

Jacqueline Léon. Atelier jeudi 31 août A. Les enjeux des langues universelles dans la période contemporaine.

Langues intermédiaires pour la traduction automatique et sous-langage des sciences

1. Petr Petrovitch Smirnov-Trojanskij (1894-1950) et sa machine à traduire

Trojanskij est né à Orenbourg dans l'Oural. Il fit ses études à St Petersburg à l'Institut krasnoj professorii puis se consacre à sa machine à traduire. En 1933 il obtint un brevet d'invention pour « la sélection et l'impression de mots dans la traduction d'une langue en une autre ou dans plusieurs autres langues simultanément ». Ce brevet fut publié le 31 janvier 1935. Mais ses travaux, qu'il présenta à l'Académie des sciences restèrent ignorés, et le modèle expérimental de machine à traduire, comprenant un dictionnaire de 1000 mots, ne fut jamais construit. Ce n'est qu'en 1959, bien après sa mort en 1950, que les pionniers de la TA en URSS, Bel'skaja, Panov et Korolev publieront une édition commentée de l'ouvrage que Trojanskij avait rédigé en 1947 : *De la machine à traduire, construite sur la base d'une méthode monolingue*.

Comme base d'automatisation de la traduction, Trojanskij postule une structure commune à toutes les langues; entre la langue-source et la langue-cible, il suggère de passer par un texte intermédiaire, qu'il dit *logique*, qui consiste en fait en une suite de mots apparaissant à leur forme grammaticale de base, comme autant d'entrées de dictionnaire qui seront ensuite traduites :

Un traducteur obtient, à partir d'un texte en langue source A, un texte en langage *logique* A1, que la machine traduira automatiquement en langage *logique* B1, grâce à un dictionnaire qui traduit automatiquement les entrées. Le deuxième texte en langage logique ainsi obtenu sera ensuite arrangé manuellement par un rédacteur en langue cible B, selon le modèle ci-dessous d'une traduction de l'allemand simultanément en russe et en français. Pour Trojanskij, le grand intérêt de la méthode réside dans le fait qu'il s'agit d'une méthode de traduction *monolingue*, chaque traducteur pouvant ne connaître qu'une langue et une seule

Les premiers symboles sont empruntés à l'esperanto : dans le tableau : *o* (substantif) *as, is, os* (temps des verbes à l'indicatif au présent, passé et futur ;

A	A1	(B1 + C1)		(B + C)	
Das Bild	Das Bild - o	картина - o	Le tableau -o	картина	Le tableau
der Welt	die Welt de - o	мир de -o	le monde de - o	мира	du monde
zeigt,	zeigen - as	показывать - as	montrer - as	показывает,	montre
wie	wie	как	comment	как	comment
die Materie	die Materie - o	материя - o	la matière - o	материя	la matière
sich bewegt	sich bewegen - as	двигаться - as	se mouvoir - as	движется,	se meut
wie	wie	как	comment	как	comment
die Materie	die Materie - o	материя - o	la matière - o	материя	la matière
denkt	denken - as	мыслить - as	penser - as	мыслит	pense

références :

Bel'skaja I.K. 1958. "Osnovnye charakteristiki slovarja i grammatičeskix sxem maöinnogo perevoda s anglijskogo jazyka na russkij", *Sbornik statej po maöinnomu perevodu*, 47-80, Moskva [Trad. angl. : " Basic Characteristics of the Dictionary and Grammatical Routines for MT from English into Russian ", JPRS/925, archives ATALA].

Andreev N.D., Ivanov V.V., Mel'cuk I.A. 1960. "Nekotorye zameanija i predloenija odnositel'no raboty po maöinnomu perevodu v SSSR", *Mainnyj Perevod i Prikladnaja Lingvistika* 4, 3-24 [Trad. angl. : " Some remarks and suggestions to work on machine translation in the USSR ", JPRS/8026, archives ATALA].

Archaimbault S. et Léon J., 1997, "La langue intermédiaire dans la Traduction Automatique en URSS (1954-1960). Filations et modèles", *Histoire Epistémologie Langage* 19-2:105-132.

John Hutchins John, Two precursors of machine translation: Artsrouni and Trojanskij [Website: <http://ourworld.compuserve.com/homepages/WJHutchins>].

2. Contexte historique et intellectuel de la conception des langues universelles en Grande-Bretagne au 17^{ème} siècle

Au 17^{ème} siècle, les projets de construction de langues universelles étaient non seulement conçus comme un projet intellectuel destiné à fournir un moyen de communication universel et une représentation philosophique (c'est-à-dire scientifique) des connaissances, mais également comme une entreprise empirique expérimentable. Ces projets étaient perçus comme une avancée scientifique et technologique dans le contrôle de l'information face aux besoins croissants en nouvelles formes de stockage et de diffusion de l'information liés à l'intensification du commerce outre-mer et en Europe. Le développement des langues universelles est lié à celui de la cryptographie, objet d'une grande demande en ces temps de guerre civile. Il est aussi lié au développement de la sténographie. C'est explicitement un des objectifs de Dalgarno et de Wilkins d'améliorer la sténographie, dont le premier système appelé 'Characterie' a été mis au point par Timothy Bright en 1588, une centaine d'années avant toute utilisation en Europe.

Dans un contexte de net déclin du latin, les auteurs de langues universelles sont des 'language planners', préoccupés de planification linguistique des vernaculaires (Salmon 1992) et de réformes en tous genres: orthographe, grammaire, dictionnaires. Wilkins était intéressé par des objectifs pratiques: concordanciers qui n'avaient pas seulement une fonction théologique: droit, éducation, scientifique (Cram 1985). Sur le plan intellectuel, si l'on suit Maap et Cram (2000), le langage était considéré au 17^{ème} siècle comme un outil qui devait être jugé à l'aune de son efficacité en fonction des buts à atteindre. C'était quelque chose d'extérieur qui pouvait être remplacé par un meilleur outil au besoin par une langue artificielle.

Références :

Cram David, 1985, "Universal Language Scheme in 17th century Britain", *Histoire Epistémologie Langage* vol 7-2: 35-44.

Maat Jaap, Cram David, 2000, " Universal Language Schemes in the 17th Century ", in *History of the Language Sciences, An International Handbook on the Evolution of the study of Language from the Beginnings to the Present* Edited by Sylvain Auroux . E. F.K. Koerner Hans-Josef Niederehe Kees Versteegh, Volume 1 Walter de Gruyter . Berlin. New York :1030-1042.

Salmon Vivian, 1992, "Caractéristiques et langues universelles", *Histoire des Idées Linguistiques* t.II , Mardaga, Liège :407-423.

3. John Wilkins (1614-1672) et l'*Essay towards a Real Character and a Philosophical Language* (1668)

John Wilkins fit ses études à Oxford et fut nommé vicaire à Fawsley, devint chapelain de nombreux personnalités, dont le prince électeur du Palatinat. En 1648, il fut nommé Warden de Wadham College à Oxford. En 1656 Wilkins épousa une soeur Oliver Cromwell, et en 1659 Richard Cromwell, le fils de Cromwell et le Lord Protector of England, le nomma Master of Trinity College à Cambridge ; mais il dut démissionner au moment de la restauration de la monarchie. En 1660, il fonda la Royal Society (the Royal Society of London for Improving Natural Knowledge) dont il devint le premier secrétaire. Finalement il fut nommé évêque de Chester en 1668.

<http://www-history.mcs.st-andrews.ac.uk/history/Biographies/Wilkins.html>

Ferme de mathématiques , il s'intéressa à la cryptographie, travailla avec John Hooke sur des projets de machines volantes, et forma le projet d'une langue universelle, conçue comme une énumération de choses et de notions comme nouveau moyen de communications. Il commença par dresser des tables de classification qui aboutira à l'*Essay towards a Real Character and a Philosophical Language* publié en 1668 (après que la première version ait été détruite dans le grand incendie de Londres de 1666).

L'Essay de Wilkins est organisé en deux parties ; la première partie est un plan de classification hiérarchique des concepts organisées en quatre points: prolégomènes, philosophie universelle, grammaire philosophique, « a Real Character and a Philosophical Language » ; la seconde partie est un dictionnaire alphabétique où les mots sont renvoyés aux tables de classifications.

Real character et Philosophical language sont deux système de signes équivalents , le second seul étant destiné à être non seulement écrit mais parlé.

Chap. V. *The first Principle of Communication.*

21

If to every thing and notion there were assigned a distinct *Mark*, together with some *provision* to express Grammatical *Derivations* and *Inflections*; this might suffice as to one great end of a *Real Character*, namely, the expression of our Conceptions by *Marks* which should signify *things*, and not *words*. And so likewise if several distinct *words* were assigned for the *names* of such things, with certain invariable *Rules* for all such Grammatical *Derivations* and *Inflections*, and such onely, as are natural and necessary; this would make a much more easie and convenient Language than is yet in being.

But now if these *Marks* or *Notes* could be so contrived, as to have such a *dependance* upon, and relation to, one another, as might be suitable to the nature of the things and notions which they represented; and so likewise, if the *Names* of things could be so ordered, as to contain such a kind of *affinity* or *opposition* in their letters and sounds, as might be some way answerable to the nature of the things which they signified; This would yet be a farther advantage superadded; by which, besides the best way of helping the *Memory* by natural Method, the *Understanding* likewise would be highly improved; and we should, by learning the *Character* and the *Names* of things, be instructed likewise in their *Natures*, the knowledg of both which ought to be conjoyned.

For the accurate effecting of this, it would be necessary, that the *Theory* it self, upon which such a design were to be founded, should be exactly suited to the nature of things. But, upon supposal that this *Theory* is defective, either as to the *Fulness* or the *Order* of it, this must needs add much perplexity to any such Attempt, and render it imperfect. And that this is the case with that common *Theory* already received, need not much be doubted; which may afford some excuse as to several of those things which may seem to be less conveniently disposed of, in the following *Tables*, or *Schemes* proposed in the next part.

The End of the First Part.

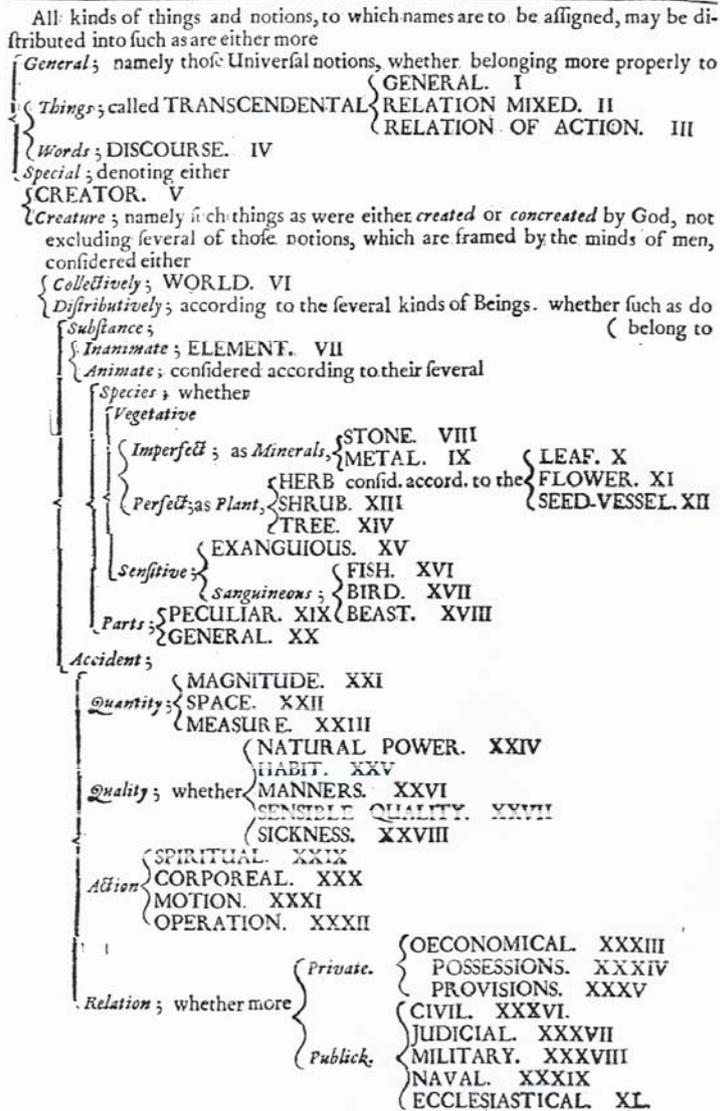
référence :

Wilkins John, 1668, *An Essay towards a real character and a philosophical language [An alphabetical dictionary, wherein all English words, according to their various significations, are either referred to their places in the philosophical tables, or explained by such words as are in those tables]* London: S. Gellibrand and J. Martin.

Dans la partie 4 de l'Essay de Wilkins, "A Real Character and a Philosophical Language", on trouve (p.415) le mot Deba: dont la décomposition reflète la classification hiérarchique des concepts :

De (élément) + b (la première différence, feu) + alpha (la première espèce, flamme)

Chap. I. *The General Scheme.* 23



In

Partie 2 de l'Essay: Universal Philosophy :The scheme of Genus's

4. Masterman Margaret (1910-1986) et le Cambridge Language Research Unit (CLRU)

Le groupe britannique CLRU (Cambridge Language Research Unit) commence ses recherches en traduction automatique en 1954-55, dans le sillage de la première démonstration de TA sur ordinateur effectuée par les américains en janvier 1954 à New York. Il est dirigé par Margaret Masterman, philosophe et logicienne, élève de Wittgenstein et mariée au spécialiste de philosophie morale à Cambridge, Richard Braithwaite. Le CLRU est composé de linguistes comme M.A.K. Halliday, alors assistant en chinois, d'un statisticien A.F. Parker-Rhodes, de mathématiciens, physiciens et philosophes. A ce groupe se joindront ultérieurement, à partir de 1958, les linguistes Martin Kay et Paul Bratley puis du spécialiste de traitement automatique des langues Yorik Wilks. L'originalité du CLRU tient au fait qu'il a mené une réflexion plurielle sur la formalisation du langage ou des langues en vue de l'automatisation de la traduction, centrée sur une méthode de TA par thésaurus. La mise en oeuvre d'un thésaurus, qui s'inscrit dans une tradition britannique des langues universelles durable à travers les siècles, prend des formes très diverses selon les chercheurs du CLRU: tour à tour langue intermédiaire ou dictionnaire de contextes, contenant des informations exclusivement lexicales ou bien grammaticales et lexicales. Par ailleurs dans les modèles de TA élaborés par le CLRU, deux conceptions du mot sont à l'oeuvre: le

mot-concept ou mot-idée propre au thesaurus et le “mot en contexte” issu d’une approche wittgensteinienne, propre à M.Masterman, associée à une approche du sens en contexte du linguiste J.R. Firth.

« It is evident that if we wish to come to a decision as to the extent to which thesaurus-theory has an empirical foundation, the vital notion to examine is that of context.

Having said this, I propose now to examine it, not concretely but abstractly ; because in the course of examining it abstractly, it will become clear how very many obstacles there are to examining it concretely. Roughly, if a language were merely a large set of texts, there would be no such difficulty ; research with computers would show to what extent these could be objectively divided up by using linguistic methods, and into how small slices ; a list of the slices of appropriate size, (i.e morphemes, rather than phonemes,) would be the contexts. Actually, however, language is not like that. Firstly, nobody knows how large a number of texts, and what texts, would be required for these to constitute a true sample. Secondly, we have to know quite a lot about any language, both as to how it functions as and to what it means in order to give the computer workable instructions as to how to slice up the text. So even if we wish to be 100% empirical “to go by the facts and nothing but the facts” - we find that a leap of the creative intellect is at present in fact needed to arrive at a purely empirical notion of collocation, or context. And that being so, there is everything to be said, for using to the full, in an essentially general situation, the human capacity to think abstractly.

The philosopher Wittgenstein, in his book *Philosophical Investigations* (1953), compared the set of uses of a word to the set of ways in which one can see a gestalt figure. At the time, he was primarily investigating, not the concept consisting of the set of uses of a word, but the concept considered as a gestalt given by perception ; of how it itself, (considered as an actual percept, given by experience, and also, by extension, as a picture, or an image) can be affected by environment, or context. » (Masterman, 1959 :21)

“ the fundamental hypothesis about human communication which lies behind thesaurus making is that, although the set of possible uses of words in a language is infinite, the number of primary extra-linguistic situations which we can distinguish sufficiently to talk to one another is finite. Given the complexity of the known universe it might be the case that we refer to a fresh extra-linguistic situation every time we create a new use of a word. In fact we do not ; we pile up synonyms, to rerefer, from various and differing new aspects, to the stock of basic extralinguistic situations which we already have. The importance of this fact for MT is obvious. If the hypothesis is right, communication and translation alike depend on the fact that two people and two cultures, however much they differ, can share a stock of extra-linguistic contexts. When they cannot come to share such a stock, communication and translation alike break down. Imagine two cultures, one, say, human, one termite. The members of the first of these sleep, and also dream, every night; the members of the second do not know what sleep is. As between these two cultures, communication on the subject of sleeping and dreaming would be impossible until acquired knowledge of sleeping and dreaming by members of the second culture sufficed to establish it.” (Masterman, 1959 :34-35)

5. Richard H. Richens (1919-1984) et la langue *Nude*

Biologiste et botaniste, Directeur du *Cambridge Bureau on Plant Breeding and Genetics* du Commonwealth Agricultural Bureau, Richens est l’auteur d’une langue intermédiaire pour la traduction automatique dans le cadre du CLRU. Cette langue intermédiaire est constituée d’un réseau sémantique d’idées nues, les ‘naked ideas’, d’où le nom de *Nude* donné à cette langue. Débarrassée de toutes les particularités lexicales et syntaxiques des langues naturelles, *Nude* se veut une langue intermédiaire algébrique purement notationnelle.

Nude: interlangue algébrique, idéographique. Chacun des cinquante éléments primitifs de cette langue artificielle dénote une idée de base (naked idea) telle que ‘pluralité’, ‘animal’, ‘négation’ et ne peut comporter qu’une seule lettre.

Liste des 19 éléments (sur une cinquantaine) utilisés dans l’exemple de traduction (du japonais vers *Nude*, puis de *Nude* vers l’anglais, l’allemand, le gallois et le latin)

résultat de la traduction en anglais: *the percentage of matured capsules and the number of grains of seeds of one capsule are different according to the time of hybridizing*

B	becoming, change	p	plant
c	straight, plane	P	plurality, group, number
C	causation, influence	Q	hard, firm
f	possibility, potentiality	S	same, equal
H	pertain	T	time, period, duration
I	in, inside	u	elongate
L	living, alive	x	textile

M	much, more, great	X	part, component
n	near, adjacent, together	z	negation, opposite, contrary
N	contact, adhere, attach		

La syntaxe de Nude consiste en deux connecteurs et une convention de parenthésage. Le premier connecteur, les deux points, représente une relation d'ajout à l'élément principal. Le slash est un connecteur verbal non commutatif qui représente la relation de sujet au verbe ou de verbe à objet. Les parenthèses relient un ajout avec son élément principal, un objet à un verbe qui précède et un sujet avec le prédicat.

un point [.] = relation monadique: un adjectif ou un verbe intransitif (sont marqués par l'attente d'une relation avec un nom)

deux points [:] = relation dyadique: verbe transitif

apostrophes sont utilisés comme des parenthèses à l'intérieur d'un mot.

Exemples de formules

.Pz = one

Xp'CL= seed

Xp'CL.Pz = one seed

Le sens d'un mot est une composition d'éléments de sens:

seed (graine) = Xp'CL X partie + p plante / C cause L vie

6. Peter Mark Roget (1779-1869) et le *Roget's Thesaurus*

P.M Roget est d'abord médecin. Il commence à exercer dès l'âge de 19 ans à Manchester, puis à Londres. Entre 1824 et 1837 contribue à de nombreuses publications dont *l'Encyclopædia Britannica* sur la physiologie. De 1827 à 1848, il est secrétaire de la Royal Society (comme Wilkins) où il cotoya Huxley et Darwin. Le *Roget's Thesaurus* est le résultat d'une compilation de vocabulaire pendant une cinquantaine d'années (1805-1852) et de son admiration pour Cuvier. Le *Roget's Thesaurus* connut un succès immédiat et constant et une diffusion à très grande échelle. (1852 : 1ère édition; 2002 : 150ème édition). Mais lors de sa parution, il n'est pas reconnu par les philologues et lexicographes, à une époque où dominent le romantisme dans la littérature et l'historicisme dans la philologie.

Le *Roget's Thesaurus* est organisé en deux parties: un thesaurus thématique et un index alphabétique. Cette organisation s'inspire de celle de *l'Essay* de Wilkins .

Le plan de classement de la partie thématique comporte 6 classes primitives (comme les catégories de Wilkins):

1. Abstract Relations
2. Space
3. Matter
4. Intellect: the exercise of the mind
 - Division 1: Formation of ideas
 - Division 2: communication of ideas
5. Volition: the exercise of the will
 - Division 1: Individual volition
 - Division 2: Social volition
- 6 Emotion, religion and morality

Ces six classes sont elles-mêmes subdivisées en sections où apparaissent les heads. Les heads sont suivies d'une liste de mots apparentés sémantiquement. Un mot peut donc apparaître dans des listes différentes, sous des heads différentes, dans des classes différentes.

7. Exemple de traduction à l'aide du *Roget's thesaurus*

établir le sens du mot 'pen' dans les deux phrases (+ analyse en constituants):

- 1) the pen is in the inkstand ((the pen) / (is (in (the inkstand))))
- 2) the inkstand is in the pen ((the inkstand) / (is (in (the pen))))

Pour les mots *pen*, *in* et *inkstand* les entrées (heads) dans le Roget sont les suivantes:

Pen: region, inclosure, limit, writing, book, hinderance, restraint, prison

44	KIND	Specie	Same :kind (being of the same specie)
45	GRAIN	Pattern(artistic , thought)	Think:(stuff:grain) (chemistry)
46	HOW	Mode , quality, adjective	(think/same) : how.
47	WHEN	Time	Count:(part:when)(unit of time)
48	WHERE	Space	Change/where (move)
00	NOT	Causes all Nude elements to mean their opposites.	

exemples:

<i>speak</i>	<i>he says</i>	<i>speaker</i>
cause / (have /sign)	man/(cause / (have /sign))	man:(cause/ (have /sign))

in the inkstand

IN! | (man /use) / (IN!: thing) -INKSTAND

Références :

- Bar-Hillel Yoshua, 1960, Appendix III *A demonstration of the Nonfeasibility of Fully Automatic High Quality Translation*, p.158-159.
- Hutchins W.J., 2000, *Early Years in Machine Translation*, John Benjamins, Amsterdam, Philadelphia.
- Léon J., “From universal languages to intermediary languages in Machine Translation : the work of the Cambridge Language Research Unit (1955-1970)” *Proceedings of ICHoLS 9, 9th International Conference on the History of Language Sciences*, Sao Paulo (Brésil), 27-30 August 2002, John Benjamins Publishing Company, (à paraître).
- Masterman Margaret, 1959, “What is a thesaurus?” in *Essays on and in Machine Translation by the Cambridge Language Research Unit*. rapport non publié, ML90
- R.H. Richens, 1956 “A general programme for mechanical translation between any two languages via an algebraic interlingua”[ML5, archives du CLRU].
- Roget Peter Mark, 1852, *Thesaurus of English words and phrases classified and arranged so as to facilitate the expression of ideas and assist in literary composition*, London: Longman [2ème édition, 1853].
- Wordley C.1959, “Establishing the contextual meaning of words” in *Essays on and in Machine Translation by the Cambridge Language Research Unit*. rapport non publié, ML91, *appendice II*

8. Les templates de Wilks

Yorick Wilks (né en 1939) est actuellement professeur d'intelligence artificielle à l'Université de Sheffield. Il a participé aux travaux du CLRU et a fait sa thèse avec Margaret Masterman.

Wilks reprend le langage Nude dans un projet de résolution d'ambiguïtés sémantiques dans un texte. Contrairement à la notion d'ambiguïté admise par tout expérimenteur de TA, puis plus tard de TAL en général, qui consiste à dire qu'un mot ambigu sémantiquement est équivalent à un mot polysémique, c'est-à-dire que tous les sens possibles de ce mot sont les sens attribués par le dictionnaire (par exemple), Wilks définit l'ambiguïté par rapport à un texte donné.

Pour résoudre ces ambiguïtés, il met au point un système de représentation sémantique des textes à l'aide de templates qui doivent capturer l'essentiel (the gist) de l'information (du message dit-il) dans les textes. Alors que la méthode par thesaurus se contentait de coder le sens des mots. Les templates, sortes de patrons du sens d'une proposition ou d'une phrase simple, sont fabriquées à l'aide de formules très proches des formules de Nude mises au point par Richens et réélaborées par Masterman.

Les formules permettent d'encoder les différents sens d'un mot à l'aide de 53 primitives, dont voici la liste :

BE	FORCE	MAN
BEAST	FROM	MAY
CAN	GRAIN	MORE
CAUSE	HAVE	MUCH
CHANGE	HOW	MUST
COUNT	IN	ONE
DO	KIND	PAIR
DONE	LET	PART
FEEL	LIFE	PLANT
FOLK	LIKE	PLEASE
FOR	LINE	POINT
SAME	THINK	WHEN

SELF	THIS	WHERE
SENSE	TO	WHOLE
SIGN	TRUE	WILL
SPREAD	UP	WORLD
STUFF	USE	WRAP
THING	WANT	

(COLOURLESS((((((WHERE SPREAD)(SENSE SIGN))NOT HAVE)KIND)
 (COLOURLESS AS NOT HAVING THE PROPERTY OF COLOUR))))

référence :

Wilks Yorick A. 1972 *Grammar Meaning and the machine analysis of language*. London Routledge and Kegan Paul.

9. Zellig Sabbetai Harris (1909-1992) et les sublangages

Harris, né à Balta en Ukraine émigra aux Etats-Unis à l'âge de quatre ans. Il fit ses études, une thèse sur la grammaire du phénicien, et enseigna à l'Université de Pennsylvanie jusqu'à sa retraite en 1979. Matthews (1999) divise l'oeuvre de Harris en trois phases qui se recouvrent en partie :

Phase 1 (1942-1955), il se consacre à définir les unités de base de la langue à partir des patrons distributionnels. La grammaire est un ensemble d'instructions qui permet de générer les phrases d'une langue. Cette recherche sera développée ultérieurement par Chomsky sous le nom de « Generative Grammar »

Phase 2 : début des années 1950 fin des années 1960, consacrée à l'analyse du discours et aux transformations.

Phase 3 : à partir des années 1970, consacrée à l' « operator grammar » et à l'analyse de l'information dans les textes.

Toujours selon Matthews, les travaux de la 1ère phase font partie de l'histoire ; ceux de la seconde phase ont été éclipsés par ceux de Chomsky. Enfin les travaux de la 3ème phase ont été largement ignorés encore jusqu'à nos jours. C'est à cette troisième phase qu'appartient l'analyse des sous-langages de la science.

« Each science-language is a sublanguage of the whole of a natural language because if we operate on the special sentences or sentence-pairs of the science-language with the operations of the whole language (eg *and*, reductions or other transformations) we obtain again a sentence of the science-language. In addition, the sublanguage for a given science in one language is similar to the sublanguage for that science in another language, so that its status as a sublanguage of English, French or the like is not its main characterization. ... the individual sentences of the science-language are then the formulas themselves. In that case it can be regarded as a new kind of linguistic system, used identically by scientists of whatever native language – even if each scientist pronounces or writes the symbols or words in a natural language, eg English, of his own. This science-language has its own vocabulary (the subclass symbols) in its own word-classes, occurring in its own sentence-types, under a more general system of conjunctions and meta-science material. It is a linguistic structure which has arisen in trying to represent the information and constraints met with in a given science. As a system it is in certain respects intermediate between natural language and mathematics.

The grammar of each science sublanguage is not a subgrammar of the grammar of its whole language, since it has entities (eg word-subclasses) and operations (eg conjunction restrictions) which do not exist in the whole grammar. It also is not a pure abstract system (such as a whole grammar is) because its classes are not defined solely by an a priori relation defined on them ; instead, its elementary arguments are collected into particular classes by their relation to particular operator-classes. These particular classes can be defined, or their members listed, in the meta-language of the science-language, ie in the whole natural language. » (Harris et al. 1989 :80).

(Harris et al. 1989 :220-221)

Pathogenic bacteria carried on the lymph stream are often arrested in the glands through which this stream passes, with result that the infection travels no further. 1, 783.1.7.

$G^w U_1 T_n^w$	pathogenic bacteria	are often arrested	in the glands
$GU^y T_l$	WH (pathogenic bacteria)	(are) carried	on the lymph stream
$T_l W_u^y T_n$	which this stream	passes through	(the glands),
$G_l U \sim$	with result that the infection	travels no further	

It seems possible that within the lymph glands there may occur some formation of antibodies to antigens arriving by the lymph stream. 1, 783.2.1.

M	It seems possible that
$GU^yT_l: A^cV_pT_n$	(antigens) arrive by the lymph stream WH ← antibodies to antigens there may occur some formation of within the lymph glands ←

A : class of antibody terms

G : class of antigen terms

T : tissue words

T_l: lymph

T_n: glands

U : verbs whose subject is G with a second argument T (or C for cell words)

Superscripts *f, t, y* : prepositions *from, to, by* introducing the second argument

V_p (*produce, form, synthesize*) : *have a role in, participate in, constitute a factor in production of*

W : verbs in the C-T sentences

~ négation

M : meta-science verb-class: *find, study, observe, investigate ... believe, examine, state, note ... (it is) possible, probable likely, significant, clear, evident ...*

| segment boundaries (reading left to right within the segment boundaries)

||| conjunctions, sentence introducers etc.

← reading right to left

WH ← proceed to the leftmost segment and begin reading left to right

pathogenic bacteria carried on the lymph stream	GU^yT_l
antigens arriving by the lymph stream	GU^yT_l

Références :

Harris Zellig, Michael Gottfried, Thomas Ryckman, Paul Mattick, Anne Daladier, T.N. Harris and S. Harris, 1989, *The Form of Information in Science*, Dordrecht, Boston, London : Kluwer Academic Publishers.

Matthews P.H., 1999 « Obituary of Zellig Sabbetai Harris », *Language* 75-1 :112-119.