

Le syntagme du nombre cardinal en français moderne

PAR

CARL VIKNER

0. *Introduction*

Le syntagme du nombre cardinal (pour plus de commodité désigné dans la suite par «le syntagme cardinal») constitue un domaine de la grammaire qui, jusqu'à une époque récente, n'avait pas particulièrement retenu l'attention des grammairiens. En effet, les manuels de grammaire se limitent ordinairement à donner une liste renfermant les formes simples (*un, deux, ..., onze, douze, ..., vingt, trente, ..., etc.*) plus quelques échantillons de formes composées (*vingt et un, deux cents, etc.*). A partir d'une telle liste, le lecteur (étranger – car il va de soi que, pour le francophone, cette liste n'offre aucun intérêt pratique) est censé pouvoir construire des cardinaux corrects et éviter les constructions incorrectes. Et il le peut effectivement. Cela revient à dire que le lecteur établit lui-même, plus ou moins consciemment, une grammaire qui engendre les cardinaux – ce qui n'est possible, sans doute, que parce qu'il existe une très grande affinité entre les systèmes des numéraux cardinaux des différentes langues indo-européennes: un Français, un Anglais, un Allemand, un Russe, un Scandinave, etc. auront plutôt tendance à trouver «naturel» qu'on exprime le nombre 2963 au moyen d'un syntagme cardinal composé de six formes simples et le nombre 3000 à l'aide de deux formes simples seulement. Qu'il n'y ait là rien de «naturel» se révèle par le fait que la représentation en système décimal utilise quatre symboles pour chacun des deux nombres.

Comme les descriptions traditionnelles des numéraux cardinaux remplissent leur tâche de manière satisfaisante d'un point de vue pédagogique, on pourrait juger superflu de tenter une description plus exhaustive. Or, il y a, au moins, deux raisons d'effectuer pareille tentative:

1°. La méthode traditionnelle, implicite, est insuffisante lorsqu'il s'agit de l'appliquer au traitement automatique des textes, et plus particulièrement en matière de traduction automatique.

2°. Une grammaire précise et exhaustive du syntagme cardinal offrirait un moyen de faire progresser la théorie de la grammaire.

rait beaucoup d'intérêt sur le plan théorique. Brainerd (p. 41)¹ note avec justesse:

«Number names ... provide a small, usually well-defined, corpus upon which to test various structural models used for the description of natural language.»

Et il enchaîne:

«The use of formal grammars in the description of number names and numerical expressions has the advantage of giving a potentially complete description of these corpuses, unlike many of the descriptions given in the classical grammars. In the case of a formal grammar for a corpus such as the numerical expressions of a given language, inadequacies in the description given by the grammar at any stage become readily apparent and can be corrected by introducing changes in the grammar.»

De fait, il se trouve justement que ces dernières années ont vu, avec l'avènement de la grammaire générative, s'accroître l'intérêt pour la structure des numéraux, accroissement dont témoigne surtout la publication récente, par H. Brandt Corstius, du livre *Grammars for Number Names*. Cet ouvrage contient une série d'articles sur la description formelle des numéraux de différentes langues (hollandais, anglais, roumain, etc.); en particulier, Barron Brainerd a inséré (pp. 47-48), dans son article «A Transformational-Generative Grammar for Rumanian Numerical Expressions», une grammaire des cardinaux français, sur laquelle je reviendrai plus loin.²

Il convient de mentionner également l'ouvrage intéressant de Hans Chr. Sørensen: *Tal – numre – navne. Glossematiske studier* (Copenhague, 1969), qui aborde le problème des noms de nombre d'un point de vue assez différent du mien.

Dans ce qui suit, je vais tâcher de décrire la structure du syntagme cardinal (dans la langue écrite). Pour ce faire, je me servirai des méthodes élaborées par les tenants de la grammaire générative. Qu'on me pardonne d'avance ce que la notation peut avoir de rebutant pour le

1: Les références renvoient à la bibliographie de la page 229.

2: Je n'ai pu, malheureusement, – parce que j'en ai eu connaissance trop tard – me procurer l'article mentionné dans Corstius, 1968 a (p. 107): R. P. G. de Rijk: «Une grammaire «context-free» pour la génération mécanique des noms de nombre français», in: Briffort and F. van Schepen, *Automation in language etc.*, Euratom CID, Brussels, 1967.

non-initié. Mais j'ai cru pouvoir me dispenser d'explications supplémentaires à ce sujet, étant donné qu'il existe déjà d'excellentes introductions à la théorie générative, où le lecteur intéressé pourra s'informer.³

J'ai laissé de côté les problèmes d'accord entre le cardinal et un éventuel substantif le suivant (*un/une*, ainsi que le flexif du nombre dans le substantif, etc.). Ces phénomènes, ne concernant que le rapport entre le syntagme cardinal et d'autres syntagmes dans la phrase, n'ont pas d'incidence sur la structure interne du syntagme cardinal. Par contre, quand il se présente des problèmes de ce genre à l'intérieur du syntagme, je me suis efforcé d'en tenir compte.

Bien que le syntagme cardinal m'intéresse d'abord et surtout en tant que membre du syntagme substantif:

le pont aérien ... a comporté mille huit cent trente-sept vols
(Monde 30/3 1969, 7),

il n'y a pas lieu de distinguer ici cet emploi de celui où le cardinal figure comme pronom indéfini:

Le nombre des hélicoptères américains abattus ... est ainsi porté, selon les statistiques américaines, à deux mille quatre cent soixante-trois (ib. 3),

puisque la structure du syntagme cardinal reste exactement la même dans les deux cas.

Ce qui est intéressant avec le syntagme cardinal, et ce qui en fait en quelque sorte une espèce de «microlangue», c'est qu'à partir d'une trentaine de formes simples, il est possible de construire – selon des règles rigoureuses – un très grand nombre de formes composées:

«Indeed, when we employ *thousand*, *million*, *billion* in American English we can make 10^{12} – 1 number names which would take about 100 years to list at 300 names per second. Yet any speaker of English can construct any given one of them in a few seconds» (Brainerd, p. 41).

Pourtant le système n'utilise qu'une infime partie des combinaisons possibles: ainsi **dix-cinq*, **deux trentes*, **trente douze*, **trente vingt-deux*, **quarante cents*, **cent deux cents*, **mille douze cents*, **deux millions mille milliards*, etc. etc. sont totalement exclus. Il convient donc d'établir avec précision une grammaire qui permette d'engendrer toutes et rien que

3: Je pense surtout aux ouvrages de Ruwet (1967) et de Bach (1964).

les combinaisons correctes, et d'examiner ensuite la structure que cette grammaire impose aux syntagmes engendrés.

1. G_1 – la grammaire de Brainerd

Commençons par examiner la grammaire que propose Brainerd (pp. 47–48) pour les numéraux cardinaux français et que je désignerai dans la suite par G_1 . Le numérotage des règles employé par l'auteur se réfère à sa grammaire des cardinaux roumains, je l'ai remplacé par des numéros consécutifs:

Vocabulary:

Terminal: *æ, yn, do, trwa, katr, s̄ek, sis, set, iqit, n̄ef, di, dis, diz, ſz, duz, trez, katorz, k̄ez, sez, diset, dizqit, dizn̄ef, v̄e, tr̄at, kar̄at, s̄ek̄at, swas̄at, s̄a, mil, miljō, milja:r, e, d̄a, N_F, N_M*.

Auxiliary: *S, S₀, R, S₁, S₂, T, T₁, T₂, Q, R, R¹, N, U*.

Boundary marker: *#*.

Initial string: #S#.

Production rules:

Context-free rules:

$$1. \quad S \rightarrow S_0 (N)$$

$$2. \quad S_0 \rightarrow \left\{ \begin{array}{l} R \text{ milja:r } (R \text{ miljō }) ((R^1) \text{ mil }) (R) \\ R \text{ miljō } ((R^1) \text{ mil }) (R) \\ (R^1) \text{ mil } (R) \\ R \end{array} \right\}$$

$$3. \quad R \rightarrow \left\{ \begin{array}{l} S_1 \\ T \\ X \\ C \end{array} \right\}$$

$$4. \quad R^1 \rightarrow \left\{ \begin{array}{l} S_2 \\ T \\ X \\ C \end{array} \right\}$$

5. $S_1 \rightarrow \begin{cases} U \\ S_2 \end{cases}$
6. $S_2 \rightarrow \begin{cases} d\emptyset \\ trwa \\ \vdots \\ n\alpha f \end{cases}$
7. $T \rightarrow \begin{cases} T_1 \\ dis \end{cases}$
8. $T_1 \rightarrow \begin{cases} T_2 \\ \tilde{z}z \end{cases}$
9. $T_2 \rightarrow \begin{cases} duz \\ trez \\ \vdots \\ dizn\alpha f \end{cases}$
10. $X \rightarrow \begin{cases} Q \begin{cases} e \tilde{a} \\ S_2 \end{cases} \\ swas\tilde{a}t \ (\begin{cases} di(s) \\ e \tilde{z}z \end{cases}) \\ \vdots \\ katr v\tilde{e} \ (\begin{cases} di(s) \\ T_2 \end{cases}) \\ S_1 \\ \vdots \\ katr v\tilde{e} \ (\begin{cases} di(s) \\ T_1 \end{cases}) \end{cases}$
11. $Q \rightarrow \begin{cases} v\tilde{e} \\ tr\tilde{a}t \\ \vdots \\ swas\tilde{a}t \end{cases}$
12. $C \rightarrow (S_2) s\tilde{a} (\begin{cases} S_1 \\ T \\ X \end{cases})$

Context-sensitive rules:

13. $U \# \rightarrow \tilde{a} \#$
14. $\# U \ N \rightarrow \# \begin{cases} \tilde{a} \ N_M \\ vn \ N_F \end{cases}$

$$15. \ #\beta N \rightarrow \#\beta \begin{Bmatrix} N_M \\ N_F \end{Bmatrix}$$

for $\beta \neq U$ and β not of the form a *miljō* or a *milja:r*.

$$16. \begin{bmatrix} miljō \\ milja:r \end{bmatrix} N \rightarrow \begin{bmatrix} miljō \\ milja:r \end{bmatrix} d\beta \begin{Bmatrix} N_M \\ N_F \end{Bmatrix}$$

G_1 appelle tout d'abord quelques remarques sur des fautes dues peut-être à l'inadveriance.

L'auteur note les formes *di*, *dis*, *diz* dans son vocabulaire terminal. Pourquoi ne pas y inclure également *sē*, *si*, *siz*, *yi*? Dans la règle 10, il marque l'alternative *di/dis*. Pourquoi pas *diz* aussi? Et pourquoi n'y a-t-il pas d'alternative dans la règle 7?

L'emploi de σ et de $\tilde{\sigma}$ semble quelque peu fortuit.

Dans la liste du vocabulaire auxiliaire (c'est-à-dire non-terminal), R figure deux fois, tandis que X et C font défaut.

Dans la règle 10, il faudra entourer l'élément $\begin{Bmatrix} e \tilde{\sigma} \\ S_2 \end{Bmatrix}$ de parenthèses, ou bien il faudra introduire Q directement dans les règles 3 et 4, sinon la grammaire ne sera pas capable d'engendrer des syntagmes comme *vingt*, *trente*, etc., mais seulement *vingt et un*, *vingt-deux*, etc.

Dans la règle 14, il faudra supprimer l'indicateur de frontière ($\#$). Qu'il ne s'agisse pas là d'un lapsus se révèle à la page 49, où l'auteur dit expressément: «In the case of French, *un* is inflected for gender, only when it occurs alone before a noun syntagma».⁴ C'est inexact: *un* s'accorde toujours en genre quand il occupe la dernière – ou la seule – position du syntagme cardinal:

vingt et un anciens amiraux (Julien, Empire 368);

mais:

vingt et une régions (Monde 30/3 1969, 23)

vingt et une réunions régionales (ib. 15/4 1969, 1)

la fermeture de vingt et une autres banques nationales (Afrique 31/3 1969, 20)

soixante et une organisations d'artisans (Monde 16/4 1969, 29)

c'est la géographie future de cent une communes (ib. 15/4 1969, 10).

Cf. aussi Grevisse (§ 405).

4: L'expression «before a noun syntagma» me semble peu heureuse. Je préférerais *inside*.

Défaut peut-être plus grave: G_1 n'engendre pas les formes d'empietement («overlapping»), c'est-à-dire que des nombres comme 1270 ou 2 942 000 000 ont aussi bien les formes *mille deux cent soixante-dix* et *deux milliards neuf cent quarante-deux millions que douze cent soixante-dix* et *deux mille neuf cent quarante-deux millions*. Or, G_1 n'engendre que les premières.

Voici quelques exemples des formes d'empietement:

- les quinze cents élèves des cinq écoles nationales de la marine marchande (Monde 16/4 1969, 15)
- Le conseil des ministres a porté à 2 942 millions le programme d'économies budgétaires (Monde Hebdomadaire 23/1 1969, 7)
- des contrats totalisant 26.700 millions de dollars (Julien, Empire 370)
- 3.429 millions de dollars ... les 29.000 millions investis en Europe ... une augmentation de 1.251 millions ... un total de 5.297 millions (ib. 226-227)
- 1 273 millions de mètres cubes (Calvet, Société 270)
- 22 942 millions de dollars (Afrique 31/3 1969, 16).

Dans le cas des nombres allant de 1100 à 1699, c'est même cette forme qui est la plus usuelle, cf. Gougenheim, p. 69, Grevisse, § 402, Rem. 1, et Togeby, 1965, p. 154.

Ce qui me gêne surtout dans G_1 , c'est la description structurale qu'elle assigne aux syntagmes engendrés. Considérons quelques dérivations. Le numéro de la règle qui a été appliquée est noté entre parenthèses:

- (1) S_0
- (2) R
- (3) C
- (12) $S_2 \ sā \ T$
- (6) $dø \ sā \ T$
- (7) $dø \ sā \ T_1$
- (8) $dø \ sā \ T_2$
- (9) $dø \ sā \ duz$

A cette dérivation correspond l'indicateur syntagmatique («phrase structure marker») suivant:

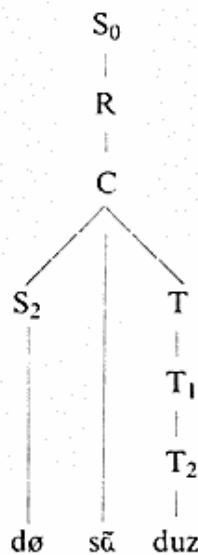


Fig. 1.

On voit que la structure associée à l'élément *duz* est plutôt «abondante». En effet, l'arbre de la figure 1 nous apporte l'information structurale que *duz* fait partie de la catégorie T_2 (c'est-à-dire la classe des éléments susceptibles d'être postposés à *swasāt* sans *e*, cf. la règle 10), laquelle fait partie de la catégorie T_1 (une des classes d'éléments susceptibles d'être postposés à *katr vē*, cf. la règle 10), qui elle-même fait partie de la catégorie T (une des classes d'éléments susceptibles d'être postposés à *sā*, cf. la règle 12). Or, pour le syntagme considéré, c'est uniquement cette dernière information qui paraît pertinente.

Cette «abondance» de structure devient encore plus évidente si l'on examine des syntagmes très simples comme *do* ou *diznæf*. A ces deux syntagmes correspondent les arbres suivants:

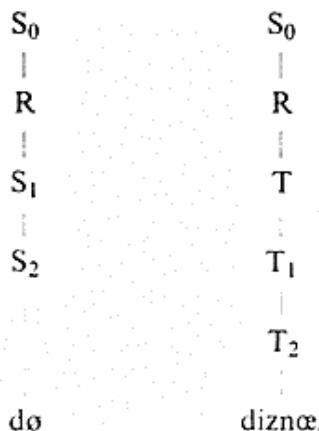


Fig. 2.

A remarquer que l'arbre associé à *diznæf* ne révèle même pas que ce syntagme est une combinaison de deux autres éléments terminaux: *diz* et *næf*.

Cette particularité provient de l'économie de la grammaire de Brainerd. En effet, à l'exception de *dis*, *ɔz* et *swasāt*, les éléments terminaux ne figurent qu'une seule fois dans les règles indépendantes du contexte. Cependant, comme l'a bien dit R. W. Langacker⁵, la description structurale qu'une grammaire associe aux suites terminales engendrées importe plus que l'élégance des règles elles-mêmes. Je vois mal à quoi correspondrait, dans l'intuition du sujet parlant, les structures de la figure 2.

C'est pourquoi je proposerai une grammaire, certes moins élégante et plus complexe, mais qui assigne des indicateurs syntagmatiques plus simples aux syntagmes engendrés.

2. *G₂ – une grammaire des cardinaux danois*

La difficulté d'établir une grammaire simple des cardinaux français tient notamment à ce que les cardinaux de *un* à *dix-neuf* se répartissent en 5 ou 6 classes différentes, à cause des différences de combinabilité de ces éléments. Tel n'est pas le cas des cardinaux danois, et il ne sera sans doute pas inutile d'esquisser d'abord une grammaire engendrant ceux-ci (*G₂*):

1. **#TALORD #**
2. **TALORD** → $\left\{ \begin{array}{l} \text{MIA} \\ \text{MIO} \\ \text{TUS} \\ \text{HUN} \\ \text{HUN}' \\ \text{TIER} \\ \text{ND1-19} \end{array} \right\}$
3. **MIA** → $\left\{ \begin{array}{l} \text{ND1-19} \\ \text{TIER} \\ \text{HUN} \end{array} \right\}$ milliarder ($\left\{ \begin{array}{l} \text{og} \left\{ \text{ND1-19} \right\} \\ \text{TIER} \end{array} \right\}$) $\left\{ \begin{array}{l} \text{HUN} \\ \text{TUS} \\ \text{MIO} \end{array} \right\}$

5: «Transformational analysis does not consist just in finding a set of rules that work; rather, what the linguist is after is the RIGHT set of rules, one that makes true structural claims about the nature of a speaker's linguistic competence» (Langacker, p. 108).

4. MIO → $\left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \\ \text{HUN} \end{matrix} \right\}$ millioner ($\left\{ \begin{matrix} \text{og} & \left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \end{matrix} \right\} \\ \text{HUN} \\ \text{TUS} \end{matrix} \right\}$)
5. TUS → $\left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \\ \text{HUN} \end{matrix} \right\}$ tusind(e) ($\left\{ \begin{matrix} \text{og} & \left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \end{matrix} \right\} \\ \text{HUN} \end{matrix} \right\}$)
6. HUN → (ND1-9) hundrede (og $\left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \end{matrix} \right\}$)
7. HUN¹ → $\left\{ \begin{matrix} \text{ND11-19} \\ \text{TIER} \end{matrix} \right\}$ hundrede (og $\left\{ \begin{matrix} \text{ND1-19} \\ \text{TIER} \end{matrix} \right\}$)
8. TIER → (ND1-9 + og) $\left\{ \begin{matrix} \text{tyve} \\ \text{tredive} \\ \text{fyrre} \\ \text{halvtreds} \\ \text{tres} \\ \text{halvfjérds} \\ \text{firs} \\ \text{halvfems} \end{matrix} \right\}$
9. ND1-19 → en, to, tre, fire, fem, seks, syv, otte, ni, ti, elleve, tolv, tretten, fjorten, femten, seksten, sytten, atten, nitten
10. ND11-19 → elleve, tolv, tretten, fjorten, femten, seksten, sytten, atten, nitten
11. ND1-9 → en, to, tre, fire, fem, seks, syv, otte, ni

J'utilise des majuscules pour désigner les éléments non-terminaux, des minuscules pour les éléments terminaux.

Cette grammaire est une grammaire de constituants indépendante du contexte («context-free phrase structure grammar»). Il faut la compléter par des règles transformationnelles obligatoires, qui tiendront compte des problèmes d'accord interne:

T.1.

$$X \text{ en } \left[\begin{matrix} \text{hundrede} \\ \text{tusind(e)} \end{matrix} \right] Y \Rightarrow X \text{ et } \left[\begin{matrix} \text{hundrede} \\ \text{tusind(e)} \end{matrix} \right] Y$$

T.2.

$$X \text{ en } \begin{bmatrix} \text{millioner} \\ \text{milliarder} \end{bmatrix} Y \Rightarrow X \text{ en } \begin{bmatrix} \text{million} \\ \text{milliard} \end{bmatrix} Y$$

G_2 répartit les éléments de *en* à *nitten* en trois classes différentes (règles 9 à 11). ND1-19 est la classe des éléments qui n'admettent pas d'antéposition ni de postposition d'autres éléments. ND1-9 est la classe des éléments susceptibles d'être antéposés à *tyve*, *tredive*, etc. (cf. règle 8). ND11-19 se justifie uniquement par le fait qu'il faut exclure *ti* de la règle 7, pour éviter le syntagme **ti hundrede*.

Le caractère compliqué des règles 3 à 7 ainsi que la répétition des éléments terminaux dans les règles 9 à 11 sont compensés par la description structurale très simple que la grammaire assigne aux suites terminales. En voici quelques exemples:

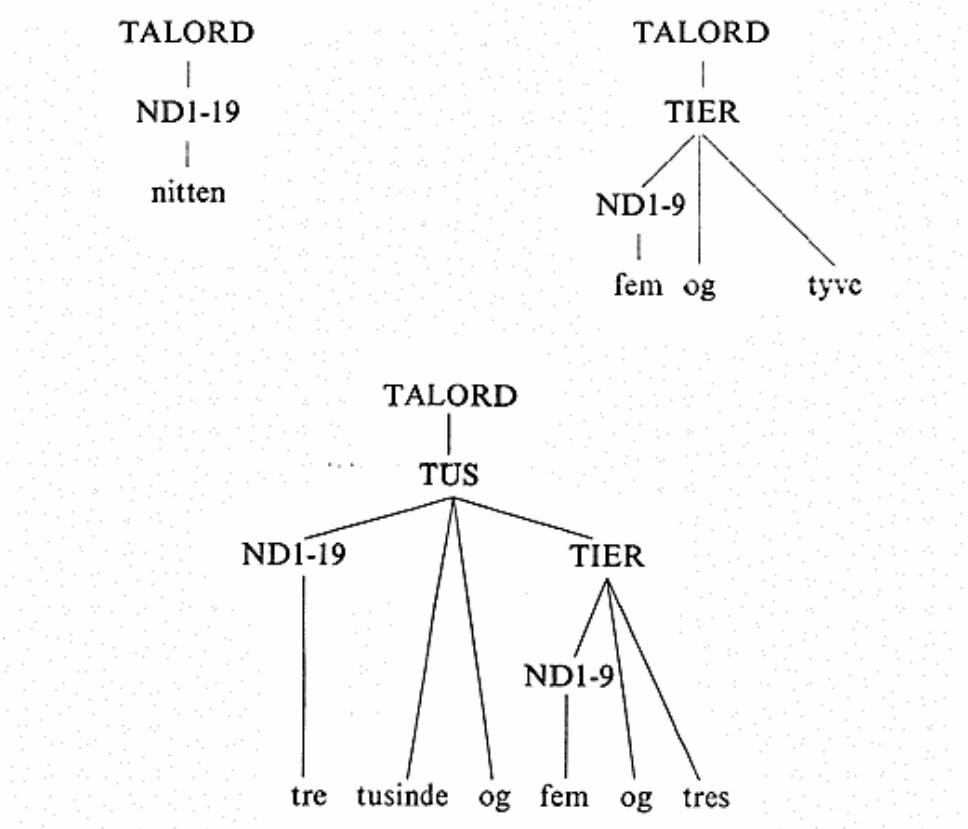


Fig. 3.

HUN¹ de la règle 7 permet d'engendrer les formes d'empietement avec *hundrede*: *toogtyve hundrede*, *femogfyrre hundrede*, etc., tandis que l'appli-

14*

cation de la règle 5 donnerait : *to tusinde to hundrede*, etc.⁶ Il semble tentant, à première vue, de réunir les règles 6 et 7 en une seule, mais ce n'est pas possible, car HUN apparaît dans les règles 3 à 5 et donnerait ainsi naissance à des syntagmes agrammaticaux, tels que **tre tusinde toogtyve hundrede*.

3. G_3 – une grammaire des cardinaux français

3.1. Structure du syntagme cardinal

La grammaire suivante (G_3) est construite en partie selon les mêmes principes que G_2 :

1. #CARD#

2. CARD → $\left\{ \begin{array}{l} \text{MIA} \\ \text{MIO} \\ \text{MIO'} \\ \text{MIL} \\ \text{CENT} \\ \text{CENT'} \\ \text{DIZAINE} \\ \text{NF1-19} \end{array} \right\}$

3. MIA → $\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{CENT'} \\ \text{MIL} \end{array} \right\}$ milliards ($\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{MIL} \\ \text{MIO} \end{array} \right\}$)

4. MIO → $\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \end{array} \right\}$ millions ($\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{MIL} \end{array} \right\}$)

5. MIO' → $\left\{ \begin{array}{l} \text{CENT'} \\ \text{MIL} \end{array} \right\}$ millions ($\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{MIL} \end{array} \right\}$)

6: La délimitation de la liste des éléments antéposables présente toutefois des difficultés. Dirait-on, par exemple, *tredivre hundrede* ou *nioghalvfems hundrede*? J'en doute.

6. MIL \rightarrow ($\begin{cases} \text{NF2-19} \\ \text{DIZAINE} \\ \text{CENT} \end{cases}$) mille ($\begin{cases} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \end{cases}$)
7. CENT \rightarrow $\begin{cases} \text{cent} \\ \text{NF2-9 + cents} \end{cases}$ ($\begin{cases} \text{NF1-19} \\ \text{DIZAINE} \end{cases}$)
8. CENTⁱ \rightarrow NF11-19 + cents ($\begin{cases} \text{NF1-19} \\ \text{DIZAINE} \end{cases}$)
9. DIZAINE \rightarrow $\begin{cases} \text{vingt} \\ \text{trente} \\ \text{quarante} \\ \text{cinquante} \\ \text{soixante} \\ \text{soixante} \\ \text{quatre-vingts} \end{cases}$ ($\begin{cases} \text{et un} \\ \text{- + NF2-9} \\ \text{et onze} \\ \text{- + NF10-19S} \\ \text{- + NF1-19} \end{cases}$)
10. NF1-19 \rightarrow un, deux, trois, quatre, cinq, six, sept, huit, neuf, dix (- + NF7-9), onze, douze, treize, quatorze, quinze, seize
11. NF2-19 \rightarrow deux, trois, quatre, cinq, six, sept, huit, neuf, dix (- + NF7-9), onze, douze, treize, quatorze, quinze, seize
12. NF11-19 \rightarrow onze, douze, treize, quatorze, quinze, seize, dix + - + NF7-9
13. NF10-19S \rightarrow dix (- + NF7-9), douze, treize, quatorze, quinze, seize
14. NF2-9 \rightarrow deux, trois, quatre, cinq, six, sept, huit, neuf
15. NF7-9 \rightarrow sept, huit, neuf

Les règles 1 à 15 ne constituent qu'une partie de G_3 , à savoir la grammaire de constituants indépendante du contexte qui assigne aux syntagmes cardinaux leur structure profonde. Les problèmes d'accord interne seront traités au § 3.2. En attendant, G_3 engendrera donc des syntagmes tels que **un millions*, **deux cents trois*, **quatre-vingts-dix*, etc., mais ceci n'a

aucune influence sur la structure profonde des syntagmes et il me paraît préférable de discuter auparavant de celle-ci.⁷

Voici d'abord, à titre d'illustration, une dérivation avec l'indicateur syntagmatique correspondant:

- (1) #CARD#
 - (2) MIO
 - (4) DIZAINE millions MIL
 - (6) DIZAINE millions NF2-19 mille CENT
 - (7) DIZAINE millions NF2-19 mille NF2-9 cents DIZAINE
 - (9) cinquante millions NF2-19 mille NF2-9 cents DIZAINE
 - (9) cinquante millions NF2-19 mille NF2-9 cents trente-NF2-9
 - (11) cinquante millions quinze mille NF2-9 cents trente-NF2-9
 - (14) cinquante millions quinze mille sept cents trente-NF2-9
 - (14) cinquante millions quinze mille sept cents trente-cinq

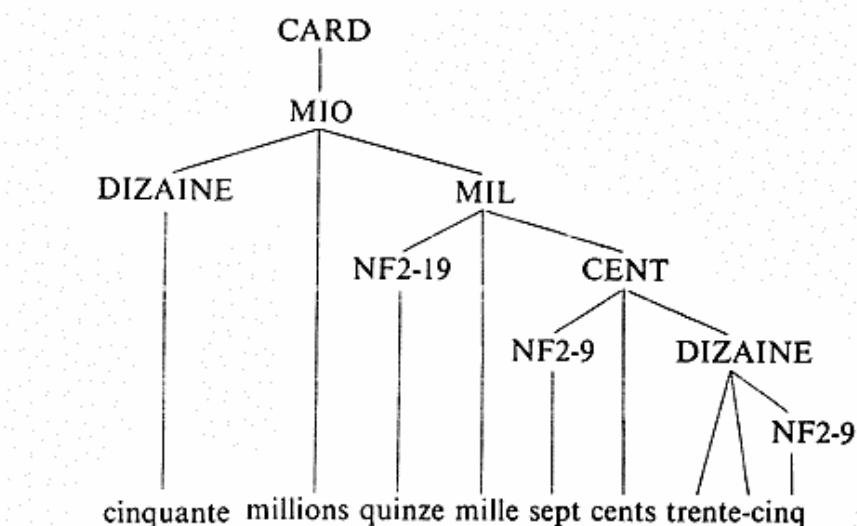


Fig. 4.

G_3 , à l'instar de G_1 et de G_2 , ne donne pas de règles pour les cardinaux supérieurs à ceux qui sont formés à l'aide de *milliard*. Théoriquement, on pourrait former des syntagmes contenant *trillion*, *quatrillion*, *quintil-*

7: Pourtant, pour ne pas trop agacer le lecteur, j'utiliserais les formes orthographiquement correctes, sauf dans les cas où il s'agit expressément de syntagmes engendrés par G_3 .

lion, etc. Mais, d'une part, ces syntagmes sont «pour ainsi dire inusités» (Grevisse, § 402, Rem. 3), d'autre part, ils ne présentent aucun intérêt syntaxique particulier puisque, à partir de *million*, les règles restent analogues. – Je n'ai pas indiqué l'élément terminal *billion* (= *milliard*) parce qu'il est également d'un emploi très rare, mais on pourrait facilement l'insérer dans la règle 3: {milliards} {billions}.

Je n'ai pas inclus *zéro* parmi les éléments terminaux, car il n'entre dans la formation d'aucun constituant. Si on tient à le considérer comme un cardinal, on pourrait l'introduire directement dans la règle 2, mais, dans ce cas-là, ne faudrait-il pas y inclure également *plusieurs*, comme le fait Togeby (1951, p. 162)? – Dans une grammaire des nombres décimaux, il faudrait naturellement placer *zéro* sur un pied d'égalité avec les autres éléments terminaux, puisque, dans une telle grammaire, il ferait partie intégrante du système.

Les éléments MIO¹ et CENT¹ donnent naissance aux formes d'empîtement. Pour *million*, il s'agit des cardinaux de *mille millions* à *neuf cent mille millions*; pour *cent*, il s'agit seulement des cardinaux de *onze cents* à *dix-neuf cents* (cf. p. 207), puisque, à l'encontre de l'usage danois, des formes telles que **vingt-deux cents* sont exclues, cf. Togeby, 1965, p. 154. – Il est nécessaire d'avoir aussi bien les éléments non-terminaux MIO et CENT que MIO¹ et CENT¹, parce que MIO figure dans la règle 3 et CENT dans les règles 3 à 6, où ils contribuent à la formation de constituants placés à un niveau supérieur. Si l'on se contentait d'un élément CENT en réunissant les règles 7 et 8 en une seule, la règle 6, par exemple, donnerait naissance à des syntagmes comme **douze cents mille douze cents*. – Qu'il ne soit pas nécessaire d'avoir également un élément MIA¹ tient au fait que MIA n'apparaît pas dans des syntagmes supérieurs. Si j'avais voulu faire engendrer par G₃ des syntagmes contenant *trillion*, il m'aurait fallu scinder la règle 3 en deux pour établir les constituants MIA et MIA¹. – Il n'y a pas d'élément MIL¹. C'est que les formes contenant *mille* ne peuvent recouvrir le domaine de *million*: ainsi 1500000 se lit *un million cinq cent mille* et non **quinze cent mille*.^{7A}

7 A: Et pourtant... Lors de la correction des épreuves de cet article, Henrik Prebensen me communique cet exemple inquiétant: *Quelque quinze cent mille habitants des provinces orientales* (A. Fontaine: *Histoire de la Guerre Froide*, I, Paris, Fayard, 1967, p. 204).

quatre-vingts figure comme un seul élément terminal et n'est pas analysé en *vingt* précédé de *quatre*. Il y a plusieurs raisons à cela: D'abord, il me semble douteux qu'une telle analyse corresponde à l'intuition du sujet parlant.⁸ Ensuite, cela compliquerait inutilement la grammaire, et le fait que *vingt* admette un élément antéposé dans ce seul cas ne me paraît pas une raison suffisante pour le traiter isolément. Enfin, selon Togeby (1965, p. 153), *quatre-vingts* se prononce [katrə vɛ̃] et non [katvɛ̃], ce qui pourrait indiquer qu'il ne s'agit pas de l'élément terminal *quatre*, puisque *quatre cents*, *quatre mille* se prononcent bien [kat sã], [kat mil]. – Cependant, la manière dont *quatre-vingts* s'attache les éléments postposés dénote une affinité avec les éléments *cent(s)* et *mille*. Cf.: *quatre-vingt-un* et *cent un*, mais: *vingt et un*. Et aussi: *quatre-vingt-onze* et *cent onze*, mais: *soixante et onze*.⁹

Comme dans G_2 , mais à la différence de G_1 , la règle 2 permet d'engendrer plus directement (c'est-à-dire avec «moins de» structure) des syntagmes tels que *trois*, *dix-neuf*, *vingt-deux*, qui auront les structures suivantes:

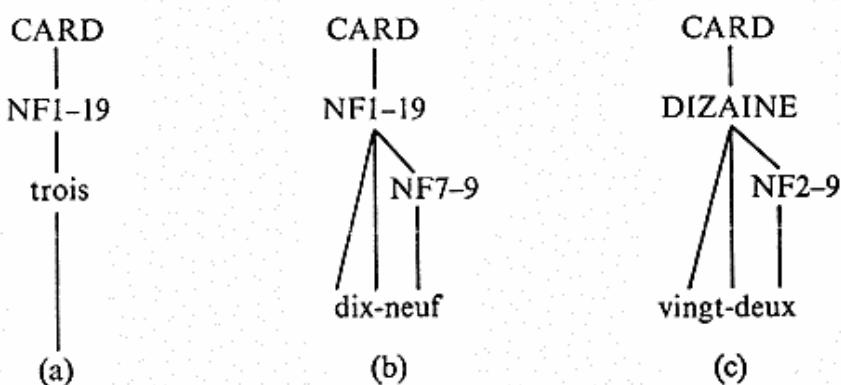


Fig. 5.

8: En danois, *firs(indstyve)* est senti comme un seul élément terminal à part entière et n'évoque pas l'idée 4×20 . Il est vrai que, là, l'élément *styve* a disparu depuis longtemps de l'usage courant.

9: *Huit cent quatre-vingt-onze trains* (Monde 28/3 1969, 10), *cent quatre-vingt-onze enseignants* (ib. 16/4 1969, 13), *cent soixante et onze votants* (ib. 28/3 1969, 20), *En 1967, deux cent soixante et onze prêtres auraient quitté la vie religieuse* (ib. 15/4 1969, 12). Cf. Gougenheim, p. 69, Grevisse, § 403, Rem. 1, Togeby, 1965, p. 153.

On remarquera que, contrairement à ce qui se produisait pour l'engendrement de *dix-neuf* par G_1 (cf. la figure 2), l'arbre de la figure 5 (b) rend compte du caractère composé de *dix-neuf*.

G_3 répartit les éléments terminaux *un*, *deux* ... *dix-neuf* en 6 classes (règles 10 à 15). $NF1-19$ comprend tous les éléments concernés. Les éléments appartenant à $NF1-19$ peuvent apparaître avant et après *milliards* (règle 3) et *millions* (règles 4 et 5), après *mille* (règle 6), *cent* (règles 7 et 8) et *quatre-vingt* (règle 9). $NF2-19$ comprend ceux des éléments de $NF1-19$ qui sont susceptibles d'être antéposés à *mille* (règle 6). $NF11-19$ comprend les éléments susceptibles d'être antéposés à *cents* dans les formes d'empîtement (règle 8). $NF10-19S$ comprend les éléments susceptibles d'être postposés à *soixante* à l'aide d'un trait d'union, mais non à *vingt*, *trente*, ... *cinquante* (règle 9). $NF2-9$ comprend les éléments susceptibles d'être antéposés à *cents* dans les formes sans empiètement (règle 7) et postposés à *vingt*, *trente*, etc. à l'aide d'un trait d'union (règle 9). $NF7-9$ comprend les trois éléments susceptibles d'être postposés à *dix* (règles 10 à 13).

On voit que $NF7-9$ constitue un sous-ensemble de $NF2-9$, et $NF2-9$ un sous-ensemble de $NF2-19$, etc. Il aurait été plus élégant de faire exprimer ce fait explicitement par la grammaire, par exemple au moyen des règles suivantes (G_{3A}):¹⁰

C'est justement une formulation analogue¹¹ qui confère aux syntagmes

$$10A. \quad NF1-19 \rightarrow \left\{ \begin{array}{l} \text{un} \\ \text{NF2-19} \end{array} \right\}$$

$$11A. \quad NF2-19 \rightarrow \left\{ \begin{array}{l} \text{NF2-9} \\ \text{dix} \\ \text{NF11-19} \end{array} \right\}$$

$$12A. \quad NF11-19 \rightarrow \text{onze, douze, treize, quatorze, quinze, seize,} \\ \quad \text{dix} + - + NF7-9$$

$$14A. \quad NF2-9 \rightarrow \text{deux, trois, quatre, cinq, six, NF7-9}$$

$$15A. \quad NF7-9 \rightarrow \text{sept, huit, neuf}$$

10: *dix* figure, isolé, dans la règle 11A et la classe $NF10-19S$ a disparu. Cela est dû au fait que $NF11-19$ n'est pas un sous-ensemble de $NF10-19S$.

11: Cf. G_1 , règles 5 à 9.

engendrés par la grammaire de Brainerd ce «surplus» de structure que je lui ai reproché. En effet, le syntagme *sept cent douze* engendré par G_3 a la structure:

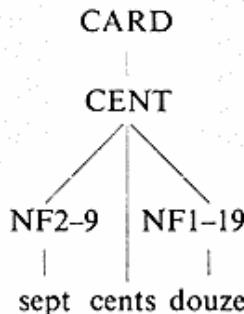


Fig. 6.

Tandis que G_{3A} lui imposerait cette structure:

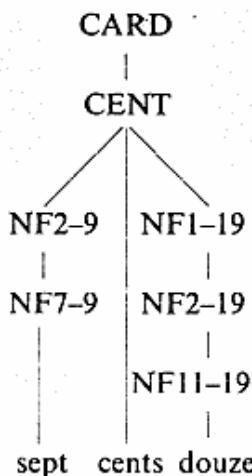


Fig. 7.

Ce dernier indicateur syntagmatique nous fournit pas mal d'informations non pertinentes (voir la critique relative à la figure 1). Voilà pourquoi j'ai écarté la version G_{3A} .

Togeby (1951, p. 162) estime qu' «il n'y a pas de différence fonctionnelle entre les neuf nombres» (c'est-à-dire entre les nombres *deux*, *trois*, ... *dix*). La discussion qui précède aura montré que je ne suis pas

d'accord avec lui. *sept*, *huit* et *neuf* se distinguent effectivement des autres en ce qu'ils sont les seuls à pouvoir apparaître après *dix*. *dix* se singularise de plusieurs manières: il ne peut pas apparaître avant *cent*, ni après *vingt*, *trente*, etc., alors que les autres le peuvent; il peut s'adjointre un élément postposé, les autres, non.

Le phénomène des sous-ensembles se manifeste également si l'on examine les règles 3 à 8 de G_3 . Par exemple, l'ensemble des constituants spécifiés dans la règle 4 après *millions* est un sous-ensemble des éléments spécifiés dans la règle 3 après *milliards*. En tenant compte de ce fait, on pourrait simplifier considérablement la grammaire (j'omets les formes d'empîtement):

- | | | |
|-------------|---------------|---|
| 3A. MIA | \rightarrow | CENT milliards (MIO) |
| 4A. MIO | \rightarrow | $\left\{ \begin{array}{l} \text{MIL} \\ \text{CENT + millions (MIL)} \end{array} \right\}$ |
| 5A. MIL | \rightarrow | $\left\{ \begin{array}{l} \text{CENT} \\ \text{(CENT) mille (CENT)} \end{array} \right\}$ |
| 6A. CENT | \rightarrow | $\left\{ \begin{array}{l} \text{DIZAINE} \\ \left\{ \begin{array}{l} \text{cent} \\ \text{NF2-9 + cents} \end{array} \right\} \end{array} \right\} \text{ (DIZAINE)}$ |
| 7A. DIZAINE | \rightarrow | $\left\{ \begin{array}{l} \text{NF1-19} \\ \text{vingt} \\ \text{trente} \\ \vdots \\ \vdots \\ \vdots \\ \dots \end{array} \right\}$ |
| 9A. DIZAINE | \rightarrow | $\left\{ \begin{array}{l} \text{etc., comme la règle 9 de G}_3 \end{array} \right\}$ |

Suivent les règles 10A, etc., de G_{3A} .

Mais ici encore la simplicité des règles s'acquiert au prix d'une description structurale qui me paraît totalement inadmissible. Voici, par exemple, l'arbre correspondant à *deux millions quinze*:

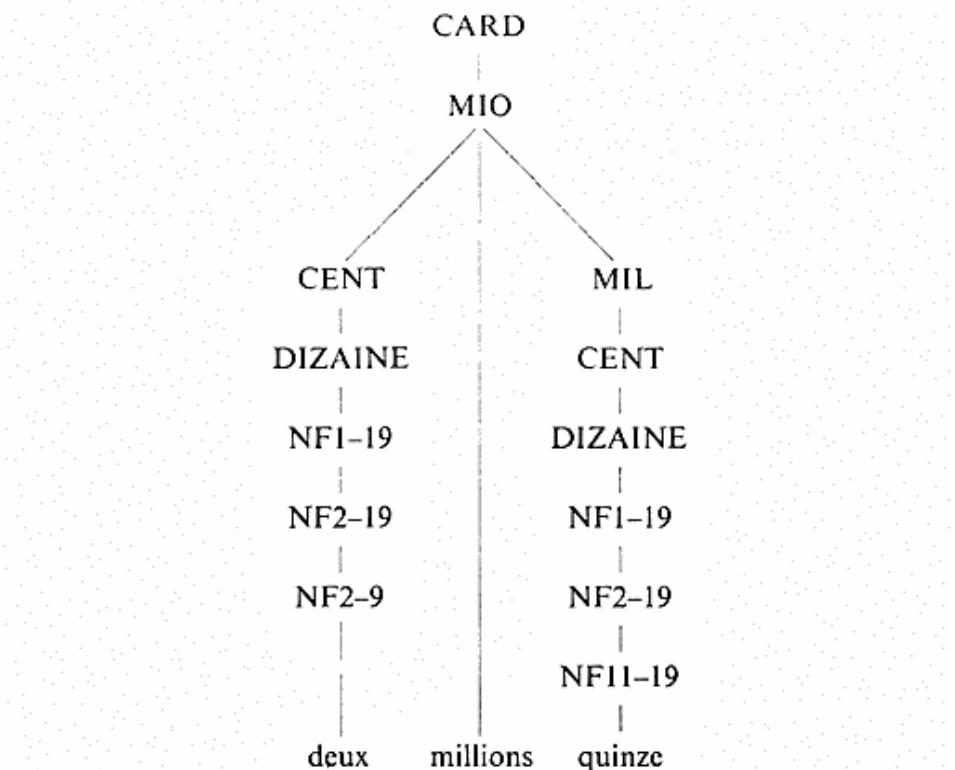


Fig. 8.

Comment les rapports de subordination sont-ils rendus par nos grammaires? Il se trouve qu'ici G_3 a au moins un avantage sur G_1 . Prenons, par exemple, le syntagme *deux millions trois mille*. G_1 lui assigne la structure suivante:

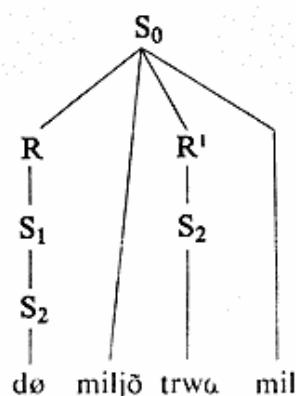


Fig. 9.

alors que G_3 produirait cet arbre:

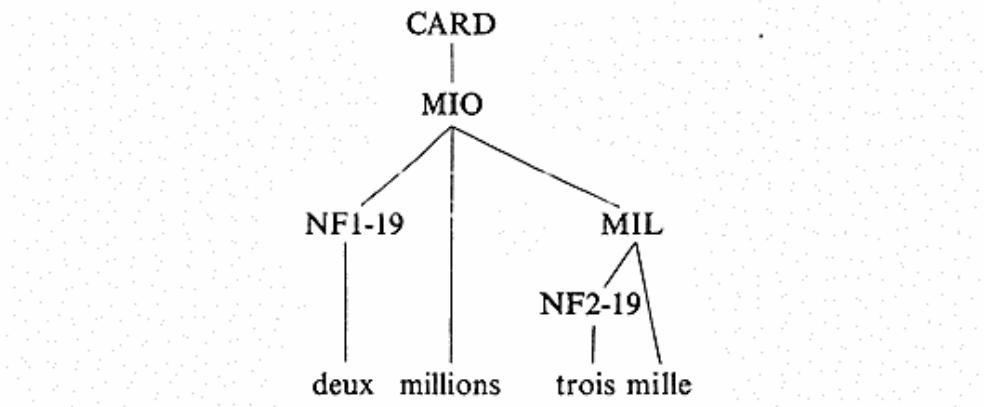


Fig. 10.

Ce qui signifie que G_1 analyse le syntagme en quatre constituants immédiats, n'indiquant pas si *trois* se rattache à *millions* ou à *mille*. G_3 exprime clairement que *trois* se rattache à *mille*, les deux éléments étant des constituants de *MIL*.

Il y a toutefois un rapport très important qui n'est indiqué explicitement ni par G_1 ni par G_3 . Ainsi, en regardant la figure 7, on se rend compte que *sept* et *douze* sont considérés comme se rapportant à *cents* d'une manière identique. Ce phénomène ressort peut-être plus clairement si l'on se sert de parenthèses étiquetées pour représenter la description structurale (cette représentation est exactement équivalente à la représentation par «arbre»). Voici le même syntagme parenthétisé :

$((\text{(sept)}_{\text{NF2-9}} \text{ cents} (\text{douze})_{\text{NF1-19}} \text{CENT}) \text{CARD}$

Pourtant, pour obtenir la valeur numérique 712 – c'est-à-dire pour procéder à une interprétation sémantique – il est clair que *sept* et *douze* n'ont pas ici la même fonction. En réalité, il faut :

- 1° multiplier cent par sept,
- 2° additionner le résultat de cette multiplication et douze.

Personne ne songerait à additionner d'abord et à multiplier ensuite.

On pourrait faire exprimer explicitement ces rapports par la grammaire en substituant CENT-A à l'élément CENT et en remplaçant la règle 7 par :

7A. CENT-A \rightarrow CENT-M ($\left\{ \begin{matrix} \text{NF1-19} \\ \text{DIZAINE} \end{matrix} \right\}$)

$$7B. \text{CENT-M} \rightarrow \left\{ \begin{array}{l} \text{cent} \\ \text{NF2-9} + \text{cents} \end{array} \right\}$$

Avec cette modification, G_3 affectera une autre description structurale à *sept cents douze*:

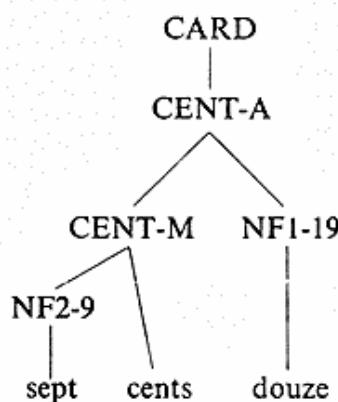


Fig. 11.

Ou en représentation parenthétisée:

$((((\text{sept})_{\text{NF2-9}} \text{cents})_{\text{CENT-M}} (\text{douze})_{\text{NF1-19}})_{\text{CENT-A}})_{\text{CARD}}$

CENT-M sera alors interprété: «les constituants immédiats doivent être multipliés»; CENT-A: «les constituants immédiats doivent être additionnés».

Toutefois, pour une application automatique de G_3 , si l'on veut, par exemple, faire traduire par un ordinateur des cardinaux français en chiffres arabes ou dans une autre langue, cette nouvelle analyse n'est pas indispensable. On pourrait se tirer d'affaire par les trois conventions suivantes:

- 1° A l'intérieur d'un même constituant, les éléments antéposés doivent être traités avant les éléments postposés;
- 2° Le rapport d'antéposition équivaut à la multiplication;
- 3° Le rapport de postposition équivaut à l'addition.

Mais ces conventions ne peuvent s'appliquer à G_1 , qui ne distingue pas dans tous les cas de manière suffisamment claire entre antéposition et postposition. Pour G_2 , il y aura des complications supplémentaires parce qu'il y a des cas où l'élément antéposé, dans un syntagme cardinal danois, doit être additionné, voir G_2 règle 8.

Les éléments terminaux de G_3 se répartissent en cinq grandes classes:

Classe 1: *un, deux, trois, quatre, cinq, six, sept, huit, neuf, onze, douze, treize, quatorze, quinze, seize.*

(Les éléments qui n'admettent pas d'autres éléments postposés ni antéposés.)

Classe 2: *dix, vingt, trente, quarante, cinquante, soixante, quatre-vingts.*

(Les éléments qui admettent un élément postposé.)

Classe 3: *cent(s), mille.*

(Les éléments qui admettent à la fois des éléments antéposés et postposés, et qui peuvent seuls faire fonction de syntagme.)

Classe 4: *millions, milliards.*

(Les éléments qui admettent un élément postposé et exigent un élément antéposé.)

Classe 5: – (trait d'union), *et*.

(Les éléments qui apparaissent uniquement entre un élément de la classe 2 suivi d'un élément de la classe 1.)

La distinction entre les classes 3 et 4 est nécessitée par le fait qu'à la différence de *cent* et de *mille*, *million* et *milliard* n'apparaissent jamais seuls comme syntagmes cardinaux, Cf.:

près de cent experts (Express 18/9 1967, 22)

Plus de mille maisons (Monde 27/3 1969, 2)

plus d'un million d'habitants (Calvet, Société 62)

plus d'un million de dollars (Observateur 5/8 1968, 4)

Ces classes ne se dégagent pas nettement de G_3 . On pourrait les faire ressortir immédiatement, par exemple, en commençant la grammaire par des règles comme:

1. #CARD#

2. CARD → {SIMPLE
COMPOSÉ}

3. SIMPLE → {CL1
CL2
CL3}

Mais on voit immédiatement que, pour formuler les règles spécifiant les

constituants de COMPOSÉ (c'est-à-dire de formes comme *vingt-deux*, *trois cent dix-neuf*, etc.), il faudrait recourir à une foule de règles dépendantes du contexte.

3.2. Problèmes d'accord interne

Pour donner aux syntagmes engendrés par G_3 une forme en accord avec l'usage orthographique courant, il faut compléter les règles 1 à 15 des pages 212-13 par trois règles transformationnelles obligatoires:

T.1.

$$\# \text{un} \begin{bmatrix} \text{millions} \\ \text{milliards} \end{bmatrix} X \Rightarrow \# \text{un} \begin{bmatrix} \text{million} \\ \text{milliard} \end{bmatrix} X$$

T.2.

$$X \text{ cents} \begin{bmatrix} \text{NF1-19} \\ \text{DIZAINE} \\ \text{mille} \end{bmatrix} Y \Rightarrow X \text{ cent} \begin{bmatrix} \text{NF1-19} \\ \text{DIZAINE} \\ \text{mille} \end{bmatrix} Y$$

T.3.

$$X \text{ quatre-vingts} \begin{bmatrix} \text{-NF1-19} \\ \text{mille} \end{bmatrix} Y \Rightarrow X \text{ quatre-vingt} \begin{bmatrix} \text{-NF1-19} \\ \text{mille} \end{bmatrix} Y$$

T.2 et T.3 transforment des syntagmes comme **deux cents vingt* et **quatre-vingts-dix* en respectivement *deux cent vingt* et *quatre-vingt-dix*.

Chevalier *et al.* (§ 403) sont d'avis que «cette bizarrerie orthographique, qui date de la fin du XVII^e siècle, semble peu à peu sortir de l'usage». Cette opinion semble – malheureusement – un peu trop optimiste. Du moins n'ai-je pas trouvé, parmi les quelques centaines d'exemples que j'ai relevés, un seul qui profite de la liberalité accordée par l'arrêté ministériel du 26 février 1901. Cf.:

Quatre-vingts postes d'assistants (Monde Hebdomadaire 28/11 1968, 8)
quatre-vingts ans (Observateur 17/3 1969, 37)

quatre-vingt-dix lits (ib. 24)
quatre-vingt-dix-neuf ans (Monde 27/3 1969, 5)

quelque deux cents barons de l'empire (Julien, Empire 351)

Quatre cents malades (Express 18/9 1967, 33)

A cinq cents mètres des portes (Zegel, Idées 47)

Treize mille demandes d'inscription ont été reçues, dont quatre mille six cents pour la première année (Monde Hebdomadaire 28/11 1968, 8)

avec ses huit cent quatre-vingt-trois filiales (Monde 28/3 1969, 6)
Deux cent quatre-vingts pièces (ib. 16/4 1969, 23)

Huit cent quatre-vingt-onze trains (ib. 28/3 1969, 10)
 deux cent cinquante mille personnes (Zegel, Idées 19)
 sept cent mille étudiants (Observateur 17/3 1969, 9)
 un million trois cent mille familles (ib. 8/5 1968, 22)
 deux millions huit cent mille (Monde 27/3 1969, 12).

Même la règle «superbizarre» qui veut que *cents* et *quatre-vingts* gardent leur *s* devant *millions* et *milliards* (mais pas devant *mille*) est respectée:

il y a deux cents millions d'années (Afrique 31/3 1969, 44)

Lorsqu'il s'agit d'*engendrer* des suites terminales, les règles transformationnelles ne posent pas de problèmes particuliers pour le traitement automatique. Il suffit d'examiner les suites engendrées par G_3 avec *leur description structurale* pour voir si une des suites à gauche de la flèche dans T.1, T.2 et T.3 s'y trouve et, dans l'affirmative, lui substituer la suite correspondante à droite de la flèche.

Le problème est autrement délicat si l'on veut faire *analyser* par la machine des suites terminales données. Dans ce cas-là, les règles de la grammaire doivent être considérées non plus comme des règles d'expansion, mais comme des règles de réduction. Autrement dit, il faut les lire à rebours, de droite à gauche. Pour appliquer une règle comme T.2, il faut donc savoir si, dans une suite donnée, se trouve un constituant DIZAINE: mais cela, on ne peut pas le savoir avant d'avoir analysé la suite, ce qu'on ne peut pas faire avant d'avoir appliqué la règle T.2, s'il le faut. Cette difficulté tient au fait que les règles transformationnelles sont dépendantes du contexte.

Il serait donc intéressant de pouvoir exprimer les faits d'accord à l'aide de règles indépendantes du contexte. Dans notre cas, cela est parfaitement possible puisque l'ensemble des syntagmes que nous considérons est un ensemble fini. On pourrait, par exemple, construire une grammaire G_4 en modifiant les règles de G_3 de manière que celles-ci engendrent l'ensemble des syntagmes orthographiquement corrects et rien que ceux-là. Ces modifications entraînent évidemment une complication assez considérable de la grammaire, mais les syntagmes engendrés par G_4 seront structuralement équivalents à ceux de G_3 . Voici les règles de G_4 (je ne donne pas celles qui restent identiques aux règles correspondantes de G_3):

3. MIA → $\left\{ \begin{array}{l} \text{un milliard} \\ \text{NF2-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{CENT} \\ \text{MIL} \end{array} \right\}$ milliards $\left(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{MIL} \\ \text{MIO} \end{array} \right\} \right)$

4. MIO → $\left\{ \begin{array}{l} \text{un million} \\ \text{NF2-19} \\ \text{DIZAINE} \\ \text{CENT} \end{array} \right\}$ millions $\left(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \\ \text{MIL} \end{array} \right\} \right)$

6. MIL → $(\left\{ \begin{array}{l} \text{NF2-19} \\ \text{DIZAINE-AM} \\ \text{CENT-AM} \end{array} \right\})$ mille $(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \\ \text{CENT} \end{array} \right\})$

7. CENT → $\left\{ \begin{array}{l} \text{cent} \left(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \end{array} \right\} \right) \\ \text{NF2-9} \left\{ \begin{array}{l} \text{cent} \left(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \end{array} \right\} \right) \\ \text{cents} \end{array} \right\} \end{array} \right\}$

7A. CENT-AM → (NF2-9) cent $(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE-AM} \end{array} \right\})$

8. CENT¹ → NF11-19 $\left\{ \begin{array}{l} \text{cent} \left(\left\{ \begin{array}{l} \text{NF1-19} \\ \text{DIZAINE} \end{array} \right\} \right) \\ \text{cents} \end{array} \right\}$

9. DIZAINE → $\left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{vingt} \\ \text{trente} \\ \text{quarante} \\ \text{cinquante} \end{array} \right\} \left(\left\{ \begin{array}{l} \text{et un} \\ \text{-} + \text{NF2-9} \end{array} \right\} \right) \\ \text{soixante} \\ \text{soixante} \left\{ \begin{array}{l} \text{et onze} \\ \text{-} + \text{NF10-19S} \end{array} \right\} \\ \text{quatre-vingts} \\ \text{quatre-vingt} + \text{-} + \text{NF1-19} \end{array} \right\}$

$$9A. DIZAINE-AM \rightarrow \left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{vingt} \\ \text{trente} \\ \text{quarante} \\ \text{cinquante} \\ \text{soixante} \end{array} \right\} \left(\left\{ \begin{array}{l} \text{et un} \\ \text{ - } + \text{NF2-9} \end{array} \right\} \right) \\ \text{soixante} \left\{ \begin{array}{l} \text{et onze} \\ \text{ - } + \text{NF10-19S} \end{array} \right\} \\ \text{quatre-vingt} \left(\text{ - } + \text{NF1-19} \right) \end{array} \right\}$$

Il y a encore une autre manière de résoudre ce problème à l'aide de règles indépendantes du contexte. Cette méthode m'a été suggérée par Peter Møller-Nielsen. D'abord, il faut entourer l's final dans les règles 3, 4, 7 et 8 de G_3 par des parenthèses: *milliard(s)*, *million(s)*, *cent(s)*. J'appellerai G_5 la nouvelle grammaire qui résulte de ces modifications. Si l'on désigne par $L(G_i)$ le langage (c'est-à-dire l'ensemble des suites terminales) engendré par une grammaire G_i , on voit que $L(G_5)$ contient plus de suites que $L(G_3)$, parce que G_5 engendre maintenant toutes les suites qui forment des syntagmes corrects, plus quelques suites incorrectes (**deux cent*, **deux cents trois*, par exemple). Il s'agit maintenant de construire une autre grammaire (G_6) qui, par des règles indépendantes du contexte, engendre toutes les suites incorrectes de $L(G_5)$, plus, éventuellement, d'autres suites incorrectes, mais aucune suite qui soit un syntagme cardinal correct. Cette «anti-grammaire» servira de «filtre» pour les suites de $L(G_5)$, en ce sens que seulement celles des suites de $L(G_5)$ qui ne seront pas acceptées par G_6 (qui n'appartiennent pas à $L(G_6)$) sont des syntagmes cardinaux corrects. Symboliquement: une suite S est un syntagme cardinal français si et seulement si $S \in L(G_5) \cap \overline{L(G_6)}$.

Voici comment pourrait être conçue G_6 :

$$1. \text{ ANTICARD} \rightarrow \left\{ \begin{array}{l} \left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{un} \\ \text{T} \left\{ \begin{array}{l} \text{cent} \\ \text{-vingt} \end{array} \right\} \end{array} \right\} \left\{ \begin{array}{l} \text{millions} \\ \text{milliards} \end{array} \right\} \\ \text{T} \left\{ \begin{array}{l} \text{million} \\ \text{milliard} \end{array} \right\} \\ \text{T} \left\{ \begin{array}{l} \text{cents} \\ \text{-vingts} \end{array} \right\} \text{T}^{\text{II}} \\ \text{T} \left\{ \begin{array}{l} \text{cent} \\ \text{-vingt} \end{array} \right\} \end{array} \right\} (\text{T}) \end{array} \right\}$$

2. $T \rightarrow \left\{ \begin{matrix} T^1 \\ T^{11} \end{matrix} \right\} (T)$
3. $T^{11} \rightarrow \left\{ \begin{matrix} \text{TERM} \\ \text{un} \end{matrix} \right\} (T^{11})$
4. $T^1 \rightarrow \left\{ \begin{matrix} \text{TERM} \\ \text{million(s)} \\ \text{milliard(s)} \end{matrix} \right\} (T^1)$
5. $\text{TERM} \rightarrow \left\{ \begin{matrix} \text{deux} \\ \text{trois} \\ \vdots \\ \text{vingt} \\ \text{trente} \\ \vdots \\ \text{cent(s)} \\ \text{mille} \\ - \\ \text{et} \end{matrix} \right\}$

Tous les éléments terminaux sauf *un*, *million(s)*, *milliard(s)*.

Ainsi **deux cents trois*, qui appartient à $L(G_5)$, n'est pas un syntagme correct puisqu'il est également engendré par G_6 . Cf. la dérivation:

ANTICARD

- (1) $T \text{ cents } T^{11}$
- (2) $T^1 \text{ cents } T^{11}$
- (3) $T^1 \text{ cents TERM}$
- (4) TERM cents TERM
- (5) deux cents TERM
- (5) deux cents trois

alors que *deux cent trois*, qui appartient également à $L(G_5)$, est correct puisqu'il ne peut pas être engendré par G_6 . En effet, *cent*, dans la règle 1 de G_6 , n'apparaît qu'immédiatement avant *millions* ou *milliards*, ou bien comme élément final d'une suite.

*

Je tiens à remercier cordialement mes amis et collègues, M. Peter Møller-Nielsen (Institut de Mathématiques Supérieures de l'Université de Copenhague), qui m'a fourni de précieuses suggestions relatives au sujet de cet article, ainsi que M. François Marchetti (Institut d'Études

Romanes de l'Université de Copenhague), qui a bien voulu me prêter son sentiment linguistique.

Carl Vikner

COPENHAGUE

BIBLIOGRAPHIE

Ouvrages de référence:

- Bach, Emmon. *An Introduction to Transformational Grammars*. New York, Holt, Rinehart and Winston, 1964.
- Brainerd, Barron. «A Transformational-Generative Grammar for Rumanian Numerical Expressions». *Cahiers de Linguistique Théorique et Appliquée* 4 (1967), pp. 35-45. Réimprimé dans Corstius, 1968b, pp. 41-52. Je renvoie à cette dernière édition.
- Chevalier, J.-C., C. Blanche-Benveniste, M. Arrivé et J. Peytard. *Grammaire Larousse du français contemporain*. Paris, Larousse, 1964.
- Corstius, H. Brandt. 1968a. «Automatic Translation between Number Names». In Corstius, 1968b, pp. 103-123.
- Corstius, H. Brandt, éd. 1968b. *Grammars for Number Names. Foundations of Language*, Supplementary series, volume 7. Dordrecht-Holland, D. Reidel, 1968.
- Gougenheim, Georges. *Système grammatical de la langue française*. Paris, d'Artrey, 1939. Cité d'après la réimpression, 1962.
- Grevisse, Maurice. *Le Bon Usage*. Gembloux, Duculot, 1964. 8^e édition revue.
- Langacker, R. W. Compte rendu de A. Koutsoudas: *Writing Transformational Grammars*. In: *Language* 44 (1968), pp. 98-108.
- Ruwet, Nicolas. *Introduction à la grammaire générative*. Paris, Plon, 1967.
- Togeby, Knud. 1951. *Structure immanente de la langue française*. Copenhague, 1951. Cité d'après la 2^e édition: Paris, Larousse, 1965.
- Togeby, Knud. 1965. *Fransk grammatik*. Copenhague, Gyldendal, 1965.

Textes cités:

- Calvet, Henri: *La Société française contemporaine*. Paris, Nathan, 1956.
- L'Express*.
- Jeune Afrique*.
- Julien, Claude: *L'Empire américain*. Paris, Grasset, 1968.
- Le Monde*.
- Le Monde - Sélection hebdomadaire*.
- Le Nouvel Observateur*.
- Zegel, Sylvain: *Les idées de mai*. Paris, Gallimard, Collection «Idées», 1968.

Romanes de l'Université de Copenhague), qui a bien voulu me prêter son sentiment linguistique.

Carl Vikner

COPENHAGUE

BIBLIOGRAPHIE

Ouvrages de référence:

- Bach, Emmon. *An Introduction to Transformational Grammars*. New York, Holt, Rinehart and Winston, 1964.
- Brainerd, Barron. «A Transformational-Generative Grammar for Rumanian Numerical Expressions». *Cahiers de Linguistique Théorique et Appliquée* 4 (1967), pp. 35-45. Réimprimé dans Corstius, 1968b, pp. 41-52. Je renvoie à cette dernière édition.
- Chevalier, J.-C., C. Blanche-Benveniste, M. Arrivé et J. Peytard. *Grammaire Larousse du français contemporain*. Paris, Larousse, 1964.
- Corstius, H. Brandt. 1968a. «Automatic Translation between Number Names». In Corstius, 1968b, pp. 103-123.
- Corstius, H. Brandt, éd. 1968b. *Grammars for Number Names. Foundations of Language*, Supplementary series, volume 7. Dordrecht-Holland, D. Reidel, 1968.
- Gougenheim, Georges. *Système grammatical de la langue française*. Paris, d'Artrey, 1939. Cité d'après la réimpression, 1962.
- Grevisse, Maurice. *Le Bon Usage*. Gembloux, Duculot, 1964. 8^e édition revue.
- Langacker, R. W. Compte rendu de A. Koutsoudas: *Writing Transformational Grammars*. In: *Language* 44 (1968), pp. 98-108.
- Ruwet, Nicolas. *Introduction à la grammaire générative*. Paris, Plon, 1967.
- Togeby, Knud. 1951. *Structure immanente de la langue française*. Copenhague, 1951. Cité d'après la 2^e édition: Paris, Larousse, 1965.
- Togeby, Knud. 1965. *Fransk grammatik*. Copenhague, Gyldendal, 1965.

Textes cités:

- Calvet, Henri: *La Société française contemporaine*. Paris, Nathan, 1956.
- L'Express*.
- Jeune Afrique*.
- Julien, Claude: *L'Empire américain*. Paris, Grasset, 1968.
- Le Monde*.
- Le Monde - Sélection hebdomadaire*.
- Le Nouvel Observateur*.
- Zegel, Sylvain: *Les idées de mai*. Paris, Gallimard, Collection «Idées», 1968.