



HAL
open science

GR, un jeu interactif pour l'étude de la grammaire du français

François Trouilleux

► **To cite this version:**

François Trouilleux. GR, un jeu interactif pour l'étude de la grammaire du français. [Rapport de recherche] Université Clermont Auvergne, LRL. 2021. hal-03339781

HAL Id: hal-03339781

<https://hal.science/hal-03339781>

Submitted on 9 Sep 2021

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

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

GR, un jeu interactif pour l'étude de la grammaire du français

Rapport de recherche
Juillet 2021

François Trouilleux
Université Clermont Auvergne, LRL, F-63000 CLERMONT-FERRAND, FRANCE

Résumé

Ce rapport de recherche présente un logiciel baptisé GR, pour « groupe et rattache », qui invite l'utilisateur à produire pas à pas la structure de phrases en français (<http://gr.19c.fr>). On présente d'abord les objectifs du jeu, puis on justifie l'approche générale et la représentation adoptée pour la structure syntaxique. Enfin, on évoque les développements possibles du prototype, du point de vue logiciel, pédagogique ou linguistique.

Mots clés

enseignement de la grammaire, grammaire de phrase, jeu sérieux, structure syntaxique.

Abstract

This research report introduces GR (groupe et rattache), a software which invites users to progressively build up the structure of sentences in French (<http://gr.19c.fr>). I first present the goals of the game, then I justify the general approach and the chosen representation for syntactic structure. Finally, I mention possible developments of the prototype, from a software, didactic or linguistic point of view.

Keywords

grammar teaching, sentence grammar, serious game, syntactic structure.

Sommaire

1. GR : pour qui et pour quoi faire ?.....	3
1.1. Historique de l'idée	4
1.2. Publics cibles	5
1.3. Objectifs	5
1.3.1. Travail général sur la notion de structure	5
1.3.2. Notions sous-jacentes à la structure.....	6
2. Justification de l'approche.....	6
2.1. Problématique du rattachement	6
2.2. Problématique de la construction incrémentale.....	7
2.3. Justification de la structure	7
2.3.1. Multiplicité des représentations.....	7
2.3.2. Equivalence des représentations.....	9
2.3.3. Petits et grands groupes	10
2.3.4. Régularité distributionnelle des petits groupes.....	10
3. Quelques points de discussion.....	11
3.1. Absence du complément de phrase	11
3.2. Pronom clitique sujet.....	12
3.3. Structure du groupe nominal	14
3.3.1. Des linguistes hésitants.....	14
3.3.2. Des commutations peu concluantes.....	15
3.3.3. Une activité cérébrale en contradiction apparente avec l'analyse commune	16
3.4. Auxiliaire et participe passé	16
4. Développements envisagés.....	17
4.1. Logiciel.....	17
4.1.1. Suggestions d'amélioration	17
4.1.2. Traitement de la phrase complexe	17
4.2. Pédagogie	18
4.3. Linguistique.....	18
5. Conclusion.....	19
6. Références	19
7. Annexe.....	20

Ce rapport de recherche présente un logiciel baptisé *GR*, pour « groupe et rattache », qui invite l'utilisateur à produire pas à pas la structure de phrases en français. La figure ci-contre montre l'analyse finale de la phrase *L'équipe nationale de hockey a remporté le championnat du monde cette année.*

Le principe de fonctionnement de *GR* est inspiré des jeux de déplacements de blocs de type Tetris ou [2048](#). Les blocs sont des mots distribués un à un ; l'utilisateur doit les arranger de façon appropriée en utilisant quelques touches du clavier. Le jeu se joue en ligne, à l'adresse <http://gr.19c.fr>.

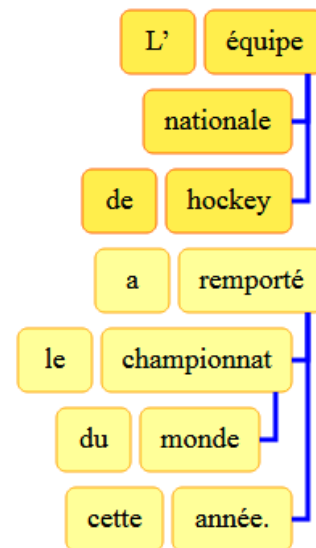


Figure 1. Analyse d'une phrase dans *GR*.



Avant de lire ce rapport, l'utilisateur est invité à tester le jeu en ligne. Ce rapport ne reprend pas les diverses informations disponibles sur le site de *GR*, notamment les mécanismes du jeu et les règles de formation des groupes. Une copie des pages du site à la date de rédaction de ce rapport est donnée en annexe.

Le rapport est organisé en quatre parties. La première présente les objectifs pédagogiques du jeu. Les deuxième et troisième parties discutent plus particulièrement de l'approche et de la structure adoptées dans le jeu, de façon générale puis particulière. La quatrième partie évoque les développements envisagés dans trois domaines : le développement logiciel, la pédagogie et la linguistique.

1. *GR* : pour qui et pour quoi faire ?

L'objectif général de *GR* est de soutenir l'étude de la grammaire / syntaxe du français. Il s'adresse à essentiellement deux publics, des « apprenants » et des « experts ». Les apprenants sont les élèves de tous niveaux, de l'école primaire à l'université, qui doivent apprendre les concepts de la grammaire de phrase. Les experts sont les linguistes et les didacticiens de la grammaire. Les premiers discutent des structures syntaxiques, développent des corpus annotés, s'interrogent sur les processus de production et compréhension des phrases, etc. Les seconds transposent les analyses des linguistes dans des séquences pédagogiques.

Cette partie s'intéresse aux objectifs du jeu pour les apprenants.

1.1. Historique de l'idée

Le découpage en petits groupes mis en œuvre dans GR vient du concept de *chunker* développé dans les années 1990 dans le cadre du traitement automatique des langues. Plutôt qu'une analyse complète, le *chunker*, ou « analyseur de surface », produit un découpage de la phrase en tronçons (*chunks*) qui se succèdent sans s'emboîter les uns dans les autres, contrairement aux constituants de l'analyse en constituants immédiats. L'article fondateur de la notion de *chunk* a été publié par Abney en 1991. Vergne (1998, 1999) a réalisé des travaux précurseurs pour l'analyse automatique du français dans un cadre proche. Les analyseurs de surface se sont multipliés dans les années 2000 (voir quelques références dans Trouilleux, 2009). En 2004 a été organisée une campagne d'évaluation des analyseurs syntaxiques (EASY), pour laquelle le standard d'annotation des corpus adoptait un découpage en tronçons, appelés « groupes syntaxiques », et des relations entre ces groupes (Gendner et Vilnat, 2004). Une deuxième campagne d'évaluation, PASSAGE, sera organisée quelques années plus tard, avec le même schéma d'annotation (Vilnat *et al.*, 2010).

L'idée du jeu m'est venue face à un public d'étudiants de L3 Sciences du langage (SDL). Dans le cadre d'un cours d'initiation au traitement automatique des langues (TAL), on envisage le développement d'un analyseur syntaxique de surface pour le français. Première étape : on présente le schéma d'annotation des campagnes d'évaluation EASY et PASSAGE et on demande aux étudiants d'annoter quelques phrases selon ce schéma (seulement les groupes). On constate alors que malgré plus de deux ans d'enseignements en SDL, les étudiants peinent à accomplir correctement la tâche. L'entraînement améliore les résultats et les étudiants en viennent à la conclusion que faire cet exercice au début de leur cursus les aurait aidés. J'ai donc proposé ce type d'activité aux étudiants de L1, mais sur papier, sans passer par une initiation au TAL.

Les groupes définis dans une analyse de surface étant non emboîtés les uns dans les autres, sur papier, on peut simplement demander de former les groupes en insérant des barres verticales entre eux :

L'équipe | nationale | de hockey | a remporté | le championnat | du monde | cette année.

Une particularité des groupes est qu'ils suivent des schémas très réguliers en termes de catégories (voir plus loin 2.3.4). L'exercice peut alors être augmenté par une tâche consistant à assigner à chaque mot une étiquette de catégorie (Déterminant, Nom, Adjectif, etc.), comme sur la figure 2. Le caractère très répétitif du schéma (PREP) D N est en particulier intéressant pour aborder la catégorie des déterminants.

L'	équipe		nationale		de	hockey		a	remporté		le	championnat		du	monde		cette	année.	
D	N		A		PREP	N		AUX	PP		D	N		PREP	D	N		D	N

Figure 2. Notation des groupes et des catégories sur papier.

L'autre aspect d'une analyse à la GR est le rattachement des groupes. La figure 3 ci-dessous montre comment les rattachements peuvent être notés sur papier. Chaque groupe reçoit un identifiant numérique (première colonne), qui est utilisé dans la troisième colonne pour indiquer à quel groupe se rattache chaque groupe. L'exercice consiste à remplir la colonne 3. Les mots *sujet* et *verbe* indiquent le début du groupe sujet et du groupe verbal, les chiffres indiquent la cible du rattachement.

1	L'équipe	<i>sujet</i>
2	nationale	1
3	de hockey	1
4	a remporté	<i>verbe</i>
5	le championnat	4
6	du monde	5
7	cette année.	4

Figure 3. Notation des rattachements sur papier.

Une proposition de stage de la part d'Alain Régnier, docteur en linguistique et qui suivait une formation d'ingénieur, a été l'occasion de développer le jeu interactif.

1.2. Publics cibles

Le public pour lequel a été conçu le jeu est celui des étudiants qui entrent à l'université. Dans une très large mesure, ils et elles ne maîtrisent pas les concepts de base de la grammaire. Une simple tâche d'étiquetage de corpus telle que celle de la figure 2 est inenvisageable de prime abord. Soit la chose a été vue et oubliée, soit elle n'a jamais vraiment été maîtrisée.

A la question « De quoi le groupe à Marie est-il complément dans la phrase *Pierre a donné un livre à Marie ?* », 3,3 % de 91 étudiants entrant en L1 SDL ont choisi la réponse « je ne sais pas du tout », 45,1 % ont choisi « complément du nom *livre* », 51,6 % ont choisi « complément du verbe *a donné* ».

Le corpus d'exercices disponible sur le site de GR au moment de la rédaction de ce rapport est sélectionné pour ce public de nouveaux bacheliers, qui ont déjà des compétences en grammaire. On pourra bien sûr envisager d'autres publics. Rien n'empêche, par exemple, que le jeu puisse intéresser des élèves de l'école primaire, pourvu que les phrases utilisées soient à leur portée et s'inscrivent dans une progression pédagogique adéquate.

Il n'est pas non plus exclu que le programme puisse apporter quelque chose aux apprenants étrangers. Des étudiants chinois nous ont dit tirer profit d'exercices sur papier de découpage en groupes ou d'assignation de catégories. Pour un étranger, le jeu permet de s'imprégner des schémas récurrents de la langue, avec ou sans le métalangage des catégories.

Les profils d'apprenants sont donc potentiellement variés. Chaque type de public demandera des corpus spécifiques, organisés dans une progression particulière. La création de nouveaux exercices est très facile. Il y a là un espace de création pour les didacticiens et pédagogues (voir ci-dessous 4.2).

1.3. Objectifs

L'objectif général est de travailler sur la notion de structure, qui est elle-même sous-tendue par des connaissances particulières.

1.3.1. Travail général sur la notion de structure

Le volet *Etude de la langue* des programmes officiels pour l'école primaire (*Bulletin officiel de l'éducation nationale*, 2015 : 23), pour le cycle 2, contient ces directives :

Les élèves apprennent progressivement à pratiquer des observations, à entrer dans des réflexions organisées sur le fonctionnement de la langue pour commencer à acquérir les notions fondamentales d'un enseignement qui se poursuit jusqu'à la fin du collège. [...] Les élèves sont conduits à centrer leur attention sur la forme de l'énoncé lui-même, à relativiser certains aspects sémantiques pour privilégier un regard sur la formation des mots (la morphologie) et sur les relations entre les mots (la syntaxe). [...] Des activités ritualisées fixent et accroissent les capacités de raisonnement sur des énoncés et l'application de procédures qui s'automatisent progressivement. Des séances courtes et fréquentes sont donc le plus souvent préférables à une séance longue hebdomadaire.

Réflexions sur le fonctionnement de la langue, regard sur les relations entre les mots, activités ritualisées, procédures qui s'automatisent : GR nous semble pouvoir trouver une application dans ce cadre et dans les cycles d'apprentissage suivants.

Dans ses *Eléments de syntaxe structurale*, Tesnière (1959 : 655) raconte comment une institutrice invite ses élèves à « dessiner la phrase » avec les « stemmas » proposés par le grammairien. Plus que les schémas de Tesnière, ce sont les représentations en arbres qui ont la faveur des grammaires aujourd'hui. Mais une telle représentation est peu pratique dès qu'une phrase est un peu longue. Genevay *et al.* (1985) précisent que pour les élèves de 7^e année, il faut l'utiliser en le présentant de façon plus ou moins développée et qu'il est recommandé aux enseignants de ne jamais demander aux élèves de dessiner eux-mêmes l'arbre correspondant à la phrase donnée (cité par Canelas-Trevisi, 2009, p. 114). La simplicité de présentation de GR permet de lever ces restrictions : on peut demander à un apprenant de produire la structure d'une phrase arbitrairement longue.

1.3.2. Notions sous-jacentes à la structure

On évoquera plus loin (partie 3) des aspects de la structure adoptée dans GR qui pourront faire débat. Indépendamment de ces aspects, on peut noter ici que GR fait travailler des notions qui sont sous-jacentes à l'analyse quelle que soit la structure qu'on adopte. Ainsi, pour former les groupes, il faut notamment :

- distinguer les mots lexicaux et les mots grammaticaux,
- identifier le noyau lexical d'un groupe (verbal, nominal, adjectival ou adverbial),
- distinguer les pronoms forts des pronoms faibles.

Ces trois points sont développés dans les règles exposées sur le site de GR. Notons que les deux premières compétences incluent la maîtrise des concepts de préposition (p. ex. *après la pluie* = un groupe et non deux), de déterminant (p. ex. *trois cents résiniers* = un groupe et non deux) et de verbe auxiliaire.

Pour ce qui est de rattacher les groupes, le site de GR ne donne aucune règle. Il suppose connues les notions de sujet, verbe et complément, ainsi que la notion de coordination.

GR vise à *aider* à travailler sur la syntaxe du français, il ne prétend pas être l'alpha et l'oméga de l'exercice sur la structure de la phrase. On peut le voir comme spécifiant un objectif à atteindre, sans qu'il contienne en lui tous les moyens pour l'atteindre. **Le jeu doit s'inscrire dans une progression pédagogique** qui, d'une part, inclue des corpus d'exercices adaptés au niveau de l'apprenant, d'autre part envisage d'autres angles d'approche. Par exemple, en parallèle aux analyses *bottom up* de GR (des mots à la structure totale), ou avant elles, on peut travailler sur des analyses *top down* : décomposition de la phrase en sujet / prédicat, identification des compléments du verbe, analyse des constituants immédiats de groupes nominaux, etc.

2. Justification de l'approche

On l'a dit plus haut (§ 1.1), le jeu GR est inspiré de travaux dans le domaine du traitement automatique des langues (TAL). Le TAL fournit aussi au jeu sa problématique, qui peut se formuler ainsi : comment construire mot après mot la structure qui correspond à l'interprétation de la phrase ? Il fournit également la structure en groupes de surface qui se rattachent les uns aux autres.

2.1. Problématique du rattachement

Une problématique apparue avec les travaux en TAL est celle de l'ambiguïté de rattachement. Le problème est bien connu en cas d'ambiguïté réelle ; on en parle depuis au moins Vaugelas, qui dans une remarque *De l'équivoque*, commente, entre autres phrases, *L'orateur arrive à sa fin, qui est de persuader, d'une façon toute particulière* (Vaugelas, 1647 : 588). « L'intention de celui qui parle ainsi, écrit Vaugelas, est que ces mots *d'une façon particulière*, se rapportent à ceux-ci *arrive à sa fin*, et néanmoins comme ils sont placés, il semble qu'ils se rapportent à *persuader*. »

Ce qu'ont montré les tentatives de traitement automatique des langues, c'est l'ambiguïté de rattachement qui se cache dans des phrases apparemment non ambiguës. La formalisation a permis le calcul et le calcul a montré un problème dont on ne soupçonnait pas l'étendue.

Avec une grammaire formelle syntagmatique produisant des arbres binaires standards où les syntagmes prépositionnels peuvent se rattacher aux nœuds P, SV, V-barre, V, N-barre ou N (tous cas attestés dans les descriptions standards), on obtient 41 128 arbres différents pour la phrase suivante.

1. Le voyageur a convenu avec la secrétaire de l'agence de voyage de l'horaire de départ du train pour Paris.

Le défi pour les systèmes de TAL est toujours à l'heure actuelle de sélectionner adéquatement la bonne structure dans le vaste ensemble de structures que produisent des descriptions encore incomplètes. GR pose les questions qui se posent au TAL : quels sont les rattachements à obtenir ? Comment fait l'utilisateur pour obtenir la bonne structure ? Quelle information utilise-t-il, que n'ont pas les systèmes de TAL ? Le jeu n'apporte pas les réponses, mais invite à se poser les questions.

2.2. Problématique de la construction incrémentale

Une idée fondamentale, dans GR, est qu'on peut obtenir les structures du type proposé en les construisant *mot par mot*. Ferreira et Henderson (1991) ont avancé l'idée que la construction de l'interprétation de la phrase se faisait progressivement et intervenait aux moments où l'on rencontre des têtes de syntagmes. Chacun a pu rencontrer au cours d'une lecture une « phrase labyrinthe », c'est-à-dire une phrase pour laquelle il ou elle a dû revoir la façon dont il fallait la comprendre. Par exemple, on peut lire au prime abord la phrase suivante en prenant le premier groupe pour le sujet de la phrase ¹ :

2. Des pétunias en bordure de l'allée montait par bouffées une odeur de cassis.

L'existence des phrases labyrinthes est un indice fort du fait que l'interprétation des phrases se fait de façon incrémentale. C'est pourquoi, plutôt que de montrer une phrase à analyser dans son ensemble, le jeu propose de faire l'analyse mot par mot.

Bien sûr, dans une progression pédagogique particulière, il pourra être intéressant d'observer et d'analyser des phrases complètes, mais l'objectif de produire l'analyse mot à mot est intéressant. Au besoin, on peut afficher la phrase complète en cours d'analyse dans GR en passant la souris sur le coin supérieur gauche.

Notons que l'idée d'une construction incrémentale de la structure de la phrase se démarque de la pratique, commune en linguistique, qui consiste à tester des déplacements ou des effacements pour déterminer le rattachement d'un syntagme. On se démarque aussi de toute analyse qui expliquerait en termes de « mouvement » une observation. On suit en cela ce que font en général les systèmes de TAL, au contraire des approches chomskyennes (cf. Culicover, 2005 : 16).

2.3. Justification de la structure

2.3.1. Multiplicité des représentations

Tout le monde s'accorde à dire que les phrases ont une structure, mais les façons de la représenter sont multiples. Certains adoptent une représentation par des dépendances entre mots (cf. par exemple Tesnière, 1959 ; Martinet, 1979 ; Mel'čuk, 1988), d'autres une représentation par des arbres de constituants (Bloomfield, 1933 ; Wells, 1947). Chomsky (1957), montrant une supposée inadéquation des arbres de l'analyse en constituants immédiats (ACI), a proposé d'analyser certaines phrases par recours à des trans-

¹ Le correcteur grammatical de mon logiciel de traitement de texte m'invite à écrire *montaient*.

formations. Par la suite, il a préconisé des analyses où les constituants se déplacent. La notion de déplacement étant absente de GR, on évoquera ici seulement les deux premiers types d'analyse.

La figure 4 montre une analyse en dépendances. Elle a été produite par l'analyseur FRMG² (La Clergerie, 2005) et est exprimée dans les termes du modèle *Universal dependencies*³. La caractéristique essentielle de ce modèle est qu'il exprime des relations entre les *mots* pour former un graphe qui connecte l'ensemble de la phrase ; chaque mot est cible d'une seule flèche. Chaque flèche part d'un mot, sauf une qui identifie le mot considéré comme le noyau de la phrase (*remporté* dans la figure 4). Les flèches reçoivent des étiquettes qui distinguent les différentes fonctions des mots. Différents modèles d'analyse en dépendances se distinguent par les flèches utilisées et les étiquettes qu'on leur associe (p.ex. certains diront que *équipe* est sujet de *a* plutôt que de *remporté*).

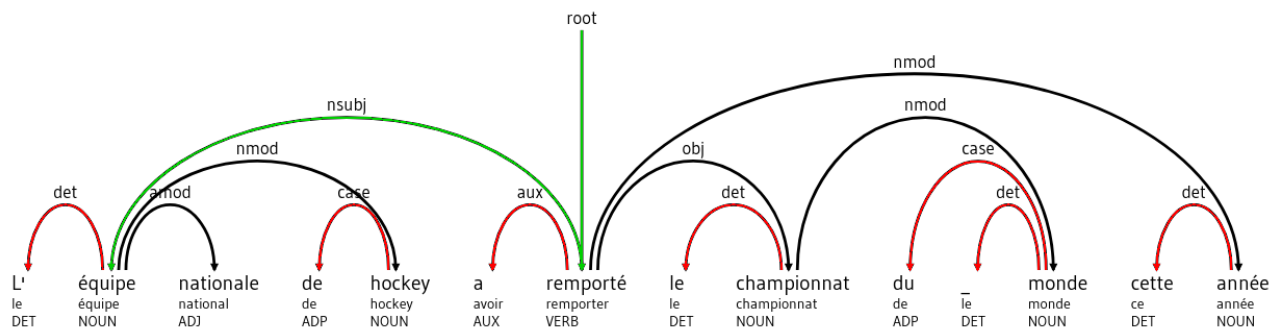


Figure 4. Analyse d'une phrase en dépendances.

La figure 5 montre une analyse de la même phrase sous la forme d'un arbre de constituants. Comme pour les dépendances, d'autres façons de faire certaines parties de l'arbre peuvent être envisagées. P. ex. certains auront le souci d'éviter des branchements ternaires et regrouperont *équipe* et *nationale* en un constituant auquel *de hockey* sera ensuite associé⁴ ; de même le constituant *cette année* pourrait être rattaché à un constituant regroupant la « phrase minimale » formée par le groupe nominal sujet et le groupe verbal. La figure 6 illustre ces choix possibles.

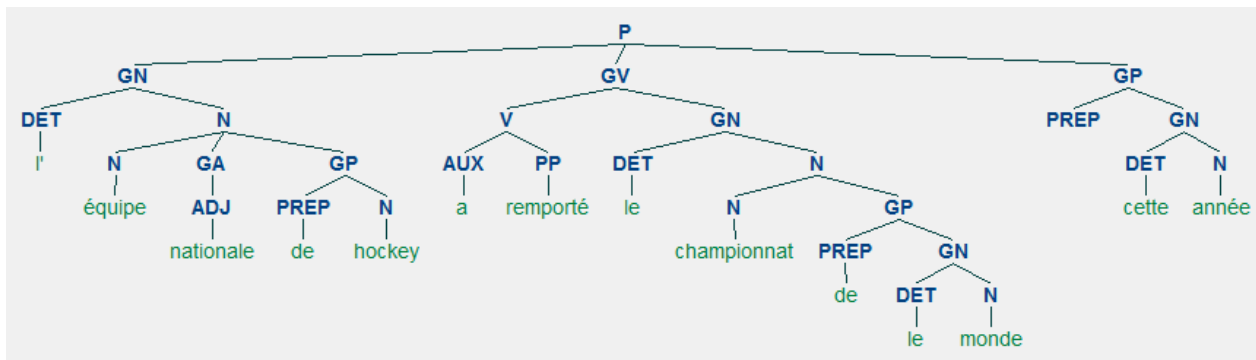


Figure 5. Analyse d'une phrase en constituants.

² http://alpage.inria.fr/frmgwiki/frmg_main/frmg_server

³ <https://universaldependencies.org/>

⁴ Les adeptes des déplacements de constituants pourront vouloir rattacher d'abord *de hockey* à *équipe*, puis *nationale* à *équipe de hockey*. Pour représenter cela avec un arbre, il faut changer l'ordre des mots en *l'équipe de hockey nationale*.

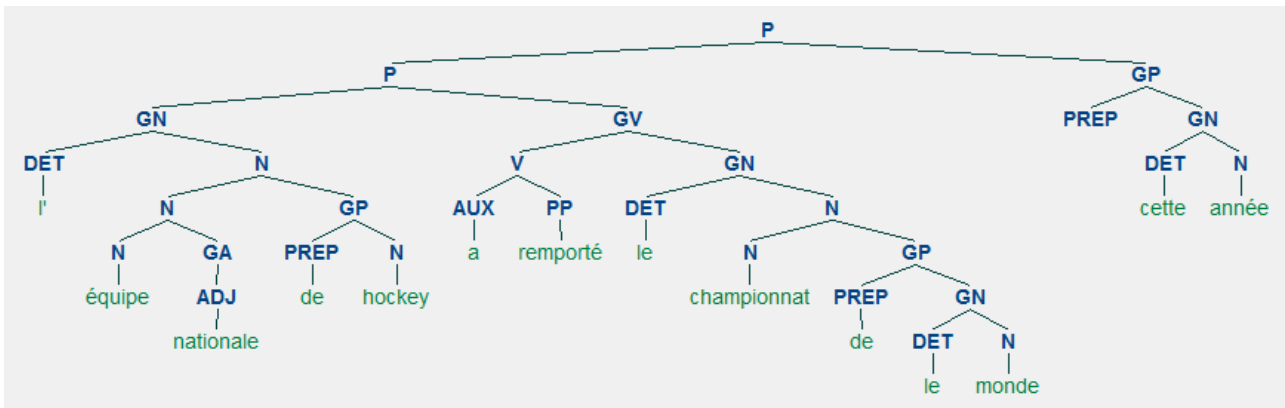


Figure 6. Analyse d'une phrase en constituants (version alternative)

La représentation utilisée dans GR s'inscrit dans la tradition des analyses en dépendances telles qu'illustrées par la figure 4. Pour la phrase en question, les groupes de GR sont définis par les relations **det**, **case** et **aux** (en rouge sur la figure 4), tandis que les relations **obj** et **nmod** (en noir sur la figure 4) s'expriment dans GR par les traits bleus qui signifient « complément de ». La relation sujet-verbe (**nsubj**) s'exprime quant à elle dans GR par une distinction de couleur entre le groupe sujet et le groupe verbal.

2.3.2. Equivalence des représentations

Chomsky (1957) a popularisé l'idée que la tâche des linguistes devait être de déterminer quelle est la structure adéquate à associer aux phrases. Le but de la théorie linguistique est de déterminer laquelle de deux grammaires est la meilleure (Chomsky, 1957 : 59) et comme une grammaire, dans les travaux de Chomsky, est un objet qui assigne des structures aux phrases, dans cette optique, il ne peut y avoir qu'une façon de représenter la structure des phrases.

A rebours de cette problématique, cependant, on peut observer que dans une large mesure, il est possible de définir un processus de traduction entre les trois langages de représentation de la structure que sont les dépendances, la représentation de GR et les arbres de constituants, et que par conséquent, il importe de distinguer l'information sous-jacente à ces représentations des représentations elles-mêmes. Des problèmes de traduction peuvent se poser dans la mesure où les représentations ne sont pas forcément aussi *riches*. Par exemple, dans GR, les étiquettes des relations ne sont pas explicites – mais elles pourraient l'être.

Prenons par exemple la notion de sujet, que codent les trois représentations :

- Un adepte des arbres de constituants dira que le sujet de la phrase est le groupe nominal *l'équipe nationale de hockey*.
- Dans GR, c'est après le groupe *l'équipe* qu'on appuie sur **s** pour dire qu'on a identifié le sujet.
- Dans la représentation en dépendances, c'est le seul mot *équipe* qui semble être sujet.

On a donc apparemment trois façons de voir les choses. Cependant, derrière ces différences de façade, il y a les mêmes idées :

- le constituant *l'équipe nationale de hockey*, c'est le groupe *l'équipe* et tous ses compléments (dans GR) ou le nom *équipe* et tous les mots qui dépendent de lui (dans la notation en dépendances) ;
- inversement, le nom *équipe* qui porte la relation **nsubj** dans la notation en dépendances ou qui termine le groupe *l'équipe* dans GR est le noyau du constituant *l'équipe nationale de hockey*, ce qui se voit dans l'arbre par le fait que le nœud N fils du nœud GN est défini récursivement par référence au nœud N qui domine *équipe*.

On retrouve donc les mêmes notions dans les différentes représentations.

2.3.3. *Petits et grands groupes*

Pour ce qui est de la représentation de GR par rapport aux arbres de constituants, on peut poser cette relation :

Un groupe de GR et tous ses compléments est un constituant dans une représentation en arbre de constituants.

La conséquence de cette relation est qu'on peut voir les opérations qu'on fait dans GR (désigner le sujet, le verbe, rattacher un complément ou un groupe coordonné) comme s'appliquant aussi bien au groupe formé au moment de l'opération qu'au groupe étendu formé de ce groupe et de ses compléments. Quand on appuie sur *s* pour dire que *L'équipe* est sujet, on dit aussi que le constituant formé de ce groupe et ses compléments est sujet.

On voit là qu'on a deux types de groupes différents, que je propose d'appeler les **petits groupes** et les **grands groupes**. Les petits groupes sont produits suivant les principes exposés dans la page *Former les groupes* du site GR (association des mots grammaticaux aux mots lexicaux, etc.) ; les grands groupes sont produits par le rattachement des petits groupes les uns aux autres. Ainsi, par exemple, dans notre phrase exemple, *l'équipe* est un « petit groupe » et *l'équipe nationale de hockey* est un « grand groupe ».

Cette terminologie permet de faire cohabiter dans une large mesure la représentation de GR avec l'analyse en constituants (voir plus loin, partie 3, pour quelques divergences). On pourra en chercher une plus savante, par exemple groupes **noyaux** vs **étendus**, groupes **simples** vs **complexes**, ou encore, en suivant Beauzée (1767, t. 2 : 55-56) groupes **grammaticaux** (les petits) vs **logiques** (les grands)⁵. Les adjectifs *petit* et *grand* ont l'avantage d'être parlants pour les enfants. En tout état de cause, la thèse sous-jacente à GR est qu'il est utile de faire la distinction : cela permet de dissocier deux dimensions de la syntaxe qui n'ont pas les mêmes propriétés.

2.3.4. *Régularité distributionnelle des petits groupes*

La particularité des petits groupes est que leur formation suit une logique distributionnelle assez stricte. L'ordre des mots dans les petits groupes n'est pas libre, tandis que la mise en ordre des petits groupes admet des variations :

3. Contre nous / de la tyrannie / l'étendard / sanglant / est levé.
L'étendard / sanglant / de la tyrannie / est levé / contre nous.

Autre exemple : on peut dire aussi bien *sortir de l'ombre les créatrices* ou *sortir les créatrices de l'ombre*)

La formation des petits groupes est aussi non récursive et peut être décrite par une grammaire à états finis, le type le plus simple dans la hiérarchie de Chomsky. En comparaison, la formation des grands groupes met en œuvre une récursivité⁶.

Pour sentir cette différence entre les deux dimensions, on peut essayer de raisonner en ne prenant en compte que la catégorie grammaticale des mots. Les quarante cinq phrases des neuf exercices du site GR version 1 contiennent au total 482 petits groupes. Le tableau 1 ci-dessous donne la répartition de ces groupes en

⁵ Je signale la terminologie de Beauzée surtout pour indiquer qu'il faisait une distinction analogue à celle qui est faite ici. Reprendre ces termes ne me semblerait pas une bonne idée. Beauzée emploie le mot *grammatical* dans le sens de « qui, dans les langues qui admettent la déclinaison, est assujéti à prendre telle ou telle forme en qualité de complément » et *logique* dans le sens de « qui exprime l'idée totale que la raison envisage comme le vrai complément ».

⁶ On peut définir un grand groupe avec ces deux règles :

- Un petit groupe sans complément est aussi un grand groupe.
- Un petit groupe ayant pour complément un grand groupe forme avec lui un grand groupe.

La deuxième règle, qui fait appeler à la notion de grand groupe pour définir un grand groupe, est récursive.

fonction du type de noyau ⁷ (colonne i) et donne une expression régulière qui, utilisant des codes de catégories grammaticales ⁸, couvre l'ensemble des groupes considérés.

<i>Expression régulière</i>	<i>Qt</i>	<i>Exemples</i>
ADV? ADV	9	<i>assez mal</i>
(ADV? ADJ) PP (<i>sans auxiliaire</i>)	54	<i>plus facile</i> (ADV ADJ), <i>sortis</i> (PP)
CC? PREP? NEG? ADV? P-* (V (AUX? AUX PP))	86	<i>à ne plus poursuivre</i> (PREP NEG ADV V) <i>je les ai dénoncés</i> (P- P- AUX PP) <i>et se coiffait</i> (CC P- V) <i>a été condamné</i> (AUX AUX PP)
CC? (ADV? PREP)? D* ADJ* (N-p+ N)	323	<i>autour des coureurs</i> (ADV PREPD N) <i>toute la journée</i> (D D N) <i>de Mme Pelletier</i> (PREP N-p N-p) <i>sur deux belles grandes scènes</i> (PREP D ADJ ADJ N)
PREP? P\+	10	<i>dans chacune</i>

Tableau 1. Expressions régulières décrivant les groupes du corpus GR-1.

Bien sûr, le corpus du site de GR n'est pas parfaitement représentatif de l'ensemble des possibles (par exemple, il peut y avoir un adverbe entre l'auxiliaire et le participe passé), mais la tendance exprimée par ces expressions régulières est forte.

Pour appréhender cette régularité et éventuellement la faire travailler à des élèves, on peut paramétrer le jeu de façon à afficher des codes de catégorie au lieu des mots. La page *Aller plus loin* du site GR propose de tester ce mode, sous le titre *L'expérience de l'asémantisme*. En perdant les mots, on a perdu l'accès au sens, et on peut dire alors qu'on fait une analyse qui n'est que syntaxique ⁹. L'expérience montre que dans ce mode, quelqu'un qui a une connaissance des parties du discours et des principes de base de la formation des groupes *peut* faire les groupes, mais *pas* bien les rattacher. On a là un indice qu'il est pertinent de distinguer les deux dimensions que distingue le jeu et une preuve que le sens intervient de façon essentielle dans la détermination des rattachements (sur ce point, voir Trouilleux, 2021, § 4.3).

3. Quelques points de discussion

Le propos de ce rapport de recherche est plus de présenter GR que de se lancer dans des discussions poussées sur l'analyse linguistique. Cependant, comme certains seront sans doute dubitatifs face à certains points de l'analyse adoptée dans le jeu, j'évoque ici quelques points de discussion.

3.1. Absence du complément de phrase

Depuis les années 1960-1970, il est d'usage de considérer que les phrases *Mon père lit le journal* et *Mon père lit le soir* n'ont pas la même structure, comme le montre la figure 7.

⁷ ADV = adverbe, ADJ = adjectif, PP = participe passé, V = verbe, AUX = verbe auxiliaire, N-p = nom propre, N = nom commun, P+ = pronom fort.

⁸ CC = conjonction de coordination, PREP = préposition, D = déterminant, NEG = particule *ne*, P- = pronom faible (clitique).

⁹ Pour avoir une analyse purement syntaxique, il faudrait que l'ensemble des catégories possibles pour chaque mot soit donné. Ici, le simple fait que l'ambiguïté adjectif-nom soit résolue a priori est un indice crucial pour la formation des groupes.

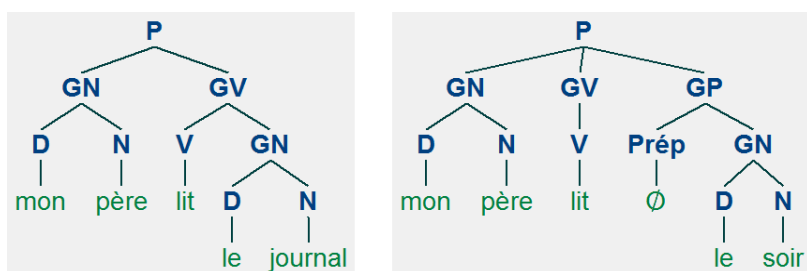


Figure 7. Deux structures différentes pour deux phrases de même forme apparente.

Dans GR, les deux phrases en question auront exactement la même représentation. Les groupes *le soir* et *le journal* sont l'un et l'autre complément du verbe *lit*. On considère que la différence entre les deux phrases résulte de la différence de sens entre *le soir* et *le journal*, qui appartiennent respectivement aux paradigmes « temps », représenté par exemple par les adverbes *quand* ou *aujourd'hui*, et « objet », représenté par les pronoms *quoi* et *ceci*.

Cette question a fait l'objet d'un article spécifique : *Critique du complément de phrase* (Trouilleux, 2021).

3.2. Pronom clitique sujet

On a dit plus haut que les grands groupes de GR correspondaient à des constituants de l'analyse en constituants immédiats. Par contre, les petits groupes ne correspondent souvent pas à un constituant. Une première divergence réside dans le traitement du pronom clitique sujet.

Classiquement, l'analyse en constituant traite le pronom clitique sujet comme le groupe nominal sujet (voir figure 8). Dans GR, le pronom sujet est inclus dans un petit groupe avec le verbe (voir figure 9). Le groupe *il lit* de la figure 9 ne correspond à rien dans l'arbre de droite de la figure 8.

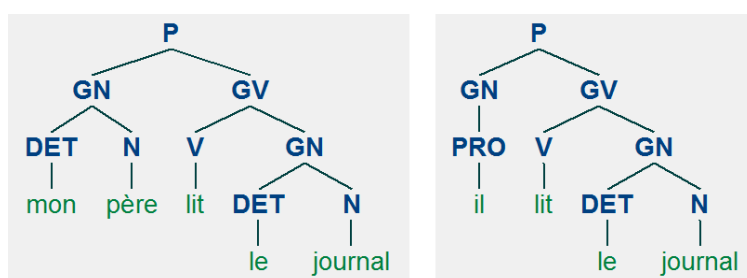


Figure 8. Analyse du pronom clitique sujet dans l'ACI.

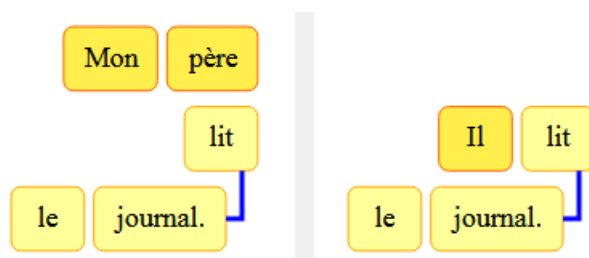


Figure 9. Analyse du pronom clitique sujet dans GR.

L'inclusion du pronom clitique sujet dans un petit groupe syntaxique avec le verbe était déjà le choix des campagnes EASY et PASSAGE, mais aucune justification à ce choix n'était donnée. On en trouvera une très parlante dans Creissels (1995 : 22-31), dont les idées se retrouvent dans le discours qui suit.

Dans les termes de Wells (1947), *mon père* est une « expansion » de *il*, et *il* est un « modèle » de *mon père*. Les deux expressions appartiennent à la même « classe focus », une classe focus étant un ensemble de séquences de mots qui peuvent apparaître dans un même environnement. Wells représente focus et

environnement ainsi : () *lit le journal*, où () représente le focus. On pourrait généraliser ce focus particulier à () *GV*.

Dans les arbres de la figure 8, le nœud GN fils de P désigne le focus qui nous intéresse. Si on suit la logique de cette notation, alors partout où il y aura le GN *mon père*, on pourra mettre le GN pronominal *il*. Or il y a des environnements très communs où *mon père* peut apparaître, mais pas *il*, par exemple :

4. mon père et ma mère * il et ma mère
pour mon père * pour il

Il se trouve qu'il existe un « modèle » de l'expression *mon père* qui est valable dans plus d'environnements que *il* ; c'est le pronom fort *lui* :

5. lui et ma mère
pour lui

Le pronom fort *lui* est aussi possible en position sujet, si bien que la phrase *Lui lit le journal* est grammaticale. Voici un exemple authentique de *lui* sujet :

6. Il y a une entraide entre arbres et champignons : les champignons apportent à l'arbre de l'eau, et lui fournit des sucres aux champignons, de l'énergie.

« Quand on choisit les classes focus principales qui seront utilisées pour décomposer chaque énoncé en constituants, écrit Wells (1947 : 88), il faut prendre en considération le système entier que ces classes focus, prises collectivement, forment les unes avec les autres. » Suivant ce principe, on doit traiter les pronoms forts comme les groupes nominaux et les pronoms faibles comme des éléments associés au verbe. Il en résulte que les deux phrases de la figure 10 se distinguent structurellement.

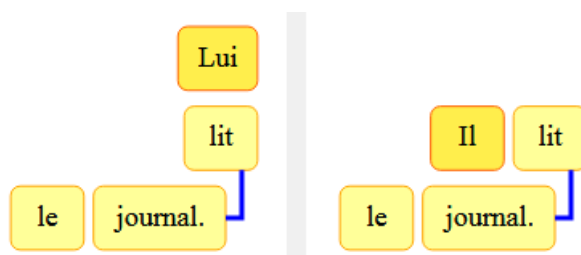


Figure 10. Traitement différencié des pronoms sujets dans GR.

On pourra objecter que la notation GN pour le sujet des phrases de la figure 8 ne doit pas être prise au sens littéral. « Au niveau de P, précise Monneret (1999 : 190), les symboles SN et SP [= GN et GP dans ce document] ont un sens fonctionnel et non morphologique, puisqu'ils représentent respectivement le sujet et le complément circonstanciel de la phrase. La forme SN constitue seulement la forme typique du sujet, comme le SP constitue la forme typique du complément circonstanciel. » Plus récemment, sont proposés des arbres où un symbole identifiant explicitement le sujet est proposé au lieu du symbole GN de la figure 8. Par exemple, Monneret et Poli (2020) décomposent la phrase en GS + GV + GC, soit groupe sujet, groupe verbal, groupe circonstanciel. Dans cette optique, le pronom *il* trouve mieux sa place sous GS, mais cela n'invalide pas la pertinence de la distinction entre pronoms forts et faibles. L'équivalence des pronoms faibles avec les groupes nominaux est fonctionnelle et non distributionnelle. On a classé les « adjectifs possessifs » (*mon, ton, son* etc.) comme des déterminants pour des raisons distributionnelles (on aurait pu les classer comme pronoms, comme faisait Vaugelas ou comme font les Anglo-Saxons pour leurs équivalents). Le même genre de raisonnement suggère de traiter *il* et les pronoms clitiques comme particuliers.

3.3. Structure du groupe nominal

Autre divergence entre les petits groupes de GR et les arbres d'analyse en constituants : le traitement du déterminant dans le groupe nominal. Prenons par exemple l'arbre de constituants suivant, repris de Riegel *et al.* (1994 : 113).

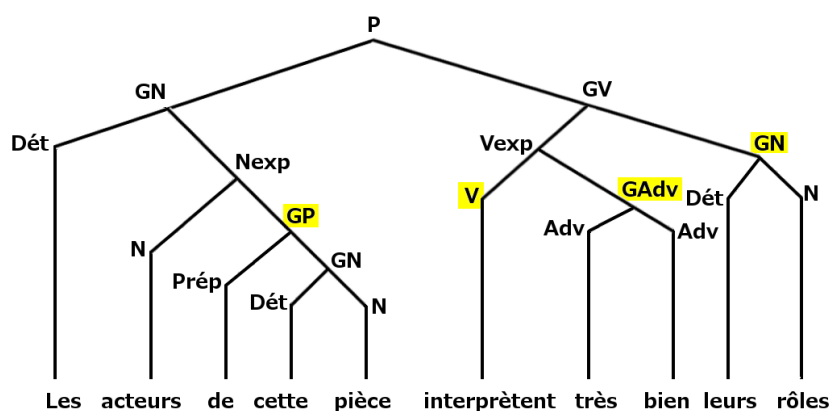


Figure 11. Exemple d'arbre d'analyse en constituants immédiats de Riegel *et al.* (1994).

La phrase analysée se décompose en cinq petits groupes :

Les acteurs / de cette pièce / interprètent / très bien / leurs rôles.

On retrouve quatre de ces groupes dans l'arbre de la figure 11 : ce sont les séquences dominées par un nœud surligné en jaune. Par contre, on ne retrouve pas de nœud dominant la seule séquence *les acteurs*, mais un nœud *Nexp* dominant *acteurs de cette pièce*.

Dans un article sur les constituants immédiats qui fait référence¹⁰, Wells (1947 : 89) pose la question : « la séquence *the king of England* doit-elle être analysée comme *the king / of England* ou comme *the / king of England* ? C'est la deuxième option qui est retenue par lui et depuis lors.

On aurait tort, cependant, d'en faire une évidence. Sans épuiser le sujet, qui mériterait des études approfondies, je donne ici trois éléments qui invitent à prendre en considération le découpage adopté dans GR – sans forcément rejeter le découpage couramment admis.

3.3.1. Des linguistes hésitants

Wells lui-même, qui pose les principes de l'ACI, hésite entre les deux analyses. Il adopte au début de son article (1947 : 89) les découpages *the / king of England* et *the / English king*, mais quand il envisage plus tard des constituants discontinus (p. 104), il écrit :

L'analyse de *the English king* en *English* et *the ... king* donne des constituants d'une plus grande indépendance et mobilité que l'analyse en *the* et *English king*. De plus, le schéma de *poor / John* et *English / literature* serait mieux imité. Mais si l'admission de constituants multiples et discontinus n'était sujette à aucune autre restriction que de produire des constituants qui soient maximalelement indépendants, l'analyse en constituants immédiats deviendrait une affaire extrêmement intriquée. Les possibilités à étudier seraient de beaucoup multipliées.

A la suite de quoi, il pose le principe qu'il faut au maximum faire des constituants continus et n'adopter des constituants discontinus qu'en dernière extrémité. Il pose une condition sur les constituants discontinus¹¹ et constate plus loin (p. 106) :

¹⁰ Il est notamment cité par Chomsky (1957). Graffi (2001 : 285) le décrit comme « la première vraie tentative systématique pour mettre en place une procédure explicite pour délimiter les constituants ».

L'analyse de *the English king* en *English* et *the ... king* satisfait la condition que nous proposons pour les constituants discontinus. Si cette analyse est acceptée, alors l'analyse de *the king of England* demande à être réexaminée. Car un des arguments en faveur de *the / king of England* était sa conformité avec *the / English king*. Le même argument, appliqué à la nouvelle analyse *English + the ... king*, tolèrerait aussi bien *the king / of England*, ce qui après tout semble être l'analyse que préfère le sens commun.

Peut-être est-ce aussi le sens commun qui fait noter à Riegel *et al.* ceci, dans leur commentaire de l'arbre de la figure 11 (1994 : 113-114) :

Comme les mots, les groupes syntaxiques appartiennent à des catégories qui se distinguent par leur structure interne et en particulier par leur catégorie-tête (ou constituant principal qui détermine la nature du groupe : nom, verbe, adjectif, adverbe). Les syntagmes *les acteurs*, *cette pièce* et *leurs rôles* sont donc classés dans la catégorie des groupes nominaux ainsi dénommés parce que leur catégorie-tête est un nom [...].

Désigner *les acteurs* plutôt que *les acteurs de cette pièce* comme GN de l'arbre de la figure 11 est une erreur. Le fait qu'elle se produise montre que le découpage *les acteurs / de cette pièce* mérite l'attention ¹².

3.3.2. Des commutations peu concluantes

En principe, l'analyse en constituants immédiats s'appuie sur des commutations. Pour analyser une séquence comme un constituant, on cherche une séquence plus courte (idéalement un seul mot) qui commute avec elle.

Il existe en français une commutation qui justifie le découpage à la GR :

- 7. les acteurs | de cette pièce
ceux | de cette pièce

Pour le nœud NExp de l'arbre de la figure 11, il paraît facile de trouver des « modèles ». Tout nom simple peut faire l'affaire :

- 8. les | acteurs de cette pièce
les | interprètes
les | enfants
les | comédiens

Le problème est que l'évidence de ces commutations n'est qu'apparente. Le syntagme *de cette pièce* peut être effacé après *acteurs*, ce qui fait que les commutations de l'exemple 8 peuvent se réduire à une commutation de *de cette pièce* avec \emptyset (la chaîne vide) + des commutations du nom *acteurs* avec d'autres noms simples, qui ne disent rien sur la structure. Seules la pronominalisation par *en* ou l'ellipse peuvent justifier le constituant NExp de la figure 11 :

- 9. Jacques connaît deux acteurs de cette pièce ; Marie en connaît trois.
Deux acteurs de cette pièce ont reçu un prix, tous les autres \emptyset ont été très mauvais.

En termes stricts de commutation, il y a donc des arguments pour les deux approches, mais ils sont assez exceptionnels.

¹¹ « Une séquence discontinue est un constituant si dans un environnement, la séquence continue correspondante apparaît comme un constituant dans une construction sémantiquement harmonieuse avec la construction dans laquelle la séquence discontinue en question apparaît. » (Wells, 1947 : 104)

¹² L'erreur figure toujours dans la 7^e édition de la *Grammaire méthodique du français* (Riegel *et al.*, 2009 : 219).

3.3.3. Une activité cérébrale en contradiction apparente avec l'analyse commune

La figure 12 ci-dessous illustre un entretien avec Stanislas Dehaene paru dans *La Recherche* d'octobre 2017. Elle plaide en faveur d'un découpage à la GR, puisque *dix grands amis* y est désigné comme une structure syntaxique dans *dix grands amis de Bill Gates*.

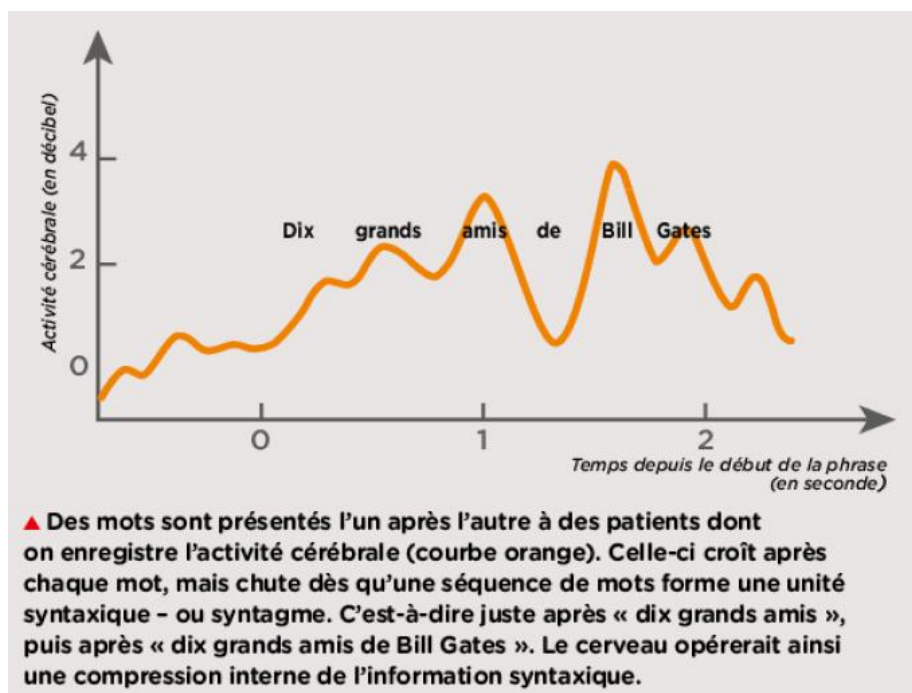


Figure 12. Au rythme des syntagmes (extrait de *La Recherche*, octobre 2017, p. 7)

Les trois éléments d'information donnés dans cette partie me semblent montrer la pertinence du découpage du groupe nominal en petits groupes à la GR. Je n'en tirerai pas la conclusion que le point de vue traditionnel est erroné, mais seulement que la question se pose de faire coexister, s'il est possible, les deux types d'analyse.

3.4. Auxiliaire et participe passé

Dans GR, la séquence *il a été arrêté* forme un seul groupe, dans le schéma des campagnes EASY et PASSAGE, elle en forme trois : *il a | été | arrêté*. Ce choix est justifié dans (Gendner et Vilnat, 2004) par la volonté de pouvoir annoter séparément ce qui peut venir entre l'auxiliaire et le participe, par exemple un adverbe :

10. <GP>En quelle année</GP> <NV>a-t-on</NV> <GR>vraiment</GR> <NV>construit</NV> <GN>la première automobile</GN> ?

Dans GR, *a-t-on vraiment construit* formerait un petit groupe. Il est vrai que des séquences très complexes peuvent s'insérer entre l'auxiliaire et le participe :

11. M. Borisav Jovic **avait**, *en sa qualité de commandant en chef des forces armées*, **convoqué** d'urgence les membres de cette institution.

Des exemples tels que celui-ci mettent en défaut le schéma de GR : on ne peut raisonnablement faire un petit groupe de *avait, en sa qualité de commandant en chef des forces armées, convoqué*.

La pertinence des groupes auxiliaire-participe passé peut cependant être défendue par les arguments suivants.

Le segment qui coupe le groupe *avait ... convoqué* dans l'exemple 11 s'apparente à une incise. Or on peut faire des incises à peu près partout :

12. Il arrive en retard, **avec**, *dans sa poche*, **un petit discours** écrit qu'il est obligé...
13. **Quelques excellents** – *ils ne le sont pas tous* – **reportages** de télévision ont enfin fait sortir le spectateur de son ghetto des bas-côtés.
14. Les artisans **de la**, *comme on dit*, **chaîne** du livre...

Il y sûrement des cas plus exceptionnels que d'autres. 13 et 14 le sont plus que 12 ou 11. Corrélativement, on peut dire que l'intégration de l'adjectif antéposé ou du déterminant avec le nom dans 13 et 14 est plus forte que celle de la préposition avec le groupe nominal dans 12 ou de l'auxiliaire avec le participe passé dans 11.

Il n'en reste pas moins qu'il y a une relation étroite entre l'auxiliaire et le participe passé, qui est d'ailleurs à l'origine du concept même d'auxiliaire, verbe devenu un mot outil. C'est cette relation qui justifie la formation de petits groupes avec auxiliaire + participe passé. La formation des petits groupes supporte tout à fait que puissent s'insérer des adverbes (*a-t-on vraiment construit*) ou des pronoms (*ils sont tous venus*) entre l'auxiliaire et le participe ; ce ne sont pas des incises. Elle est mise en défaut par l'insertion de groupes prépositionnels ou de propositions conjonctives, mais ceux-là, s'ils ne sont pas des locutions adverbiales figées, tendent à être des incises.

4. Développements envisagés

4.1. Logiciel

4.1.1. Suggestions d'amélioration

Le logiciel est un premier prototype qui pourrait bénéficier de nombreuses améliorations. Par exemple :

- Le système de score pourrait être plus positif et aller de 0 à 100 pour un exercice.
- On devrait pouvoir enregistrer ou faire défiler le graphe terminé.
- Les erreurs pourraient être plus clairement signalées.
- Permettre à l'utilisateur d'hésiter sur un rattachement. Ne pas donner immédiatement la réponse en cas d'erreur.
- Sur un écran tactile, le pilotage pourrait être fait par glissement de gauche à droite pour la formation des groupes, et pointage vers la cible pour les rattachements des compléments.
- En l'état, on peut demander l'affichage des mots, des mots + étiquettes de catégorie ou seulement des étiquettes de catégorie (voir *L'expérience de l'asémantisme* dans la page *Aller plus loin*). On pourrait imaginer remplacer seulement certains mots par une étiquette de catégorie, par exemple seulement les prépositions, ou tous les mots lexicaux.

4.1.2. Traitement de la phrase complexe

En l'état actuel, GR ne peut traiter que des phrases sans propositions subordonnées. Le comportement du jeu et la représentation graphique des subordonnées est prévu mais non encore implanté. L'idée est que quand arrivera un subordonnant (pronom relatif, conjonction de subordination ou interrogatif), l'utilisateur devra indiquer la nature particulière de ce mot par un appui sur la touche **i** (comme « introducteur »), touche qui génèrera une coloration particulière du mot et un cadre destiné à délimiter la proposition subordonnée. La figure 13 ci-dessous donne un exemple de la représentation envisagée, avec la phrase *D'autres l'accuseront d'entretenir ce climat qu'elle déplore puisqu'elle refuse d'entendre les revendications des syndicats*.

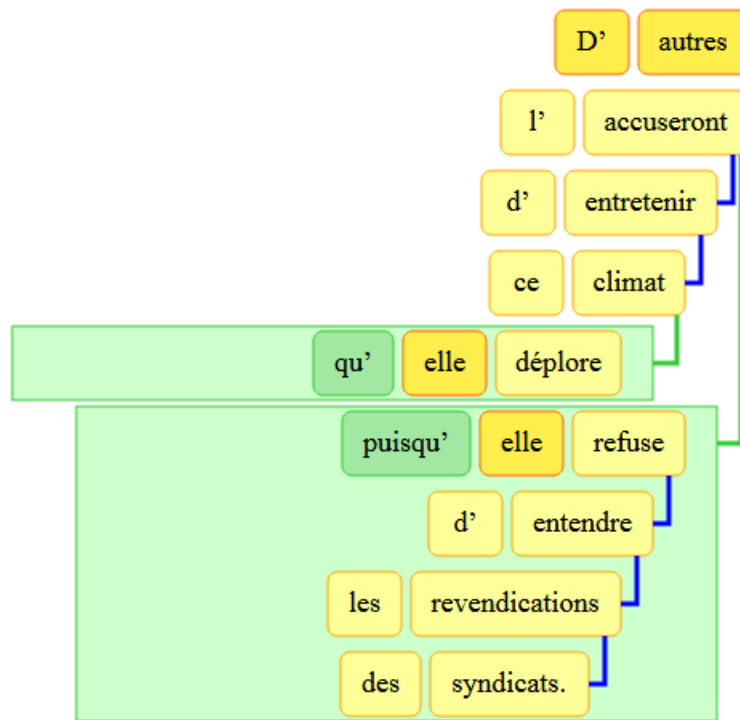


Figure 13. Analyse d'une phrase avec deux subordonnées.

Outre les propositions subordonnées, si le jeu rencontre un intérêt chez les linguistes, on pourra aussi chercher à traiter des structures particulières (voir ci-dessous § 4.3).

4.2. Pédagogie

On l'a dit plus haut (§ 1.2 et 1.3), l'utilisation de GR doit s'inscrire dans une séquence pédagogique plus large, qui utilise des corpus adaptés au niveau des apprenants. Fort heureusement, créer un corpus d'exercice est assez facile. Il me paraîtrait intéressant de développer des corpus pour des jeunes enfants apprenant les notions de base de la grammaire.

4.3. Linguistique

Pour utiliser une phrase dans GR, il faut en noter la structure. Cela se fait avec un tableur (voir ci-dessous figure 14) :

- un mot par ligne (col. B),
- avec sa catégorie (col. C),
- un identifiant de groupe (col. A),
- la notation du rattachement (col. D)
- et quelques informations complémentaires (col. E), s'il y a lieu : pronom clitique sujet, lien de coordination.

	A	B	C	D	E
1		L'	D		
2		1 équipe	N	s	
3		2 nationale	ADJ		1
4		de	PREP		
5		3 hockey	N		1
6		a	AUX		
7		4 remporté	PP	v	
8		le	D		
9		5 championna	N		4
10		du	PREPD		
11		6 monde	N		5
12		cette	D		
13		7 année.	N		4

Figure 14. Exemple de corpus au format tableur.

Un programme permet de convertir le corpus du format tableur au format *json*, utilisé par le logiciel. Un autre programme permet de traduire le format tableur en XML.

A la base de GR se trouve donc un schéma d'annotation de corpus. La nouveauté est qu'une représentation graphique originale lui est associée. Il serait intéressant d'enrichir en parallèle le schéma d'annotation et la représentation graphique. Par exemple, comment représenter des phrases avec ellipse (ex. *Pierre mange des pommes, Marie des poires*) ? Faudrait-il une relation particulière pour les attributs ? Dans la ligne du point de vue défendu dans (Trouilleux, 2018, 2021), ne serait-il pas pertinent de typer sémantiquement les relations marquées par les traits bleus ? etc.

5. Conclusion

Ce rapport a présenté GR, un logiciel dont l'objectif est d'aider à travailler sur la grammaire ou syntaxe du français. Si le jeu est certainement nouveau, les représentations sur lesquelles il s'appuie ne le sont pas. Pourtant, à divers égards, GR va contre la doxa – sur l'absence de compléments de phrase, sur la structure du groupe nominal, sur l'absence de mouvement ou de justification des structures par le test de déplacement... Autrement dit, GR soulève des questions. Conclure, dans ces conditions, est impossible, si ce n'est sur un espoir que le jeu lui-même contribuera à y apporter des réponses.

6. Références

Abney, S. 1991 « Parsing By Chunks », in R. Berwick, S. Abney, C. Tenny (eds), *Principle-Based Parsing*, Dordrecht: Kluwer Academic Publishers.

Beauzée, N. 1767. *Grammaire générale, ou Exposition raisonnée des éléments nécessaires du langage : pour servir de fondement à l'étude de toutes les langues*.

Bloomfield, Leonard. 1933. *Le langage*.

Bulletin officiel de l'éducation nationale, 26 novembre, 2015,
http://www.education.gouv.fr/pid285/bulletin_officiel.html?pid_bo=33400.

Canelas-Trevisi, S. 2009. *La grammaire enseignée en classe – Le sens des objets et des manipulations*. Berne : Peter Lang

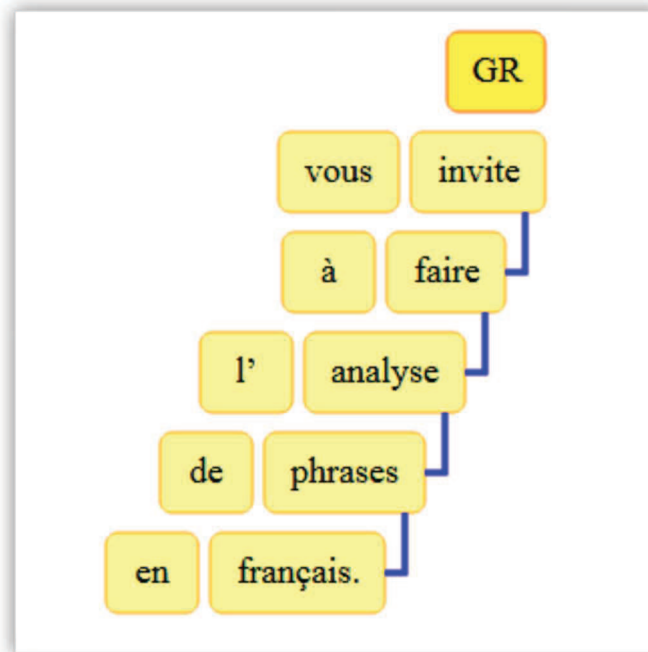
Creissels, D. 1995. *Éléments de syntaxe générale*. Paris : Presses universitaires de France.

Culicover, P. W. et Jackendoff, R. 2005. *Simpler Syntax*. Oxford : Oxford University Press

- Ferreira F. et Henderson, J. M. 1991. "Recovery from Misanalyses of Garden-Path Sentences". In *Journal of Memory and Language* 30, 725-745.
- Gendner, V. et Vilnat, A. 2004. *Les annotations de référence PEAS*.
https://perso.limsi.fr/anne/Guide/PEAS_reference_annotations_v2.2.html
- Genevay, E., Lipp, B. et Schoeni, G. (éd.). 1985. *activités en grammaire Français 7e, Division pré-gymnasiale. Collection Langue et parole*. Lausanne : Fournitures et éditions scolaires du canton de Vaud, Editions Loisirs et Pédagogie.
- Graffi, G. 2001. *200 years of syntax : a critical survey*. Benjamins.
- La Clergerie, É. de. 2005. "From metagrammars to factorized TAG/TIG parsers." In *Proceedings of IWPT'05*, p. 190–191, Vancouver, Canada.
- Martinet, A. 1979. *Grammaire fonctionnelle du français*. Didier.
- Mel'čuk, I. A. 1988. *Dependency Syntax*. Albany: State University of New York Press.
- Monneret, Ph. 1999. *Exercices de linguistique*. Paris : Presses universitaires de France.
- Riegel, M., Pellat, J.-C. et Rioul, R.. 1994 [1^{re} édition] et 2009 [7^e édition]. *Grammaire méthodique du français*. Presses universitaires de France.
- Tesnière, Lucien. 1959. *Éléments de syntaxe structurale*. Klincksieck.
- Trouilleux, F. 2009. « Un analyseur de surface non déterministe pour le français. » *TALN*, Senlis, 24–26 juin 2009.
- Trouilleux, F. 2021. « Critique du complément de phrase. » *Corela*, 19-1 | 2021,
<http://journals.openedition.org/corela/13003>
- Universal dependencies*. <https://universaldependencies.org/>
- Vaugelas, C. F. de. 1647 *Remarques sur la langue française : utiles à ceux qui veulent bien parler et bien écrire*. Paris.
- Vergne, J. et Houben, F. 1998. « L'analyseur syntaxique Vergne-98 présenté aux actions d'évaluation GRACE et EASy. » In *Actes de TALN 2005*. Dourdan, 6-10 juin 2005.
- Vergne J. (1999). *Etude et modélisation de la syntaxe des langues à l'aide de l'ordinateur. Analyse syntaxique automatique non combinatoire*. Dossier d'habilitation à diriger des recherches. Université de Caen.
- A. Vilnat , P. Paroubek , E. De La Clergerie , G. Francopoulo , M. L. Guénot. 2010. "PASSAGE Syntactic Representation: a Minimal Common Ground for Evaluation." *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.
- Wells, R. S. 1947. "Immediate constituents." *Language*, Vol. 23, No. 2, pp. 81-117. Linguistic Society of America.

7. Annexe

Les pages suivantes donnent une copie des pages du site GR à la date de rédaction de ce rapport.



GR – groupe et rattache – est un jeu interactif qui permet d'étudier la grammaire du français.
L'utilisateur reçoit un par un les mots d'une phrase et doit en construire la structure pas à pas.

Entrer

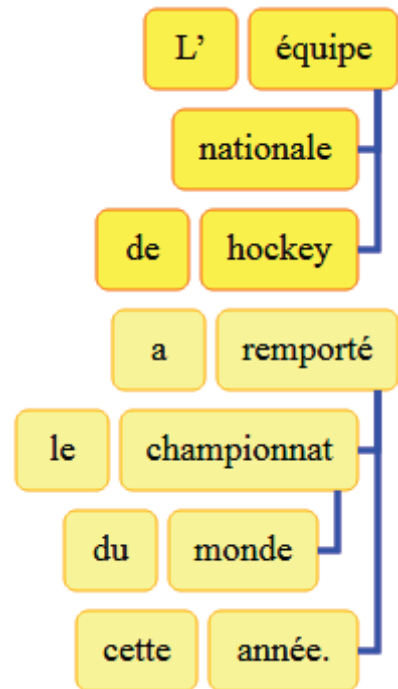
GR invite ses utilisateurs à produire des analyses structurales de phrases en français.

Par exemple, pour la phrase

L'équipe nationale de hockey a remporté le championnat du monde cette année.

il s'agira de produire la structure ci-contre.

La phrase est divisée en **sept petits groupes**. Les **deux teintes** caractérisent respectivement le groupe sujet et le groupe verbal. Les **traits** bleus indiquent le rattachement des compléments (par exemple les groupes *nationale* et *de hockey* sont l'un et l'autre compléments du nom *équipe*).



Pour analyser une phrase, on effectue deux opérations assez différentes :

1. On forme des petits **groupes**.
2. On **rattache** chaque groupe au graphe de la phrase en cours d'analyse.

GR permet d'aborder ces deux opérations ensemble ou séparément.

Pour apprendre à utiliser GR, nous vous invitons à d'abord rattacher des groupes.

Rattacher les groupes

Former les groupes

Grouper et rattacher

Aller plus loin







Après un exercice de prise en main, cette page vous propose trois exercices de **rattachement des groupes** : un exercice de découverte, un exercice avec des compléments antéposés et un exercice traitant du cas particulier de la coordination.

Remarque : il ne s'agit pas ici d'une séquence didactique qui permettrait à un novice d'apprendre à quoi rattacher les groupes. Cette page s'adresse à un public qui a déjà une connaissance des notions de compléments et de coordination et qui souhaite découvrir comment elles sont traitées dans GR.

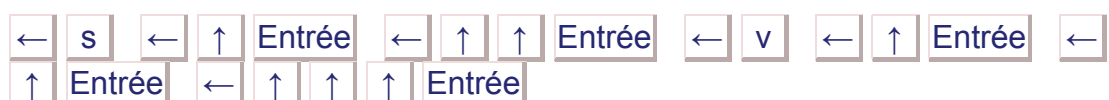
Prise en main

Cliquez sur le bouton **Rattacher : Prise en main** ci-dessous pour produire l'analyse de la phrase *L'équipe nationale de hockey a remporté le championnat du monde cette année.*

Pour piloter le jeu, utilisez les **touches du clavier** ou, si vous avez un écran tactile, les **boutons** à gauche et à droite de la page du jeu. Valeurs des touches et boutons :

	:	fait apparaître le nouveau groupe (il s'agit de la flèche gauche, pas de la touche d'effacement),
	:	déclare le groupe en cours comme sujet,
	:	déclare le groupe en cours comme verbe,
	:	sélectionne la cible du rattachement,
 ou 	:	valide la cible du rattachement.

Ainsi, la phrase *L'équipe nationale de hockey a remporté le championnat du monde cette année* s'analyse avec cette séquence de touches :



Avant de lancer l'exercice, [cliquez ici pour ouvrir une fenêtre rappelant les commandes](#).

A la fin de l'exercice, cliquez sur le rectangle bleu en bas à droite pour revenir sur cette page.

Rattacher : Prise en main

1. Exercice de découverte

L'exercice 1 vous invite à analyser cinq phrases de longueurs croissantes (de 6 à 19 groupes).

Chaque erreur vous fait perdre un point sur les 10 attribués au départ.

Vous pouvez afficher la phrase complète à analyser en plaçant le curseur de la souris en haut à gauche du cadre blanc.

► *Voir le corpus avant de faire l'exercice*

2. Rattachement des compléments antéposés

L'exercice 2 vous propose des phrases contenant des compléments détachés en tête de phrase.

Soit la phrase suivante :

Dissimulée dans des boîtes de fruits exotiques, la drogue était acheminée par un réseau international très perfectionné.

Cette phrase commence par un groupe formé du mot *Dissimulée*, suivi de son complément. Qu'est-ce qui est dissimulé dans des boîtes de fruits exotiques ? La drogue. Le groupe de surface *Dissimulée* se rattache au groupe *la drogue*, le **sujet** de la phrase, qui vient après lui.

Autre exemple :

Sur le parcours de ce véhicule se trouve le quai des Orfèvres.

Le groupe *Sur le parcours* dénote le lieu où se trouve le quai des Orfèvres. Il se rattache au groupe *se trouve*, le **groupe verbal** qui vient après lui. **GR** n'utilise pas le concept de complément de phrase de la nouvelle grammaire.

Seuls ces deux cas de figure sont pris en compte par GR. Ils ne sont pas les seuls possibles dans la réalité, mais sont les plus fréquents. En présence d'un complément en tête de phrase, il faudra donc choisir entre les deux commandes suivantes, qui s'ajoutent aux commandes de **GR** déjà vues :

1. Si le groupe se rattache au sujet qui suit : suivi de .
2. Si le groupe se rattache au verbe qui suit : suivi de .

► Voir le corpus avant de faire l'exercice

3. Rattachement des groupes coordonnés

La phrase suivante se décompose en cinq groupes :

Pour cent francs / par an, / elle faisait / la cuisine / et le ménage.

Comment rattacher le groupe *et le ménage* ?

Pour rattacher *et le ménage*, une option pourrait être de dire qu'il est complément de *elle faisait*, comme *la cuisine*. Cette option, cependant, ferait peu de cas du sens de la conjonction de coordination *et* qui ouvre ce groupe. On a là un rattachement particulier, qui est une **coordination de groupe**. Le groupe *et le ménage* est coordonné au groupe *la cuisine*.



Règle de la coordination de groupes. Lorsqu'un groupe est introduit par une conjonction de coordination, on le rattache au groupe auquel il est coordonné.

Remarque : les conjonctions de coordination pourraient aussi coordonner des propositions, mais ce cas ne se présente pas dans GR, qui ne traite que des propositions indépendantes.

Comme ce rattachement n'est pas du même type que le rattachement des compléments, **GR** le matérialise graphiquement d'une façon un peu différente.

En fait, la femme dont il est question ici (Félicité, un personnage de Flaubert) faisait beaucoup plus que la cuisine et le ménage. Le texte de Flaubert se poursuit en énumérant d'autres activités. Même si elle est faite sans conjonction de coordination explicite, une énumération est analysée dans GR comme une coordination :

Pour cent francs par an, elle faisait la cuisine et le ménage, cousait, lavait, repassait, savait brider un cheval, engraisser les volailles, battre le beurre, et resta fidèle à sa maîtresse...

Chaque groupe coordonné se rattache au terme coordonné qui précède. Cliquez sur le bouton qui suit pour voir comment **GR** analyse cette phrase (le logiciel fait tout, appuyez alternativement sur  et ) :

Exemple de groupes coordonnés

► *Voir le corpus avant de faire l'exercice*




Rattacher 3 : groupes coordonnés

Former les groupes

Après un exercice de prise en main, cette page vous invite à découvrir et appliquer les règles de **formation des groupes**, en trois parties : le cas général, le cas particulier des pronoms faibles et forts et la règle du noyau du groupe nominal.

Prise en main

A vous de produire le découpage en groupes de la phrase *L'équipe nationale de hockey a remporté le championnat du monde cette année*. Quand le logiciel s'occupe des rattachements, deux touches (ou boutons, avec un écran tactile) suffisent pour piloter le jeu :

 : commence un groupe ou ajoute le mot au groupe en cours de construction,
 ou  : termine le groupe.

Ainsi, la phrase s'analyse avec cette séquence de touches :

A la fin de l'exercice, cliquez sur le rectangle bleu en bas à droite pour revenir sur cette page.

Grouper : prise en main

1. Cas général

La première règle de formation des groupes s'appuie sur la distinction entre deux types de mots : les **mots grammaticaux** (ou « mots outils », ou « mots fonctionnels ») et les **mots lexicaux**. Les premiers s'associent aux seconds.

▼ Lire les explications

Prenons par exemple la phrase suivante :

Les tigres nous attendent à Trieste.

Cette phrase contient trois mots grammaticaux : *Les, nous, à*, et trois mots lexicaux : *tigres, attendent, Trieste*. Les mots grammaticaux s'associent aux mots lexicaux pour former trois groupes : *Les tigres / nous attendent / à Trieste*.

Les mots lexicaux d'une langue se comptent par milliers. Les mots grammaticaux se comptent seulement par dizaines.

Les mots lexicaux sont riches en information. Si on ne garde qu'eux, on comprend encore le message, ou à peu près : *tigres attendent Trieste* : ça pourrait faire un télégramme. Les mots grammaticaux, par contre, sont pauvres en information. Si on ne garde qu'eux, on est sûr de ne rien comprendre du message : *les nous à* : ça ne veut rien dire.

Si vous prenez un texte en français, vous constaterez que les mots qui le

composent se répartissent à peu près moitié-moitié : 50 % de mots grammaticaux, 50 % de mots lexicaux. Mais les mots grammaticaux sont très souvent les mêmes, alors que les mots lexicaux sont souvent différents. Dans n'importe quel texte un peu long en français, les mots les plus fréquents seront *de*, *le-la-les*, la préposition *à*, la conjonction de coordination *et*, *un* ou *une*, *que*, *il*, *elle*, etc. Des mots grammaticaux.

Par exemple, le texte qui suit contient 80 mots. Sur 80 mots, il y a 41 mots grammaticaux (en orange) et 39 mots lexicaux (en bleu) :

Il y avait *à* *Montmartre*, *au* *troisième étage* *du* *75 bis*
de *la* *rue* *d'* *Orchampt*, *un* *excellent* *homme* *nommé* *Dutilleul*
qui *possédait* *le* *don* *singulier* *de* *passer* *à* *travers* *les* *murs*
sans *en* *être* *incommodé*. *Il* *portait* *un* *binocle*,
une *petite* *barbiche* *noire* *et* *il* *était* *employé*
de *troisième* *classe* *au* *ministère* *de* *l'* *Enregistrement*.
En *hiver*, *il* *se* *rendait* *à* *son* *bureau* *par* *l'* *autobus*
et *à* *la* *belle* *saison*, *il* *faisait* *le* *trajet* *à* *pied*,
sous *son* *chapeau* *melon*.



(Marcel Aymé, *Le Passe-muraille*. Ed. Gallimard)

41 mots grammaticaux, 39 mots lexicaux : à peu près 50-50, donc. Mais pour les 41 mots grammaticaux, on compte seulement 17 mots différents (il y a 7 occurrences de *le-la-les*, 6 de *de-du*, 6 de *à-au*, 5 de *il* et ainsi de suite), alors que pour 39 mots lexicaux, on compte 38 mots différents (seul *troisième* apparaît deux fois).

On remarque que les mots grammaticaux sont plutôt courts (2,34 caractères en moyenne dans ce texte ; ici, parmi les mots grammaticaux, seule la locution prépositionnelle *à travers* est un peu longue). Les mots lexicaux, en revanche, sont plutôt longs (6,74 caractères en moyenne dans ce texte).

Sur ces données, formons les groupes : cliquez ici pour les matérialiser. Le texte contient 34 groupes, soit un peu moins que le nombre de mots lexicaux. Si on observe bien ces groupes, on voit se dessiner un schéma général qui se résume à la règle qui suit.

Principe de la formation des groupes. Dans le cas général, les mots grammaticaux s'associent avec un mot lexical qui les suit pour former un groupe. On désigne ce mot lexical comme le **noyau** du groupe.

Le bouton qui suit vous propose un premier entraînement à la formation des groupes. Le logiciel se charge des rattachements ; appuyez simplement sur ,  ou **Entrée** à la fin de chaque groupe.

► Voir le corpus avant de faire l'exercice

Grouper 1 : cas général

2. Pronoms forts et pronoms faibles

Relativement à la règle précédente, qui s'appuie sur la répartition des mots en mots grammaticaux et mots lexicaux, la catégorie des pronoms pose problème.

En fait, certains pronoms, bien qu'ils soient des mots grammaticaux, vont compter comme des mots lexicaux : il s'agit des pronoms **forts**, par opposition aux pronoms **faibles**.

▼ Lire les explications

La différence entre pronoms forts et faibles est d'ordre prosodique et distributionnelle, c'est-à-dire qu'elle relève de la façon dont les pronoms se combinent avec les autres mots de la phrase.

Voyons d'abord les pronoms faibles. Le qualificatif *clitique*, en linguistique, caractérise des mots grammaticaux qui s'associent dans la prononciation à leur voisin de telle façon que le clitique et son voisin tendent à ne former qu'une seule unité phonétique. En français, l'élision et la liaison sont des indices qui permettent de repérer les clitiques. Exemples :

- *l'homme* (élision), *les enfants* (liaison) => le déterminant *le* est un clitique ;
- *ils en ont* => les pronoms *ils* et *en* sont des clitiques ;
- *dit-on* => le pronom *on* est un clitique ;
- *d'Avignon* => la préposition *de* est un clitique ;
- *dans un an* => la préposition *dans* et le déterminant cardinal *un* sont des clitiques ;
- *n'en parlons plus* => la particule de négation *ne* est un clitique.

A l'exception des conjonctions de coordination et des verbes auxiliaires, toutes les catégories de mots grammaticaux contiennent au moins un mot qui s'élide obligatoirement devant le mot qui le suit ou entraîne une liaison avec lui. Dans une large mesure, en français, cette notion de formes clitiques est un corollaire de l'association des mots grammaticaux aux mots lexicaux telle qu'on la voit dans la règle précédente.

Une partie des pronoms sont donc des pronoms clitiques : ils sont étroitement liés à un verbe, en général après eux (*il lui en parle*), parfois avant, dans l'interrogation (*où va-t-il ?*) ou l'impératif (*prends-en*). Ces pronoms, qu'on appelle aussi pronoms « faibles », obéissent à la règle générale d'association des mots grammaticaux au mot lexical voisin.

Une autre partie des pronoms, en revanche, peuvent prendre la place de n'importe quel groupe nominal. On les appellera ici les pronoms « forts ». Ce sont tous ceux qu'on peut mettre après une préposition (*pour rien, selon lui, à eux, d'autrui, avec quelqu'un*), en position détachée (*moi, j'aime le music-hall*) ou en position de complément direct après le verbe (*il veut celui-là*).

Règle du noyau pronominal. Les pronoms forts comptent comme des mots lexicaux ; ils forment le noyau d'un groupe.

Chacune des cinq phrases de l'exercice qui suit contient au moins un pronom fort. Pensez à terminer le groupe quand vous rencontrez ces pronoms. Le logiciel se charge des rattachements.

► Voir le corpus avant de faire l'exercice

Grouper 2 : pronoms forts et pronoms faibles

3. Noyau du groupe nominal

Troisième principe du découpage en groupes : le noyau lexical, ou pronom fort, termine le groupe (sauf cas très particuliers).

▼ Lire les explications

On a dit dans la première règle de formation des groupes que « dans le cas général, les mots grammaticaux s'associent avec un mot lexical à leur suite ». Cela ne veut pas dire que le mot lexical en question se trouve *immédiatement* à la suite des mots grammaticaux. Dans certains cas, un groupe pourra contenir plusieurs mots lexicaux.

Prenons cette séquence de trois mots : *un petit inconvenient*. On peut envisager de l'analyser en deux groupes (*un petit / inconvenient*, avec *inconvenient* rattaché à *petit*) ou en un seul groupe. On préférera l'analyse en un groupe, pour deux raisons : la liaison obligatoire après *petit* suggère de ne pas le séparer du nom *inconvenient*, et surtout le nom *inconvenient* est le noyau du groupe qui commence par le mot grammatical *un*, c'est lui le mot le plus important du groupe ; c'est à lui que *un* s'associe. L'adjectif *petit* est secondaire, on pourrait s'en passer.

Considérons maintenant cette séquence de quatre mots : *un magnifique cheval alezan*. En fera-t-on un groupe ou deux ? En vertu de l'analyse faite pour *un petit inconvenient*, on ne peut pas envisager de couper entre *un* et *cheval*, qui est le noyau de ce groupe nominal ; les deux options sont donc : un seul groupe ou deux groupes *un magnifique cheval / alezan*. On retient dans GR la seconde option, en posant la règle qui suit.

Règle du noyau comme limite de groupe. Sauf cas très particuliers, le noyau lexical d'un groupe (ou pronom fort) termine le groupe.

Les cas particuliers sont ceux où des pronoms clitiques (pronoms faibles, des mots grammaticaux) se trouvent après le verbe, comme dans *dit-il* ou *donne-le-moi*. Après un nom, on peut trouver les particules *ci* ou *là*, ex. *ces jours-ci*.

Attention : cette règle du noyau comme limite de groupe a un caractère un peu arbitraire. Il nous semble intéressant de structurer la phrase autour des noyaux de groupes, mais ce découpage ira parfois à l'encontre des groupes formés dans la prononciation. Par exemple, la séquence *le puissant groupe américain* se décompose en *le puissant groupe / américain*, alors que la prononciation découperait plutôt entre *puissant* et *groupe américain*.

Chacune des cinq phrases de l'exercice qui suit contient au moins un problème d'identification du noyau d'un groupe nominal. Pensez à terminer le groupe avec le noyau. Le logiciel se charge des rattachements.

► Voir le corpus avant de faire l'exercice

Grouper 3 : noyau du groupe nominal

Grouper et rattacher



Vous avez groupé et rattaché séparément, voici maintenant trois corpus pour lesquels vous devez grouper ET rattacher. Voyez si vous pouvez analyser les phrases mot après mot, sans utiliser l'affichage du coin supérieur gauche.

Grouper et rattacher 1

Grouper et rattacher 2

Grouper et rattacher 3

Vous pouvez aussi retrouver en mode *groupe et rattache* les six corpus utilisés pour découvrir les deux tâches du jeu :

Rattacher 1 : découverte

Rattacher 2 : compléments antéposés

Rattacher 3 : groupes coordonnés

Grouper 1 : cas général

Grouper 2 : pronoms forts et pronoms faibles

Grouper 3 : noyau du groupe nominal

[Aller plus loin](#)



Ce site GR est une démonstration à l'attention de linguistes et didacticiens du français. En l'état, on ne prétend pas permettre à un utilisateur lambda d'apprendre la syntaxe avec lui. En fonction du public avec lequel on voudra travailler et des problèmes ciblés, il faudra développer des corpus spécifiques et inscrire l'utilisation du jeu dans des séquences pédagogiques plus larges.

Un rapport de recherche accompagne la publication du site. Y figurent notamment des **justifications de l'approche proposée** et une évocation des **développements envisagés**.

[Lire le rapport de recherche.](#)

Les personnes intéressées par une participation au développement de GR (séquence pédagogique, structures linguistiques particulières, développement informatique...) sont vivement invitées à me contacter :

[francois.trouilleux](mailto:francois.trouilleux@uca.fr) à uca.fr.

L'expérience de l'asémantisme

Pour alimenter la réflexion sur la pertinence des petits groupes formés dans GR, on propose ici une expérience : en cliquant sur le bouton **Analyser sans les mots** ci-dessous, vous serez amenées à analyser un corpus de phrases dans lesquels tous les mots ont été remplacés par une étiquette de catégorie (par exemple, **D** pour déterminant, **N** pour nom, etc. : voir le [tableau des catégories](#)).

En perdant les mots, on a perdu l'accès au sens ; en ce sens, l'analyse est « asémantique » ou, si l'on veut, uniquement syntaxique. La problématique de cette expérience est évoquée à l'alinéa 2.3.4 du rapport de recherche (voir ci-dessus) et dans un article intitulé *Critique du complément de phrase* ([Trouilleux, 2021](#)).

L'expérience montre que dans ce mode, quelqu'un qui a une connaissance des parties du discours et des principes de base de la formation des groupes *peut* faire les groupes, mais *pas* bien les rattacher. On a là un indice qu'il est pertinent de distinguer les deux dimensions que distingue le jeu et une preuve que le sens intervient de façon essentielle dans la détermination des rattachements (sur ce point, voir [Trouilleux, 2021](#), § 4.3).

Analyser sans les mots (grouper et rattacher)

Comme le rattachement est impossible à bien faire, le bouton ci-dessous vous propose d'analyser le même corpus en ne faisant que les groupes (le logiciel fera les rattachements).

Analyser sans les mots (grouper seulement)

[Retour à l'accueil](#)