

*Version de travail*

*Paru dans*

Léon Jacqueline, 2004, “*Lexies, synapsies, synthèmes: le renouveau des études lexicales en France au début des années 1960*” *History of Linguistics in Texts and Concepts ~ Geschichte der Sprachwissenschaft in Texten und Konzeptionen*, Gerda Hassler (ed.) Münster: Nodus Publikationen : 405-418.

*Lexies, synapsies, synthèmes: le renouveau des études lexicales en France au début des années 1960.*

### ***Abstract***

This paper aims to show how language automatization, more precisely machine translation, contributed to renew researches on lexicon and syntax in France in the 1960s. The new inquiries focused on the study of word groups, traditionally handled by grammarians by the means of composition and derivation, and ignored by structuralist linguists, as compound lexical units. The idea of translating texts by electronic machines lead linguists to question the status of units which have to be simultaneously segmentation, syntactic, semantic and translation units. In other words, graphic forms, the only units recognizable by machines, had to coincide with syntactic and semantic units. Criteria and procedures then developed had to be relevant to identify units in texts and to stock them in dictionaries.

In France, linguists, starting from Saussure’s syntagm, had to reconsider ‘word’ and word groups as units and were lead to define syntactic units greater than words and to consider them not only from their internal construction but from their relations inside utterances. The terms which were coined for this purpose, *lexies* by Pottier, *synapsies* by Benveniste, *synthèmes* by Martinet, expressed original views on the criteria of identification, construction and classification of these units. A new conception of lexicology emerged from questioning the morphological and/or syntactical grounds of compounds, and from considering their status as discursive units.

### ***Introduction***

Cet article a pour objectif de montrer comment les débuts de l’automatisation du traitement du langage, et en particulier les travaux en traduction automatique, ont suscité, en France, un renouveau de réflexion sur le lexique et la syntaxe dans les années 1960. Ce renouveau s’est notamment fondé sur une approche des groupes de

mots graphiques en termes d'unités lexicales composées. Ces groupes de mots, traités au travers des opérations traditionnelles de dérivation et composition par les grammairiens, n'avaient guère suscité jusqu'alors l'intérêt des linguistes structuralistes. L'objectif consistant à vouloir traduire des textes par une machine électronique a conduit les linguistes à s'interroger de façon originale sur le statut des unités traitées. Ils ont été ainsi amenés à définir des unités syntaxiques supérieures au mot et à les appréhender non seulement du point de vue de leur mode de construction interne mais aussi de leurs rapports avec le reste de l'énoncé. Les termes forgés à l'occasion de ces réflexions, *lexies* chez Bernard Pottier (né en 1924), *synapsies* chez Emile Benveniste (1902-1976) et *synthèmes* chez André Martinet (1908-1999), traduisent des interrogations inédites sur les critères d'identification, de construction et de classement ces unités. On voit émerger au travers de questionnements sur l'ancrage morphologique et/ou syntaxique des mots composés, et sur leur statut discursif, un renouveau de la lexicologie.

## **1. Le mot et les linguistes français dans les années 1950-60**

### *1.1 Mot et syntagme*

Curieusement, c'est par une interrogation sur le 'mot' en tant que tel, notion bannie comme éminemment suspecte par les structuralistes, que débute cette réflexion spécifiquement française sur le lexique.

Il faut rappeler que, chez les structuralistes, la conception du groupe de mots, ou unités lexicales composées, est issue de la notion de syntagme chez Ferdinand de Saussure (1857-1913). Dans le Cours de linguistique générale, (chap 5: les rapports syntagmatiques et rapports associatifs), Saussure désigne par syntagme "des combinaisons de mots dans le discours, fondées sur le caractère linéaire de la langue, qui exclut de prononcer deux éléments à la fois." (pp.170-171) Le syntagme désigne des unités consécutives (mots composés, dérivés, membres de phrase, phrases entières) allant du terme *re-lire* à une phrase entière à plusieurs propositions comme *s'il fait beau* ou *nous sortirons*. Donc la définition est vaste et le critère de définition repose essentiellement sur le caractère syntagmatique des rapports entretenus entre ces unités consécutives dans son opposition aux rapports associatifs.

Ce qui intéresse Saussure c'est moins de définir la structure et les frontières du syntagme que de statuer sur son appartenance à l'ordre de la langue ou à celui de la parole. Ce qui milite en faveur de la parole, c'est le fait qu'une phrase, unité de parole par excellence, est un syntagme. Contre la parole, c'est que les unités de parole sont, par définition, libres, et qu'on trouve des syntagmes qui ne sont pas libres. En

particulier un certain nombre de locutions, comme *prendre la mouche* sont figées et sont des unités de langue, de même que les unités construites de façon régulière comme *impardnable, infatigable* etc. L'identification de ces syntagmes non libres comme unités de langue, constitue donc tout ce qui est dit sur les unités composées, dans ces débuts structuralistes.

## 1.2 *Mot et mécanisation du vocabulaire*

Si l'on revient au début des années 1960, il faut noter que le renouveau de la réflexion sur le mot et les mots composés, est issu des premiers traitements automatiques du langage, traitements qui ont pris en France une forme tout à fait singulière par rapport aux autres pays. En effet, alors que la France se rallie tardivement à la TA, en 1959-60 - les premiers travaux avaient commencé plus de dix ans auparavant en Grande-Bretagne et aux USA - il est intéressant de noter que les premiers traitements automatiques ont concerné le traitement du vocabulaire et cela dès les années 50. Cet intérêt continu pour le vocabulaire<sup>1</sup>, initié par Mario Roques (1875-1961) et son *Inventaire Général de la Langue Française* en 1936 (voir Chevalier & Encrevé, 1984) a été jalonné par un colloque à Strasbourg en 1957 intitulé *Lexicologie et lexicographie françaises et romanes* qui a abouti à la création du TLF en 1960. Lors de ce colloque, a été évoqué explicitement l'apport prometteur des machines mécanographiques et électroniques dans l'accélération des dépouillements et des classements du lexique.

Cet essor de la lexicologie par sa mécanisation est également marqué par la création, en 1959 à Besançon, du Laboratoire d'analyse lexicologique et des deux revues, *Les Cahiers de Lexicologie* et *Les Etudes de Linguistique Appliquée*, tous trois sous la direction de Bernard Quemada (né en 1926), suivie par un colloque international sur la mécanisation des recherches lexicologiques qui a eu lieu en 1961, toujours à Besançon, et d'un second colloque à Strasbourg en 1964 intitulé *Statistiques et analyse linguistique*.

Ainsi, peut-on dire qu'en France, le traitement automatique du langage commence par la mécanisation du lexique qui s'inscrit en droite ligne dans les préoccupations de nombre de linguistes de l'époque, à savoir l'étude du vocabulaire français et la stylistique.<sup>2</sup>

---

<sup>1</sup> Dans la *Grammaire Larousse du français contemporain* de Chevalier, Blanche-Benveniste et Peytard de 1964, les processus de formation, composition et dérivation sont traités dans le chapitre intitulé 'le vocabulaire français'.

<sup>2</sup> Certains ont voulu se démarquer de cette filiation. En 1969, Bernard Vauquois (1929-1985), alors directeur du Centre de Traduction Automatique de Grenoble et président de l'ATALA, jugeant trop vaste le domaine couvert par la *computational linguistics* américaine, décide d'exclure "les travaux de linguistes conduits par des linguistes et qui utilisent des ordinateurs à titre d'outil: analyse de concordances, calculs de fréquence de mots, recherches lexicographiques, confection de dictionnaire à consultation automatique, et surtout grammaires génératives (sic)" et de renommer "traitement

## 2. Les lexies comme unités syntaxiques pour la traduction automatique (Pottier, 1962 a, b, c)

Toutefois c'est bien avec la traduction automatique (TA infra) qu'une certaine lexicologie 'scientifique' a pris son essor. Avec la TA, les linguistes structuralistes vont être amenés à reconsidérer le statut des groupes de mots. L'objectif consistant à faire traduire un texte par la machine les contraint à définir des unités, qui soient à la fois des unités de segmentation du texte, unités de langue, unités de sens et unités de traduction.<sup>3</sup> Il s'agit de faire coïncider forme graphique, unité syntaxique et unité sémantique. Les critères de définition doivent permettre de reconnaître ces unités dans un texte et de les stocker dans un dictionnaire.

C'est bien à cause de la machine qui ne reconnaît que les mots graphiques, à savoir des suites de lettres entre deux blancs, que la réflexion sur le mot et les groupes de mot va commencer en France. En témoigne le colloque intitulé 'le mot' organisé en 1962 par l'ATALA, association pionnière en France dans le domaine du traitement automatique du langage, et qui regroupe des linguistes, des mathématiciens, des ingénieurs, des documentalistes et des traducteurs. Parmi les linguistes, on trouve Bernard Pottier, seul linguiste de formation littéraire, avec Guy Bourquin, qui se soit investi réellement dans des projets concrets de TA en fondant notamment le Centre de Nancy en 1960.<sup>4</sup>

Dans ses premiers articles sur la traduction automatique, Pottier (1962, a, b, c) propose la *lexie* comme unité lexicale et syntaxique pour l'analyse automatique des textes.

Tableau 1. Pottier (1962 a b c)

Lexies: éléments fondamentaux en langue de la construction syntaxique					
Lexie simple	Lexie composée	Lexie complexe		lexie?	groupe de mots
Disponible en langue	Disponible en langue	Disponible en langue		Zone d'incertitude	Réunion fortuite en discours
		Groupements motivés	Groupements arbitraires		

---

automatique des langues" – à vrai dire sans succès à l'époque - les seules recherches menées conjointement par des linguistes et des mathématiciens traitant de traduction automatique, communication homme-machine, enseignement programmé, documentation automatique ou études de formalisation sémantique.

<sup>3</sup> Voir Mounin (1964) pour un état des lieux des unités d'analyse utilisées dans les méthodes de traduction automatique jusqu'en 1961.

<sup>4</sup> Toutefois, il faut noter que les linguistes qu'on pourrait qualifier "d'institutionnellement établis", s'ils ne sont pas directement acteurs du domaine, s'y intéressent de très près. Certains sont membres fondateurs de l'ATALA comme Marcel Cohen (1884-1974). D'autres font partie des comités d'évaluation des centres de TA. C'est le cas, pour ne citer que les plus prestigieux, d'Emile Benveniste, Georges Gougenheim (1900-1972), Michel Lejeune (1907-2000), André Martinet, Bernard Quemada ou Jean Fourquet (1898-2001). Sur les débuts de la traduction automatique en France, voir Léon (1998).

<i>chaise</i>	<i>Sous-chef</i> <i>Cheval-vapeur</i>	<i>Pomme de terre</i> <i>Fondé de pouvoir</i> <i>Cheval de frise</i>  <i>prendre la mouche</i>	<i>Il n'y a [] qui</i> <i>tienne</i>  <i>Œil pour œil, dent</i> <i>pour dent</i>	<i>Plaque</i> <i>minéralogique</i>  <i>Crise de</i> <i>croissance</i>	<i>Le cheval de</i> <i>Jean</i>
	Valeur de l'unité = valeur fonctionnelle de la hiérarchie la plus haute	Valeur de l'unité = valeur fonctionnelle de la hiérarchie la plus haute [notion de tête ?] <i>Plaque tournante: S + A = S</i>			
		Critère sémantique: unicité de l'objet désigné sous une forme complexe			
		critère statistique: degré de lexicalisation			
		Critères fonctionnels: non- séparabilité: <i>*pomme ronde de terre</i> non qualification et non quantification: <i>* prendre la grosse mouche</i> <i>* une plaque peu minéralogique</i> pas de prédication: <i>* cette plaque est minéralogique</i> traduction: dans une autre langue, il peut y avoir un seul mot pour traduire une lexie complexe <i>cheval de course = corcel (esp.)</i>		<i>*Le cheval</i> <i>alezan de</i> <i>course</i>  <i>? plaque en fer</i> <i>minéralogique</i>  <i>? crise aiguë</i> <i>de croissance</i>	<i>Le cheval</i> <i>alezan de Jean</i>

A la fois unités de langue, unités sémantiques, unités syntaxiques et unités de traduction, les lexies peuvent être simples (*chaise*), composées (*sous-chef*, *cheval-vapeur*) ou complexes (*pomme de terre*, *prendre la mouche*). Elles ne sont jamais inférieures à l'unité empirique que constitue le mot-graphique.

Ce sont les lexies complexes qui posent problème à la TA, puisque les autres, simples et composées sont graphiquement identifiables par la machine. Il est en effet souvent difficile de les distinguer des groupes ou associations libres de mots dont la réunion, dit Pottier, est fortuite en discours. On voit bien ici que c'est Saussure et sa définition du syntagme qui fait référence.<sup>5</sup> De fait, les critères (voir tableau 1) permettant de définir les lexies complexes laissent des zones d'incertitude. Outre le critère sémantique référant à une désignation stable et à une unicité de l'objet assurée par ce type d'unité lexicale, déjà également proposé par Bally, les critères les plus importants, que Pottier désigne comme fonctionnels, sont d'ordre syntaxique. En effet ils signent l'impénétrabilité de l'unité par toute forme d'insertion. Ainsi les critères de non séparabilité, de non qualification et non quantification auxquels il faut ajouter une contrainte logico-sémantique, le critère de non-prédicativité\* *cette plaque est minéralogique*.

<sup>5</sup> Il est à noter que déjà Charles Bally (1865-1947), en 1932, parlait du syntagme libre et parlait de fractionnement de faux syntagmes pour aborder les phénomènes de composition.

Un autre critère, lié à l'objectif de traduction, vient de la comparaison inter-langues: une lexie complexe est susceptible de se traduire par une lexie simple dans une autre langue. C'est le cas de *cheval de course* et son équivalent *corcel* en espagnol.<sup>6</sup>

Entre les groupements de mots dont les éléments sont réunis fortuitement dans le discours, les syntagmes libres de Saussure, comme *le cheval de Jean* et une lexie composée comme *pomme de terre*, il y a des zones d'incertitude comme *crise de croissance* où le critère de non-séparabilité n'est pas aussi évident: on peut avoir *crise aiguë de croissance*.

Pottier appréhende cette zone d'incertitude par des critères statistiques qui déterminent un degré de lexicalisation: *pomme de terre* est plus lexicalisé que *crise de croissance*. Bally (1932) avait déjà abordé la question de la lexicalisation, d'une part en invoquant les différents degrés de blocage des composés et en tant que procès de "dégrammaticalisation" qui sape la grammaire au profit du lexique.

Dans la procédure de TA, Pottier propose de stocker les lexies complexes dans un dictionnaire bilingue qui sera appliqué au texte à traduire. Ce qui suppose que ces lexies complexes ne sont pas reconnues automatiquement dans le texte et donc qu'on ne fait pas appel à l'analyse automatique de leur structure interne pour les identifier. Pourtant, Pottier mentionne la structure interne de ces lexies complexes en termes de catégories (subst, adj ou V) où il est intéressant de voir émerger un début de traitement syntaxique. Pottier introduit en effet une catégorisation qui s'apparente à la notion de tête en grammaire syntagmatique: *plaque tournante* est catégorisée comme substantif, parce que c'est la catégorie hiérarchiquement supérieure (dans le syntagme). Ainsi S + A = S. Ainsi, comme Bally, Pottier se situe dans la lignée de Saussure pour définir les mots composés à partir de la notion de syntagme libre. Comme lui, il utilise le critère sémantique d'unicité de l'objet. Comme lui enfin, il distingue différents degrés de lexicalisation entre syntagmes complètement libres et unités figées. Ce qui est nouveau chez Pottier, en partie grâce à la TA, c'est une conception unificatrice des unités lexicales. Les lexies, simples, composées ou complexes, sont à la fois des unités syntaxiques, des unités de sens et des unités de traduction. Ceci explique pourquoi Pottier s'intéresse moins aux aspects morphologiques des lexies qu'à leur mode de construction et de catégorisation syntaxique.

---

<sup>6</sup> Cette question, essentielle pour la TA, a été évoquée la première fois par Yehoshua Bar-Hillel (1915-1975) dans un article intitulé "Idioms" et paru dans le premier recueil de travaux en TA en 1955 aux USA. Inédite dans un contexte structuraliste américain où le mot n'existe pas, la question des unités lexicales composées se pose en TA quand le nombre d'unités formant le composé varie selon les langues et quand le sens ne peut être traduit de façon univoque. Par exemple *red herring* en anglais doit être traduit par *fausse piste* en français et par *Finte* en allemand. Il est à noter que Algirdas Greimas (1917-1992) dans un article de 1960, se réfère à Bar-Hillel et reprend la question de la traduction des expressions idiomatiques, les idiotismes, en évoquant la nécessité d'élaborer un modèle sémantique pour comparer les langues au niveau de leurs signifiés (voir Léon, 2001).

## 2. Synapsies et structure syntaxique interne des unités composées (Benveniste, 1966, 1967)

Dans des articles postérieurs de quelques années à ceux de Pottier, Benveniste (1966, 1967) identifie des *synapsies*, qu'il définit comme des groupes de lexèmes formant une unité nouvelle à signifié unique et constant.

Tableau 2. Benveniste (1966, 1967)

conglomérés	synapsie
<i>va-nu-pied</i> <i>meurt-de-faim</i>	<i>pomme de terre</i> <i>modulation de fréquence</i>
<i>désormais</i> <i>naguère</i>	<i>gardien d'asile de nuit</i>
<i>justaucorps</i> <i>gendarme</i>	<i>hélice à pas variable</i> <i>moteur à refroidissement par air</i> <i>aigle pêcheur à tête blanche</i>
Construction complexe qui se soude en bloc	groupe entier de lexèmes, reliés par divers procédés, et formant une désignation constante et spécifique. ≠ du syntagme (n'importe quel groupement opéré par des moyens syntaxiques)
	1. <b>nature syntaxique</b> (non morphologique) de la liaison entre les membres 2. emploi de joncteurs <i>de</i> et <i>à</i> 3. ordre déterminé / déterminant des membres 4. forme lexicale pleine et choix libre de tout substantif ou adjectif 5. absence d'article devant le déterminant 6. possibilité d'expansion pour l'un ou l'autre membre 7. caractère unique et constant du signifié
	1. et 6. ( <i>gardien (d'asile de nuit)</i> ) est la conversion nominale de l'énoncé prédicatif <i>il garde un asile de nuit</i>
	2. fonction des joncteurs <i>à</i> et <i>de</i> Destination : <i>salle à manger, service à café</i> caractéristiques distinctive : <i>bête à cornes</i> agent moteur : <i>moulin à vent, machine à vapeur</i> partie d'un tout virtuel : <i>verre de montre, pied de biche</i> circonstance : <i>chemise de nuit, fusil de chasse</i>
	Aspects technologique et discursif: terminologie scientifique Articulation syntaxe / discours

Les synapsies sont clairement distinguées des formes graphiquement soudées, qu'il nomme "conglomérés" (*va-nu-pied, justaucorps*) et des syntagmes qui désignent n'importe quel groupement opéré par des moyens syntaxiques. Les définitions de ces différentes unités paraissent sensiblement proches des lexies composées et complexes de Pottier.

Toutefois, moins préoccupé de catégorisation dans l'énoncé que Pottier, Benveniste s'intéresse davantage à la structure interne des synapsies qu'il considère comme d'ordre exclusivement syntaxique. Ce statut syntaxique est développé au travers des critères 2 à 6 qui définissent les synapsies (voir tableau 2). On notera toutefois que le critère 7, concernant le caractère unique et constant du signifié, est de nature sémantique et qu'il est aussi utilisé par Bally et Pottier.

L'argument proposé par Benveniste pour fonder la nature syntaxique des synapsies repose sur leur rapport avec le syntagme propositionnel. La fonction du composé

consiste à transférer au virtuel le rapport actuel de prédication énoncé par la proposition (Benveniste 1967). Une synapsie est issue de la conversion nominale d'un énoncé prédicatif : *gardien d'asile de nuit* est issu de l'énoncé prédicatif *il garde un asile de nuit*.

Cette proposition n'est toutefois pas nouvelle. C'est l'argument proposé par Arsène Darmesteter (1846-1888) dans son *Traité de la formation des mots composés* dont la première édition date de 1875. Darmesteter propose une conception syntaxique de la composition, qu'il oppose à la formation des mots par juxtaposition: " c'est à la syntaxe qu'appartient la composition, et sa théorie rentre tout entière dans celle de la construction de la phrase " (1875 [1967], p.5).<sup>7</sup>

Tableau 3. *Darmesteter (1875, 1877)*

Composition apparente juxtaposition	composition à l'aide de particules (adv ou prep)	Composition propre, de dépendance
réunion de deux ou plusieurs termes qui ont été joints l'un à l'autre suivant les règles ordinaires de la syntaxe, sans ellipse, et qui, avec le temps, et par la force de l'usage, ont fini par se souder		la composition est une union intime de mots dont la combinaison présente à l'esprit une idée nouvelle que ne fournissent pas les éléments composants pris à part
Sans ellipse		Mode de construction syntaxique avec ellipse
<i>Plafond</i> <i>Vinaigre</i>	Verbe + adv <i>sur-prendre</i>	Par apposition <i>Chou-fleur (chou qui est une fleur)</i>
<i>Arc-en-ciel</i> <i>Char-à-banc</i>	Adv + adj <i>bien-heureux</i> Adv + subst <i>arrière-cour</i>	Prep + subst/ inf régi par cette prep <i>A compte, pourboire</i>
Subst + subst /verbe L'un régissant l'autre : <i>Pomme de terre</i> <i>Sergent de ville</i> <i>Machine à coudre</i>	Partic + v + suffixe + suffixe v <i>Encourag-er</i> = parasynthétiques verbaux partic + subst + suffixe nom <i>em-place-ment</i> = parasynthétiques nominaux	Génitif ou datif ; subst + subst où l'un régit l'autre <i>Timbre-poste</i> Subst + v qui le régit <i>Colporteur</i> V (à l'impératif + subst régi) <i>Porte-plume, serre-livre</i>

Les mots composés sont des propositions en raccourci. L'opération principale de la composition est l'ellipse des rapports syntaxiques (subordination, coordination, complément avec préposition etc.). Par exemple, *timbre-poste* est l'ellipse de la préposition dans *timbre de poste* où les noms sont réunis par un rapport de subordination. Rappelons que la formation la plus célèbre et la plus discutée à l'époque est celle opérée par un verbe à l'impératif et son complément. Par exemple *portefeuille* est formé sur le modèle du *tiens-toi-bien* (petit chariot dans lequel l'enfant apprend à marcher dans le Jura) et du *revenez-y*.

<sup>7</sup> Toute définition sémantique n'est toutefois pas complètement exclue: "La composition est une union intime de mots dont la combinaison présente à l'esprit une idée nouvelle que ne fournissent pas les éléments composants pris à part." (1877, chap 10)

Pour Benveniste, ce que garantit la nature syntaxique des synapsies c'est leur expansion et donc leur productivité. Cette nature syntaxique autorise en effet les expansions (critère 6), comme (*gardien d'asile de nuit*) ce qui est tout à fait novateur dans les processus de formation par composition (voir David 1993).

C'est pourquoi les synapsies constituent la forme de composition de base dans les nomenclatures techniques. En mettant en oeuvre les mécanismes réguliers de la syntaxe, les synapsies constituent un mode très productif de création de nouveaux composés, et ceci contrairement à ce qui passe pour les autres composés en français qui se forment à un rythme très lent. Tous les vocabulaires techniques y font appel dans la mesure où, grâce à leur extrême flexibilité paradigmaticque, elles permettent la spécification détaillée du déterminé et la classification des séries par leur trait distinctif. Benveniste cite comme exemple des termes appartenant au vocabulaire de l'aviation et tirés de l'ouvrage de Guilbert (*Formation du vocabulaire de l'aviation*, 1965): *volet de courbure à fente, hélice à pas variable, moteur à refroidissement par air*.

On retrouve là encore la position de Darmesteter, pour qui la composition de dépendance est un mode de formation nouveau qu'imposent les nécessités du commerce et de l'industrie : "Ceux-ci ont besoin d'expressions courtes, dégagées autant que possible de prépositions gênantes, et le plus souvent inutiles. Ils trouvent l'apposition qui leur fournit un cadre tout préparé, et se l'approprient au point de la transformer." (Darmesteter, 1877, p.160).

Il faut noter que Benveniste ne propose pas une organisation précise du groupe nominal dans la phrase et que les critères de définition des synapsies fonctionnent davantage comme des traits de reconnaissance a posteriori que comme des règles de construction de ces unités. Comme pour Pottier, ce sont des développements technologiques, traduction automatique pour l'un, nomenclature technique pour l'autre qui ont contribué à l'identification de ces nouvelles unités. Dans les deux cas, on est en présence d'une approche non morphologique de certains processus de composition.

#### **4. Mot, syntagme, synthèse (Martinet, 1960, 1967, 1968)**

Martinet a été conduit, lui aussi, à partir du milieu des années 1960, à reconsidérer les problèmes de la composition, et à penser des unités supérieures à ses unités minimum de première articulation, les monèmes, et qui soient distinctes du syntagme saussurien. On considère généralement que Pottier a forgé son terme de lexie à partir du lexème de Martinet (Fruyt, 1992). On peut montrer qu'il n'en est rien et qu'au contraire, s'il s'agit d'établir des priorités, l'idée de s'intéresser aux unités lexicales composées vient de Pottier et plus généralement de la TA.

En 1965, Martinet publie l'article "Mot" dans la revue Diogène. Sous ce titre qui se dit explicitement provocateur, Martinet introduit, associée à la notion de mot, l'idée de critères permettant pour toute langue et dans tous les cas, d'identifier et de délimiter un segment de la chaîne comme un mot déterminé; le vrai problème étant celui de savoir si les segments isolables qu'on désigne comme des mots correspondent à une réalité linguistique bien déterminée. Ces questions, on l'a vu, sont celles que se posent à l'époque les linguistes impliqués dans la TA et on sait que Martinet n'est pas ignorant du domaine (voir note 4). Il évoque d'ailleurs les méthodes utilisées dans la TA, notamment la méthode de détection statistique des morphèmes préconisée par Harris (1955) issue de la théorie de l'information. Enfin, il critique explicitement la TA qui ne s'intéresse qu'au texte écrit, dit-il, et à sa segmentation en oubliant l'énoncé oral.

Si, par ailleurs, on remonte aux premiers travaux de Martinet qui ne traitent pas exclusivement de phonologie, c'est-à-dire aux *Eléments de Linguistique Générale*, dont la première édition date de 1960, on s'aperçoit que Martinet ne manifeste aucun intérêt particulier ni pour le mot, ni pour les mots composés.

**Tableau 4: Martinet (1960)**

monème	syntagme autonome		syntagme de composition et de dérivation
- lexème - radical - affixe - désinence	combinaison de deux ou plus de deux monèmes dont la fonction ne dépend pas de sa place dans l'énoncé	syntagme autonome formé de monèmes non séparables et étroitement unis par le sens = "mot"	
	<i>en voiture</i> <i>avec mes valises</i>	<i>rosarum</i> syntagme amalgamé de 3 monèmes : . <i>ros-</i> : monème lexical . <i>-arum</i> : amalgame des sants gén et plur	les modalités ne peuvent porter sur un seul élément: <i>chaise-plus-longue</i> <i>sacs à main /sakamè/</i>

Martinet désigne par syntagme autonome les regroupements de monèmes, groupes de mots (*en voiture, avec mes valises*) ou formes fléchies (*rosarum*). Cette dernière forme est composée du regroupement d'un radical lexical *ros-*, qu'il nomme monème lexical, et d'un amalgame *-arum* de deux monèmes, le génitif et le pluriel. Cette analyse en amalgame permet, selon Martinet, de distinguer le monème du morphème des structuralistes américains dont la définition reste linéaire.

A côté du syntagme autonome, regroupant des unités disparates, Martinet signale l'existence de syntagmes formés par composition ou par dérivation. Il utilise le critère d'inséparabilité *\*chaise plus longue* pour déterminer qu'il s'agit d'une seule unité ou bien des critères formels: *sac à main* n'est qu'un seul mot parce qu'au pluriel *sacs à main* se dit /sakamè/ et non /sakzamè/. *Cheval de bois* en revanche contient trois mots puisqu'au pluriel on dit *chevaux de bois*.

Ce n'est que dans les éditions des *Éléments* postérieures à 1970 que le syntème, proposé dans l'article de 1967, apparaît. Martinet rebaptise d'ailleurs " syntème " le chapitre IV sur la composition et la dérivation.

Tableau 5. *Martinet (1967, 1968)*

monème	syntagme / syllème	syntème
- lexème - radical - affixe - désinence	combinaison de monèmes libres  syntagme formé d'un noyau et de ses modalités fonctionnelles	combinaison de monèmes conjoints  tout signe formé de deux ou de plus de deux éléments sémantiquement identifiables et qui se comportent syntaxiquement comme les signes minima avec lesquels il commute
	1986 : ensemble formé par un noyau, ses déterminants et, éventuellement, le connecteur qui relie cet ensemble au reste de l'énoncé. La phrase avec son noyau prédicatif est un syntagme sans connecteur.	tout ce qui relève de la dérivation, de la composition et du figement  Le terme 'syntagme' selon Saussure recouvre les syntagmes/syllèmes et les syntèmes
	<i>donnerons</i> <i>dominorum</i> <i>sur la table</i> <i>il le lui dira</i>	<i>lavage</i> <i>entreprendre</i> <i>indésirable</i> <i>pomme de terre</i>  <i>*chaise-plus-longue</i> mais <i>ministre italien du commerce</i>

Le syntème, unité syntaxique, regroupe toutes les formes construites par dérivation, composition ou figement. Ce sont des groupements de monèmes conjoints, alors que le syntagme, appelé aussi syllème par Martinet, est une combinaison de monèmes libres. Le critère fonctionnel de non-séparabilité n'est pas obligatoire pour définir le syntème. Ainsi est accepté comme syntème *le ministre du travail* qui accepte l'insertion *le ministre italien du travail*.

Au travers de cette hétérogénéité de critères, morphologiques, flexionnels et dérivationnels, et des critères d'ordre syntaxique non obligatoires, on voit que Martinet ne s'intéresse pas véritablement au mode de construction des unités que recouvre le terme de syntème. En particulier, les syntèmes ne permettent pas d'identifier un niveau spécifique du lexique distinct du niveau morphologique. Il reste proche de la conception structuraliste selon laquelle toutes les unités de langue (mots, syntagmes, phrases) sont organisées de manière homogène et uniforme: ce sont des monèmes ou des combinaisons de monèmes. En revanche comme les unités de Bally et Pottier, les syntèmes sont des unités syntaxiques à signifiant stable dans l'énoncé.

## **Conclusion**

Pour conclure, on notera que ces travaux, stimulés par des développements technologiques comme la traduction automatique et la terminologie scientifique et technique, sont à la base de développements inédits dans le domaine de la syntaxe et du lexique. Le fait que ces unités sont souvent des termes de spécialité et que leur reconnaissance fait appel à un savoir extra-linguistique les rend sensibles à une approche discursive qui s'inscrit dans une tradition spécifique d'études du vocabulaire. C'est ainsi qu'on peut dire que ces études débouchent sur une lexicologie ancrée dans le discours, où un début de recherche sur les modes de construction interne des unités sont associées à leur fonction syntaxique dans l'énoncé et à l'étude de leur fonction référentielle dans le discours.

Il faut toutefois reconnaître que les traits permettant d'identifier les différents types d'unités lexicales sont loin de faire consensus chez les auteurs comme en témoigne le tableau 6 : alors que Martinet a une définition large du syntème, les critères de Pottier pour la lexie, d'Henri Mitterand (né en 1928) pour les composés et surtout de Benveniste pour la synapsie sont beaucoup plus restrictifs. De plus les unités définies sont loin d'être opérationnelles.

Tableau 6

	<i>Pomme de terre</i> <i>*pomme ronde de</i> <i>terre</i>	<i>Vendeur de voitures</i> <i>*vendeur sérieux de</i> <i>voitures</i>	<i>Mur du son</i>	<i>Ministre du commerce</i> <i>ministre italien du</i> <i>commerce</i>
Pottier	Lexie complexe	Lexie complexe	Lexie complexe	-
Benveniste	synapsie	synapsie	-	-
Martinet	syntème	syntème	syntème	syntème
Mitterand	Composé	-	composé	-

A l'heure actuelle, la question de la définition des unités composées et de leur reconnaissance par la machine est loin d'être résolue et le débat est encore vif dans les travaux en morphologie, qu'ils soient ou non orientés vers le traitement automatique sur la question de l'ancrage morphologique et/ou des unités lexicales composées : par exemple, à côté du tout syntaxique préconisé par Maurice Gross (1934-2001) et son équipe (Gross, 1988), d'autres (David, 1993) défendent l'idée que seules les unités polylexicales relevant de la composition sont du ressort de la syntaxe.

### **Bibliographie**

- Bally Charles. 1965 (1ère édition 1932). *Linguistique générale et linguistique française*. Berne: Francke.
- Bar-Hillel Yehoshua. 1955. "Idioms". *Machine Translation of Languages, 14Essays*. ed. by W.N.Locke & A.D. Booth, 183-193. MIT & John Wiley.
- Benveniste Emile. 1966. "Formes nouvelles de la composition nominale". *BSL* 61.82-95.
- Benveniste Emile. 1967. "Fondements syntaxiques de la composition nominale". *BSL* 62.15-31.
- Chevalier Jean-Claude & Encreve Pierre. 1984. "La création de revues dans les années 60. Matériaux pour l'histoire récente de la linguistique en France". *Langue Française* 63.57-102.

- Darmesteter Arsène. 1875 (2ème ed 1893) [1967]. *Traité de la formation des mots composés dans la langue française comparée aux autres langues romanes et au latin*. Paris: Honoré Champion.
- Darmesteter Arsène. 1877. *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent*. Genève: Slatkine reprints 1972.
- David Sophie. 1993. *Les unités nominales polylexicales*. Thèse de Doctorat, Université Paris 7.
- Fruyt Michèle. 1992. "Le mot: aperçu théorique et terminologique. Peut-on donner une définition théorique du mot? La polysémie de mot et ses conséquences". *Lalies* 10.113-124.
- Greimas A.J. 1960. "Idiotismes, proverbes, diction". *Cahiers de lexicologie* 2.41-61.
- Gross Maurice. 1988. "Les limites de la phrase figée". *Langages* 90:7-22.
- Harris Z.S. 1955. "From phoneme to morpheme". *Language* 31. 190-222.
- Léon Jacqueline. 1998. "Les débuts de la traduction automatique en France (1959-1968): à contretemps?". *Modèles Linguistiques* 19-2.55-86.
- Léon Jacqueline. 2001. "Conceptions du mot et débuts de la traduction automatique". *Histoire Epistémologie Langage* 23-1.81-106.
- Martinet André. 1960. *Elements de linguistique générale*. Paris:Armand Colin.
- Martinet André. 1965. "Le mot". *Diogène* 51.39-53.
- Martinet André. 1967. "Syntagme et syntème". *La Linguistique* 2.1-14.
- Martinet André. 1968. "Mot et syntème". *Lingua* 21.294-302.
- Martinet André. 1986, "Que faire du mot?". *Mot et parties du discours, la pensée linguistique 1*, ed. by Pierre Swiggers & Willy Van Hoecke, 75-84. Leuven & Paris: Peeters.
- Mitterand Henri. 1965. *Les mots français*. Que-sais-je? n° 270 Paris: PUF.
- Mounin Georges. 1964. *La machine à traduire. Histoire des problèmes linguistiques*. La Haye: Mouton.
- Pottier Bernard. 1962a. "Le mot, unité de comportement", colloque ATALA *Le mot pour la Traduction Automatique et la linguistique appliquée*, 8 décembre 1962.
- Pottier Bernard. 1962b. "Introduction à l'étude des structures grammaticales fondamentales". *La Traduction Automatique* III-3.63-91.
- Pottier Bernard 1962c. "Les travaux lexicologiques préparatoires à la traduction automatique". *Cahiers de lexicologie* 3. 200-206.
- Vauquois Bernard. 1969. "Dix ans d'ATALA: de la traduction automatique au traitement automatique des langues". *TA Informations* 1969-2.57-61.