

Morphologie et syntaxe : Lexicalisme

1. Chomsky 1970
2. Jackendoff 1975
3. Halle 1973 et Aronoff 1976

1. Chomsky 1970 : *Remarks on Nominalizations*

Origine du lexicalisme

Transformations nominalisantes

Aux débuts de la grammaire générative, la syntaxe était responsable de combiner les éléments issus du lexique (mots simples et morphèmes) pour former des structures complexes comme des mots et des phrases ('c'est ce qu'on a vu chez Del').

Par exemple :

syntaxe : *john+refuse+PASSÉ+INF+come*

phono. (épellation) : *john+refuse+ed+to+come*

synaxe : *john+GÉN+refuse+NOM+INF+come*

phono. (épellation) : *john+s+refuse+ed+to+come*

L'insertion des morphèmes se faisait en structure profonde, avant l'application des transformations qui étaient responsables, entre autres, de la formation des phrases passives et interrogatives à partir des actives et déclaratives correspondantes :

Structure Profonde

[You]₁ [bought]₂ [something]₃ →

[What]₁ did [you]₁ [buy]₂?

Transformation: 1 2 3 →

3 *did* 1 2

[John]₁ [eats]₂ [the apple]₃ →

[The apple]₃ is [eat]₂-en by [John]₁

Transformation: 1 2 3 →

3 is 2-en by 1

En constatant les ressemblances entre certaines phrases et les syntagmes nominaux correspondants, on a fait l'hypothèse que certains phénomènes morphologiques, en particulier les nominalisations, résultaient de transformations semblables à celles qui forment les phrases passives.

Structure Profonde

Structure de Surface

[The romans]₁ [destroy]₂ [the city]₃ → The [destroy]₂-tion of [the city]₃ by [the romans]₁

Transformation: 1 2 3 → *the 2-tion of 3 by 1*

Un des arguments pour le traitement transformationnel des nominalisations est le fait que les nominalisations comme *refusal* et *destruction* ont sensiblement la même structure argumentale et peuvent sous-catégoriser le même type de complément que les verbes correspondants (*refuse* et *destroy*).

Sujet	V/N	Proposition infinitive
[John]	[refused]	[to come]
[John's]	[refusal]	[to come]
Sujet	V/N	COD
[The romans']	[destruction]	[of the city]
[The romans]	[destroy]	[the city]

Les transformations nominalisantes pouvaient donc insérer des morphèmes comme le génitif ('s), des prépositions (*by*) et des suffixes dérivationnels (*-tion* et *-al*). Ces éléments n'appartenaient que dans la structure de surface.

Remarks on nominalizations (Chomsky 1970)

Dans un article important (qui est la source de ce qu'on appelle le lexicalisme), Chomsky a soulevé plusieurs problèmes avec le traitement transformationnel des nominalisations et a suggéré qu'il est plus adéquat de considérer que les nominalisations sont effectuées dans le lexique et non par la syntaxe.

- Problème de sous-catégorisation:

Les nominalisations doivent hériter de la structure argumentale et du cadre de sous-catégorisation du verbe dont elles sont dérivées : si *refuse* doit avoir un sujet et peut prendre une proposition infinitive comme complément, alors *refusal* aussi.

Cependant, Chomsky remarque que ce n'est pas toujours le cas :

John is eager to please → *John's being eager to please* → *John's eagerness to please*

John is easy to please → *John's being easy to please* → **John's easyness to please*

Ici, *eager* et *easy* peuvent tous deux être suivis d'une proposition infinitive (*to please*) et conservent cette propriété quand on met la phrase au gérondif (*being* → *to please*), mais *easyness*, une nominalisation qui est issue de *easy*, ne peut plus être suivi de *to please*.

Il semble donc que les nominalisations ont un cadre de sous-catégorisation qui est parfois différent de celui du mot (adjectif ou nom) dont on les dérive grâce à une transformation.

Ce comportement différent ne peut pas non plus être attribué aux propriétés de *-ness* puisque ce suffixe n'a pas le même effet sur *eager*.

Comme les transformations ne sont pas supposées modifier le cadre de sous-catégorisation des mots qu'elles manipulent, il faut que les nominalisations (*refusal*, *destruction*, *easyness*, *eagerness*) soient présentes en structure profonde et non dérivées par transformation.

Chomsky dit essentiellement que *easy* et *easyness* ne sont pas reliés par une transformation mais sont reliés d'une manière ou d'une autre dans le **lexique**.

Il existe aussi d'autres types d'arguments contre un traitement transformationnel des nominalisations. Ceux-ci reposent aussi sur le fait que les transformations ne doivent pas affecter le sens et doivent être parfaitement régulières.

- Problème sémantique

Le sens de la nominalisation est souvent plus restreint (ou imprévisible) que le sens du verbe :

try 'essayer' ~ *trial* 'procès'

reside 'être situé' ~ *residence* 'logis'

- Manque de régularité

Plusieurs verbes n'ont pas de nominalisation : *eat* ~ ?

Certains verbes ont plusieurs nominalisations :

commit ~ *commitment*, *commial*, *commission*

Certains nominalisations apparentes n'ont pas de base :

**agress* ~ *agression*, **perdi* ~ *perdition*

Signification de Remarks on nominalizations

L'intervention de Chomsky sert principalement à clarifier un problème pour la théorie syntaxique. Si on laisse la syntaxe (les transformations) s'occuper des nominalisations, ses règles n'auront plus la même généralité puisque qu'elles devront modifier le sens des mots d'une manière arbitraire (changer les cadres de sous-catégorisation et, par exemple, stipuler que *residence* ne veut pas dire 'action de résider' mais 'logis').

Par contre, si on exclue les nominalisations du domaine de la syntaxe, celle-ci s'en trouve simplifiée et conserve sa portée générale.

Il faut donc comprendre que les remarques de Chomsky ne concernent pas en premier lieu la théorie morphologique. Ceci est d'autant plus vrai qu'à cette époque il n'y a pas de module dédié à la morphologie dans le modèle générativiste.

Chomsky énonce ce que la syntaxe ne peut pas faire. Logiquement, le seul endroit où on peut repousser ce que la syntaxe ne peut pas faire et qui concerne néanmoins la sémantique ne peut être que le lexique.

Extension à l'ensemble de la morphologie

L'hypothèse qui découle de cette constatation n'est pas formulée explicitement par Chomsky, mais elle en est la conséquence logique.

- L'hypothèse lexicaliste

La syntaxe ne manipule pas la structure interne des mots.

Donc, elle ne modifie pas les entrées lexicales et ne construit pas les mots.

Cette hypothèse va plus loin que les remarques de Chomsky puisqu'il n'avait examiné que certains cas de morphologie dérivationnelle.

Cependant, le genre d'irrégularités qu'il avait constatées se retrouve aussi dans la plupart des phénomènes dérivationnels et dans certains phénomènes flexionnels. On a donc très vite supposé que la morphologie ne pouvait pas être faite par la syntaxe, mais devait se trouver dans le lexique.

- Extension du lexicalisme

nominalisation > dérivation > flexion > ensemble de la morphologie

L'hypothèse lexicaliste dit que la syntaxe ne peut pas faire de morphologie, suggère qu'elle devrait se faire « avant », donc dans le lexique, mais elle ne dit absolument pas à quoi ressemble la morphologie dans la grammaire générative.

La morphologie générative lexicaliste reste donc à inventer. C'est ce que feront, entre autres, Morris Halle dans son *Prolegomena to a theory of word formation* (1973) et Mark Aronoff dans *Word formation in Generative Grammar* (1976).

2. Jackendoff 1975 : *Morphological and Semantic Regularities in the Lexicon*

Lexique comme une liste/prison/sacoche : Économie

On suppose donc que l'on doit minimiser le nombre de choses à mémoriser en maximisant l'utilisation de règles (généralisations).

Donc, s'il existe 300 mots (un chiffre arbitraire) en français qui se terminent *-eur* et dont une partie du sens correspond à 'personne qui fait...', on doit éviter de mémoriser 300 fois le même *-eur* et le même sens qui lui est rattaché. De la même façon, s'il y a un élément de forme et de sens qui reste constant à travers tout le paradigme verbal, pourquoi le mémoriser (on dit souvent *stocker*) plusieurs fois?

mang-eur, march-eur, lanc-eur, etc.

mange, mange-ons, mange-ait, etc.

Intuitivement, il serait plus simple de stocker une seule fois *-eur* et *mange-* et d'avoir une règle générale qui permette de les combiner pour obtenir les autres mots.

Stockage Généralisation

-eur se combine à un V pour faire un N qui veut dire 'personne qui fait V'

mange- sert de base aux flexions verbales (*-ons*, *-ait*, etc.)

Dans cette approche, le lexique ne doit contenir que l'information qui n'est pas prévisible, c.-à-d. l'information qui ne peut pas être fournie par une règle. Ainsi, le fait que ce soit la forme *-eur* et pas *-ouf* qui soit associée au sens 'personne qui fait...' est imprévisible, mais le fait que ce *-eur* se combine à des verbes est prévisible au sens où on peut le formuler comme une généralisation qui se confirme par l'examen de la liste des mots en *-eur* en français.

Le lexique est donc une liste des éléments imprévisibles de la langue. Tous les mots morphologiquement complexes (par ex. *mang-eur* et *mange-ons*) en sont donc exclus et, quand on en a besoin, on les fabrique à l'aide des morceaux dont ils sont composés et des règles pertinentes.

Cette vision du lexique exclue donc toute forme d'organisation ou de traitement (« computation ») à l'intérieur du lexique. Entre autres, pas d'associations paradigmatiques de forme, de sens ou de forme et de sens. Les généralisations font partie de la grammaire et non du lexique, c'est pourquoi on dira parfois que le lexique est comme une prison : il ne contient que des hors-la-loi.

Le lexique comme base de données structurée

Une conséquence inattendue de *Remarks on nominalisations*, c'est que l'article conduit un autre chercheur, Ray Jackendoff, à reconsidérer la notion traditionnelle d'économie dans le lexique.

Si le lexique peut contenir certains mots morphologiquement complexes, pourquoi ne les contiendrait-il pas tous?

La réponse classique : Cela ne serait pas économique puisque qu'il faudrait stocker les mêmes informations un grand nombre de fois. En effet, la séquence *-tion*, qui fait des noms à partir des verbes, serait présente dans toutes les entrées lexicales des nominalisations comme *description* ou *intervention*.

C'est le problème de la redondance de l'information lexicale.

Normalement, on considère qu'il faut minimiser la redondance pour ne pas surcharger la mémoire d'un paquet d'informations qui sont prévisibles par des règles générales.

Jackendoff propose au contraire de considérer que la redondance *aide* à mémoriser de nouveaux mots et fait l'hypothèse que tous les mots que nous connaissons sont stockés.

Si on reprend l'exemple des mots français en *-eur*, on aurait une entrée lexicale pour chacun de ces mots ainsi que pour les verbes dont ils sont dérivés:

[chante; V; 'chante']	[chanteur; Nmasc.; 'personne qui chante']
[lance; V; 'lance']	[lanceur; Nmasc.; 'personne qui lance']
[mange; V; 'mange']	[mangeur; Nmasc.; 'personne qui mange']
etc.	etc.

Il y aurait évidemment un grand nombre de paires d'items lexicaux comme celles-ci, chacune contenant toute l'information phonologique, syntaxique et sémantique nécessaire.

Pour rendre compte de ce qui est prévisible dans ces paires d'items lexicaux, Jackendoff suggère que nous utilisons des **règles de redondance** (*redundancy rules*).

[X; V; 'X'] ↔ [Xeur; Nmasc.; 'personne qui fait X']

Selon lui, ce qui colle le plus cher à mémoriser c'est tout ce qui n'est pas mentionné par une règle de redondance.

Ainsi, si je rencontre un nouveau mot dérivé de *blague(r)*_V, disons *blagueur*, je crée une entrée lexicale pour lui. Celle-ci correspond à l'entrée lexicale pour *blague(r)*_V.

Vieux mot

Nouveau mot

[blague(r); V; 'faire des blagues'] [blagueur; Nmasc.; 'personne qui fait des blagues']

Pour calculer ce que ça colle d'ajouter ce nouveau mot, il suffit de compter la quantité d'information qui n'est pas présente dans la règle de redondance ou dans le vieux mot auquel correspond le nouveau.

L'information sur le radical de *blagueur* se trouve déjà dans l'entrée lexicale de *blague(r)*_V.

[blagu ... 'fait des blague'] [blague(r); V; 'faire des blagues']

L'information sur le suffixe se trouve dans la partie de droite de la règle de redondance.

[-eur; Nmasc.; 'personne qui fait'] [Xeur; Nmasc.; 'personne qui fait X']

Donc ce mot ne me colle absolument rien à mémoriser. La seule chose qui est contenue pour la mémoire, c'est de mémoriser de l'information qui ne figure nulle part, c'-à-d. qui n'est pas mentionnée dans une règle de redondance où dans une autre entrée lexicale.

Ainsi, si je rencontre le verbe *mougrir*, je devrai payer le prix de l'information phonologique qui ne figure nulle part ailleurs ([muʒiʁ]), pour l'information sémantique ('faire le son « mou » très fort), et pour la catégorisation du mot comme un V.

Cette nouvelle façon de voir l'économie dans le lexique est une mesure de l'*independent information content* : seule l'information qui n'est pas prévisible par une règle de redondance constitue un coût pour la mémorisation, le reste est gratuit.

Donc, même en mettant tous les mots dans le lexique, on ne surcharge pas la mémoire. En fait, la somme de l'information indépendante demeure la même que dans un lexique où on ne met que l'information qui n'est pas prévisible par une règle (comme chez Bloomfield et dans la conception standard du lexique).

3. Halle 1973 et Aronoff 1976

Jackendoff et Chomsky innovent en proposant que les mots complexes sont stockés dans le lexique.

La morphologie lexicaliste de Halle innove en proposant que le lexique contient plusieurs composantes qui travaillent ensemble pour créer de nouveaux mots. Il veut rendre compte du fait qu'on peut inventer de nouveaux mots et comprendre des mots que nous n'avons jamais entendus avant tout en incorporant l'hypothèse lexicaliste.

En effet, Halle dit qu'une grande partie des mots que nous utilisons sont mémorisés tels quels dans le dictionnaire et que nous n'utilisons la morphologie que dans les cas où nous entendons ou prononçons des nouveaux mots.

C'est pourquoi il dit que les règles de formation de mot, les règles morphologiques, ne sont pas utilisés souvent :

... It is possible to suppose that a large part of the dictionary is stored in the speaker's permanent memory and that he needs to invoke the word formation component only when he hears an unfamiliar word or uses a word freely invented. While this is by no means an exceptional occurrence, its frequency is quite low.

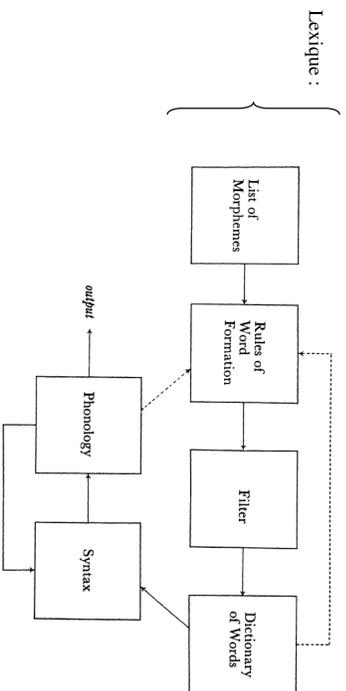


Figure 1

Son lexique a quatre parties :

La liste des morphèmes : tous les morphèmes de la langue (mots simples, affixes et radicaux) avec leur représentation phonologique, leur catégorie syntaxique (les affixes sont de catégorie Aff, pas N, V ou A) et leurs propriétés grammaticales.

Les Règles de Formation de Mots : spécifient la catégorie de la base et la catégorie du dérivé, l'ordre des morphèmes, et toutes les propriétés syntaxiques et sémantiques qui ne sont pas propres à la base.

[Noun]_{pl, human} hood]N *boyhood, priesthood, *treehood*
 [Noun-ish]A *boyish, *ishboy, *tallish*
 [en-Adj]V *enlarge, enrich, *richen, *enboy*

La liste des morphèmes et des RFM définissent l'ensemble des mots possibles.

Le Filtre : bloque les mots non attestés et modifie le sens des mots idiosyncrasiques. Les mots bloqués sont étiquetés [-Lexical Insertion].

Le Dictionnaire : contient tous les mots qui passent le filtre avec succès.

Le filtre réduit l'ensemble des mots possibles à l'ensemble des mots attestés (dictionnaire). La syntaxe va chercher les mots dans le dictionnaire.

Remarque que Halle inclut la morphologie flexionnelle dans le lexique, une extension de l'hypothèse lexicaliste de Chomsky. Il soutient en effet que les mots flexionnels présentent aussi des irrégularités qui doivent être prises en charge par le filtre. Ces mots flexionnels sont tous présents dans le dictionnaire.

Rôle du filtre

- Permet d'identifier les mots possibles mais non attestés.

Pour des raisons sémantiques/pragmatiques :

décéder, dériver, désosser *céréder, *nèver, *ossor
tuer, mourir *reuer, *remourir

Bases non attestées :

*repter reptation
 *deflecter déflection
 *insul insulaire
 *électre électrique

Blocage des mots qui sont synonymes de mots existants :

banalité, générosité
 *neurosité, peurrosité >> bonheur, peur
 *arrivation, *confusid >> arrivad, confusion

Lacunes accidentelles :

fillette, pommette, jupette, brochette
 *femmette, *pimmette, *robette, *igette

Les mots non attestés doivent passer par le composant conventionnel avant d'être insérés dans la syntaxe. S'ils sont marqués comme non attestés (trait [-Lexical Insertion]), ils sont stoppés avant le dictionnaire et donc ne peuvent pas être insérés dans une structure syntaxique.

- Ajoute certaines informations idiosyncrasiques aux mots complexes

transmission, chantage, réception

Ne veulent pas nécessairement dire :

'action de transmettre', 'action de chanter', 'action de recevoir'

Le composant conventionnel les « ajuste » pour leur donner le sens :

'pièce de voiture', 'extorsion', 'événement social'

Problème :

Si on se sert de la morphologie seulement pour créer ou comprendre de nouveaux mots, comment peut-on savoir que certains de ces mots ne doivent pas être insérés (sont possibles mais non existants) ou que certains ont un sens particulier qui ne peut être déduit de la RFM qui les produit?

Une fois que l'on est corrigé et que l'on a pris note d'un mot à un sens imprévu ou n'existe pas en français, la morphologie (Liste de morphèmes + RFM + Filtre) ne sert plus à rien. En effet, il serait totalement inutile d'avoir mémorisé les morphèmes *chant* et *-age* séparément, de produire *chantage* grâce à une règle puis de vérifier dans le filtre que ce mot a un sens spécial.

Logiquement, le filtre ne peut fonctionner que si l'on connaît le mot d'avance ou que l'on sait que le mot n'existe pas. Donc, la morphologie va continuer à surgénérer jusqu'à temps que l'on soit corrigés. Peut-on dire que l'on sait que *défecter* n'existe pas?

Le paradoxe est que le filtre contient des informations sur les mots que nous ne connaissons pas.

Éliminer le composant conventionnel (Aronoff 1976).

Il n'y a pas de raison de filtrer les mots possibles si on ne connaît pas déjà leurs caractéristiques idiosyncratiques

Si on les connaît, on n'a pas besoin d'utiliser des RFM pour les produire puisque tous les mots que nous connaissons se trouvent dans le dictionnaire.

Donc, les mots qui sont non attestés pour des raisons pragmatiques demeurent peu probables mais n'ont pas à figurer dans une liste spéciale pour spécifier que l'on ne peut pas les produire (à l'aide du trait [-LI]). D'ailleurs, si le contexte se présente, on doit pouvoir les produire.

Mots possibles

retuer, remourir

Comme la morphologie de Aronoff est basée sur le mot (l'input d'une RFM doit être un mot, donc quelque chose qui se trouve dans le dictionnaire), les mots qui sont apparemment complexes mais qui ne correspondent pas à des mots existants ne sont pas, par définition, complexes.

**repler* *repiation*

**défecter* *défection*

**insul* *insulature*

**électre* *électrique*

Cependant, il n'est pas exclu que l'on puisse utiliser une règle à l'envers (cf. *edit ~ editio*), ce que l'on appelle parfois formation rétrograde (*back-formation*). Ce qui pourrait produire des mots comme les précédents et comme :

cérbrer, nerver, assser

En général, les morphologues croient que l'utilisation à l'envers d'une RFM est un phénomène sporadique, presque accidentel.

Les lacunes accidentelles ne doivent pas non plus être filtrées.

Mots possibles

femmeite, pruneite, robeite, tigeite

Soit on sait que l'usage préfère « *petite N* », on s'abstient de les produire, sinon on les produira et probablement que l'on sera parfaitement compris.

Si on suppose que l'absence d'un mot possible (grammatical) est une partie de notre connaissance de la langue, comment peut-on en même temps expliquer le fait que ces

mots seront compris, c.-à-d. qu'on ne peut pas à la fois expliquer notre capacité à faire des néologismes et notre connaissance de leur inexistence préalable.

Le résultat des RFM est par définition prévisible. C'est le temps qui fait que les mots changent de sens, qu'ils se spécialisent.

cf. *transmission, chantage, réception*

Si on ne sait pas que ces mots ont des sens particuliers, on leur donnera le sens que la RFM leur donne (action de V') au moment où on les crée. Si on le sait, alors on n'a pas besoin de RFM pour les produire.

Le blocage n'est pas un principe absolu, il dépend de notre connaissance lexicale. Si on connaît le mot qui en bloque un autre (par ex. *confusion* qui bloque *confusid*), alors le blocage peut jouer. Si on ne le connaît pas, il ne peut pas y avoir de blocage. On peut aussi décider de créer un synonyme si on en a besoin.

Mots possibles

heurostie, peurostie

arrivation, confusid

D'ailleurs il existe beaucoup de doubles, entre autres dans les nominalisations :

abattement, abattage, claquement, claquage

En fait, le blocage est seulement un principe de moindre effort : il est plus facile d'aller chercher un mot tout fait que de le fabriquer à l'aide d'une règle. Mais, si le mot que l'on cherche n'a pas le sens attendu, on risque de le recréer.

antiquité 'période historique'

antichité 'caractère de ce qui est antique'

Qu'est-ce que la morphologie?

... *the task of morphology should be restricted to describing directly only the possible but non occurring words of language, a set which is presumably regular, so that then the regular properties of the actual words in the language(those which are listed in the lexicon) might be described derivatively.*

Comparée à la vision traditionnelle de la morphologie (étude de la structure interne des mots d'une langue), la proposition de Aronoff ressemble à une petite révolution.

Il suggère en fait que la partie de la grammaire que l'on nomme morphologie (l'ensemble des Règles de Formation de Mots) ne sert qu'à déterminer l'ensemble des mots possibles, tout comme la syntaxe, dans l'optique générative, sert à déterminer l'ensemble infini des phrases possibles (grammaticales).

Comment ça fonctionne :

L'enfant qui apprend sa langue reçoit comme input une fraction du lexique social (extème).

Ces mots forment son lexique mental.

Dans son lexique mental, il découvre des alternances morphologiques (forme et sens) régulières et des alternances moins régulières.

Il formule des généralisations (règles) qui rendent compte des alternances régulières et qui lui permettent de créer de nouveaux mots sur le même modèle.

Ces règles déterminent un ensemble de mots possibles qui sont, par définition, compositionnels.

Il produit des mots qu'il n'a jamais entendus et il constate que certains sont approuvés par son entourage, alors que d'autres ne le sont pas (on ne dit pas *ils jouzent*, *ils ruzent* mais *ils jouent*, *ils tuent*⁽¹⁾). Parmi les néologismes qu'il produit, certains sont déjà des mots du français et d'autres ont des chances, si d'autres les produisent aussi, d'être acceptés dans la norme.

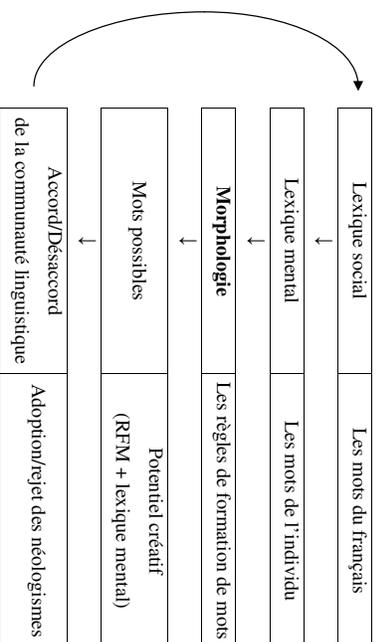
Une fois inclus dans le lexique social, ces mots peuvent acquérir avec le temps des particularités sémantiques, mais ils peuvent aussi demeurer compositionnels.

La morphologie ne sert donc pas à décrire l'ensemble des mots du français, mais à décrire le potentiel de **créativité lexicale** des individus qui parlent le français.

C'est l'aboutissement logique de la perspective mentaliste et générativiste de la morphologie : décrire la connaissance internalisée du système de la langue et non décrire la norme en vigueur dans une communauté linguistique donnée.

Rôle de la morphologie

Décrire les mots possibles, mais pas nécessairement les mots existants.



Références :

- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, Mass.: MIT Press.
- Chomsky, Noam. 1970. Remarks on nominalizations. In *Readings in English Transformational Grammar*, eds. Roderick A. Jacobs and Peter S. Rosenbaum, 184-221. Waltham, Mass.: Ginn. [version française : Chomsky, Noam. 1975. *Questions de sémantique*. Paris: Editions du Seuil, p. 73-131.]
- Halle, Morris. 1973. Prolegomena to a theory of word-formation. *Linguistic Inquiry* 4: 3-16.
- Jackendoff, Ray. 1975. Morphological and Semantic Regularities in the Lexicon. *Language* 51:639-671.