



**Linx**

Revue des linguistes de l'université Paris X Nanterre

**80 | 2020**

**L'héritage de Jean Dubois et Françoise Dubois-Charlier**

---

## Jean Dubois, un passeur ?

*Jean Dubois's role in the introduction of American linguistic theories in France*

Jacqueline Léon

---



### Édition électronique

URL : <http://journals.openedition.org/linx/5561>

DOI : 10.4000/linx.5561

ISSN : 2118-9692

### Éditeur

Presses universitaires de Paris Nanterre

### Référence électronique

Jacqueline Léon, « Jean Dubois, un passeur ? », *Linx* [En ligne], 80 | 2020, mis en ligne le 10 juillet 2020, consulté le 05 août 2020. URL : <http://journals.openedition.org/linx/5561> ; DOI : <https://doi.org/10.4000/linx.5561>

---

Ce document a été généré automatiquement le 5 août 2020.

Département de Sciences du langage, Université Paris Ouest

---

# Jean Dubois, un passeur ?

*Jean Dubois's role in the introduction of American linguistic theories in France*

Jacqueline Léon

---

- 1 Dans le contexte intellectuel français du début des années 1960, où la linguistique américaine, en particulier sa mathématisation, est encore largement méconnue, Jean Dubois a largement contribué à l'introduction en France de l'analyse distributionnelle et de l'analyse de discours harrisiennes, de la théorie de l'information, enfin de la grammaire générative<sup>1</sup>. Dans cet article, nous retracerons les premiers travaux de Jean Dubois ayant contribué à cette diffusion, dans les années 1960-70, en tentant d'évaluer son rôle en tant que passeur ou bien plus simplement de facilitateur.
- 2 Si l'on définit un passeur par sa capacité à faire circuler les concepts et les méthodes d'une discipline à l'autre dans les deux sens et, pour les recherches américaines, des deux côtés de l'Atlantique, on peut dire que Marcel-Paul Schützenberger, Benoît Mandelbrot et Maurice Gross, ou bien de sociétés savantes comme l'ATALA (Association pour la Traduction Automatique et la Linguistique Appliquée) et le Centre Favard, ont joué un rôle de passeurs pour les travaux en traduction automatique, en linguistique formelle et computationnelle.
- 3 Nous examinerons dans cette perspective les premiers travaux de Jean Dubois : sa thèse sur le lexique où l'analyse distributionnelle est appliquée pour la première fois en France, ses travaux inspirés des travaux de la conception de Roman Jakobson de la théorie de l'information, enfin ses travaux sur l'analyse de discours largement inspirés par ceux de Zellig S. Harris ; travaux alimentés et complétés par son important travail d'édition des traductions des travaux américains, contribuant largement à leur diffusion.
- 4 Une des façons d'évaluer son rôle de passeur consistera à appréhender la façon dont ses disciples immédiats ont repris ou poursuivi ses travaux dans le domaine de l'analyse de discours. Pour ce faire, nous nous appuierons sur l'examen des numéros de *Langue Française* n° 9 et *Langages* n° 23, tous deux publiés en 1971 et dirigés par des élèves de Jean Dubois. Nous nous limiterons à l'analyse de discours, et nous n'envisagerons pas la postérité de ses travaux en lexicographie qui se sont poursuivis bien après les années

1970, notamment dans le cadre de sa collaboration avec Maurice Gross et son équipe du LADL (Laboratoire d'Analyse Documentaire et Linguistique).

## 1. Réception de la linguistique américaine en France : le rôle des passeurs

- 5 La réception de la linguistique américaine est particulièrement tardive en France. Alors qu'aux États-Unis, les linguistes sont déjà engagés dans un processus de mathématisation et d'automatisation, les linguistes français ne sont familiers avec aucun des domaines qui en font le fondement. En effet, un certain nombre de bases mathématiques et méthodologiques de la linguistique structurale américaine, en particulier son ancrage dans la mathématisation du langage des années 1930, et qui leur ont assuré une certaine familiarité avec l'axiomatisation, les méthodes algébriques, les chaînes de Markov et les automates à états finis via la théorie de l'information, ont été totalement absentes de l'environnement et de l'ancrage intellectuel des linguistes français (sur ce point voir Léon 2015).
- 6 Parallèlement, et pour les mêmes raisons, auxquelles s'ajoute un retard conséquent en informatique, la France a une dizaine d'années de retard sur les premières expériences en traduction automatique. Les premiers centres sont créés en France à la fin de l'année 1959, alors qu'elles avaient débuté à la fin des années 1940 aux États-Unis et en Grande-Bretagne.
- 7 Face à cette situation, le rôle des passeurs, chercheurs et institutions, a été capital. On peut en citer quelques-uns : Marcel-Paul Schützenberger (1920-1996) a assuré le passage entre mathématiques, informatique et langages formels d'une part, et entre les travaux américains et les travaux français d'autre part. Maurice Gross (1934-2001) peut également être considéré comme un passeur entre travaux français et américains, entre langages formels et linguistique. Gross contribuera largement à la diffusion des grammaires formelles et des travaux de Harris et à leur intégration à la linguistique française.
- 8 Quant au Danois Knud Togeby (1918-1974), disciple de Hjelmslev, il sera le premier passeur de la linguistique structurale américaine en France. Son ouvrage, *Structure immanente de la langue française*, paru en français en 1951 dans *Les cahiers de linguistique de Copenhague*, établit une véritable typologie des méthodes structuralistes, européennes et américaines. Il connaît un succès certain en France dès sa parution, fait l'objet de nombreuses recensions dans des revues françaises et est réédité chez Larousse en 1965 par Claude et Jean Dubois. C'est grâce à Togeby notamment, que Harris sera plus connu et mieux compris que Chomsky en France.
- 9 Ce n'est qu'en 1965-66, que les travaux américains vont être diffusés de façon systématique auprès des linguistes voire d'un public plus large, au-delà des cercles des sociétés savantes comme l'ATALA et le Centre Favard. L'ATALA, créée en 1959, qui réunit des linguistes, des informaticiens et des documentalistes, se veut un forum de discussion sur les langages formels, la traduction automatique, la documentation automatique et la linguistique appliquée. Le Centre Favard, créé en mars 1960, assure, au sein de son Séminaire de Linguistique Quantitative à l'Institut Henri Poincaré, un enseignement de linguistique pour mathématiciens auquel participent Jean Dubois et André Martinet, et un enseignement de mathématiques, logique mathématique et

théorie de l'information par les mathématiciens René Moreau et Daniel Hérault. En 1962-1963, un cours supplémentaire sur la théorie des langages est assuré par Jacques Pitrat et Maurice Gross. Ces enseignements donneront lieu à la publication en 1967 de l'ouvrage de Maurice Gross et André Lentin *Notions sur les grammaires formelles*, qui restera longtemps la seule véritable introduction au domaine.

## 2. Jean Dubois : introduction et traduction de la linguistique américaine en France

- 10 Outre son activité d'enseignement au Centre Favard et la réédition de Togeby, Jean Dubois joue un rôle fondamental dans la diffusion de la linguistique américaine par sa position aux éditions Larousse où travaille également son frère Claude. La politique de traduction systématique qu'il met en œuvre dans la revue *Langages* qu'il contribue à créer (en 1966) en est un vecteur important (Chevalier 2006). Cette pénétration des travaux linguistiques américains en France coïncide, il faut le souligner, avec l'explosion éditoriale de 1966 qualifiée par Dosse (1991) d'« année structurale » et dominée par la linguistique érigée en science pilote.
- 11 *Langages* a consacré nombre de ses premiers numéros aux premières traductions en français de Chomsky. Rappelons que *Syntactic Structures*, paru aux Etats-Unis en 1957, n'a été traduit en français qu'en 1969. Dans *Langages* 2, consacré au thème 'Logique et linguistique' et confié à Oswald Ducrot, paraît le second article traduit de Chomsky<sup>2</sup>. Il sera suivi d'une série de quatre numéros où seront traduits des articles exposant les différentes facettes de ses travaux : *Langages* 4 (1966) 'La grammaire générative' est dirigé par Nicolas Ruwet ; *Langages* 9 (1968) 'Les modèles en linguistique', sous la direction de Maurice Gross, traite des aspects formels ; *Langages* 16 (1969) 'Psycholinguistique et grammaire générative', sous la direction de Georges Mehler, est consacré au volet cognitif. Il faut également signaler, bien qu'ils ne comportent pas d'articles de Chomsky, la parution en 1967 de *Langages* 8 consacré à la phonologie générative, dirigé par Sanford Shane, et en 1969 celle de *Langages* 14, de nouveau dirigé par Ruwet, sur les tendances nouvelles de la syntaxe générative.
- 12 Quant aux traductions de Harris en français, c'est Maurice Gross qui, le premier, traduit un article de Harris « From Morpheme to Utterance » dans le numéro de *Langages* de 1968 consacré aux modèles en linguistique. Succèdent l'article « Discourse Analysis » (1952) traduit par Françoise Dubois-Charlier dans *Langages* 13 de 1969 ; l'ouvrage de Harris de 1968, *Mathematical Structures of Language* est traduit en 1971 par Catherine Fuchs ; enfin la traduction de « Les deux systèmes de la grammaire : prédicat et paraphrase » paraît dans *Langages* 29 dirigé par Danielle Leeman en 1973.

## 3. Jean Dubois : le courant lexicologique de l'Analyse de discours française

### 3.1. Premières utilisations de la méthode distributionnelle de Harris et lexicologie structurale

- 13 Dans un premier article lié à son travail de thèse sur le vocabulaire politique et social en France de 1869 à 1872, qui sera publié en 1962, Dubois (1960) examine des unités

complexes comme *l'émancipation des travailleurs*. Le rôle que joue ce type d'unité dans le vocabulaire révolutionnaire lui fait perdre sa compositionnalité au profit d'un nouveau trait signifiant qui s'ajoute ou se substitue aux autres, tel que *établissement d'un nouvel ordre économique favorable aux travailleurs*. C'est par neutralisation, terme qu'il emprunte à la phonologie, que les unités composées se constituent comme unités sémantiques (on pourrait dire discursives) dans un texte. A partir du vocabulaire social de l'époque considérée, il constitue une série d'unités sémantiques à l'aide de classes d'équivalence :

émancipation des travailleurs  
des masses  
des ouvriers  
du prolétariat  
de la classe ouvrière

- 14 On voit là une application de la méthode distributionnelle de Harris que Dubois est probablement le premier linguiste français à utiliser. Il connaît cette méthode, et dans sa thèse, il cite Togeby et *Methods in Structural Linguistics* de Harris. Il se démarque des travaux de lexicologie et de stylistique de l'époque fondés sur des études statistiques (voir les travaux de Pierre Guiraud et de Charles Muller par exemple) en amorçant les débuts d'une lexicologie structurale où l'univocité signifiant / signifié n'est pas tenue pour acquise.
- 15 Comme dans l'article de 1960, il appréhende les groupes nominaux complexes comme des unités sémantiques qui ont une fonction discursive pour un corpus donné. Il établit des unités sémantiques à l'aide de classes d'équivalence qui lui permettent d'étudier discursivement le vocabulaire social de l'époque considérée.

Tableau figurant dans la thèse de Jean Dubois (p. 186)

Affranchissement	Du prolétaire Du prolétariat Des travailleurs	Avènement politique	De la classe ouvrière Des travailleurs
Emancipation	Des masses Du prolétariat De la classe ouvrière Des travailleurs	Aspirations	Des classes ouvrières Du prolétariat Du peuple Ouvrières
Avènement	De la classe ouvrière Du prolétariat Des travailleurs	Amélioration	Du sort des travailleurs Des conditions des classes ouvrières
		Cause	Des travailleurs Des masses Du peuple

## 3.2. Analyse du discours et énonciation

- 16 Dans son article « Lexicologie et analyse d'énoncé » paru dans les *Cahiers de Lexicologie* en 1969, Dubois déclare que l'ambition du lexicologue est avant tout d'élaborer une science de l'énoncé. Son projet est certes toujours lexicologique, et le point de départ est l'analyse du vocabulaire, mais celle-ci doit être comprise comme une partie de l'analyse du discours.
- 17 Sur le plan méthodologique, Dubois applique de nouveau l'analyse distributionnelle de Harris, complétée par une manipulation transformationnelle aboutissant à la constitution de classes d'équivalence. Les manipulations transformationnelles sont destinées à mettre dans les mêmes classes d'équivalence passifs et actifs, compléments de nom et adjectifs. De même, les phrases complexes sont transformées en propositions à deux arguments. L'analyse consiste à étudier les répartitions des mots co-occurents dans les mêmes types de propositions, par exemple le terme « étudiant » dans les textes de mai 1968. Ces opérations linguistiques une fois effectuées, on met en correspondance les modèles linguistiques (modèles d'analyse d'énoncé sur la base des propositions lexicales et modèles de discours) et les modèles sociologiques, définis ici comme des structures idéologiques.
- Lorsque, par analyse lexicale, on choisit dans ce corpus un certain nombre de vocables, on émet du même coup l'hypothèse que les propositions réunies autour de ces termes sont représentatives du corpus et permettent d'établir une relation avec le modèle idéologique de l'auteur (Dubois 1969a, p. 117).
- 18 L'inspiration est multiple, d'abord l'analyse distributionnelle de Harris mais aussi les approches énonciatives de Jakobson et Benveniste. C'est pourquoi l'analyse du discours pour Dubois nécessite une typologie des discours qui détermine des rapports de l'énonciation à l'énoncé ; il distingue notamment les discours polémiques des discours didactiques.
- 19 La même année, en 1969, Dubois co-dirige avec Joseph Sumpf un numéro de *Langages* consacré à l'analyse du discours où il poursuit son intérêt pour une approche mixte, transformationnelle et énonciative<sup>3</sup>. Cette association entre grammaire transformationnelle et énonciation s'illustre parfaitement dans la dernière phrase de son article « Énoncé et énonciation » figurant dans ce numéro de *Langages* 13 :
- Sans doute fallait-il qu'une théorie linguistique modifie entièrement les modes d'analyses en renversant certains des axiomes les mieux établis : celui des niveaux et des rangs, des unités discrètes et de la combinatoire, pour lui substituer l'ordonnement, la suite des transformations et que place soit laissée à une continue intervention du sujet dans l'objet en voie de réalisation, pour que l'énonciation retrouve une place fondamentale dans l'étude linguistique. (Dubois, 1969b : 110)

## 4. Théorie de l'information

### 4.1. Linguistes structuralistes et théorie de l'information

- 20 On comprend que les linguistes structuralistes américains aient été séduits par la théorie de l'information qui offre une large panoplie de nouvelles méthodes et de nouvelles notions. L'approche behavioriste liée à une conception du langage en tant qu'ensemble d'événements physiques (*sound waves and ink marks*)<sup>4</sup> est tout à fait

compatibles avec le traitement des 'signaux' de la théorie de l'information. Les deux approches partagent l'idée que le sens n'est pas pertinent, et que l'utilisation de méthodes quantitatives et probabilistes est tout à fait adaptée pour l'analyse linguistique. C'est ainsi que Hockett (1953 entre autres) et Harris (1955) adoptent les notions de redondance et de non-équiprobabilité des éléments linguistiques, les probabilités transitionnelles à l'œuvre dans les chaînes de Markov et le traitement de la parole et de la chaîne écrite en tant qu'événements physiques, qu'ils appliquent au découpage des morphèmes en phonèmes et des énoncés en morphèmes.

- 21 En ce qui concerne l'analyse du discours, Harris place la répétition et l'idée que les langues ne sont pas équiprobables (idée déjà formulée par Shannon en 1948) au fondement de la construction des classes d'équivalence<sup>5</sup>. En 1957 cependant, l'information devient une notion sémantique et se définit comme invariant sémantique entre deux structures unies par une transformation, idée qu'il développera tout au long des années 1960. Dans la dernière partie de son œuvre centrée sur 'Language and Information', il se consacre à la construction de sous-langages des sciences : la représentation de l'information n'est possible que pour un sous-langage d'une science donnée à un moment particulier. En incorporant progressivement la notion d'information sémantique au cœur même de la théorie, un nouvel objectif, celui d'une grammaire de l'information pour l'étude des sous-langages des sciences<sup>6</sup>, se constitue peu à peu. Tout en se détachant des aspects méthodologiques de la théorie de l'information shannonienne (chaînes de Markov et probabilités) qu'il adapte à la méthode distributionnelle de repérage et de classement des unités, Harris développe une conception sémantique de l'information, destinée à construire des sous-langages des sciences.
- 22 Le glissement de l'interprétation de Harris entre conception quantitative et conception sémantique de l'information constitue une illustration exemplaire de l'utilisation de cette notion par les linguistes. Il faut signaler que, dès son apparition comme terme scientifique, une confusion existe entre information-contenu et information-signal. C'est pourquoi Carnap et Bar-Hillel (1952) proposent alors deux termes pour rendre compte des deux dimensions de l'information, la « théorie de la transmission du signal » dont l'objet est le traitement des messages en tant que séquences de signaux, et la « théorie de la transmission de l'information » pour le traitement des messages en tant que porteurs de sens. C'est vers ce dernier sens, sémantique, que vont dériver ou converger les différentes approches linguistiques. C'est le cas notamment de Jakobson et à sa suite Jean Dubois qui vont concevoir l'information comme une entité sémantique, et parlent de contenu d'information.

## 4.2. Jakobson : d'une conception quantitative à une conception sémantique de l'information

- 23 Les développements de Jakobson à partir de la théorie de l'information ont eu un impact certain sur la linguistique européenne comme américaine. En poursuivant le développement de la DFT (Distinctive Features Theory) avec le théoricien britannique de l'information Colin Cherry, tenant d'une orientation communicationnelle de la théorie de l'information, Jakobson et Halle font l'hypothèse que la théorie mathématique de la communication va fournir un fondement scientifique rigoureux pour l'interprétation et l'analyse des systèmes phonologiques (Cherry et al. 1953). Pour

Jakobson, la théorie de l'information est l'occasion d'édifier une mathématisation du continu, à savoir formaliser le passage du continu (le signal de la parole observable en production et en perception) au discret (les unités phonologiques du message).

- 24 Toutefois, lorsqu'il ne traite pas de phonologie, Jakobson n'hésite pas à confondre le sens d'information au sens mathématique et celui d'information comme contenu. Ainsi, dans son article de 1959 « *Boas' view of grammatical meaning* », afin de montrer que toutes les marques morpho-syntaxiques sont marquées dans le code de façon variable selon les langues et que l'information grammaticale est une information sémantique, Jakobson utilise le terme *information* à la fois au sens mathématique (*bits of information*) et au sens sémantique (*grammatical information* et *semantic information*). C'est cette conception de l'information par Jakobson, éminemment sémantique, que choisira Jean Dubois dans ses travaux.

The choice of a grammatical form by the speaker presents the listener with a definite number of *bits of information*. The compulsory character of this kind of *information* for any verbal exchange within a given speech community and the considerable difference between the *grammatical information* conveyed by diverse languages were fully realized by Franz Boas, thanks to his astonishing grasp of the manifold semantic patterns of the linguistic world... It was clear to Boas that any difference of grammatical categories carries *semantic information*<sup>7</sup> (Jakobson 1971 [1959], p. 490-493).

### 4.3. Réception en France de la théorie de l'information

- 25 Il faut noter que c'est la réception de Jakobson par André Martinet d'une part, et par Jean Dubois d'autre part qui a ouvert la voie à l'introduction de la théorie de l'information dans la linguistique structurale française (voir Léon 2008).
- 26 Pour les linguistes stylisticiens (en particulier Pierre Guiraud), c'est un mathématicien, Benoît Mandelbrot, qui effectue le passage, ce qui conduit au renouvellement des études traditionnelles de vocabulaire dans les années 1950-1960. Une autre voie encore est frayée par des ingénieurs-cryptanalystes français, chargés du « chiffre » pendant la guerre d'Indochine, comme René Moreau, associant théorie de l'information et tournant de l'automatisation-mathématisation en France<sup>8</sup>.
- 27 La conception de la notion d'information de Martinet est somme toute assez classique. Il considère qu'il existe un rapport constant et inverse entre la fréquence d'une unité et l'information qu'elle apporte. Plus une unité est probable, moins elle est informative. Son apport spécifique consiste en une définition de l'information en termes de moindre effort. Outre la phonologie, il applique la notion d'information à la littérature et à d'autres unités linguistiques, comme les monèmes ou les formes lexicales. Martinet utilise les chaînes de Markov pour traiter ces séquences d'unités d'une façon très proche de la méthode mise au point par Harris (1955) pour segmenter les énoncés en morphèmes. Dans les *Éléments de linguistique générale* (Martinet 1960), il consacre un long développement à sa propre vision de l'information, ramenée à l'idée d'énergie et de coût.



#### 4.4. Dubois : une version harrisso-jakobsonienne de la théorie de l'information

- 28 L'intérêt de Jean Dubois pour la théorie de l'information apparaît dans des textes postérieurs à ses travaux en lexicographie, d'abord dans un article sur l'aspect et le temps paru en 1964 dans *le Français Moderne*, puis dans sa *Grammaire structurale du français* en trois tomes parus respectivement en 1965, 1967 et 1969. Il est à noter que Dubois a publié, dans un numéro du *Français moderne* toujours en 1964 (Dubois 1964b) un compte-rendu des *Essais de linguistique générale* de Jakobson paru en français en 1963.
- 29 Dans cet article, « La traduction de l'aspect et du temps dans le code français (structure du verbe) », Dubois applique les théories de Jakobson aux problèmes de temps et d'aspect, notamment le schéma de communication et les *grammatical meanings*. De plus il utilise les distinctions de Benveniste en temps/récit, etc.
- 30 L'aspect et le temps sont marqués dans le code (on notera l'utilisation du terme 'code' pour structure linguistique, emprunté à la théorie de l'information<sup>9</sup>) de façon variable selon les langues. On retrouve ici la notion de signification grammaticale (*grammatical meanings*) de Boas via Jakobson (1959), selon laquelle les marques morpho-syntaxiques sont marquées dans le code de façon variable selon les langues et que l'information grammaticale est une information sémantique.

Les notions d'aspect et de temps appartiennent non au code lui-même mais au message. Il ne peut exister un système temporel ou un système aspectuel du français. On doit seulement envisager à l'aide de quelles structures formelles sont traduites dans un code donné les oppositions sémantiques d'accompli/ non accompli ou d'antériorité / non-antériorité qui définissent ce que l'on a pris l'habitude d'appeler, d'une part, l'aspect, et de l'autre le temps ; on doit en effet constater qu'il existe des langues qui sont appelées à traduire non ces trois oppositions, mais parfois seulement deux d'entre elles, le développement de la postériorité / non postériorité se faisant quelquefois tardivement. (Dubois, 1964a, p. 1)

- 31 C'est surtout dans la *Grammaire structurale du français* que Dubois va utiliser la théorie de l'information. Il faut préciser que le terme *information* apparaît massivement dans le tome 1 de 1965 pour quasiment disparaître dans les tomes 2 et 3.
- 32 Dans l'introduction du tome 1, Dubois précise qu'il a pour objectif une analyse distributionnelle des marques de genre et de nombre du nom et du pronom, qui se veut complémentaire de l'analyse distributionnelle de Harris. C'est en termes d'information, redondance et bruit qu'il se propose d'analyser les contraintes croisées entre marques de genre et de nombre pour montrer que ce sont des règles et non des exceptions en français. Alors que, dit-il - faisant sans doute référence à la notion de *rendement fonctionnel* de Martinet - les linguistes ont surtout emprunté à la théorie de l'information la notion de coût, il se propose de mettre en œuvre celles de redondance et de bruit.
- 33 Ces notions permettent de rendre compte, dans le schéma de communication, de la réception qui ne serait pas symétrique à l'émission. Il introduit la mémoire et les règles. Autrement dit, on ouvre la boîte noire : le cerveau exécute des programmes de reconnaissance (des opérations), applique des règles pour comparer les signaux reçus avec les éléments en mémoire. On notera que cette conception anti-behavioriste est probablement inspirée de l'ouvrage de George A. Miller *Language and communication* (1951, traduction française 1956) que Dubois cite dans sa thèse. Il n'y a pas de symétrie

entre émission et réception, parce qu'il y a des bruits qui viennent masquer la communication. L'auditeur n'entend qu'une partie du message ; celui-ci a donc besoin d'être redondant pour être compris.

- 34 La notion de quantité d'information utilisée par Dubois (de même que celles de conservation, perte et coût de l'information) est bien celle de la théorie de l'information : elle est calculée en termes de probabilités et de fonction inverse : plus la probabilité d'apparition d'une marque est grande, plus la quantité d'information est faible. Ainsi la quantité d'information apportée par le code écrit est faible, puisque celui-ci est très redondant :

Les marques du code parlé et du code écrit ne sont donc pas équiprobables : la *quantité d'information* apportée par une marque redondante dans le code parlé est plus grande que celle apportée dans le code écrit, puisque la probabilité de la rencontrer après la marque initiale est moindre. La *quantité d'information* présentée en français écrit par la seconde marque est très faible, puisqu'il est très rare que le code écrit ne présente pas de redondance. (Il n'en est évidemment pas de même pour le code parlé) (Dubois 1965, p. 21).

- 35 Cette utilisation quantitative et shannonienne de l'information par Dubois est probablement issue de la lecture de Mandelbrot (1954) figurant dans sa bibliographie mais aussi de celle de Harris, dont les principaux travaux des années 1950 sont cités, et de son utilisation des chaînes de Markov. Plus loin dans le texte, le terme *information* et les collocations dans lesquelles il apparaît (*une information, la première information, l'information du pluriel, l'information féminin, l'information de genre, celle de nombre, les /ces deux informations*), sont utilisés de façon très différente :

On peut se demander ce que deviennent *ces deux informations* lorsqu'elles sont cumulées, c'est-à-dire lorsque les énoncés au pluriel sont aussi porteurs de *l'information féminin* (Dubois 1965, p. 82).

... il arrive que *l'information de genre* se maintienne, tandis que celle de nombre s'efface. ... *l'information féminin* a disparu des deux phrases (Dubois 1965, p. 83).

- 36 Cet emploi d'*information* fait glisser le terme vers *contenu* ou *sens*. Alors qu'il désignait une quantité abstraite, calculée à l'aide de probabilités et excluant le sens, le terme *information* définit maintenant une signification grammaticale d'ordre morpho-syntaxique et donnée par les marques. Ce glissement de sens correspond à la conception jakobsonienne des *grammatical meanings*.

## 5. Transmission des conceptions harrissiennes dans l'analyse de discours « française »

- 37 On observera qu'aucune des deux interprétations de l'information, qu'elle soit quantitative ou sémantique, ne figure dans les travaux ultérieurs d'analyse de discours ou de lexicographie, bien que les références à Jakobson (aux *Essais* et à la *Distinctive Features Theory*) aient été transmises. Il nous reste à examiner ce qu'il en est de la théorie harrissienne et de sa transmission.
- 38 Dubois ne cite *Discourse Analysis* (Harris 1952) qu'une seule fois dans son article de 1969 « Lexicologie et analyse d'énoncé » alors qu'on peut penser qu'il s'agit d'un texte fondateur pour son approche des textes. De plus, Dubois ne s'en inspire que comme méthode et non comme cadre théorique.

- 39 C'est aussi le cas de ses disciples, comme en témoignent les numéros de *Langages* 23 et *Langue Française* 9, tous deux parus en 1971 et consacrés aux travaux en analyse de discours.
- 40 Comme l'indique la note 1 de l'introduction de Louis Guespin au numéro de *Langages* 23, on compte parmi les disciples de Jean Dubois les chercheurs réunis autour de Jean Dubois et Louis Guilbert à l'occasion de la parution du numéro 9 de *Langue française* : A.-A. Archibald, F. Dubois-Charlier, B. Gardin, J. Guespin-Michel, L. Guespin, J. Guilhaumou, M. Houssin, D. Maldidier, J.-B. Marcellesi, B. Perrier, R. Robin, J. Rony, C. Serrano, J. Touar-Estrada.
- 41 Par ailleurs Marcellesi, dans l'introduction du numéro de *Langue Française* 9, signale que quatre articles sur les cinq que comporte le numéro sont issus d'un séminaire de linguistique ayant eu lieu à Nanterre en 1968-69 ; tous sont inspirés des travaux de Dubois sur les transformations.
- 42 Comme l'indiquent les articles figurant dans ces numéros, les sources de l'analyse de discours de Dubois sont multiples. Outre la méthode distributionnelle, Dubois s'inspire des théories de l'énonciation de Jakobson et de Benveniste. Les auteurs des deux numéros ne citent d'ailleurs Harris que de façon sporadique et sa méthode est souvent médiatisée par la grammaire structurale de Dubois. C'est une méthode qui n'est d'ailleurs pas utilisée seule, mais associée à la méthode lexicologique de Dubois et à un cadre énonciatif inspiré de Jakobson et Benveniste.
- 43 Ainsi, dans son introduction au numéro de *Langue Française* n° 9, intitulé « Linguistique et société » consacré à l'analyse de corpus politiques, Jean-Baptiste Marcellesi cite la thèse de Dubois et sa grammaire structurale à propos des transformations, mais ne cite pas Harris ; Lucile Courdesses analyse les discours de Blum et Thorez en mai 1936 en étudiant le procès d'énonciation (Weinreich et Dubois) et cite les embrayeurs de Jakobson et Benveniste. Elle se situe dans la perspective de la grammaire transformationnelle de Dubois en utilisant la méthodologie de Harris ; dans son analyse des discours de De Gaulle en mai 1958, Denise Maldidier cite les phrases de base de Harris mais pas les classes d'équivalence ; Régine Robin aborde la question de l'apport de la linguistique aux historiens avec « l'analyse du discours fondée sur le modèle harrissien et élargie par la connaissance des règles transformationnelles telle [...] qu'elle a été utilisée par les élèves de Jean Dubois [...] » (Robin 1971, p. 47).
- 44 Dans la bibliographie générale du numéro, trois textes de Harris sont cités, contre neuf de Dubois, trois de Weinreich et quatre de Sapir. L'article de Chauveau est particulièrement intéressant dans la mesure où il justifie la nécessité du recours à une théorie de l'énonciation chez les Européens au nom de la différence de statut et de traitement entre phrase et discours :
- Dans la tradition européenne, les limites imposées à la phrase sont d'ordre théorique, la différence entre la phrase et l'au-delà de la phrase est qualitative : avant, il s'agit de la langue (système de signes) ; au-delà, il s'agit d'autre chose (parole, discours) et les procédures d'analyse ne sont plus identiques (théorie de l'énonciation). Rien de tel dans la tradition américaine où l'énoncé, présenté dans les termes du behaviorisme, est considéré, quelle que soit sa longueur, comme relevant en droit de la linguistique, et descriptible selon des procédures similaires à tous les niveaux (Chauveau 1971, p. 12).

- 45 La discussion porte sur la compatibilité entre méthode harrissienne (Ecole américaine) et énonciation (Jakobson, Benveniste et plus généralement Ecole européenne). Pour Chauveau, elles sont incompatibles.
- 46 Dans le numéro de *Langages* 23 « Le discours politique », sous la direction de Louis Guespin, Jean-Baptiste Marcellesi, Denise Maldidier et Denis Slatka, les auteurs revendiquent également le double parrainage de Harris (et de « l'école américaine ») et de « l'école européenne de l'énonciation » (Roman Jakobson, Emile Benveniste). Ce numéro se veut plus théorique et propose des aménagements de l'analyse distributionnelle et de la conception du discours de Harris. Dans son introduction, Guespin souligne l'importance de l'analyse de contenu, issue de la psychologie sociale et de la documentation automatique, dans les sources de l'analyse de discours française et l'élaboration des concepts de conditions et de processus de production du discours. Le lien avec les Althusériens, Michel Pêcheux et Paul Henry notamment, et leurs conceptions de l'analyse de discours sont ici revendiqués (voir Léon 2014).
- 47 La méthode d'analyse de discours de Harris doit être corrigée par l'analyse d'un corpus constitué par la sélection de vocables en vue de la constitution d'un énoncé fortement récurrent, comme le propose Denise Maldidier (les deux phrases de base qui sous-tendent le fonctionnement du discours politique de la guerre d'Algérie : *l'Algérie est la France* et *l'Algérie dépend de la France*). Enfin il faut souligner que, dans ce numéro - contrairement à celui de *Langue française* - la notion de transformation est davantage référée à Chomsky qu'à Harris. Ainsi, dans la bibliographie générale du numéro, seulement trois textes de Harris sont cités pour six textes de Chomsky (et trois textes de Dubois).

## Conclusion

- 48 Pour revenir à notre questionnement de départ consistant à déterminer dans quelle mesure Jean Dubois peut être identifié comme un passeur des théories linguistiques américaines, on peut avancer les éléments suivants. Par son activité éditoriale, de traduction et de recension, Jean Dubois a certes joué un grand rôle de diffuseur de la linguistique américaine (Harris, Chomsky, Jakobson). Il peut être considéré comme le premier linguiste français à avoir utilisé la méthode distributionnelle harrissienne, ainsi que les significations grammaticales de Boas via Jakobson. Il fait partie également de quelques linguistes, avec Guiraud et Martinet, à avoir utilisé la théorie de l'information. Mais cette dernière semble avoir été abandonnée par ses disciples alors que ce qui semble avoir été transmis de façon claire et unanime, est son intégration de l'énonciation et de la méthode harrissienne au sein d'une conception originale de l'analyse de discours. Chacun à sa façon, Pêcheux en intégrant les conditions de production d'Althusser, et Dubois celles de l'énonciation de Benveniste et de Jakobson, a fondé ce que Louis Guespin (1971, p.15) a dénommé « l'analyse de discours française ».

---

## BIBLIOGRAPHIE

- Carnap, R., Bar-Hillel, Y. (1952). « An outline of a theory of semantic information », Massachusetts Institute of Technology, Research Laboratory of Electronics, Technical Report No 247.
- Chauveau, G. (1971). « Problèmes théoriques et méthodologiques en analyse du discours », in *Langue française*, n° 9, pp. 6-21.
- Cherry, C., Halle, M., Jakobson, R. (1953). « Toward the Logical Description of Languages in their Phonemic Aspect », in *Language*, n° 29, pp. 34-46.
- Chevalier, J-C. et P. Encrevé (2006). *Combats pour la linguistique de Martinet à Kristeva*. Lyon : ENS Éditions.
- Dosse, F. (1991). *Histoire du Structuralisme. I. Le champ du signe 1945-1966*. Paris : Éditions La Découverte.
- Dubois, J. (1960). « Les notions d'unité sémantique complexe et de neutralisation dans le lexique », in *Cahiers de lexicologie*, n° 2, pp. 62-66.
- Dubois, J. (1962). *Le vocabulaire politique et social en France de 1869 à 1972. A travers les œuvres des écrivains, les revues et les journaux*. Paris : Larousse.
- Dubois, J. (1964a). « La traduction de l'aspect et du temps dans le code français (structure du verbe) », in *Le Français moderne*, n° 32-1, pp. 1-26.
- Dubois, J. (1964b). « Compte-rendu de Roman Jakobson *Essais de linguistique générale* Les Editions de Minuit », in *Le Français moderne*, n° 32-4, pp. 303-307.
- Dubois, J. (1965). *Grammaire structurale du français, nom et pronom*. Paris : Larousse.
- Dubois, J. (1967). *Grammaire structurale du français, le verbe*. Paris : Larousse.
- Dubois, J. (1969). *Grammaire structurale du français, la phrase et les transformations*. Paris : Larousse.
- Dubois, J. (1969a). « Lexicologie et analyse d'énoncé », in *Cahiers de Lexicologie*, n° 15, pp. 115-126.
- Dubois, J. (1969b). « Énoncé et énonciation », in *Langages*, n° 13, pp. 100-110.
- Guespin, L. (1971). « Problématique des travaux sur le discours politique », in *Langages*, n° 23, pp. 3-24.
- Hommage à Jean Dubois, *LINX*, n° 34-35 (1996).
- Harris, Z. S. (1951). *Methods in Structural Linguistics*. Chicago: The University of Chicago Press.
- Harris, Z. S. (1952). « Discourse Analysis », in *Language*, n° 28, pp. 18-23. Traduction française : Françoise Dubois-Charlier. 1969. « Analyse du discours », in *Langages*, n° 13, pp. 8-45.
- Harris, Z. S. (1955). « From phoneme to morpheme », in *Language*, n° 31, pp. 190-222.
- Harris, Z. S. (1959). « Linguistic Transformations for Information Retrieval ». in *Proceedings of the International Conference on Scientific Information*. Washington DC 16-21 nov 1958, Washington, National Academy of Sciences, National Research Council, pp. 937-950.
- Hockett, C. F. (1953). « Review: The Mathematical Theory of Communication by Claude L. Shannon and Warren Weaver », in *Language*, n° 29-1, pp. 69-93.

- Jakobson, R. (1971) [1959]. « Boas'view of grammatical meaning » [1959], in *Selected Writings II*. Gravenhage, Mouton, pp. 489-496. Traduction française « La notion de signification grammaticale selon Boas », in *Essais de Linguistique générale*. [1963], pp. 197-206.
- Léon, J. (2008). « Théorie de l'information, information et linguistes français dans les années 1960. Un exemple de transfert entre mathématiques et sciences du langage », in J. Durand, B. Habert, B. Laks (éds), *Actes du Congrès Mondial de Linguistique Française*, Paris, 9-12 juillet 2008, pp. 923-938.
- Léon, J. (2014). "Questioning the Idea of 'Founding Text': Harris's Discourse Analysis and French Analyse du discours" in V. Kasevich, Y. Kleiner & P. Sériot (éds) *History of Linguistics 2011*. Selected Papers from the 12th International Conference on the History of the Language Sciences (ICHoLS XII), Saint-Petersburg, 28 August - 2 September 2011 [SiHoLS 123]. John Benjamins Publishing Company, pp. 117-125.
- Léon, J. (2015). *Histoire de l'automatisation des sciences du langage*. Lyon : ENS Editions.
- Maingueneau, D. (1996). « Jean Dubois et l'analyse du discours en France : quelques réflexions » in *Hommage à Jean Dubois*, in *LINX*, n° 34-35, pp. 27-33.
- Maldidier, D. (1971a). « Lecture des discours de De Gaulle par six quotidiens parisiens : 13 mai 1958 », in *Langue française*, n° 9, pp. 34-46.
- Maldidier, D. (1971b). « Le discours politique de la guerre d'Algérie : approche synchronique et diachronique », in *Langages*, n° 23, pp. 57-86.
- Mandelbrot, B. (1954). « Structure formelle des textes et communication », in *Word*, n° 10-3, pp. 1-27.
- Marcellesi, J-B. (1971). « Présentation », in *Langue française*, n° 9, pp. 3-5.
- Marcellesi, J-B. (1971). « Éléments pour une analyse contrastive du discours politique », in *Langages*, n° 23, pp. 25-56.
- Martinet, A. (1960). *Elements de linguistique générale*. Paris : Armand Colin.
- Miller, G-A. (1956) [1951]. *Langage et communication*. Paris : PUF.
- Robin, R. (1971). « Histoire et linguistique : premiers jalons », in *Langue française*, n° 9, pp. 47-57.
- Shannon, C., Weaver, W. (1949). *The mathematical Theory of Communication*. University of Illinois.
- Sumpf, J., Dubois, J. (1969). « Problèmes de l'analyse du discours », in *Langages*, n° 13, pp. 3-7.
- Togoby, K. (1951). *Structure immanente de la langue française*. Travaux du cercle linguistique de Copenhague.

## NOTES

1. L'entretien de Jean Dubois dans Chevalier (2006) montre bien le rôle important de Dubois dans l'introduction des théories linguistiques américaines en France dans les années 1960.
2. Le premier article de Chomsky publié en France paraît dans une double parution anglais-français de la revue *Diogène* (n° 51 juin-septembre 1965) sur le thème « Problèmes du langage ».
3. Maingueneau (1996) considère que Dubois se montre même critique à l'égard de l'analyse distributionnelle en souhaitant l'enrichir par une étude des processus énonciatifs.
4. Harris (1959), p. 937.

5. Pour Harris, la répétition, au fondement de la construction des classes d'équivalence, apparaît en 1952 dans son article « Discourse Analysis ».
  6. Chez Harris les sous-langages des sciences sont des langages artificiels construits formellement et identiques pour toutes les langues pour une science donnée (voir chapitre 9).
  7. Traduction française [J. Léon] : Le choix d'une forme grammaticale par le locuteur présente à l'auditeur un nombre défini de *bits d'information*. Le caractère obligatoire de cette sorte d'*information* dans tout échange verbal au sein d'une communauté linguistique donnée et la différence considérable d'*information grammaticale* transmise selon les diverses langues ont été appréhendées par Franz Boas, grâce à sa stupéfiante compréhension des configurations sémantiques multiformes du monde linguistique ... Il était clair pour Boas que toute différence de catégorie grammaticale porte une *information sémantique*.
  8. Pour plus de détails, voir Léon (2015) chap. 4.
  9. Dans l'article, Dubois emprunte d'autres termes au schéma de communication de la théorie de l'information : message, encodage par le locuteur, décodage par l'auditeur.
- 

## RÉSUMÉS

Dans le contexte intellectuel français du début des années 1960, où la linguistique américaine, en particulier sa mathématisation, est encore largement méconnue, Jean Dubois a largement contribué à l'introduction en France de l'analyse distributionnelle et de l'analyse de discours harrisiennes, de la théorie de l'information, enfin de la grammaire générative. Dans cet article, nous retracerons les premiers travaux de Jean Dubois des années 1960-70 ayant contribué à cette diffusion, en tentant d'évaluer son rôle en tant que passeur ou bien plus simplement de facilitateur. Dans cette perspective, nous examinerons sa thèse sur le lexique où l'analyse distributionnelle est appliquée pour la première fois en France, ses travaux inspirés des travaux de la conception de Roman Jakobson de la théorie de l'information, enfin ses travaux sur l'analyse de discours largement inspirés par ceux de Zellig S. Harris ; travaux alimentés et complétés par son important travail d'édition des traductions des travaux américains. Afin d'évaluer son rôle de passeur, nous étudierons la façon dont ses disciples immédiats ont repris ou poursuivi ses travaux dans le domaine de l'analyse de discours.

In the French intellectual context of the early 1960s, where American linguistics, in particular its mathematization, is still largely unknown, Jean Dubois largely contributed to the introduction in France of Harris's distributional analysis and discourse analysis, of information theory and generative grammar. In this article, we will examine how Jean Dubois's early works in the 1960-70s contributed to this dissemination, and try to assess his role in the introduction of American linguistic theories in France. In this perspective, we will examine his PhD dissertation on the lexicon where distributional analysis was applied for the first time in France, his work inspired by Roman Jakobson's conception of information theory, finally his work on discourse analysis largely inspired by Zellig S. Harris; those works were supplemented by his intense activity of publishing translations of American works. In order to assess his role in the introduction of American linguistic theories in France, we will study the way in which his immediate disciples resumed or continued his work in the area of discourse analysis.

## INDEX

**Mots-clés** : histoire de la linguistique, linguistique américaine, analyse du discours, analyse distributionnelle, théorie de l'information, Harris, Jakobson

**Keywords** : History of linguistics, American linguistics, Discourse analysis, distributional analysis, information theory, Harris, Jakobson

## AUTEUR

**JACQUELINE LÉON**

CNRS UMR7597 Histoire des Théories Linguistiques, Université de Paris