

Grammaire distributionnelle

M. Jean Dubois

Citer ce document / Cite this document :

Dubois Jean. Grammaire distributionnelle. In: Langue française, n°1, 1969. La syntaxe. pp. 41-48;

doi : <https://doi.org/10.3406/lfr.1969.5396>

https://www.persee.fr/doc/lfr_0023-8368_1969_num_1_1_5396

Fichier pdf généré le 04/05/2018

GRAMMAIRE DISTRIBUTIONNELLE

La linguistique distributionnelle apparaît aux États-Unis vers 1930 à la fois comme une réaction contre une grammaire fondée sur une vue introspective de la langue et un psychologisme élémentaire, et comme un approfondissement et une systématisation d'une analyse combinatoire (des combinaisons possibles des éléments de la langue) connue depuis le XIX^e siècle. L'ouvrage de L. Bloomfield publié en 1935 (*Language*, Henderson et Spalding, Londres) en marque le point initial et le livre de Z. Harris (*Methods in structural linguistics*, Chicago, 1951) le point d'aboutissement et le développement théorique et méthodologique le plus rigoureux.

La linguistique distributionnelle par ses hypothèses de base et ses postulats fondamentaux rejoint la linguistique européenne, issue de F. de Saussure (*Cours de linguistique générale*, Payot), de N. Troubetskoï, R. Jakobson et de l'école de Prague, qui se développera après la Première Guerre mondiale dans différentes directions : ceux-ci, tout en s'opposant sur des questions méthodologiques (déductif/inductif) ou sur l'accent mis sur certains aspects du langage (code formel/rerelations avec les institutions sociales/incidence des facteurs psychologiques), ont de larges points d'accord et ont fondé ce que l'on appelle la grammaire structurale, la grammaire fonctionnelle et la grammaire distributionnelle.

Les principes théoriques et la méthodologie adoptées, peuvent être ainsi définis :

1. L'objet de la linguistique est la *description* d'une langue, considérée comme un système formel, un ensemble de règles capables par leur fonctionnement de rendre compte des phrases produites. Le linguiste définit donc un *corpus*, c'est-à-dire un ensemble d'énoncés qui répond à la caractéristique d'être compris par les locuteurs de cette langue. Ce corpus est un échantillon de la *langue*, et c'est à partir de lui que seront définies les règles.

La détermination du corpus devra être faite selon un certain nombre de critères assurant qu'il est *représentatif*. Ainsi on devra viser à l'*homo-*

généité des énoncés constituant l'ensemble, l'élimination des écarts ne se faisant pas a posteriori, mais au contraire en définissant des invariants de situation : l'échantillon doit être intelligible par un ensemble de locuteurs de la langue, mais il doit en général être issu d'un seul locuteur-informateur, d'un groupe socialement défini, etc.

2. Si les règles dégagées permettent de rendre compte du fonctionnement d'une langue à un moment donné du temps, ceci signifie que la description du système qu'elles constituent sera *synchronique*. L'analyse *diachronique* se fera dans deux perspectives différentes : on pourra envisager le passage dans le temps d'un système à un autre système : autrement dit, on définira la structure des règles fonctionnant en 1660 (d'après un ensemble de textes) et celle des règles fonctionnant en 1960 et on envisagera les règles de transformation qui permettent de passer de l'une à l'autre, selon une méthode proche de celle que la grammaire comparée connaissait mais avec cette différence essentielle que l'évolution est celle du système entier et non d'une de ses parties. On peut envisager aussi un lieu du système où la diachronie sera une explication de la forme existant en cette place : on rejoint ainsi certaines des explications historiques traditionnelles, mais avec cette différence, là aussi essentielle, que cette explication n'a pas valeur de fonctionnement et ne rend compte que d'un détail anecdotique.

3. La linguistique distributionnelle se place donc devant un *texte* fini, représentatif, synchronique (le mot *texte* étant pris ici en un sens général). Ce dernier est devenu l'objet unique d'analyse. En empruntant la terminologie moderne, on dira que le linguiste distributionnaliste limite son objet aux *performances*, aux énoncés réalisés. Il va alors essayer d'en donner une *description formalisée* ; il en définira la structure par l'architecture, l'interdépendance des éléments internes et sans faire appel aux facteurs que constituent le sujet et la situation. Ce principe d'*immanence du texte* conduit la linguistique structurale à séparer alors la *langue* qui répond à ce système d'organisation et la *parole* où la mise en œuvre des règles de la langue implique l'incidence du sujet parlant et du monde. Abandonnant à une étude ultérieure la linguistique de la parole, du discours, le structuralisme se donne pour tâche la seule description des règles du code que constitue la langue.

4. Chaque langue constitue alors un *système spécifique*, irréductible aux autres systèmes. La place de chaque terme dans une structure se définissant par rapport aux autres termes, il n'y a pas de relation terme à terme entre les langues. Chaque langue découpe le continuum réel (que celui-ci soit les bruits vocaux ou le monde réel perçu) selon des grilles arbitraires toutes différentes : ainsi le /p/ du français n'est pas comparable au /p/ anglais, car il entre dans un système de consonnes (occlusives/fricatives/affriquées) totalement différent d'une langue à l'autre. Chaque langue décrit diversement le spectre des couleurs, les limites de chaque dénomination variant d'un système à l'autre.

5. Ceci implique que l'on ne peut faire intervenir le *sens* dans la définition des unités de la langue et dans leur fonctionnement : le mot sens recouvrant alors la somme des expériences communes d'un ensemble de locuteurs devant le stimulus que le *signe* verbal constitue. La linguistique distributionnelle implique une *psychologie behavioriste* à laquelle elle renvoie le problème du sens. Mais la *logique du discours* qui soutend la langue comme déroulement d'énoncés est implicite. Ceci signifie que la grammaire distributionnelle se donne un certain nombre de postulats fondamentaux qui renvoient à cette logique de base : celui de sujet et de prédicat, celui d'identité, etc. Cette distinction entre le sens (ou, en discours, la signification) et les relations logiques fondamentales n'a pas toujours été reconnue et elle est la source de confusions ou d'incompréhensions.

Encore le structuralisme, définissant la langue comme un système immanent, considérera-t-il que le fonctionnement des signes verbaux s'articule sur deux structures différentes, l'une est celle qui prend pour objet les *signifiants*, les formes expressives, l'autre intéresse les *signifiés*, les formes du contenu. Ces deux structures formelles ne se confondent pas avec la *substance*, celle de l'*expression* (le continuum phonique découpé par la structure de l'expression) et celle du *contenu* (l'expérience continue du monde). La relation existante entre le *signifié* et l'*objet signifié* est *arbitraire*; il n'existe pas de rapport entre la forme du signe et l'objet dénoté. Les interjections et les onomatopées qui se réfèrent à un matériel non verbal (cris) sont en fait différentes d'une langue à l'autre, même si les matrices de formation ont un caractère semi-universel.

Mais ces deux structures, celle de l'expression et celle du contenu, ressortissent au même type d'analyse; on dira qu'il y a *isomorphisme* entre les « deux formes ». Toute syllabe sera découpée sur le plan de l'expression parce qu'elle est constitutrice d'un morphème qui renvoie, lui, à un découpage du contenu.

6. Quelle méthodologie va être appliquée à ce texte fini que l'on analysera sans recourir au sens, mais en présupposant une logique fondamentale?

a) La linguistique distributionnelle considère que tout énoncé est formé d'une *combinaison* d'éléments. La langue présente une série de *rangs* hiérarchisés (phonologique, morphologique, phrastique) où chaque unité (*segment*) est définie par ses combinaisons dans le rang supérieur. Les phonèmes se repèrent par leurs combinaisons en syntagmes constituants de la phrase. La linguistique distributionnelle établit donc une *théorie des niveaux* et, que l'on parle de l'unité supérieure, la phrase, ou de l'unité inférieure, le phonème, on aboutit toujours, soit par analyses, soit par synthèses successives, à un ensemble de niveaux. Les différences entre les écoles sont secondaires; qu'il s'agisse de considérer que l'unité ultime est le phonème ou le trait distinctif, composante du phonème, que l'on marque une différence qualitative entre les rangs du phonème et du

morphème sous la dénomination d'une *double articulation* (Martinet) ou que la terminologie soit légèrement modifiée (monème à la place de morphème), il s'agit toujours d'une combinatoire.

b) Les niveaux supposent donc que la langue se présente comme une combinaison d'éléments. On élaborera ainsi une théorie des *constituants immédiats*. Ceci signifie que deux ou plus de deux unités de rang inférieur entrent en composition pour former une unité de rang supérieur (parfois il n'y a qu'une unité) : ainsi les phonèmes /m/ et /o/ se combinent pour former /mo/ morphème de rang supérieur. Le déterminant /la/ et le nom /bar/ se combinent pour former le syntagme nominal /la bar/, constituant immédiat de la phrase : *la barre est haute*.

Inversement si l'on est devant un énoncé, on doit l'analyser en se fondant sur cette hypothèse des niveaux et des combinaisons successives. L'établissement des règles de combinaison repose essentiellement sur la *distribution des éléments* aux divers rangs : ainsi on constatera que le déterminant /la/ et le substantif se distribuent selon un certain ordre dans le syntagme nominal ou plutôt on déterminera les segments /la/ et les substantifs selon la constante de cette distribution. Le déterminant /la/ sera ainsi distingué du pronom /la/ dont la distribution est totalement différente dans le syntagme verbal.

On constatera aussi, pour reprendre l'exemple de *la barre*, que *la* ne se rencontre qu'avec les substantifs féminins et *le* avec les substantifs masculins : il y a donc des contraintes de distribution entre les deux segments. Autrement dit, on déterminera deux classes de morphèmes (noms) qui se distribueront d'une manière complémentaire les déterminants *la* et *le*, et on ne pourra pas employer *la* avec un masculin sans que l'énoncé ainsi formé soit agrammatical.

c) Cette méthode combinatoire qui définit ainsi les *environnements* de chaque segment est longue; aussi une méthode plus économique, mais qui repose sur les mêmes principes, consiste à utiliser ce que l'on appelle la *commutation*. On dira que deux phonèmes sont distincts si, lorsqu'on les commute, les unités de rang supérieur ainsi formées sont significativement différentes : *mot* et *pot*. De même on déterminera les substantifs de telle ou telle classe selon les combinaisons possibles.

d) On aboutit alors à déterminer des *classes* de phonèmes, de morphèmes, de syntagmes selon leurs combinaisons dans la succession des éléments dans la chaîne parlée ou selon leurs possibilités de commutation en un point de la chaîne parlée. Ainsi on définira une classe des déterminants (*le, mon, ce, etc.*) par les combinaisons possibles avec des noms masculins; on déterminera aussi une classe des noms masculins par les possibilités de se trouver après *le*.

On aura donc deux types de classes : les premières sont *syntagmatiques* puisqu'elles envisagent les segments sur la ligne de la chaîne parlée : A est défini par rapport à X et Y qui le précèdent ou le suivent (X A Y); les secondes sont les classes *paradigmatiques* puisque les

segments y sont définis par des environnements identiques. A et B font partie de la même classe si tous les deux se trouvent dans l'environnement de X et de Y (XAY, XBY). Il s'agit toujours en fait de la même méthode.

On a donc parlé d'axe syntagmatique et d'axe paradigmatique, mais dans les deux cas la méthode de détermination des unités est la même : l'analyse de la distribution des segments dans un énoncé.

e) Cette méthode n'implique pas que certaines manipulations grammaticales soient exclues. En effet la distribution systématique des mêmes segments constatée entre les deux phrases :

Pierre lit ce livre/Ce livre est lu par Pierre

conduit à rapprocher les deux énoncés et à établir un parallélisme systématique; le fait que la deuxième phrase soit moins fréquente que la première amène à la considérer comme une transformation de la première. Le concept de transformation est issu de la constatation de ces distributions, les énoncés restant équivalents.

f) La méthode implique aussi que soient prises en considération les *co-occurrences*. Ainsi lorsque le syntagme nominal est un pluriel, les marques du pluriel se trouvent aussi sur le verbe dont ce syntagme est le sujet. On dira qu'il y a co-occurrence de marques. Cette notion se substitue à celle d'accord et parfois de rection dans une analyse où la distribution des éléments discontinus est essentielle. On aboutit alors à la notion de *morphème discontinu*, autre dénomination de la co-occurrence de marques. Ainsi on dira que les morphèmes /s/ et /t/ sont des membres de la classe des morphèmes discontinus du pluriel dans :

Les vagues battent le rivage
/lə vag bat lə rivaz/

7. C'est cette définition des segments d'un énoncé par l'environnement qui explique la rencontre de l'analyse distributionnelle et de la *théorie de l'information*. La communication est analysée selon le schéma connu d'une source et d'un récepteur, la transmission se faisant sur des canaux spécifiques. Le *bruit* inhérent à la transmission et aux facteurs divers qui entourent cette communication a pour contrepartie une *redondance* aux divers niveaux analysés. Le langage est donc vu comme une succession d'événements, le choix d'un segment étant déterminé par le segment qui précède. Autrement dit, en un point de la chaîne parlée, le choix entre divers éléments est variable, mais relativement déterminé par ce qui précède.

Ces considérations ont justifié l'étude de la distribution statistique des éléments dans un texte, première et grossière approximation de leur distribution linguistique spécifique.

8. L'établissement des structures phonologiques, syntaxiques ou lexicales se fait à partir des principes méthodologiques exposés plus haut. Il s'agit de construire un objet dont les termes entretiennent un certain nombre de relations entre eux, ou plutôt de définir les *oppositions* qui, elles-mêmes, définissent les éléments. Ainsi on aura, en français une série d'occlusives sonores et une série d'occlusives non-sonores qui s'opposent entre elles par la présence ou l'absence d'un trait qualifié ici arbitrairement de sonorité : /p/, /t/, /k/ opposés à /b/, /d/, /g/. L'utilisation des concepts logiques dans la construction de cet objet (Troubetskoï, Jakobson, Cantineau), oppositions bilatérales ou multilatérales, oppositions privatives, etc., permettent l'établissement de *structures*, le plus souvent au niveau phonologique, car le nombre des unités y est limité. Dans une conception où tout est ramené à l'opposition présence/absence, on introduit le concept de *cas marqué*; ce dernier se signalera dans un tel type d'opposition par la présence d'un trait qualifié de *marque*. Ainsi la série sonore est marquée par rapport à la série non-sonore dans les occlusives indiquées.

9. La linguistique distributionnelle a définitivement établi la linguistique comme *science des langues*; sa méthodologie précise est fondée sur des distinctions essentielles (synchronie/diachronie; règles du code/réalisations individuelles variables; traits pertinents/traits redondants, etc.) et sur une combinatoire qui, selon les écoles, s'appuie rigoureusement sur la logique ou la théorie des probabilités. La description des langues dans leur fonctionnement synchronique s'en est trouvée considérablement améliorée. Elle devait perdre deux des caractères les plus nocifs inhérents à la méthodologie antérieure. Le *psychologisme*, le *mentalisme* outrancier, qui faisait substituer les impressions subjectives à l'étude systématique et contrôlable des faits de langue s'en est trouvé ruiné. Et ce psychologisme élémentaire, qu'on ne saurait confondre avec le recours à l'intuition du sujet parlant, faisait reposer l'analyse des énoncés sur des critères non scientifiques. La stylistique devenait un art de dissenter, la syntaxe prenait le tour de « variations sur le thème » de l'emploi des temps et des modes ou de celui de l'ordre des mots chez un auteur; les études lexicales se bornaient à des notations prises au courant de la plume ou à l'extraction hasardeuse de pseudo-néologismes. La grammaire s'appauvrissait entre les mains des gens plus attentifs à rendre compte des faits grammaticaux par le recours à un prétendu « bon sens » très éloigné de la combinatoire logique ou à fonder une norme sur des considérations morales ou pseudo-sociologiques. Le mérite des écoles structuralistes, c'est d'avoir définitivement supprimé tout ce qui entravait le progrès de la linguistique vers son statut de science humaine. Mais le structuralisme devait aussi débarrasser la description des langues de l'*ethnocentrisme* occidental qui avait rendu vaines toutes les notations antérieures. En considérant chaque système comme autonome, spécifique, on évitait de ramener chaque idiome à la grille catégorielle utilisée pour les langues

indo-européennes. Les catégories linguistiques utilisées, même si elles se fondaient sur des sortes d'universaux méthodologiques, variaient notablement d'une langue à l'autre. Et, par un retour heureux, le structuralisme a permis de reprendre la description des grandes langues de communication et de culture sur de nouvelles bases propres à rendre compte de leur fonctionnement réel.

10. Le structuralisme a posé aussi les principes d'une *théorie du langage*. S'attachant à décrire le fonctionnement d'un système de communication, il a vu que, ce faisant, il devenait aussi la science du langage. Toute analyse linguistique suppose en effet une certaine théorie de la communication et de la production du langage. Sans doute cette constatation devait-elle à son tour produire un renouvellement dans l'histoire même de la linguistique. En effet, le *distributionnalisme* prenait appui, nous l'avons vu, sur deux hypothèses d'ailleurs corréliées. D'une part il impliquait avec la linéarité de la chaîne parlée la prise en considération des contraintes contextuelles : le langage est produit selon le schéma des chaînes de Markov, chaque segment entraîne d'une certaine manière le segment suivant qui ne peut être *choisi* que dans un ensemble limité de termes. D'autre part, l'analyse exclusive des réalisations trouvées dans un corpus et le rejet systématique de toute interprétation fondée sur une théorie du sujet et de la situation montraient le lien entre la psychologie behavioriste et le structuralisme, surtout américain : les études de Skinner sur l'apprentissage attestent le rapport qui existe entre le courant bloomfieldien et les psychologues soucieux de décrire tout comportement symbolique en termes de stimuli-réactions. Après le structuralisme, il est devenu difficile de penser toute nouvelle théorie des langues sans en inférer immédiatement une théorie du langage. Avec la linguistique structurale on a eu une première théorie, élémentaire certes, mais achevée.

11. Dans la mesure où il est une théorie de la langue et où il s'efforce de construire une grammaire explicite qui rende compte du fonctionnement du code linguistique, tant sémantique que formel, le structuralisme prête le flanc aux critiques qui mettent en évidence l'*insuffisance du modèle du langage qu'il implique*. Le premier, N. Chomsky a indiqué dans *Syntactic structures* que le distributionnalisme et la théorie des constituants immédiats qui en est une version plus achevée (Bloomfield), ou les grammaires « context-sensitive » (grammaires de dépendances contextuelles) n'expliquent pas la totalité des phénomènes de production d'énoncés. En particulier, la *créativité du langage* ne peut être incluse dans la linguistique structurale, sinon sous la notion de combinatoire ouverte (possibilité de combinaisons nouvelles et infinies à partir d'éléments ou morphèmes en nombre fini). En fait, cette combinatoire reste étroitement liée au traitement de l'échantillon d'un corpus limité. Or, dans la créativité du langage il y a beaucoup plus : cela signifie que n'importe quel sujet parlant est capable de comprendre ou de produire des messages qui

n'ont pas été formulés avant lui et il convient que les grammaires construites puissent rendre compte de ce phénomène.

12. Le structuralisme a rencontré le *problème de l'histoire*. Mais ce qui s'est posé à lui n'est pas en fait la question de la perspective historique, ni celle de la réintroduction de la diachronie, mais plutôt de sa place exacte dans la linguistique. Car il existe un structuralisme historique et génétique qui apparaît avec la notion de transformation pour expliquer l'évolution des structures. On l'a d'abord vu comme un développement à partir des lieux de déséquilibre d'un système. Puis les structures décrites ont été considérées comme susceptibles de transformations sous l'impact des phénomènes extérieurs; en ce cas elles tiennent leurs modalités transformationnelles des structures internes et leur mouvement de la conjoncture; les deux types d'explication ont été ainsi combinés avec le concept de surdétermination. Ce qu'apporte alors la linguistique générative, c'est une perspective nouvelle dans la hiérarchie des évolutions, dans la mesure où sont distinguées dans les structures de surface la part des modèles de compétence et celle des modèles de performance. Ces derniers comportent l'intégration d'une théorie du locuteur et d'une théorie de la situation, et, en ce sens, ils mettent à leur vraie place les phénomènes qui étaient rejetés autrefois comme extralinguistiques. Par exemple, dans la mesure où on inclut maintenant les phénomènes prosodiques dans les règles de syntaxe, l'importance des modifications d'intonation dans l'évolution des structures est revalorisée.

Il n'en reste pas moins qu'il y a pour le structuralisme une difficulté à se mouvoir sur un axe différent de celui sur lequel il s'est placé préférentiellement, celui de la synchronie. Cette difficulté tient à ce que son analyse combinatoire et son analyse sémique sont liées à une manière d'appréhender les phénomènes linguistiques. Le structuraliste construit des *modèles de récepteur* bien plus que des modèles d'émetteur; il esquisse inductivement le modèle de langue (de compétence) à travers et à partir des performances; il ne le pose pas comme le principe à partir duquel on peut déduire (ou générer) les phrases. Le structuralisme est synchronique parce qu'il est *inductif* d'une manière cohérente et que sa référence va à des textes *homogènes*. Même quand il envisage l'histoire, il établit un rapport entre un texte et sa signification dans notre monde actuel. Il serait donc vain de croire que la négligence de l'histoire est un moment de structuralisme; elle lui est inhérente.

JEAN DUBOIS
Paris