

Sur la notion harrissienne de transformation et son application au français

In: Langages, 25e année, n°99, 1990. pp. 39-56.

Citer ce document / Cite this document :

Gross Maurice. Sur la notion harrissienne de transformation et son application au français. In: Langages, 25e année, n°99, 1990. pp. 39-56.

doi : 10.3406/lgge.1990.1591

http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_1990_num_25_99_1591

Maurice GROSS

Université Paris 7

Laboratoire d'Automatique Documentaire et linguistique ¹

Centre d'Etudes et de Recherches en Informatique Linguistique ²

SUR LA NOTION HARRISSIENNE DE TRANSFORMATION ET SON APPLICATION AU FRANÇAIS

Le concept de transformation n'est pas étranger à la grammaire traditionnelle. On le retrouve à différents niveaux de la description grammaticale avec les mêmes motivations qu'en grammaire formalisée. Par exemple, on peut considérer les voix du verbe comme des transformations de phrases élémentaires (i.e. des transformations de la structure sujet-verbe-compléments essentiels). Les voix comme le [Passif], le [Moyen] ou le [Pronominal] étendent les paradigmes verbaux de conjugaison. Comme ces formes ne peuvent se conjuguer qu'en tenant compte du sujet et des objets, les voix regroupent et rendent équivalentes des formes syntaxiques variées. De la même façon, l'analyse logique de la grammaire scolaire est une analyse transformationnelle de phrases complexes en phrases élémentaires.

Le progrès introduit par la notion harrissienne de transformation est avant tout un progrès logique : des définitions rigoureuses et minimales, autrement dit, une limitation à la seule description combinatoire de la langue, permettent de construire des grammaires cohérentes qui rendent compte de l'aspect mécanique du comportement syntaxique des locuteurs et de le simuler sur ordinateur (Z. S. Harris, 1952). Nous illustrons le système transformationnel de Harris par l'application qui en a été faite au français. Une équipe de linguistes du L.A.D.L. a en effet décrit de larges segments de la grammaire du français en recherchant une couverture lexicale importante, ce qui a conduit à préciser divers composants des grammaires transformationnelles.

1. Les phrases élémentaires

L'objectif des travaux du L.A.D.L. était d'effectuer une description détaillée des emplois des verbes du français dans des phrases simples. Par ailleurs, les transformations binaires ou généralisées rendent compte des phrases complexes. L'un des premiers problèmes rencontrés a consisté à délimiter la phrase élémentaire, c'est-à-dire la phrase à un verbe avec ses compléments essentiels. Si, en français, tout verbe a un sujet, le nombre des compléments qui dépendent du verbe est variable et leur séparation n'a jamais été claire. On notera *P* les phrases et on écrira :

1. U.A. 819 du C.N.R.S.

2. F.I.R.T.E.C.H. Industries de la langue, C.N.A.M. et Université Paris 7.

$$P = : N_0 VW$$

où N_0 est le sujet et W une suite quelconque de compléments. Un problème classique de la grammaire se pose alors : dans la séquence W de compléments attachés à un verbe, quels sont les compléments intrinsèques (aussi nommés compléments obligatoires, essentiels, objets, etc.) et quels sont les compléments circonstanciels ? Une réponse n'a pu être obtenue au L.A.D.L. qu'au prix de l'examen individuel d'un noyau de 12 000 verbes. En effet, chaque verbe possède une séquence variable de compléments essentiels. En même temps que l'on examinait les compléments de chaque verbe, il a fallu préciser les entrées verbales, car les homographies étaient nombreuses. Un exemple simple est celui de la forme verbale *voler* décrite par deux verbes ou entrées indépendantes :

voler (to fly), intransitif, de structure phrastique $N_0 V = : N_0 voler$, $W = : E$, soit « zéro »,

voler (to steal), à deux compléments (*voler quelque chose à quelqu'un*) de structure notée : $N_0 voler N_1 à N_2$.

Les cas les plus difficiles sont ceux de couples dits à sens propre et figuré (ou à métaphore). Pour un couple comme :

Luc arrose la plante

Luc arrose la nouvelle année

la métaphore étant accidentelle³, on considère que l'on a deux verbes indépendants, mais il n'en va pas de même avec des métaphores productives comme celle qui substitue à des objets concrets des concepts abstraits :

Luc a (bourré + rempli + etc.) le sac de brioches

Luc (entassé + accumulé + mis + glissé + etc.) des brioches dans le sac

Luc a (bourré + rempli + etc.) sa cervelle de connaissances inutiles

Luc a (mis + entassé + accumulé + glissé + etc.) des connaissances inutiles dans sa cervelle

Les deux ensembles de phrases ont des structures identiques, seules les distributions couplées de N_1 et de N_2 diffèrent. Qu'en conclut-on : avons-nous affaire dans chaque cas à un verbe ou à deux verbes indépendants ? La réponse, quelle qu'elle soit, est difficile à justifier.

À partir d'un ensemble de 6 000 formes verbales⁴, un ensemble de 12 000 entrées a été défini (M. Gross, 1975 ; J.-P. Boons, A. Guillet, C. Leclère, 1976a, 1976b ; A. Guillet, C. Leclère, 1990), et pour chaque entrée on a défini une séquence de compléments essentiels. Ces séquences de compléments ont servi de

3. Il n'est pas toujours simple de déterminer ce caractère accidentel des cas productifs. Pour cet exemple, l'observation résulte de l'interdiction de la même métaphore pour des phrases voisines :

Luc (soigne + nourrit) la plante, Luc fête la nouvelle année

* *Luc (soigne + nourrit) la nouvelle année, Luc fête la plante.*

4. Dans le cadre de la description syntaxique, nous ne faisons pas de différence entre les termes :
— verbe (accompagné de son sujet et de ses éventuels compléments essentiels),
— entrée (de dictionnaire ou du lexique-grammaire) ou entrée verbale,
— phrase élémentaire.

base à une classification des 12 000 entrées en une cinquantaine de classes : chaque classe contient en général des verbes ayant une même séquence de compléments. Cette description constitue le lexique-grammaire des verbes français.

Dans de nombreux cas, pour séparer les compléments essentiels des circonstanciels qui ne sont pas incorporés au lexique-grammaire, il a fallu prendre des décisions individuelles, c'est-à-dire raisonner cas par cas, sur des verbes particuliers. L'ensemble de ces décisions est si complexe et varié qu'il n'est pas possible de donner une définition de phrase élémentaire. Nous sommes donc amené à définir les compléments essentiels des verbes du français comme étant ceux qui sont représentés dans le lexique-grammaire construit au L.A.D.L. (cf. 3). Les argumentations ayant conduit à l'état publié du lexique-grammaire sont données dans la documentation du système (M. Gross, C. Leclère, à paraître 1990), et il existe un catalogue (non exhaustif) de difficultés dont la solution pourrait conduire à modifier le statut de la phrase élémentaire, donc l'ensemble des compléments circonstanciels. Ainsi, les compléments d'aspect comme ceux de *durée* sont considérés comme *circonstanciels*, ils n'apparaissent donc pas dans le lexique-grammaire en compagnie des verbes qu'ils peuvent modifier. Ils posent pourtant une question théorique : nous représentons dans le lexique-grammaire les compléments indirects en *à* par exemple, pour la raison qu'ils sont attachés à des verbes particuliers :

Max parle à Luc * *Max parle Luc*
* *Max agace à Luc* *Max agace Luc*

or la situation est la même pour le complément de *durée pendant une heure* :

La bombe a roulé pendant de longues minutes
* *La bombe a explosé pendant de longues minutes*

Autrement dit, l'acceptation du complément *pendant N* doit être marquée lexicalement, ce qui peut être fait de la même manière que pour le complément *à N*. Toutefois, on notera que l'on accepte la phrase :

Les bombes ont explosé pendant de longues minutes

où le sujet est au pluriel, ce qui permet de lever l'interdiction d'occurrence avec le verbe *exploser*. Cette situation n'existe pas avec le complément *à N*, et c'est pour cette raison que nous laisserons (provisoirement) le complément *pendant N* en dehors du lexique-grammaire.

Une conséquence empirique de nos choix a fait que la séquence *W* des compléments essentiels est apparue comme limitée à la longueur 2. Nous n'avons trouvé que très peu d'exemples où nous avons été amenés à considérer trois compléments essentiels :

— dans le cas des verbes simples, il n'existe que de rares exemples comme :

Max a parié (mille francs)₁ (avec Léa)₂ (que Luc ne reviendrait pas)₃

— dans le cas des verbes supports et des verbes composés, les exemples sont plus nombreux (mais seulement quelques centaines en regard de 40 000 ou 50 000 expressions) :

Max a pris (prétexte)₁ (de la venue de Léa)₂ (pour partir)₃

2. Les transformations

Il reste essentiel de considérer deux types principaux de transformations :
— les transformations **unaires** (e.g. le Passif) qui opèrent chacune sur une structure élémentaire $N_0 VW$, et
— les transformations **binaires** (la coordination et la subordination de phrases) qui combinent deux structures élémentaires pour redonner une structure élémentaire.

2.1. Les transformations unaires

Le [Passif], les [Pronominalisations], l'[Extraposition] du sujet sont des transformations unaires bien connues, elles conservent le sens des phrases auxquelles elles s'appliquent, elles permettent donc de constituer des classes sémantiques d'équivalence comme pour l'exemple qui suit :

- (1) *Max a lu trois de ces livres*
= *Il a lu trois de ces livres*
= *Max en a lu trois*
= *Il en a lu trois*
= (2) *Trois de ces livres ont été lus par Max*
= *Trois ont été lus par Max*
= *Trois de ces livres ont été lus par lui*
= *Trois ont été lus par lui*
= (3) *Il a été lu trois de ces livres par Max*
= *Il en a été lu trois par Max*
= *Il a été lu trois de ces livres par lui*
= *Il en a été lu trois par lui*⁵

Dans ces exemples, les pronominalisations s'appliquent à la structure $Dnum$ de $N_i = : trois de ces livres$ où $de ces livres$, appelée complément d'inclusion, subit des changements de forme.

Notons les transformations utilisées dans ces exemples de la façon suivante :

- [Passif], [Extrap], [$N_1 = LUI$], [$N_0 = IL$], [$N_1 = LE$]
[Pron (de Ndéf) z.] : pronominalisation *zéro* du complément d'inclusion,
[Pron (de Ndéf) en] : pronominalisation *en* du complément d'inclusion.

Ces pronominalisations se classent en deux types [$Pron (N_0)$] et [$Pron (N_1)$], selon qu'elles affectent le sujet N_0 ou le complément N_1 . Pour noter l'application des transformations à une phrase comme (1), utilisons une notion suffixée, l'expression :

(3) = (1) [Passif] [Extrap]
correspond à la phrase (3) obtenue par application de [Passif] d'abord et de [Extrap] ensuite. Les produits de transformation ne sont pas commutatifs en

5. Les formes *par lui* ne sont pas toujours naturelles, en tout cas sans emphase sur *lui*. Les formes extraites dans *C'est... Qu* le sont nettement plus : *C'est par lui que trois de ces livres ont été lus*.

général. Parfois, les applications de deux transformations s'excluent, c'est le cas par exemple du [Passif] et du [se moyen] :

- (4) *On lit trois de ces livres (facilement)*
(4) [se moyen] = *Trois de ces livres se lisent (facilement)*

on écrira alors :

- (4) ([Passif] + [se moyen])

L'opération « + » (« ou » exclusif) est commutative. On notera par [E] la transformation neutre : (1) [E] = (1), ce qui permet d'écrire pour une transformation facultative, des expressions comme :

- (1) ([E] + [Passif])

Avec ces notations de monoïde ⁶, la classe d'équivalence de (1) s'écrira :

- (1) ([E] + ([Passif] ([E] + [Extrap])) ([E] + [Pron (N_0)] ([E] + [Pron (N_1)]))

D'autres pronominalisations donnent lieu à d'autres formes, comme par exemple :

— la pronominalisation interrogative :

- Qui a lu trois de ces livres ?*
Max a lu lequel de ces livres ?
Trois de ces livres ont été lus par qui ?
Lesquels de ces livres ont été lus par Max ?

Une transformation de permutation en tête de phrase (l'Extraction) peut modifier certaines de ces phrases :

- Lequel de ces livres Max a-t-il lu ?*
Par qui trois de ces livres ont-ils été lus ?

— la pronominalisation relative :

- (la personne) qui a lu trois de ces livres*
(la personne) par qui trois de ces livres ont été lus
(les livres) dont Max a lu trois
(les livres) dont trois ont été lus par Max

La transformation d'Extraction est obligatoire ici et elle n'est pas associée à l'introduction d'un pronom de reprise du sujet, comme dans le cas de l'interrogation.

Il existe des contraintes complexes entre les diverses pronominalisations, ainsi l'interrogation et la relativation s'excluent, ce dont on peut rendre compte en limitant à une application par phrase élémentaire de la transformation d'[Extraction]. Les autres pronominalisations sont par contre compatibles avec l'interrogation et la relativation dans une même phrase :

6. Cf. N. Chomsky et M. P. Schützenberger, 1963, Z. S. Harris, 1968. Mais il n'est pas certain que l'opération de composition des transformations soit associative. Il semble naturel de rendre compte de certaines ambiguïtés de phrases au moyen de produits de transformations non associatifs.

*Qui en a lu trois ?
Lequel a-t-il lu ? etc.*

Les transformations unaires ont été notablement étendues par la prise en compte de la notion de **verbe support**. Ainsi, Z. S. Harris, 1964 propose la transformation unaire de nominalisation :

Bob walked = Bob took a walk

Ce type de relation peut être généralisé (M. Gross, 1981) à couvrir syntaxiquement la plupart des relations dites de morphologie dérivationnelle, comme dans les exemples suivants qui étendent la classe d'équivalence de (1) :

- | | |
|---------------------------|--|
| (2) [pouvoir i.] | <i>Trois de ces livres peuvent être lus par Max</i> ⁷ |
| [Vsup être-ible] | <i>Trois de ces livres sont lisibles par Max</i> |
| (2) [pouvoir i.] [Nég i.] | <i>Trois de ces livres ne sont pas lisibles par Max</i> |
| [Nég = Pfx] | <i>Trois de ces livres sont illisibles par Max</i> |
| [Vsup avoir-ité] | <i>Trois de ces livres ont une certaine lisibilité</i> |

On a encore des nominalisations comme :

- | | |
|-------|--|
| (1) = | <i>Max a fait une certaine lecture de ces trois livres</i> |
| (2) = | <i>Ces trois livres sont en (E + cours de) lecture</i> |

La formation de nombreux adverbes s'opère de façon analogue (C. Molinier, 1982 ; M. Gross, 1990) :

- | | |
|-----------------|---|
| [Nég i] | <i>Le fait que le lit glisse est perceptible</i> |
| [Nég = Pfx] | <i>Le fait que le lit glisse n'est pas perceptible</i> |
| [le fait → E] | <i>Le fait que le lit glisse est imperceptible</i> |
| [Vsup se faire] | <i>Que le lit glisse est imperceptible</i> |
| [Montée] | <i>Que le lit glisse se fait de manière imperceptible</i> |
| [manière -ment] | <i>Le lit glisse de manière imperceptible</i> |
| | <i>Le lit glisse imperceptiblement</i> |

On notera bien que cet exemple comporte deux phrases élémentaires et que le complément circonstanciel (ou adverbe) est introduit dans une phrase élémentaire à partir d'une autre phrase élémentaire. Autrement dit, la transformation morphosyntaxique unaire de formation de l'adverbe en *-ment* opère sur une phrase constituée par transformation binaire.

Des insertions dans (1), comme celles du verbe modal *pouvoir* ou de la négation *Nég*, font que les phrases liées par transformations ne sont plus synonymes. Toutefois, les différences de sens introduites par les transformations sont indépendantes du sens de base de la phrase (1) et ce sens reste constant. Les classes d'équivalence transformationnelles sont donc définies comme laissant invariante une unité de sens, celle-ci est représentée par la phrase (1). Comme dans toute relation d'équivalence, il est intéressant de distinguer un élément arbitraire de la classe pour nommer la classe. Dans le cas présent, le choix de la

7. Le verbe *lire* est considéré comme différent dans : *Max a lu ces trois textes à Léa.*

phrase déclarative simple (1) est naturel, on peut alors même parler de la classe d'un verbe donné, ici *lire*. Ce choix est traditionnel, même s'il ne correspond pas toujours à la forme la plus explicite de l'élément de sens.

2.2. Les transformations binaires

Elles opèrent sur deux structures élémentaires. On rappelle les principales :

— les coordinations :

Luc a conseillé à Max douze livres et il en a lu trois

Dans le second membre, on retrouve (1) pronominalisé ;

— les subordinations circonstancielles :

Max a lu trois de ces livres parce que Luc les lui avait conseillés

— les relativations :

Les trois livres que Max a lus lui ont été conseillés par Luc

— les subordinations complétives :

Luc a conseillé à Max qu'il lise trois de ces livres

Ce dernier type est intermédiaire entre phrases élémentaires simples et phrases complexes, puisque dans la position d'un *N* attaché au verbe principal on peut observer une phrase *P*.

3. Le lexique-grammaire

Nous avons mentionné à plusieurs reprises que nous considérons la phrase élémentaire comme l'unité minimale de sens, notion intuitive que recoupe la séparation formelle entre compléments essentiels et circonstanciels : les compléments circonstanciels sont introduits à partir de phrases élémentaires (cf. 2.1).

Nous appelons lexique-grammaire l'ensemble des classes d'équivalence des phrases élémentaires, ensemble qui sert de générateur pour les phrases complexes. D'un point de vue strictement lexical, on a vu qu'il était possible de caractériser, ou plutôt de nommer une classe par un verbe, en fait, par la ou les racines d'un mot. En effet, à l'intérieur de la classe d'équivalence, les différentes parties du discours n'interviennent pas, puisque les différences de préfixe et de suffixe ne modifient pas l'invariant de sens de la classe. Les racines sont indifféremment liées aux diverses parties du discours ; on a des séries sémantiquement voisines avec des « trous » morphologiques arbitraires, comme par exemple :

Ce texte est (clair + lumineux)
= *Ce texte a une certaine clarté*
? * *Ce texte a une certaine luminosité*
= *On a clarifié ce texte*

* *On a (enluminé + illuminé) ce texte* ⁸.

Plus généralement, on distinguera les racines de mots et les mots grammaticaux qui sont introduits par transformation :

(i) les racines de mots ou plus précisément les classes d'équivalence fournissent le sens de base,

(ii) les mots grammaticaux ne sont pas toujours porteurs de sens. Quand ils introduisent des notions de sens, celles-ci semblent bien fixées, au contraire du sens de base qui varie avec le contexte et les distributions des *N* et des *P* intervenant dans les phrases élémentaires. Nous avons rencontré un certain nombre de mots grammaticaux porteurs de sens :

- négation et préfixes équivalents,
- modaux (e.g. *pouvoir*) et préfixes équivalents,
- verbes supports,
- connecteurs binaires (conjonctions de coordination et de subordination).

Ces considérations sur la notion de racine et de mot ou constante grammaticale ont été présentées à partir d'exemples mettant en jeu des mots simples, verbes simples essentiellement. Or l'examen systématique du lexique a mis en évidence l'existence d'un autre composant majeur, celui des phrases élémentaires à verbes composés. Nous appelons verbes composés des expressions verbales :

- **idiomatiques** comme *prendre le taureau par les cornes*,
- **stylistiquement neutres** comme *prendre N en compte*,
- **grammaticalement rares** comme *tenir compte de N*,
- **métaphoriques** comme *son regard nous suit*,
- **techniques** comme *élever une perpendiculaire à un plan*,
- etc.

Dans ces expressions, le verbe et un substantif au moins n'ont pas de covariation distributionnelle et le sens de ces expressions n'est pas compositionnel.

Comme nous l'avons signalé, le lexique-grammaire des verbes simples du français contient environ 12 000 verbes simples. Le lexique-grammaire des verbes composés (M. Gross, 1982) en contient aujourd'hui plus de 20 000. Les verbes (i.e. phrases élémentaires) simples ou composés ont été classés en fonction de la nature de leurs compléments. Chaque classe correspond à une configuration générale de phrases élémentaires. Une cinquantaine de classes ont été définies pour les verbes simples, une trentaine pour les verbes composés. Chaque classe est représentée par une matrice dont les lignes sont les entrées du lexique-grammaire et les colonnes sont les structures de phrases élémentaires obtenues par les transformations unaires. Quand une entrée accepte une structure, un signe « + » figure à l'intersection de la ligne et de la colonne correspondante, un signe « - » dans le cas contraire. Ainsi, la classe d'équivalence d'une entrée

8. Les interdictions portent sur l'invariance sémantique des transformations en jeu. Les phrases sont acceptées avec d'autres interprétations.

Figure 1 *Tables syntaxiques extraites de M. Gross, 1975 pour le français, de A. Elia 1984 pour l'italien, de C. Subirats, 1987 pour l'espagnol.*

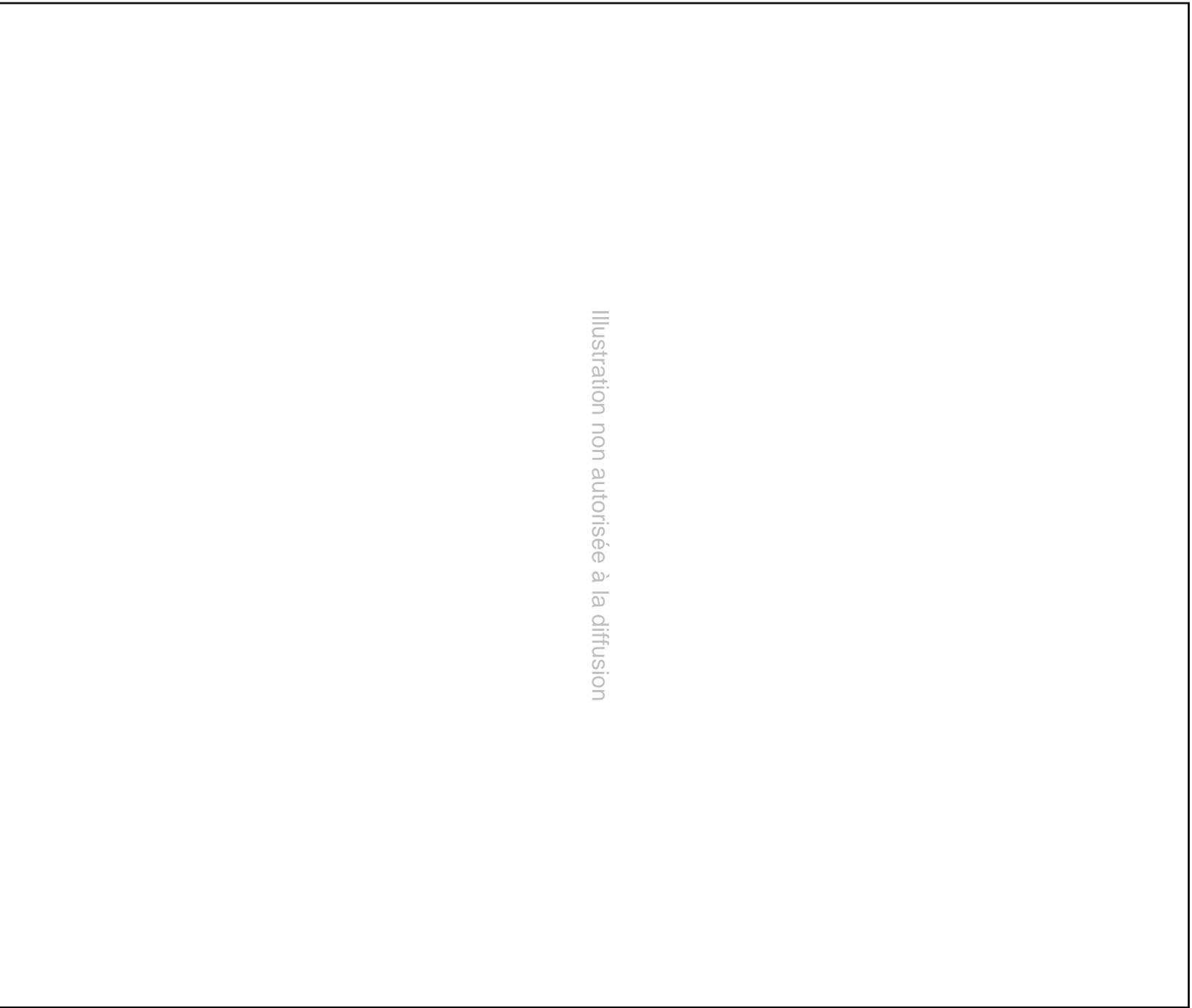


Illustration non autorisée à la diffusion

Illustration non autorisée à la diffusion

donnée est constituée des formes marquées « + » dans la ligne (cf. figure 1). La représentation des formes composées est la même, mais il n'est plus possible de parler de racine caractéristique d'une entrée, puisque deux mots au moins la définissent (cf. figure 2 ci-après).

Diverses opérations intéressantes ont pu être faites qui éclairent la notion de transformation :

— la comparaison deux à deux des lignes de matrices syntaxiques du lexique-grammaire montre qu'elles sont toutes différentes. En d'autres termes, les classes d'équivalence ont toutes des contenus différents, ou bien encore les transformations unaires sont toutes déterminées lexicalement ;

— les transformations s'appliquent de façon identique aux phrases libres et aux phrases figées, ce qui confirme la nature combinatoire des opérations syntaxiques qui sont indépendantes du sens des éléments lexicaux figurant dans les structures auxquelles elles s'appliquent ;

— une grande partie du vocabulaire (simple ou composé) peut être décrit dans le lexique-grammaire, c'est-à-dire en termes de phrases élémentaires à partir desquelles dérivent toutes les propriétés syntaxiques du vocabulaire.

4. La représentation des substantifs

Des études syntaxiques variées portent sur plusieurs dizaines de milliers de noms (pour le français, nous ne citerons que les principales : J. Giry-Schneider, 1978, 1989 ; G. Gross, 1989 ; J. Labelle, 1974 ; A. Meunier, 1977, 1981 ; D. de Négroni, 1978 ; R. Vivès, 1983). Il en existe également d'analogues pour les adjectifs et les adverbes (C. Molinier, 1982). Tous ces travaux démontrent la possibilité de représenter ces mots dans un lexique-grammaire (cf. figure 3).

Seuls semblent échapper à cette description les substantifs **concrets** (e.g. *chaise, notaire, homme des bois, pomme de terre*), qui ne donnent pas lieu à des phrases élémentaires syntaxiquement significatives, et présentent d'autres problèmes de représentation.

Les substantifs constituent la plus grande partie du vocabulaire et la plus productive :

— productivité par dérivation, souvent décrite par transformation unaire :

Ce parasite détruit la végétation

= *Ce parasite est un destructeur de la végétation*

Max et Luc coorganisent la réunion

= *Max et Luc sont les coorganisateur de la réunion*

— productivité par création de nouveaux concepts et donc de substantifs les nommant, c'est le cas de la création du vocabulaire scientifique et technique qui comporte plusieurs millions de mots pour une langue comme le français. La quasi totalité des mots créés sont des mots composés, comme *marché des changes, microscope à contraste de phase* ou *enseignement assisté par ordinateur*.

Figure 2 (tables syntaxiques de phrases figées du français.)

Nombre d'expressions: 26

$N_0 = :$	Nhum				
$N_1 = :$	N-hum				
+	-	se fiche de N	comme de le		an quarante
-	+	entrer Loc N	comme dans du		beurre
+	-	changer de N	comme de -		chemise
+	-	se moquer de N	comme de -		colin-tampon
+	-	se fiche de N	comme de Poss ⁰		dernière chemise
+	+	disparaître	comme par -		enchantement
+	-	serrer N	comme dans un		étai
-	-	faire froid	comme dans une		glacière
+	+	tomber	comme à -		Gravelotte
+	-	se soucier de N	comme de une		guigne
+	-	agir	comme à Poss ⁰		habitude
+	+	disparaître	comme par -		magie
+	-	entrer Loc N	comme dans un		moulin
+	-	y aller	comme à la		noce
+	-	se conduire Loc N	comme en -		pays conquis
+	-	se fiche de N	comme de la		peau de Poss ⁰ couilles
+	-	se méfier de N	comme de la		peste
+	-	s'installer Loc N	comme chez -		Poss ⁰
+	-	aimer	comme à le		premier jour
+	-	se fiche de N	comme de Poss ⁰		première chemise
+	-	faire	comme chez -		Pron ⁰
+	-	tenir à N	comme à la		prunelle de Poss ⁰ yeux
+	+	repartir	comme en -		quarante
+	+	repartir	comme en -		quatorze
-	+	marcher	comme sur des		roulettes
-	-	faire noir	comme dans le		trou du cul d'un nègre

Figure 4 L'expression et ses variantes s'obtiennent en parcourant le graphe de l'état initial I à l'état final F. Le symbole E correspond à une séquence vide.

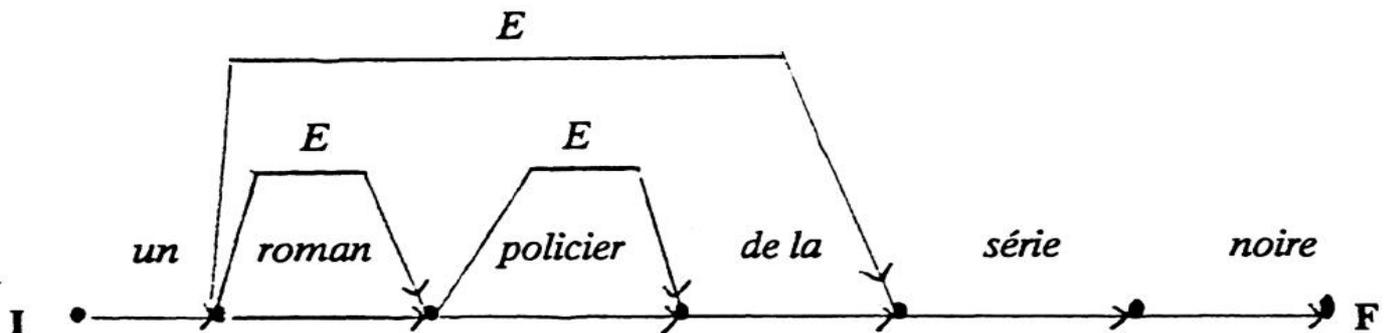


Figure 3 *Tables syntaxiques extraites de P. Machonis, 1988 pour l'anglais, de E. Ranchhod, 1983 pour le portugais, de D. de Négroni, 1978 pour le français.*

Illustration non autorisée à la diffusion

La collecte et la représentation d'importants volumes de tels termes posent divers problèmes (B. Courtois et M. Silberztein, 1990, G. Gross, 1990, M. Mathieu-Colas, 1990, R. Jung, 1990). Nous en signalerons un, important par une application informatique : la reconnaissance automatique des termes dans des textes (M. Silberztein, 1990). Nous illustrerons ces questions sur un exemple. Considérons l'expression :

un roman policier de la série noire

Illustration non autorisée à la diffusion

Illustration non autorisée à la diffusion

Elle peut être abrégée de différentes manières, sans que son interprétation change :

un roman de la série noire
un policier de la série noire
un série noire

Une façon naturelle de représenter l'expression et ses variantes consiste à utiliser un automate fini (D. Perrin, 1989), comme celui de la figure 4.

On peut traiter les abréviations comme des transformations de pronominalisation ; celles-ci remplacent des noms par des pronoms ⁹, mais elles forment aussi des expressions elliptiques. Ce point de vue est naturel lors de l'analyse de

9. Le graphe de la figure 4 comporte la forme : *un de la série noire* qui existe par exemple dans le contexte pronominal *Max en a lu*. L'automate combine des séquences elliptiques et pronominales.

textes. On constate en effet le phénomène suivant. Soit un texte technique traitant des *accélérateurs linéaires d'électrons*. Ce terme technique est explicite sous cette forme longue et on pourra le trouver dans le titre et dans le début du texte. Mais on constate que l'auteur, rapidement, abrège le terme en *accélérateur linéaire* ou *accélérateur d'électrons*, voire en *accélérateur*, s'il n'y a pas de confusion possible avec d'autres types accélérateurs. Or les quatre formes (longues ou courtes) réfèrent à une même notion du texte (référence lexicale). Autrement dit, une occurrence de forme courte peut référer à une occurrence de forme plus longue dans le texte, on peut convenir de dire qu'elle entre dans une relation de coréférence à la manière des pronoms et de leurs antécédents. Souvent les abréviations d'expressions longues sont de simples raccourcissements qui sont en général des suppressions des modificateurs de droite de l'élément nominal le plus à gauche de l'expression, comme on peut le noter par la formule :

accélérateur ((linéaire) (d'électrons))

où les parenthèses indiquent les possibilités d'omission. Mais les abréviations ne sont pas toujours aussi régulières comme on le voit sur l'exemple de la figure 4, aussi l'emploi généralisé de représentations par automates finis apparaît-il comme nécessaire pour représenter ce type de variations, ainsi d'ailleurs que d'autres variations lexicales ou syntaxiques.

Nous avons ainsi présenté l'essentiel de l'approche combinatoire de Z. S. Harris, 1968 et l'application qui en a été faite au français. Il faut en retenir que les mécanismes et représentations utilisés sont entièrement dégagés du sens, mais que toute l'approche consiste à réduire les variations de forme qui cachent la distribution du sens, but (encore éloigné) de la description linguistique. Les méthodes lexicales et syntaxiques disponibles sont loin d'avoir épuisé leur domaine d'application. En fait, aucune langue à ce jour n'a été décrite de façon raisonnablement complète au moyen de ces méthodes. Il apparaît néanmoins que des progrès ont été réalisés quand la description du sens a pu être séparée de la description des formes, alors que rien de cohérent ou de général ne peut être affirmé aujourd'hui pour la moindre unité de sens. On peut seulement espérer que des méthodes de traitement du sens pourront être développées à partir des systèmes lexico-syntaxiques au cours de construction.

RÉFÉRENCES

- BOONS, Jean-Paul ; GUILLET, Alain ; LECLÈRE, Christian, 1976a. *La structure des phrases simples en français. I Constructions intransitives*, Genève : Droz, 377 p.
- BOONS, Jean-Paul ; GUILLET, Alain ; LECLÈRE, Christian, 1976b. *La structure des phrases simples en français. II Constructions transitives*, Paris : Rapport de recherches du L.A.D.L., n° 6, 85 p., tables et index, 58 p.
- CHOMSKY, Noam ; SCHÜTZENBERGER, Marcel-Paul, 1963. *The Algebraic Theory of Context-Free Languages*, *Computer Programming and Formal Systems*, Brafford and Hirschberg eds., North Holland Pub. Co. En français, in *Langages* n° 9, 1968.

- COURTOIS, Blandine et SILBERZTEIN, Max éd., 1990. Dictionnaires électroniques du français, *Langue française*, Paris : Larousse.
- ELIA, Annibale, 1984. *Le verbe italien. Les complétives dans les phrases à un complément*, Fasano di Puglia : Schena-Nizet, 305 p.
- GIRY-SCHNEIDER, Jacqueline, 1978. *Les nominalisations en français. L'opérateur FAIRE dans le lexique*, Genève : Droz, 353 p., tables : 61 p.
- GIRY-SCHNEIDER, Jacqueline, 1987. *Étude de prédicats nominaux en français. Les constructions faire N*, Genève : Droz.
- GROSS, Gaston, 1989. *Les constructions converses du français*, Genève : Droz, 513 p.
- GROSS, Gaston, 1990. Définition des noms composés dans un lexique-grammaire, *Langue française*, Paris Larousse.
- GROSS, Maurice, 1975. *Méthodes en syntaxe*, Paris : Hermann, 412 p.
- GROSS, Maurice, 1981. Les bases empiriques de la notion de prédicat sémantique, *Formes syntaxiques et prédicats sémantiques*, A. Guillet et C. Leclère éd., *Langages*, n° 63, Paris : Larousse, p. 7-52.
- GROSS, Maurice, 1982. Une classification des phrases figées du français, *Revue québécoise de linguistique*, Vol. 11, n° 2, Montréal : Presses de l'Université du Québec à Montréal, p. 151-185.
- GROSS, Maurice, 1990. *Grammaire transformationnelle du français 3 : Syntaxe de l'adverbe*, Paris : A.S.S.T.R.I.L.
- GUILLET, Alain ; LECLÈRE, Christian, 1990. *La structure des phrases simples en français. Verbes à complément direct et complément locatif*, Genève : Droz.
- JUNG, René, 1990. Remarques sur la constitution du lexique des noms composés, *Langue française*, Paris : Larousse.
- HARRIS, Zellig S., 1952. Discourse Analysis, *Language* 28, Baltimore : The Waverly Press, p. 1-30.
- HARRIS, Zellig, 1964. Elementary Transformations, Philadelphie : University of Pennsylvania, *T.D.A.P.* n° 54, réimprimé in *Papers in Structural and Transformational Linguistics*, 1970, Reidel : Dordrecht.
- HARRIS, Zellig, 1968. *Mathematical Structures of Language*, New York : Wiley, 230 p. ; traduction française par C. Fuchs, 1971 : *Structures mathématiques du langage*, Paris : Dunod, 248 p.
- LABELLE, Jacques, 1974. *Etude de constructions avec opérateur AVOIR (nominalisations et extensions)*, Université Paris VIII : L.A.D.L., thèse de doctorat, 315 p.
- MACHONIS, A. Peter, 1988. Support Verbs : An Analysis of *be Prep X* idioms, *The SECOL Review*, 122, p. 95-125.
- MATHIEU-COLAS, Michel, 1990. Orthographe et informatique : remarques sur l'établissement d'un dictionnaire électronique de variantes graphiques, *Langue française*, Paris : Larousse.
- MEUNIER, Annie, 1977. Sur les bases syntaxiques de la morphologie dérivationnelle, *Linguisticae Investigationes*, Vol. I, n° 2, Amsterdam : J. Benjamins B.V., p. 287-332.
- MEUNIER, Annie, 1981. *Nominalisation d'adjectifs par verbes supports*, thèse de doctorat, Université Paris 7 : L.A.D.L., 282 p.
- MOLINIER, Christian, 1982. *Etude syntaxique et sémantique des adverbes en -ment*, thèse de doctorat, Université de Toulouse-Le Mirail, 416 p.
- NEGRONI-PEYRE, Dominique de, 1978. Nominalisation par être en et réflexivation (*admiration, opposition, révolte et rage*), *Linguisticae Investigationes*, Vol. II, n° 1, Amsterdam : J. Benjamins B.V., p. 127-164.

- PERRIN, Dominique, 1989. Automates et algorithmes sur les mots, *Annales des Télécommunications*, Tome 44, n° 1-2, p. 20-33.
- RANCHHOD, Elisabete, 1983. On the Support verbs *ser* and *estar* in Portuguese, *Lingvisticae Investigationes*, Vol. VII, n° 2, Amsterdam : J. Benjamins B.V., p. 317-353.
- SILBERZTEIN, Max, 1990. Le dictionnaire électronique des mots composés, *Langue française*, Paris : Larousse.
- SUBIRATS-RUDDEBERG, Carlos, 1987. *Sentential Complementation in Spanish, A lexicogrammatical study of three classes of verbs*, *Lingvisticae Investigationes Supplementa*, No 14, Amsterdam-Philadelphie : J. Benjamins Pub. Co., 290 p.
- VIVÈS, Robert, 1983. *Avoir, prendre, perdre : constructions à verbe support et extensions aspectuelles* : thèse de doctorat, Université Paris VIII, L.A.D.L., 194 p., tables : 40 p.