions comme similaires à des coordinations ont été encodés comme des appositions (Abeillé et al. 2015 : 36-39).

5.3. Récursivité directe

L'extraction des cas de récursivité directe et l'évaluation de leur profondeur révèle que si le phénomène a l'air de concerner toutes les classes syntagmatiques non terminales, les cas de récursivité au niveau de NV (1 cas), SAdj (7 cas), SAdv (3 cas) et SV-part (35 cas) sont en réalité extrêmement rares¹⁸. Les quatre classes davantage concernées sont reprises dans le Tab. 1.

SX \ Profondeur	1	2	3	4
SN	7146	327	3	0
SP	1210	36	0	0
P	6438	518	25	1
SV-inf	969	58	3	0

Tableau 1. – Profondeur de l'auto-inclusion directe par classe syntagmatique.

Pour que le nombre d'attestations et la profondeur de la récursivité commence à atteindre un palier relativement élevé, il faut qu'elle implique, par ordre croissant, des SV-inf, des SP, des P et des SN. Or, il semble que ces structures correspondent dans la plupart des cas à des constructions très précises. Les SV-inf imbriqués correspondent aux structures où l'infinitif est introduit par une préposition (cette dernière étant analysée comme un constituant immédiat de SV-inf), comme dans (36a), dont la hiérarchie est explicitée dans (36b).

- (36) (a) pour leur demander d'accorder des dérogations provisoires à la chaîne [...] afin de lui permettre de survivre (FTB, texte id225123, phrase 717)
- (b) [pour leur demander [d'accorder des dérogations provisoires à la chaîne [afin de lui permettre [de survivre]]]]

Les SP imbriqués de la sorte correspondent bien à des SP qui sont eux-mêmes le régime d'une préposition, comme (37), ainsi qu'à des constructions du type *de... à...* (38).

- (37) de chez Lanvin (FTB, texte 1112, phrase 71)
- (38) (a) Du 1er septembre 1988 au 31 août 1989 (FTB, texte 1115, phrase 120)
(b) [Du 1er septembre 1988 [au 31 août 1989]]

Les deux autres types de cas, SN et P, correspondent aux appositions nominales et aux cas de subordination évoqués précédemment. Ils représentent l'écrasante majorité des cas.

Par ailleurs, la profondeur de la récursivité est toujours très peu élevée : on ne dépasse que très rarement deux niveaux d'auto-inclusion directe, même avec une annotation peu stratifiée. Le niveau maximal est quatre (P[P[P[P]]]); il est représenté par l'unique attestation (39), qui comporte un discours direct (l'annotation ne distinguant pas l'énoncé de la proposition au niveau des constituants).

18 L'observation de ces cas révèle d'ailleurs qu'on pourrait les interpréter comme des coordinations quand il ne s'agit pas de cas exceptionnels dont l'analyse n'est pas évidente.

(39) (a) Il a assuré que l' afflux des juifs soviétiques [...] démoralisait les Arabes : « autour de nous, les Arabes sont en plein désarroi, en état de panique [...] ; ils sont submergés par un sentiment de défaite, parce qu' ils voient que l'Intifada ne sert à rien [...] ». (FTB, texte 859, phrase 8641)

(b) [Il a assuré [que l' afflux démoralisait les Arabes [« les Arabes sont en plein désarroi [parce qu' ils voient [que l'Intifada ne sert à rien]] »]]]

5.4. *Récurtivité indirecte*

L'évaluation de la profondeur de la récurtivité indirecte n'a de sens qu'au regard des patrons indirects « réguliers », c'est-à-dire constitués sur un patron de base qui se répètent au moins une fois dans le même ordre d'imbrication. Ce genre de cas est illustré par (40), où le patron SN[SP[SN]] se répète trois fois.

(40) (a) le lancement d'une formation des enseignants du supérieur (FTB, texte 1116, phrase 137) [= (6)]

(b) SN[SP[SN[SP[SN[SP[SN]]]]]]

Ce genre d'imbrication pose le problème que certains patrons sont imbriqués les uns dans les autres, ce qui gonfle les statistiques. D'après nos extractions, le corpus FTB contient 39 configurations différentes de constructions présentant une récurtivité indirecte régulière de 26 patrons de base, dont les cas les plus fréquents sont présentés dans le Tab. 2. Tous les autres cas sont d'une grande rareté.

Patron récurrent \ Profondeur	1	2	3	4	5
SN[SP]	8194	1152	164	19	2
SN[P]	137	3	0	0	0

Tableau 2. – Profondeur de l'auto-inclusion indirecte par patron récurrent.

L'écrasante majorité des cas correspondant à la récurtivité indirecte concerne les compléments prépositionnels (répétition du patron SN[SP]). Les exemples de relatives imbriquées (répétition du patron SN[P]), qui arrivent en deuxième position, sont nettement moins fréquents. Par ailleurs, la profondeur de la récurtivité y est très limitée.

Ce rapide examen d'un corpus dont les étiquettes sont très générales montre que si les descriptions formelles permettent effectivement d'encoder de la récurtivité dans un grand nombre de constructions, il faut constater qu'elle ne se manifeste pas de manière très variée, et très rarement avec un niveau de complexité (profondeur) important. Un jeu d'étiquettes plus fin mènerait automatiquement à des fréquences encore moins élevées, ce qui est interpellant pour une propriété souvent présentée comme une caractéristique fondamentale du langage humain. L'étude d'un seul corpus ne permet pas de se positionner définitivement. Néanmoins, même au départ de ces observations particulières sur un corpus français, il nous paraît raisonnable de nous interroger : l'importance que la récurtivité revêt au niveau formel n'influence-t-elle pas outre mesure les conceptions générales sur le langage ?

6. CONCLUSION

En guise de conclusion, nous rappellerons la progression de l'exposé (§ 6.1) et ferons un dernier point sur la terminologie (§ 6.2).

6.1. Progression

Nous avons présenté le concept de récursivité appliqué à la syntaxe. Nous sommes parti d'exemples non linguistiques (§ 2), pour nous concentrer ensuite sur la manière dont la récursivité se manifeste en ACI (§ 3.1) et dans les analyses dépendanciennes (§ 3.2). Nous avons vu que malgré les différences, ces deux grandes options descriptives présentent des formes de récursivités similaires, liées notamment à l'existence de rapports d'*inclusion* dans les deux modèles. Toutefois, nous avons souligné que la théorie syntaxique donne du sens aux formalismes et leur ajoute des contraintes qui impactent la manière dont se manifeste la récursivité (§ 4.1). Ainsi, il est apparu clairement que les contraintes formelles de l'ACI donnent un caractère *stratifié* aux analyses. Ce dernier provoque l'apparition de structures récursives artificielles (§ 4.2). Nous avons également vu que la différence entre récursivité directe et récursivité indirecte est parfois liée à des choix de modélisation (§ 4.3). Nous avons terminé nos remarques sur les implications de structures formelles choisies en comparant la récursivité à l'*itérativité*. La distinction entre ces propriétés permet de distinguer des structures de nature et de niveau différents : les *entassements*, qui se combinent potentiellement de manière récursive et la combinaison des énoncés (§ 4.4). Sur la base de l'examen d'un corpus particulier, la dernière partie de notre exposé a souligné que dans les faits, la récursivité ne se manifeste pas de manière très diversifiée et complexe, ce qui invite à s'interroger sur l'importance qu'on accorde à cette propriété formelle au niveau de la conception du langage (§ 5).

6.2. Point terminologique final

Cette progression nous a amené à mobiliser un certain nombre de termes, dont certains sont donnés pour quasiment synonymes de *récursivité* selon une interprétation naïve. Nous avons nommé *récursivité* (angl. *recursion*) la propriété formelle selon laquelle un élément linguistique de nature X contient, par *enchâssement*, un autre élément de même nature X. La description de l'élément englobant de nature X ne peut être effectuée qu'en mentionnant cette nature X, si bien qu'il y a *auto-référence* et, par extension, *auto-enchâssement* ou *auto-inclusion* (§ 2 et § 3.1.1). L'enchâssement peut être effectué de manière *directe* ou *indirecte* (§ 3.1.2). Si l'on prend en compte l'ordre des mots, il faut distinguer la récursivité à *gauche* ou à *droite*, qui se réalisent à la périphérie des constructions, de l'*auto-emboîtement* (angl. *nesting* ou *center-embedding*), qui se réalise au milieu de la construction auto-référentielle (§ 3.1.3). Enfin, d'un point de vue formel, il convient de distinguer la récursivité de l'*itérativité* : les cas d'*itérativité* correspondent au réemploi de règles identiques sans auto-référence, c'est-à-dire indépendamment les unes des autres (§ 4.4.1).

RÉFÉRENCES

Abeillé (A.), Clément (L.) & Toussnel (F.), 2003, « Building a treebank for French », in *Treebanks : building and using parsed corpora*, A. Abeillé éd., Dordrecht, Kluwer, 165-187.

Abeillé (A.), Toussenet (F.) & Chéradame (M.), 2015, *Corpus arboré pour le français (FTB). Annotations en constituants : les syntagmes*, Paris : LLF, Université Paris 7. En ligne : <<http://ftb.linguist.univ-paris-diderot.fr/fichiers/public/guide-constit.pdf>>

Antoine (F.), 1899, « De la parataxe et de l'hypotaxe dans la langue latine », *Revue des langues anciennes* 1, 27-46.

Askedal (J. O.), 2003, « Das Valenz- und Dependenzkonzept bei Lucien Tesnière », in *Dependency and Valency : An International Handbook of Contemporary Research, Vol. 1*, V. Ágel et al. édés, Berlin, De Gruyter, 80-99

Bar-Hillel (Y.), 1953, « On recursive definitions in empirical sciences », in *Proceedings of the 11th international congress of philosophy*, 160-165.

Benetti (L.) & Berrendonner (A.), 2012, « Chapitre XIII. Appositions », in Groupe de Fribourg 2012 : 283-301.

Benzitoun (C.), Dister (A.), Gerdes (K.), Kahane (S.), Pietrandrea (P.) & Sabio (F.), 2010, « *tu veux couper là faut dire pourquoi*. Propositions pour une segmentation syntaxique du français parlé », in *Congrès Mondial de Linguistique Française - CMLF 2010*, F. Neveu et al. édés, 2075-2090.

Berrendonner (A.), 1990, « Pour une macro-syntaxe », *Travaux de linguistique* 21, 25-36.

Berrendonner (A.), 2012a, « Chapitre IX. Structure *préparation + action* », in Groupe de Fribourg 2012 : 185-205.

Berrendonner (A.), 2012b, « Chapitre XV. Qu'est-ce qu'une période ? », in Groupe de Fribourg 2012 : 319-351.

Berrendonner (A.) & Béguelin (M.-J.), 2012, « Chapitre VIII. Typologie des "dislocations à gauche" », in Groupe de Fribourg 2012 : 155-186.

Bickerton (D.), 2009, « Recursion: Core of complexity or artifact of analysis », in Givón & Masayoshi 2009 : 531-544.

Bilger (M.), 1999, « Coordination : analyses syntaxiques et annotations », *Recherches sur le français parlé* 15, 255-272.

Blanche-Benveniste (C.), 1990, « Grammaire première et grammaire seconde : l'exemple de *en* ». *Recherches sur le français parlé* 10, 51-73

Blanche-Benveniste (C.), Bilger (M.), Rouget (Ch.) & van den Eynde (K.), 1990, *Le français parlé. Études grammaticales*, Paris, Éditions du CNRS.

Bresnan (J.), Asudeh (A.), Toivonen (I.) & Wechsler (S.), ²2016, *Lexical-Functional Syntax*, Malden, Oxford, Chichester, Blackwell.

Booij (G.), 2007, *The grammar of words : an introduction to morphology*, Oxford, New York, etc., Oxford University Press.

- Carnie (A.), 2011, *Modern syntax. A course book*, Cambridge, Cambridge University Press.
- Chomsky (N.), 1964, « La notion de “règle de grammaire” », *Langages* 4, 81-104.
- Chomsky (N.), 1965, *Aspects of the theory of syntax*, Cambridge, M.I.T. Press.
- Corbin (D.), 1987, *Morphologie dérivationnelle et structuration du lexique*, Tübingen, Niemeyer, 2 vols.
- Coulomb (C.), 1991, « La syntaxe X barre (X') », *L'information grammaticale* 50, 6-12.
- Croft (W.), 2001, *Radical construction grammar. Syntactic theory in typological perspective*, Oxford, Oxford University Press.
- Desclés (J.-P.), 2018, *Brève généalogie des grammaires applicatives*, *Verbum* 40(2), 143-172
- Deulofeu (J.), 2016, « La macrosyntaxe comme moyen de tracer la limite entre organisation grammaticale et organisation du discours », *Modèles linguistiques* 74, 135-166.
- Everett (D.), 2005, « Cultural constraints on grammar and cognition in Pirahã : another look at the design features of human language », *Current Anthropology* 46(4), 621-646.
- FTB = *French Treebank*, version. 1.0 2016. En ligne <<http://ftb.linguist.univ-paris-diderot.fr/>> ; v. Abeillé et al. 2003.
- Gerdes (K.) & Kahane (S.), 2006, « L'amas verbal au cœur d'une modélisation topologique du français », *Linguisticae investigationes* 29(1), 75-89.
- Gerdes (K.) & Kahane (S.), 2009, « Speaking in Piles. Paradigmatic Annotation of a Spoken French Corpus », in *Proceedings of the fifth Corpus Linguistics Conference, Liverpool (Coling 2009)*. En ligne <<https://kahanedotfr.files.wordpress.com/2017/01/piles-liverpool2009.pdf>>
- Givón (T.) & Masayoshi (S.) éds, 2009, *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*, Amsterdam, Philadelphia, John Benjamins.
- Grice (H. P.), 1975, « Logic and conversation », in *Syntax and semantics 3: Speech arts*, P. Cole & J. L. Morgan éds, New York, Academic Press, 41-58.
- Groupe de Fribourg, 2012, *Grammaire de la période*, Bern, Berlin, etc., Peter Lang.
- Haegeman (L. M. V.), 1994, *Introduction to government and binding theory*, Oxford, Cambridge (USA), Blackwell.
- Halford (G. S.), Baker (R.), McCredden (J.) & Bain (J. D.), 2005, « How many variables can humans process ? », *Psychological Science* 16, 70-76.

Hausser (M. D.), Chomsky (N.) & Fitch (W. T.), 2002, « The faculty of language : What is it, who has it, and how did it evolve ? », *Science* 298, 1569-1579.

Hofstadter (D.), 2000, *Gödel, Escher, Bach. Les brins d'une guirlande éternelle*, Paris, Dunod.

Hudson (R.), 2010, *An introduction to Word Grammar*, Cambridge, London, New York, etc., Cambridge University Press.

Imrényi (A.) & Mazziotta (N.) éd(s), 2020, *Chapters of dependency grammar: a historical survey from antiquity to Tesnière*. Amsterdam : John Benjamins,

Iordanskaja (L.) & Mel'čuk (I.), 2009, « Establishing an inventory of Surface-Syntactic relations : valence-controlled surface-syntactic dependents of the verb in French », in Polguère et Mel'čuk 2009 : 151-234.

Kahane (S.), 1997, « Bubble trees and syntactic representations », in *Proceedings of Mathematics of Language MOL5*, T. Becker & H.-U. Krieger éd(s), 70-76.

Kahane (S.), 2001, « Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel », in *Actes de TALN*, vol. 2. En ligne : <https://kahanedotfr.files.wordpress.com/2017/06/dg-taln2001.pdf>

Kahane (S.), 2009, « On the status of phrases in Head-driven Phrase Structure Grammar. Illustration by a totally lexical treatment of extraction », in Polguère & Mel'čuk 2009 : 111-150.

Kahane (S.), 2012, « De l'analyse en grille à la modélisation des entassements », in *Penser les langues avec Claire Blanche-Benveniste*, in S. Caddeo et al. éd(s), Aix-en-Provence, Presses de l'université de Provence, 101-116.

Kahane (S.), 2020, « How dependency syntax appeared in the French Encyclopedia : from Buffier (1709) to Beauzée (1765) », in Imrényi & Mazziotta 2020 : 85-131.

Kahane (S.), Deulofeu (J.), Gerdes (K.), Nasr (A.) & Valli (A.), 2017, *Guide d'annotation syntaxique du corpus Orfeo*. En ligne : <https://repository.ortolang.fr/api/content/cefc-orfeo/10/documentation/site-orfeo/guide-dannotation-syntaxique-du-corpus-orfeo>>

Kahane (S.) & Mazziotta (N.), 2015, « Syntactic Polygraphs. A Formalism Extending Both Constituency and Dependency. », in *Proceedings of Mathematics of Language 2015*, 152-164.

Kahane, S. & Gerdes, K. 2022. *Syntaxe théorique et formelle. Vol. 1. Modélisation, unités, structures*. Berlin : Language Science Press. En ligne : <https://zenodo.org/record/6446067>>]

Knuth (D. E.), 2013 (= ³1997). *The art of computer programming. Volume 1. Fundamental algorithms*, Mathematical Sciences Publishers.

Laenzlinger (C.), 2003, *Initiation à la syntaxe formelle du français. Le modèle Principes et Paramètres de la Grammaire Générative Transformationnelle*, Berne, Peter Lang.

- Lowenthal (F.), 2014, « Are there any relations between recursion and language ? », in Lowenthal & Lefebvre 2014 : v-vii.
- Lowenthal (F.) & Lefebvre (L.) éd(s), 2014, *Language and recursion*, New York, Heidelberg, etc., Springer.
- MacWhinney (B.), « The emergence of linguistic complexity », in Givón & Masayoshi 2009 : 405-432.
- Martins (M. D.) & Fitch (W. T.), 2014, « Investigating Recursion Within a Domain-Specific Framework », in Lowenthal & Lefebvre 2014 : 15-26.
- Matthews (P. H.), 1981, *Syntax*, Cambridge, London, New York, etc., Cambridge University Press.
- Maxwell (D.), 2013, « Why so many nodes ? », in *Proceedings of Depling 2013*, 197-206.
- Mazziotta (N.) 2014. « Coordination of verbal dependents in Old French : coordination as specified juxtaposition or specified apposition », in *Dependency linguistics. Recent advances in linguistic theory using dependency structures*, K. Gerdes et al., éd(s), Amsterdam, Philadelphia, Benjamins.
- Mazziotta (N.), 2020, « Chapter 4. Dependency in Early Sentence Diagrams : Stephen W. Clark », in Imrényi & Mazziotta 2020 : 134-62.
- Mazziotta (N.) & Kahane (S.), 2015, « Dependency-based analyses for function words. Introducing the polygraphic approach », in *Proceedings of Depling 2015*, 181-190.
- Mel'čuk (I.), 2009, « Dependency in natural language », in Polguère & Mel'čuk 2009 : 1-110.
- Mel'čuk (I.) & Milićević (J.), 2014, *Introduction à la linguistique*. Paris, Hermann, 3 vols.
- Miller (G. A.) & Chomsky (N.), 1963, « Finitary models of language users », in *Handbook of mathematical psychology*, R. D. Luce et al. éd(s), New York, Wiley.
- Monneret (P.), 1999, *Exercices de linguistique*, Paris, PUF.
- Mouret (F.), 2007, *Grammaire des constructions coordonnées. Coordinations simples et coordinations à redoublement en français contemporain*, Paris, Université Paris 7 [Thèse de doctorat].
- Neeleman (A.) & van de Koot (H.), 2006, « On syntactic and phonological representations », *Lingua* 116, 1524-1552.
- Neveu (F.), 2013, « Frontières fonctionnelles et niveaux syntaxiques des segments détachés », in *Les fonctions grammaticales. Histoire, théories, pratiques*, A. Ouattara éd., Bruxelles, Peter Lang, 223-239.

- Osborne (T.), 2018, « Tests for constituents : What they really show about the nature of linguistic structure », *Language under discussion* 5(1), 1-41.
- Osborne (T.) & Gerdes (K.), 2019, « The status of function words in dependency grammar : A critique of Universal Dependencies (UD) », *Glossa* 4(1), 17.1-28.
- Polguère (A.) & Mel'čuk (I.) éds, 2009, *Dependency in linguistic description*, Amsterdam, Philadelphia, John Benjamins.
- Pollard (C.) & Sag (I. A.), 1994, *Head-Driven Phrase Structure Grammar*, Chicago, University of Chicago Press.
- Picabia (L.), 1975, *Éléments de grammaire générative : application au français*, Paris, Armand Colin.
- Riegel (M.), Pellat (J.-C.) & Rioul (R.), 2003, *Grammaire méthodique du français*, Paris, PUF.
- Roussarie (L.), 2017, *Sémantique formelle. Volume 1 : Introduction à la grammaire de Montague*, Berlin, Language Science Press. En ligne <<http://langsci-press.org/catalog/book/143>>
- Ruwet (N.), 1968, *Introduction à la grammaire générative*, Paris, Plon.
- Soutet (O.), 1995, *Linguistique*, Paris, PUF.
- SUD-FR = corpus *SUD_French-GSD*, version 2.7. En ligne : <<https://surfacesyntacticud.github.io/data/>>
- Tesnière (L.), 1959, *Éléments de syntaxe structurale*, Paris, Klincksieck.
- Touratier (C.), 2005, *Analyse et théorie syntaxiques*, Aix-en-Provence, Publications de l'Université de Provence.