

Grammaires formelles - Historique

À la croisée des disciplines

Karën Fort

karen.fort@sorbonne-universite.fr / <https://www.schplaf.org/kf/>

25 septembre 2020



Quelques sources d'inspiration (avec leur accord)

par ordre d'importance décroissant

- ▶ B. Habert (ENS Lyon) : enjeux, langues et langages
- ▶ le Trésor de la Langue Française informatisé (TLFi) *via* le CNRTL
- ▶ S. Wauquiez (Paris 8) pour ses explications et ses liens concernant Chomsky
- ▶ F. Yvon (LIMSI), dont j'ai repris les diapos sur l'apprentissage

Introduction

Biais

Définitions

Au croisement de la linguistique et de l'informatique

Un « personnage », adulé et haï

Théorie(s)

Pour finir

Attention !

Gardez à l'esprit qui je suis et d'où je viens, car mon enseignement en est imprégné :


- ▶ je travaille sur les corpus
- ▶ je ne fais pas de syntaxe
- ▶ je ne suis pas linguiste ...
- ▶ ...encore moins générativiste


⇒ mon point de vue est forcément **partiel** et **partial**



Comment expliquer comment se construit la langue ?

Grammaires formelles ?

<http://www.cnrtl.fr/definition/grammaire>


**CNRTL**
Centre National de Ressources Textuelles et Lexicales


**Ortolang**
Outils et Ressources pour un Traitement Optimisé de la LANGue





[Accueil](#) [Portail lexical](#) [Corpus](#) [Lexiques](#) [Dictionnaires](#) [Métalexigraphie](#) [Outils](#) [Contact](#)


[Morphologie](#) **[Lexicographie](#)** [Etymologie](#) [Synonymie](#) [Antonymie](#) [Proxémie](#) [Concordance](#) [Aide](#)

**TLFi**

**Académie**
9^{ème} édition





**Académie**
8^{ème} édition

**Académie**
4^{ème} édition

**BDLP**
Francophonie

Entrez
■ une
forme

[options d'affichage](#)

Chercher    

catégorie :

■ **GRAMMAIRE**, subst. fém.

A. – [La notion de grammaire évoque l'exercice d'une langue et est associée à celle de normes caractérisant diverses manières de parler et d'écrire] „Art de parler et d'écrire correctement" (Ac. 1932); ensemble de règles conventionnelles (variables suivant les époques) qui déterminent un emploi correct (ou bon usage) de la langue parlée et de la langue écrite. Grammaire normative, traditionnelle; discuter un point de grammaire; faute, leçon, règle de grammaire; apprendre, faire de la grammaire. Il n'y avait dans ces lettres ni commencement, ni fin, ni milieu, ni grammaire, ni rien de ce qu'on entend ordinairement par style (LAMART., Raphaël, 1849, p. 239). Eh bien, mon cher, le seul tort que j'ai eu, c'a été de donner à ma femme un professeur de français. Tant qu'elle a martyrisé le dictionnaire et supplicié la grammaire, je l'ai chérie (MAUPASS., Contes et nouv., t. 1, Découverte, 1884, p. 960) :

Grammaires formelles ?

<http://www.cnrtl.fr/definition/grammaire>

B. – [La notion de grammaire évoque une langue considérée en tant qu'objet d'une étude sc.] .

1. **LING. DESCRIPTIVE CLASS.** Étude objective et systématique des éléments (phonèmes, morphèmes, mots) et des procédés (de formation, de construction, d'expression) qui constituent et caractérisent le système d'une langue naturelle; *en partic.*, étude de la morphologie et de la syntaxe d'une langue (à l'exclusion de la phonologie, de la lexicologie et de la stylistique). *En théorie, il faudrait faire la part égale à la phonologie, au vocabulaire et à la grammaire* (BALLY, *Lang. et vie*, 1952, p. 62). *Indépendamment du dictionnaire et de la grammaire, on peut faire des conventions plus ou moins libérales sur la structure de la phrase par laquelle on définit un nombre au moyen de dix mots* (E. BOREL, *Paradoxes infini*, 1946, p. 155) :

- 2. D'autre part, est-il logique d'exclure la lexicologie de la **grammaire**? À première vue les mots, tels qu'ils sont enregistrés dans le dictionnaire, ne semblent pas donner prise à l'étude grammaticale, qu'on limite généralement aux rapports existants entre les unités. SAUSSURE, *Ling. gén.*, 1916, p. 186.

. . .


2. **LING. MOD.**



a) **Grammaire générative, transformationnelle.** Mécanisme, constitué par un ensemble de règles, qui permet de produire et de décrire toutes et rien que les phrases grammaticales d'une langue. *De même qu'on peut écrire plusieurs grammaires traditionnelles pour une même langue, de même il est facile de voir que les grammaires transformationnelles sont elles aussi probabilistes* (COYAUD, *Introd. ét. lang. docum.*, 1966, p. 76). Le mérite essentiel de la grammaire générative a été, au contraire, de montrer les insuffisances de la linguistique distributionnelle et d'offrir une méthode où il fût beaucoup plus facile de détecter les erreurs (J. STÉFANINI, *Sur la grammaire historique du français* ds *Lang. fr.*, n°10, 1971, p. 24).

Grammaires formelles ?

<http://www.cnrtl.fr/definition/formel>


CNRTL
Centre National de Ressources Textuelles et Lexicales


**Ortolang**
Outils et Ressources pour un Traitement Optimisé de la LANGUE





■ Accueil ■ Portail lexical ■ Corpus ■ Lexiques ■ Dictionnaires ■ Métalexicographie ■ Outils ■ Contact


Morphologie **Lexicographie** Etymologie Synonymie Antonymie Proxémie Concordance Aide


**TLFi**


**Académie**
9^{ème} édition

**Académie**
8^{ème} édition

**Académie**
4^{ème} édition




**BDLP**
Francophonie

**BHVF**
attestations

**DMF**
(1330 - 1500)

Entrez une forme

Formel

Chercher   

options d'affichage

catégorie : toutes

■ **FORMEL, ELLE, adj.**

A. – Qui existe de façon déterminée et, p. ext., qui est énoncé de façon déterminée, claire, sans équivoque. *Avis, ordre, règlement formel; apporter, donner, opposer un démenti formel à qqc.; essayer, prononcer un refus formel. La duchesse court à la forteresse; le général Fabio Conti fut enchanté d'avoir à lui opposer le texte formel des lois militaires : personne ne peut pénétrer dans une prison d'état sans un ordre signé du prince (STENDHAL, Chartreuse, 1839, p. 418). M. Stangerson déclare que, désormais, il ne quittera plus l'appartement de sa fille. C'est en vain que celle-ci veut s'opposer à cette volonté formelle (LEROUX, Myst. ch. jaune, 1907, p. 91). V. abbaye ex. 9 :*

• 1. Cependant quand Marat lui eut expliqué toute l'affaire, il se refusa d'admettre comme certitude que Mathilde eût trahi et qu'un guet-apens fût dressé contre lui le lendemain matin. Il manquait, estimait-il, une preuve **formelle**, « une preuve par neuf ». VAILLAND, Drôle de jeu, 1945, p. 234.

SYNT. Aveu, diagnostic, engagement, témoignage, ultimatum formel; consigne, contre-indication, décision, déclaration, défense, interdiction, prescription, opposition, résolution formelle.

♦ Être formel employé absol. ou suivi de sur qqc., à cet égard. *La loi est formelle; il a été formel sur ce point. Devant le tribunal, ils [les agents] furent formels et très affirmatifs sur le coup de revolver (VERLAINE, Œuvres posth., t. 3, Prose, 1896, p. 205). Mon cousin Gustin Sabayot, à cet égard il est formel : je devrais bien changer mon genre (CÉLINE, Mort à crédit, 1936, p. 13):*

Grammaires formelles ?

<http://www.cnrtl.fr/definition/grammaire>

2. LING. MOD.

a) **Grammaire générative, transformationnelle.** Mécanisme, constitué par un ensemble de règles, qui permet de produire et de décrire toutes et rien que les phrases grammaticales d'une langue. De même qu'on peut écrire plusieurs grammaires traditionnelles pour une même langue, de même il est facile de voir que les grammaires transformationnelles sont elles aussi probabilistes (COYAUD, *Introd. ét. lang. docum.*, 1966, p. 76). Le mérite essentiel de la grammaire générative a été, au contraire, de montrer les insuffisances de la linguistique distributionnelle et d'offrir une méthode où il fût beaucoup plus facile de détecter les erreurs (J. STÉFANINI, *Sur la grammaire historique du français* ds *Lang. fr.*, n°10, 1971, p. 24).

♦ **Grammaire formelle.** Le but que se propose la linguistique moderne est de construire des grammaires pour des langues naturelles, qui soient entièrement explicitées sous forme d'automate, ou de construction algébrique (M. GROSS, A. LENTIN, *Notions sur les gramm. formelles*, Paris, Gauthier-Villars, 1967, p. 185).

♦ **Grammaire en chaîne.** „Une grammaire en chaîne est un exemple de programme d'analyse de toutes les phrases d'une langue donnée; elle fournit une analyse exprimée dans une métalangue abstraite par rapport à la langue objet” (M. SALKOFF, *Mathématiques et Sciences humaines*, 1971 ds *Terminol. ling.*).

b) **Grammaire des fautes.** „Méthode dont le but est de décrire sur un corpus, pour une population déterminée, les anomalies de fonctionnement d'un système linguistique” (MOUNIN 1974, p. 158).

Grammaires formelles ?

<http://www.cnrtl.fr/definition/grammaire>

2. LING. MOD.

a) *Grammaire générative, transformationnelle.* Mécanisme, constitué par un ensemble de règles, qui permet de produire et de décrire toutes et rien que les phrases grammaticales d'une langue. De même qu'on peut écrire plusieurs grammaires traditionnelles pour une même langue, de même il est facile de voir que les grammaires transformationnelles sont elles aussi probabilistes (COYAUD, *Introd. ét. lang. docum.*, 1966, p. 76). Le mérite essentiel de la grammaire générative a été, au contraire, de montrer les insuffisances de la linguistique distributionnelle et d'offrir une méthode où il fût beaucoup plus facile de détecter les erreurs (J. STÉFANINI, *Sur la grammaire historique du français* ds *Lang. fr.*, n°10, 1971, p. 24).

♦ *Grammaire formelle.* Le but que se propose la linguistique moderne est de construire des grammaires pour des langues naturelles, qui soient entièrement explicitées sous forme d'automate, ou de construction algébrique (M. GROSS, A. LENTIN, *Notions sur les gramm. formelles*, Paris, Gauthier-Villars, 1967, p. 185).

♦ *Grammaire en chaîne.* „Une grammaire en chaîne est un exemple de programme d'analyse de toutes les phrases d'une langue donnée; elle fournit une analyse exprimée dans une métalangue abstraite par rapport à la langue objet” (M. SALKOFF, *Mathématiques et Sciences humaines*, 1971 ds *Terminol. ling.*).

b) *Grammaire des fautes.* „Méthode dont le but est de décrire sur un corpus, pour une population déterminée, les anomalies de fonctionnement d'un système linguistique” (MOUNIN 1974, p. 158).

De quoi parle-t-on ?

Syntaxe

Parenthèse linguistique : théories

avec votre aide

Vous êtes les experts en linguistique

Exercice

- ▶ quelle(s) théorie(s) connaissez-vous ?
- ▶ est-ce qu'elle(s) impliquent une formalisation de la langue ?
- ▶ si oui, de quel type ?

Deux débats :

1. large : a-t-on besoin de formaliser au sens large les langues ?
Tous les linguistes n'en ressentent pas le besoin. Certains s'y opposent.
2. restreint : ~~une fois qu'~~ si on a accepté la nécessité de formaliser
 - ▶ quel type de formalisation ?
 - ▶ Fuchs et Victorri sont dans un cadre continuiste qui doit peu aux grammaires formelles
 - ▶ Mel'čuk (théorie sens-texte) fait porter la formalisation sur le lexique seul
 - ▶ si l'on est dans le cadre des grammaires formelles, on rentre dans un débat technique délimité : de quelle puissance expressive a-t-on besoin pour rendre compte d'une langue naturelle comme le français, l'anglais ?

Parenthèse linguistique : multilinguisme

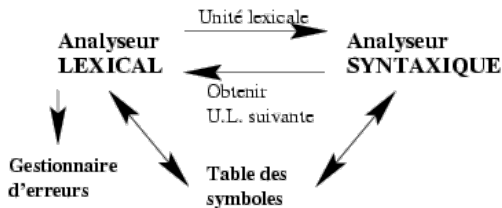
avec votre aide

Vous êtes des locuteurs de nombreuses langues

Exercice (groupes de locuteurs)

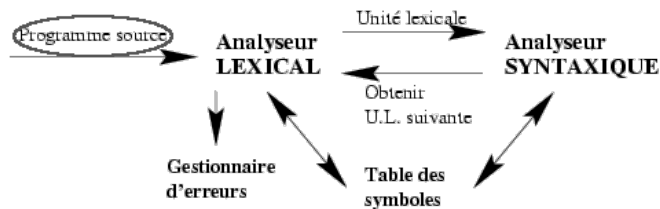
- quelles sont les caractéristiques linguistiques de ces langues ?
- quels problèmes vont-elles poser en TAL (pour l'analyse par ordinateur) ?

De quoi parle-t-on ?



avec l'accord d'Emmanuel Desmontils (U. de Nantes)

De quoi parle-t-on ?



avec l'accord d'Emmanuel Desmontils (U. de Nantes)

Langue et langage (on y reviendra)



<http://www.cnrtl.fr/definition/langue>

II. – [La langue comme système]

A. –

1. **Système de signes vocaux et/ou graphiques, conventionnels, utilisé par un groupe d'individus pour l'expression du mental et la communication.** *Le plus grand des crimes, c'est de tuer la langue d'une nation avec tout ce qu'elle renferme d'espérance et de génie* (NODIER, *Fée Miettes*, 1831, p. 139). *La langue est le signe principal d'une nationalité* (MICHELET, *Tabl. Fr.*, 1833, p. 3). *Elle parla tout à coup dans une langue que je n'avais pas encore entendue. C'était des syllabes sonores, gutturales, des gazouillements pleins de charme, une langue primitive sans doute; de l'hébreu, du syriaque, je ne sais* (NEVAL, *Filles feu*, Octavie, 1854, p. 644). *Elle disait : « Mamma! Oh! Mamma, mamma! » C'est un mot qui est le même dans presque toutes les langues de la terre* (MILLE, *Barnavaux*, 1908, p. 206). *Aventuriers pas très forts sur la grammaire, chancelant sur l'orthographe d'une langue encore instable, mais qui écrivaient comme ils parlaient, les bougres, parce qu'ils étaient des grands vivants, ne faisaient de rhétorique, mais avaient quelque chose à dire et le monde entier à raconter* (CENDRARS, *Bourlinguer*, 1948, p. 15):

. . .

– **LING.** **Système abstrait de signes (par opposition au discours, à l'énoncé ou à la parole, qui en sont l'actualisation).** *La langue est un système de pures valeurs que rien ne détermine en dehors de l'état momentané de ses termes* (SAUSSURE, *Ling. gén.*, 1916, p. 116). *La conception de la langue comme système, conduit à l'affirmation que « dans la langue il n'y a que des différences » et que « la langue est une forme et non une substance »* (F. de Saussure) (PERROT, *Ling.*, 1953, p. 116):

- 9. L'étude du langage comporte donc deux parties : l'une, essentielle, a pour objet la **langue**, qui est sociale dans son essence et indépendante de l'individu (...); l'autre, secondaire, a pour objet la partie individuelle du langage, c'est-à-dire la parole (...). Sans doute, ces deux objets sont étroitement liés et se supposent l'un l'autre : la **langue** est nécessaire pour que la parole soit intelligible et produise tous ses effets; mais celle-ci est nécessaire pour que la **langue** s'établisse; historiquement, le fait de parole précède toujours. SAUSSURE, *Ling. gén.*, 1916p. 37.

Langue et langage (on y reviendra)

<http://www.cnrtl.fr/definition/langage>

I. – [Le langage comme faculté et comme système]

A. – Faculté que les hommes possèdent d'exprimer leur pensée et de communiquer entre eux au moyen d'un système de signes conventionnels vocaux et/ou graphiques constituant une langue; p. *méton.* le langage comme réalisation de cette faculté. *Langage humain; pathologie du langage; philosophie du langage; structure du langage; universaux du langage.*

. . .

LING. [Dans la théorie fonctionnelle d'André Martinet] *Langage articulé, doublement articulé.* Langage considéré comme analysable en unités minimales significatives (les monèmes), elles-mêmes analysables en unités minimales distinctives (les phonèmes), un des caractères communs à toutes les langues. *Le langage humain est non seulement articulé, mais doublement articulé, articulé sur deux plans, celui où, pour employer les termes du parler de tous les jours, les énoncés s'articulent en mots, et celui où les mots s'articulent en sons* (A. MARTINET, *La Double articulation*, Paris, P.U.F., 1965, p. 2).

Langue et langage (on y reviendra)

<http://www.cnrtl.fr/definition/langage>

B. – Système de signes vocaux et/ou graphiques (cf. langue II A).

1. [Langages naturels : les langues parlées dans le monde] *Langage écrit, parlé.* Les mots qui composent cette classe [la classe des mots invariables], ont tous les mêmes raisons d'en être (...); c'est pourquoi ils sont les mêmes dans tous les langages (DESTUTT DE TR., *Idéol.* 2, 1803, p. 128). Les amnésiques du verbe oublient d'abord ce qu'il y a de plus particulier dans le langage, les noms propres, les substantifs, les adjectifs; les parties du langage qui ont la vie la plus dure sont les phrases toutes faites, les locutions usuelles (GOURMONT, *Esthét. lang.* fr., 1899, p. 285). J'ai reçu des lèvres de ma vieille servante le bon langage français (FRANCE, *Pt Pierre*, 1918, p. 196). La grande influence qu'il semble que Descartes ait exercée sur nos Lettres; l'événement dont il est l'auteur, de la première production en langage français d'un ouvrage de philosophie (VALÉRY, *Variété IV*, 1938, p. 209).

2. [Langages artificiels, établis en fonction d'axiomes, de règles d'écriture] *Système de symboles. Langage documentaire (v. ce mot B 2); langage formel, logique.* II [Leibniz] concevait la notion de langage formalisé, pure combinaison de signes dont seul importe l'enchaînement, de sorte qu'une machine serait capable de fournir tous les théorèmes, et que toutes les controverses se résoudraient par un simple calcul (BOURBAKI, *Hist. math.*, 1960, p. 16). Construire des langages artificiels à composantes logiques pour servir aux tâches essentielles de la documentation : analyse, enregistrement des documents, et recherche documentaire (COYAUD, *Introd. ét. lang. docum.*, 1966, p. 67).

– *INFORMAT., PROGRAMMATION.* Ensemble de symboles et de règles permettant de combiner ces symboles afin de donner des instructions à un ordinateur.

♦ *Langage de programmation.* „Langage préétabli utilisé pour écrire les programmes d'un ordinateur déterminé” (MESS. *Télem.* 1979). *Langage symbolique.* „Langage de programmation utilisant des codes mnémoniques pour représenter les instructions machines” (*Informat.* 1972). [Langage symbolique :] code intermédiaire entre le langage machine et les langages externes (langues naturelles ou langages documentaires), et permettant au programmeur de communiquer aisément avec l'automate. C'est pourquoi les langages symboliques sont aussi appelés langages de programmation (COYAUD, *Introd. ét. lang. docum.*, 1966, p. 14).

♦ *Langage machine.* „Langage dans lequel est exprimé un programme au moment de son exécution par l'ordinateur” (*Informat.* 1972).

Quelques réflexions. . .

Aujourd'hui, dans tous les cursus d'informatique (en France), les étudiants suivent un cours de « théorie des langages », qui n'est autre (au moins en partie) qu'un cours de grammaires formelles

Les « grammaires formelles » ne sont en revanche pas enseignées dans tous les cursus de linguistique (en France)



Cette formalisation fonctionne très bien sur des langages informatiques, elle reste sujette à débats sur les langues naturelles, sans doute aussi du fait du « personnage » Chomsky. . .

Introduction

Un « personnage », adulé et haï

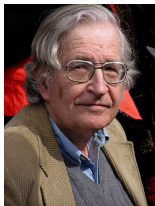
Noam Chomsky

Le « personnage » Chomsky

Théorie(s)

Pour finir

Un intellectuel, charismatique et engagé



© Duncan Rawlinson

- ▶ professeur émérite au MIT
- ▶ philosophe et linguiste
- ▶ thèse à 23 ans (!) sur la morphophonologie de l'hébreu
- ▶ intellectuel engagé

- ▶ écrit *Syntactic structures* dans les années 50 (!)
 - ▶ approche modulaire dans les années 80
 - ▶ « programme minimaliste » dans les années 90
- évolue sans cesse | ne sait pas ce qu'il dit

Un linguiste, formel

[...] la radicalité de Chomsky a validé et a imposé la linguistique comme une réflexion autour du langage comme objet formel, axiomatisable

S. Wauquier (U. Paris 8), com. perso. du 29/07/2014

Un linguiste, formel

[...] la radicalité de Chomsky a validé et a imposé la linguistique comme une réflexion autour du langage comme objet formel, axiomatisable

S. Wauquier (U. Paris 8), com. perso. du 29/07/2014

*Mais, contrairement à ce qui se dit, **il n'était pas l'inventeur de cela** : les structuralistes européens pour la phonologie et Harris pour la syntaxe avaient déjà pas mal déblayé la question*

S. Wauquier (U. Paris 8), com. perso. du 29/07/2014

Un « personnage »...

- ▶ intellectuellement très séduisant (voir ses vidéos, par exemple)
- ▶ intellectuellement clivant :
« je traite de X et le reste n'est pas intéressant pour moi »
- ▶ qui a des « adeptes » et des « ennemis »

Chomsky peut se mettre à toutes les sauces et il y a ce que dit Chomsky et puis ce qu'on dit qu'a dit Chomsky et puis ce qu'on a lu chez ceux qui sont des spécialistes de ce qu'a dit Chomsky sans l'avoir jamais regardé de très très près...

S. Wauquier (U. Paris 8), com. perso. du 29/07/2014

Introduction

Un « personnage », adulé et haï

Théorie(s)

Principaux postulats

Des conséquences à l'échelle d'un domaine

Grammaires formelles et TAL

Pour finir

(Certains des) postulats de départ

- ▶ la langue est le propre de l'homme
 - ▶ la capacité d'acquisition des langues est innée
 - ▶ pauvreté du stimulus
 - **grammaire universelle** qui s'instancie selon les langues
 - puis « programme minimaliste » : la faculté de langage au sens strict est dans la **réursion**
 - ▶ « avec un jeu réduit de règles de grammaire et un ensemble fini de termes, les humains peuvent produire un nombre infini de phrases »
- [Wikipedia, Noam Chomsky, consultée le 17/09/2014]

Formalisation

Désormais, nous appellerons 'langue' un ensemble (fini ou infini) de phrases, chacune d'entre elles étant de longueur finie et composée d'un ensemble fini d'éléments. Toutes les langues naturelles, sous leur forme écrite ou parlée, répondent à cette définition, puisque toute langue naturelle comporte un nombre fini de phonèmes (ou de lettres dans son alphabet) et que chaque phrase peut être représentée comme une suite finie de ces phonèmes (ou lettres) bien que le nombre de phrases soit infini. De même, l'ensemble d'un système formalisé de mathématiques peut être considéré comme une langue. [Chomsky and Braudeau, 1969]

Formalisation

*Puisque toute langue L susceptible de nous intéresser est un ensemble infini, nous pouvons étudier la structure de L **seulement** par le biais de l'étude de dispositifs finis (des grammaires) capables d'énumérer les phrases de L .
[Chomsky, 1959]*

Formalisation

*Puisque toute langue L susceptible de nous intéresser est un ensemble infini, nous **pouvons** étudier la structure de L seulement par le biais de l'étude de dispositifs finis (des grammaires) capables d'énumérer les phrases de L .
[Chomsky, 1959]*

Compétence vs performance : un vieux débat

- ▶ rationalistes vs empiristes
- ▶ empirisme : observation de données naturelles (corpus)
- ▶ rationalisme : jugements introspectifs conscients.

... un TRÈS vieux débat [Rastier, 2004]

- ▶ Humboldt (1839) : “in itself (language) is no product (**ergon**) but an activity (**energeia**)”
- ▶ Aristote (350 **avant JC**) : **puissance** vs **acte**

- ▶ compétence vs performance
- ▶ **compétence** : connaissance tacite, internalisée d'une langue
- ▶ **performance** : expression externe de la compétence langagière (qui peut être influencée par d'autres facteurs)

“We thus make a fundamental distinction between competence (the speaker-hearer’s knowledge of his language) and performance (the actual use of language in concrete situations). [...] A record of natural speech will show numerous false starts, deviations from rules, changes of plan in mid-course, and so on. The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that have been mastered by the speaker-hearer and that he puts to use in actual performance” [Chomsky, 1965].

Exemple [Morrill, 2000]

Peut-on dire ça en anglais ? est-ce correct ?

*The dog that chased the cat that saw the rat that ate the cheese
barked*

Exemple [Morrill, 2000]

Peut-on dire ça en anglais ? est-ce correct ?

The cheese that the rat that the cat that the dog chased saw ate stank

Exemple [Morrill, 2000]

Peut-on dire ça en anglais ? est-ce correct ?

*Oui, je **pourrais** dire ça – mais je ne le ferai **jamais***

Chomsky vs corpus

- ▶ modélisation de la compétence plutôt que de la performance
 - ▶ les langages naturels ne sont **pas** finis, les corpus ne peuvent donc pas produire une description appropriée du langage
 - ▶ l'introspection permet de détecter (aussi) les structures non grammaticales et ambiguës
- ⇒ rejet des méthodes basées sur les corpus en linguistique et mise en place d'une **orthodoxie** rationaliste

L'orthodoxie de Chomsky [Hill, 1962]

Chomsky : *The verb 'perform' cannot be used with mass word objects : one can 'perform a task' but one cannot 'perform labour'*

Anna Granville Hatcher : *How do you know, if you don't use a corpus and have not studied the verb 'perform'?*

Chomsky : *How do I know? Because I am a native speaker of the English language*

But, from BNC corpus, it is possible to 'perform magic' !

Si les grands débats linguistiques vous intéressent : [Harris, 1993]

Le retour de la linguistique de corpus [Johansson, 1991]

Date	Nb d'études
Jusqu'à 1965	10
1966–1970	20
1971–1975	30
1976–1980	80
1981–1985	160
1985–1991	320

En TAL, les corpus sont omniprésents, *via* l'apprentissage

F. Yvon, citant P. Norvig (<http://norvig.com/chomsky.html>)

- ▶ **Moteurs de recherche** : 100 % des principaux acteurs sont probabilistes et à base d'apprentissage
- ▶ **Reconnaissance de parole** : 100 % des principaux systèmes sont probabilistes, à base d'apprentissage, et reposent principalement sur des modèles de Markov cachés
- ▶ **Traduction automatique** : 100 % des concurrents leaders dans les compétitions de type NIST utilisent des méthodes statistiques (par l'exemple)
- ▶ **Question-réponse** : cette application est moins développée, mais de nombreux systèmes dépendent lourdement d'approches statistiques. Le système Watson, d'IBM, qui a récemment gagné au Jeopardy devant des humains est largement basé sur de l'apprentissage, et le système START de Boris Katz est hybride.

En TAL, les corpus sont omniprésents, *via* l'apprentissage

F. Yvon, citant P. Norvig (<http://norvig.com/chomsky.html>)

Qu'en est-il des composants utiles au linguiste computationnel et non à l'utilisateur final ?

(Est-ce la fin des souris ? J'en souris, bien qu'elles me soient souvent utiles.)

En TAL, les corpus sont omniprésents, *via* l'apprentissage

F. Yvon, citant P. Norvig (<http://norvig.com/chomsky.html>)

- ▶ **Désambiguisation lexicale** : 100 % des concurrents leaders dans la compétition SemEval-2 ont utilisé des techniques statistiques. *['souris' / animal ou 'souris' / outil ?]*
- ▶ **Résolution d'anaphore** : la grande majorité des systèmes actuels sont de type statistiques (à l'exception de celui de Haghighi et Klein, qui est hybride). *['elles' fait référence à ?]*
- ▶ **Étiquetage morpho-syntaxique** : la plupart des systèmes actuels sont de type statistiques. Le Brill tagger est une exception notable (hybride). *['souris' / N ou 'souris' / V ?]*
- ▶ **Analyse syntaxique** : approches multiples. *[suj('utile') = ?]*

Le « retour » de Chomsky ?

(...) we may have been too successful. Not only have we succeeded in making room for what we were interested in, but there is no longer much room for anything else [Church, 2011]

Introduction

Un « personnage », adulé et haï

Théorie(s)

Pour finir

CQFR : Ce Qu'il Faut Retenir

Un peu de lecture

Bibliographie



- ▶ en quoi les grammaires formelles sont à la croisée des chemins entre linguistique et informatique
- ▶ compétence vs performance
- ▶ dangers de l'adulation et de la trop grande domination d'une seule théorie (quelle qu'elle soit)

À lire, maintenant

- ▶ [Hauser et al., 2010] The faculty of language: What is it, Who has it and How did it evolve?



Chomsky, N. (1959).

On certain formal properties of grammars.

Information and Control, 2(2) :137 – 167.



Chomsky, N. (1965).

Aspects of the Theory of Syntax.

MIT, Cambridge, MA, USA.



Chomsky, N. and Braudeau, M. (1969).

Structures syntaxiques : "Syntactic structures" . Traduit de l'anglais par Michel Braudeau.

Éditions du Seuil.



Church, K. (2011).

A pendulum swung too far.

Linguistic Issues in Language Technology - LiLT, 6.



Harris, R. (1993).

The Linguistics Wars.

Oxford University Press, USA.



Hauser, M. D., Chomsky, N., and Fitch, W. T. (2010).

The faculty of language : what is it, who has it, and how did it evolve ?

In Larson, R. K., Déprez, V., and Yamakido, H., editors, The Evolution of Human Language, pages 14–42. Cambridge University Press.

Cambridge Books Online.



Hill, A. A., editor (1962).

Proceedings of the Third Texas Conference on Problems of Linguistic Analysis in English, Austin. University of Texas Press.



Morrill, G. (2000).

Incremental processing and acceptability.

Computational Linguistics, 26, 3 :319–338.



Rastier, F. (2004).

Enjeux épistémologiques de la linguistique de corpus.

In Texto !