

## *Une théorie linguistique élémentaire*

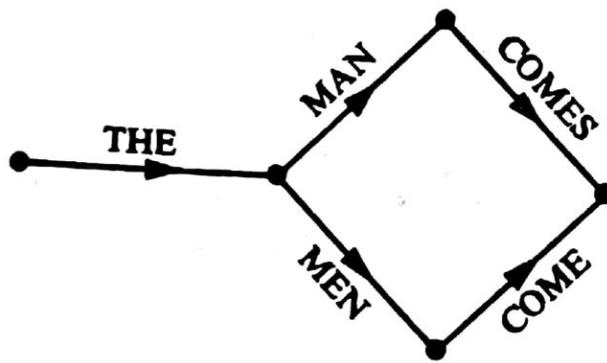
3. 1. Étant donné par hypothèse l'ensemble des phrases grammaticales de l'anglais, recherchons quel type de mécanisme peut produire cet ensemble (ou, ce qui est équivalent, quel type de théorie rend compte de manière adéquate de la structure de cet ensemble d'énoncés). Nous pouvons considérer chaque phrase de cet ensemble comme une suite de phonèmes de longueur finie. Une langue est un système extraordinairement compliqué et il est assez évident que toute tentative pour présenter directement l'ensemble des suites grammaticales de phonèmes aboutirait à une grammaire si complexe qu'elle serait pratiquement inutilisable. Pour cette raison, parmi d'autres, la description linguistique utilise un système de « niveaux de représentation ». Plutôt que d'établir directement la structure phonologique des phrases, le linguiste met en place des éléments de « niveau supérieur » tels que les morphèmes, et établit séparément la structure morphologique des phrases et la structure phonologique des morphèmes. On voit aisément que la description conjointe de ces deux niveaux sera beaucoup plus simple qu'une description directe de la structure phonologique des phrases.

Envisageons à présent divers moyens de décrire la structure morphologique des phrases. Cherchons quel type de grammaire est nécessaire pour engendrer toutes, et rien que, les suites de morphèmes (ou de mots) qui constituent des phrases grammaticales en anglais.

Une des exigences qu'une grammaire doit certainement satisfaire, est d'être finie. Partant, elle ne saurait être une simple liste de toutes les suites de morphèmes (ou de mots), puisque celles-ci sont en nombre infini. Un modèle théorique de la communication assez connu offre une solution à cette difficulté. Supposons une machine pouvant passer par un nombre fini d'états et supposons que cette machine émette

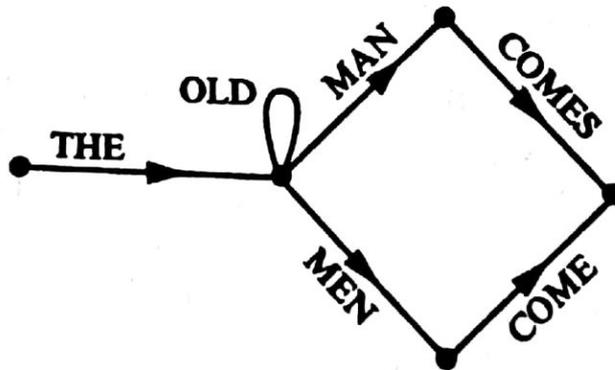
à chaque transition d'un état à un autre un certain symbole (disons « the old man » un mot anglais). Un de ces états est *initial*, un autre est *final*. Supposons que la machine parte de l'état initial, passe successivement par une série d'états en émettant un mot à chaque transition, et aboutisse à l'état final. Nous appellerons alors la suite de mots produite « phrase ». Une telle machine définit ainsi un certain langage, à savoir l'ensemble des phrases qui peuvent être produites de cette façon. Tout langage pouvant être produit par une machine de ce type sera dit *langage à états finis* et la machine elle-même pourra être considérée comme une *grammaire à états finis*. Une grammaire à états finis peut être représentée par un « diagramme d'états »<sup>1</sup>. Par exemple la grammaire qui ne produit que les deux phrases « the man comes » et « the men come » peut être représentée par le diagramme d'états suivant :

(7)



Nous pouvons étendre la production de cette grammaire à un nombre infini de phrases en lui ajoutant des boucles. Ainsi la grammaire à états finis de la sous-partie de l'anglais qui contient, en plus des phrases ci-dessus, « the old man comes », « the old old man comes »,

(8)



1. C. E. Shannon et W. Weaver, *The mathematical theory of communication* (Urbana, 1949), p. 15 s.  
 2. (7) L'homme vient.  
 Les hommes viennent. (N.d.T.)

« the old men come », « the old old men come »... peut être représentée par le diagramme d'états (8) <sup>3</sup> :

Étant donné un diagramme d'états, on produit une phrase en traçant un chemin allant du point initial à gauche au point final à droite, en suivant toujours le sens des flèches. Quand on a atteint un certain point du diagramme, on peut suivre n'importe lequel des chemins qui partent de ce point, que ce chemin ait été ou non déjà parcouru auparavant lors de la construction de la phrase en question. Chaque nœud d'un tel diagramme correspond ainsi à un état de la machine. On peut admettre diverses sortes de transition d'un état à un autre et autant de boucles qu'on voudra, aussi longues qu'on voudra. Les machines qui produisent des langues de cette manière sont connues en mathématiques sous le vocable de « processus de Markov à nombre fini d'états ». Pour compléter ce modèle théorique élémentaire de communication, nous pouvons assigner à chaque transition d'état à état une probabilité déterminée. Nous pouvons ensuite calculer l'« incertitude » liée à chaque état, et définir le « contenu informatif » du langage comme l'incertitude moyenne, mesurée par la probabilité de se trouver dans les états associés. Mais l'objet de notre étude étant la structure grammaticale du langage et non sa structure statistique, cette généralisation ne nous concerne pas.

Cette conception du langage est extrêmement puissante et générale. Si nous l'adoptons, nous considérons le locuteur comme étant essentiellement une machine de ce type. En produisant une phrase, le locuteur part de l'état initial, produit le premier mot de la phrase, passant ainsi dans un deuxième état qui limite le choix du second mot, etc. Chaque état par lesquels il passe représente les restrictions grammaticales qui limitent le choix du mot suivant à ce point de l'énoncé <sup>4</sup>. Étant donné la généralité de cette conception du langage et son application dans des disciplines connexes, telle la théorie de la communication, il est important d'explorer les conséquences qui découlent de l'option d'un tel point de vue pour l'étude syntaxique d'une langue comme l'anglais ou d'un système mathématique formalisé. Tout essai de construction d'une grammaire à états finis de l'anglais se heurte

(8) Le vieil homme vient.

Les vieux hommes viennent. (N.d.T.)

C'est essentiellement le modèle de langage développé par Hockett dans *Journal of phonology* (Baltimore, 1955), p. 10.