

Terence Russon Wooldridge, University of Toronto

# A Diachronic Study of Dictionary Structures and Consultability

*ABSTRACT: The CopuLex Project involves a study of the evolution of the dictionary as a reference work from the point of view of its systems of data retrieval, with particular attention being given to the microstructure, which over the centuries has evolved far more than the macrostructure. The codifying of information within entries has been gradual, moving from the discursive and unpredictable towards the recursive, typographically rich and conventionally concise discourse which we know today. My study takes into account two perspectives: that of intention (the lexicographer's prefatory material), and that of effect, the consultability of the dictionary text from the user's point of view. This paper addresses particularly the question of explicit structuring (and thence machine readability) and interpretation (textual ambiguity and the user's threefold competence: linguistic, dictionary, pragmatic).*

## 1. The CopuLex Project

The dictionary is a reference tool whose modes of functioning are attracting more and more the attention of both writer and user. This interest manifests itself in many ways: in efforts to market a better product, to understand the workings of lexicographical discourse in order to construct electronic dictionary databases, or to describe the dictionary as a textual genre.

The objective of the CopuLex Project is to analyse, within the general functioning of the dictionary – the simple, clear form of the macrostructure was established in the Middle Ages and has changed little since –, the history and evolution of the systems of copulas linking the address to the different items of information (orthography, pronunciation, part of speech, meaning, usage, etymology, etc.). Dictionary use can be considered as an equation going from the unknown (the user's question) to the known (the answer); the dictionary's answer can be formulated as a statement whose subject is the question and the predicate the answer: "The word *dictionary* means 'reference work containing alphabetically ordered words together with their definitions, pronunciations, etymologies, etc.'." The worth of the dictionary as reference work resides in large part in the user's understanding and speed of location of the copula connecting the two terms of the equation, the two parts of the statement.

The study corpus comprises twenty of the principal general dictionaries of the period of French monolingual lexicography, starting with Jean Nicot's "*Thresor de la langue françoise*" (1606) and ending with the "*Trésor de la langue française*" (in the process of being

published); more particularly it concentrates on a number of sample articles, the same for each text<sup>1</sup>.

In the study of the dictionary as an information system with its own discourse, articulating several different linguistic and semiotic levels, dictionary discourse is declared as being situated at the level of metalanguage – or discourse on language – dealing with various linguistic properties of lexical units. A distinction is made between first metalanguage – that of the definition expressed in plain terms like a natural language utterance –, second metalanguage – which deals with the sign and signifier (part of speech, etymology, orthography, pronunciation) using conventional abbreviations –, and base metalanguage – which articulates the different items of information contained in the microstructure (Rey-Debove 1971; Rey 1977; Wooldridge 1977). The base metalanguage is expressed essentially by the connectors, or copulas, linking the headword (or other autonomous item contained within the article) to the items of information. Whereas in early dictionaries the copula is often a natural language word, in modern dictionaries it is usually expressed by position, typography or form. Thus the definition copula, for example, may at first be realized by a word like *signifie* "ABOLIR, Signifie proprement oster, effacer, et mettre du tout à neant quelque chose." (Nicot 1606); in most modern dictionaries the definition is signaled by a combination of position (after the orthography-pronunciation-part-of-speech-etymology sequence, for example), typography (romans) and form (the hyperonym of the definition belongs to the same part of speech as the word defined): "ABOLIR [aboliR] v.tr. (1417, « détruire »; lat. *abolere*. Réduire à néant, supprimer." (Petit Robert 1977).

The aim of the project is to analyse the different types of copula of the base metalanguage of French dictionaries, and to trace the path leading from the first attempts to structure the dictionary article to the complex systems of contemporary lexicons. The function of the copula depends on the type of information: classeme, definition, hyperonym, etymology, homophone, homonym, part of speech, flexion, hierarchical or sequential semantic marker, usage label, antonym, synonym, orthography, pronunciation, example, quotation, cross reference, source, sub-headword, variant, syntactic functioning. Four types of copula are distinguished: autonomous (the information is signified by its signifier: e.g. *m.* = masculine noun); linguistic (e.g. *signifie*); positional (the place of the information in the microstructure); typographic (typeface, punctuation, paragraphing, delimiters, symbols, etc.). Data is obtained characterizing the type of information, the type of headword (noun, adjective, verb, adverb, function word), the type of dictionary (extensive vs. restrictive) and diachronic change (different editions of the same dictionary). The prefatory material is usually indicative of the consultability of the dictionary text: Émile Littré concentrates on describing the lexicographer's typical day, but his dictionary is far from user-friendly; contemporary dictionary-makers usually offer both a detailed set of directions for use and a highly structured and differentiated text.

Among initial findings, one can mention the following: a) diachronically there is a progressive formal codification, where linguistic copulas are replaced by positional, typographic and autonomous ones; b) the more complex the article, the greater the accumulation of copulas for a given item of information; c) ambiguity (absence or deficiency of the copula) is to be found in all periods and, to a greater or lesser degree, in all dictionaries. This last is of particular relevance to the question of the computerization of existing printed dictionaries. The case of the "Oxford English Dictionary" is significant: the

task of the programmers at the University of Waterloo was made relatively easy by the clear, homogeneous structure put in place by James Murray. On the other hand, studies done to evaluate the feasibility of putting the "*Trésor de la langue française*" into electronic form have demonstrated that the microstructure contains certain inconsistencies and that the articulation of information is at times ambiguous.

## 2. User competence

There is just so much that the maker of the printed dictionary is able or willing to make explicit. It is always assumed that the user has a triple competence – linguistic, dictionary and pragmatic. The linguistic competence is necessary, for example, to make sense of the definition. Dictionary competence is required at all levels: an understanding of alphabetical ordering, of the use and function of abbreviations and symbols, of the status of a definition given in the etymology sequence, and so on. Pragmatic competence enters particularly in the understanding of cultural references, such as proper names. An extreme example of this last is the use in the dictionary text of source names without the provision of a table of authorities. In his French-Latin dictionary of 1609, Guillaume Poille labels several hundred items "T. d. H.", which refers, without the user being told anything, to an account of a triumphal entry into Rouen by Henri II published 58 years previously; similarly, Pierre Marquis in his French-Latin dictionary published in the same year as Poille's marks several hundred items "H. F.", which only patient research reveals as referring to a historical chronicle that had appeared 36 years before. Tables of cited authorities are also missing from more modern dictionaries, such as Bescherelle (1858) or "*Lexis*" (1975).

A good illustration of the requirement of linguistic and dictionary competence on the part of the user is provided by the case of a textual string serving in turn as predicate and subject. In the following sequences (taken from "*Lexis*") the segment in italics is first an example of usage (predicate) of the headword, then, in whole or in part, the address (subject) of a definition:

"GAGNER [...] *C'est toujours ça de gagné* (= c'est toujours ça de pris)"

"GAGNER [...] *Il ne gagne pas lourd* (= peu)"

There is no copula to allow the user to delimit what in the italicized sequence is the subject of the gloss (the whole thing in the first, "*ne [...] pas lourd*" in the second); linguistic and metalinguistic competence, simple in the one case, more difficult in the other, have to make up for the ambiguity of the dictionary text<sup>2</sup>. The demonstration can be repeated in Robert (1985):

"GAGNER [...] *Gagner de vitesse* [...] V. Dépasser"

"GAGNER [...] *Gagner toute la cave d'un joueur*. V. Décaver"

The two second-stage subjects here are “gagner de vitesse” and “gagner toute la cave d””.

The language elements mentioned in the dictionary’s metalinguistic discourse cover the whole range of the continuum going from lemmatized headwords to signed ideolectal quotations. Headwords are conventionally set in boldface or capitals, signed quotations are usually enclosed in quotation marks. Much of the range of items in between (including lexicalized expressions, canonical and actualized examples) is printed in italics. The question of degree of variability – an item’s membership of an open or closed paradigm – is in any case beyond the competence of the lexicographer to make entirely explicit. Theoretically only an exhaustive corpus of realized and virtual natural language could provide this information. In practical terms any corpus is governed by selection, and any dictionary based on such a corpus is governed by further selection and normalization. “*Collins COBUILD Dictionary*” (1987), based on a large corpus of naturally-occurring utterances, contains examples that “are taken from actual texts *wherever possible*”; “[s]ome of the examples are *slightly adapted* from the citations in our text files”<sup>3</sup>. All dictionary examples are potential or real language items taken out of context and put into the context of metalinguistic discourse; the only thing the user can be sure of is that they are dictionary examples.

Cases of copula deficiency are numerous. I shall mention those of paragraph breaks and toggle typeface. In Trévoux (1743) the decision to start a new paragraph often seems to owe as much to a perceived need to break up the text, like in normal prose or as when one draws a breath in speech, as to an informational articulation of the dictionary microstructure<sup>4</sup>. Paragraphing in “*Trésor*” (1971-) frequently results in clean breaks rather than articulated joints; in the following segment (taken from the article IMAGE):

“— *Péj. Tableau peint avec des couleurs criardes.*”

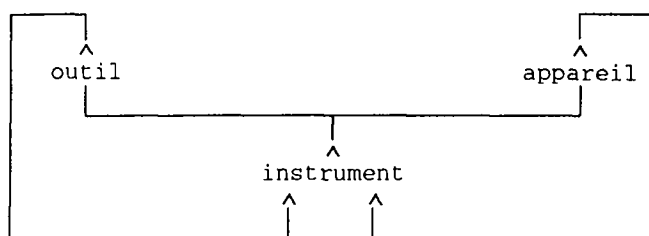
it is not clear whether this paragraph is meant to be connected to the previous two on *image d’Épinal* or to the word *image*. The prefatory material of the dictionary (preface, list of abbreviations) does not assign a function to the long dash.

Toggle typeface refers to the choice of typeface being dependent on neighbouring text. In the following example from the article LOIN in “*Petit Robert*” (1967), the romans of the first occurrence of “*Fig.*” are dictated by the italics of the locution that follows it<sup>5</sup>:

“— *Fig. Revenir de loin*”

“2° *Fig. LOIN, BIEN LOIN QUE*”

I shall finish by mentioning cases of escherization of the dictionary text, situations in which the copula is fully present but the equation itself is impossible<sup>6</sup>. A given of definitional practice is that the defining term is a hyperonym, so that a defining term is in its turn defined by a defining term from the next level up: “A is a kind of B, which is a kind of C, which is a kind of D [...]”. Davau (1972)<sup>7</sup> defines *appareil* as a kind of *objet* or *instrument*, *instrument* as a kind of *outil* or *appareil*, *outil* as a kind of *instrument*:



Lexis (1975) indicates degrees of intensity for synonyms: “aplomb, syn. ↓assurance” means that *assurance* is a synonym of *aplomb* but implies a lesser degree of self-assurance. This is contradicted by the equation “assurance, syn. *aplomb*”:

aplomb  
 ↓  
 assurance = aplomb

Lexis here provides a variation on the well-known theme of the vicious circle (A = B; B = A).

### Endnotes

- 1 The twenty dictionaries are: J. Nicot, *Thresor de la langue françoise*, 1606; P. Richelet, *Dictionnaire françois*, 1680; A. Furetière, *Dictionnaire universel*, 1690; *Dictionnaire de l'Académie françoise*, 1694; *Dictionnaire de Trévoux*, 1743; *Dictionnaire de l'Académie françoise*, 1762; *Abrégé du Dictionnaire de l'Académie*, 1786; Abbé Féraud, *Dictionnaire critique*, 1787; *Dictionnaire de l'Académie françoise*, 1835; M. Bescherelle, *Dictionnaire national*, 1858; É. Littré, *Dictionnaire de la langue françoise*, 1863-72; A. Hatzfeld and A. Darmesteter, *Dictionnaire général*, 1890-1900; P. Robert, *Dictionnaire alphabétique et analogique*, 1951-64; J. Dubois et al., *Dictionnaire du français contemporain*, 1967; P. Robert, *Petit Robert*, 1967; *Grand Larousse de la langue françoise*, 1971-8; C.N.R.S., *Trésor de la langue françoise*, 1971-; Lexis, 1975; P. Robert, *Dictionnaire alphabétique et analogique*, 1985; A. Rey et al., *Le Micro-Robert*, 1988. The sample articles are: the section LOIN... to LOIS...; the individual articles GAGNER and QUE.
- 2 See Wooldridge et al. 1992.
- 3 Sinclair et al. 1987, xv. Emphasis added.
- 4 See Wooldridge et al. 1992.
- 5 Cf. Wooldridge 1988.

- 6 I take my analogy from M.C. Escher's distortions of perspective, typified by his "Klimmen en dalen" staircase (reproduced, for example, in M.C. Escher: "L'Oeuvre graphique". Éditions Solin, Malakoff, 1973).
- 7 M. Davauet *al.*: Dictionnaire du français vivant. Bordas, Paris, 1972.

## Bibliography

- MARQUIS, Pierre (1609): Le Grand Dictionnaire françois-latin. Jean Pillehotte, Lyon.
- POILLE, Guillaume (1609). Le Grand Dictionnaire françois-latin. Jean Du Carroy, Paris.
- REY, Alain (1977): Le Lexique: images et modèles. Colin, Paris.
- REY-DEBOVE, Josette (1971): Étude linguistique et sémiotique des dictionnaires français contemporains. Mouton, The Hague and Paris.
- SINCLAIR, J.M. *ALet al./ITAL* (1987): *Collins COBUILD English Language Dictionary*. Collins, London.
- WOOLDRIDGE, Terence Russon (1977): Les Débuts de la lexicographie française. University of Toronto Press, Toronto.
- WOOLDRIDGE, Terence Russon (1988): "Les vocabulaires et fréquence métalinguistiques du discours lexicographique". In: Travaux de linguistique et de philologie 26. Klincksieck, Paris.
- WOOLDRIDGE, Terence Russon, A. IKSE-VITOLS and T. NADASDI (1992): "Le Projet CopuLex". In: Historical Dictionary Databases. CCH Working Papers 2. Centre for Computing in the Humanities, Toronto.