



La Société Historique et Littéraire Polonaise
avec le soutien de
l'Académie Polonaise des Sciences à Paris

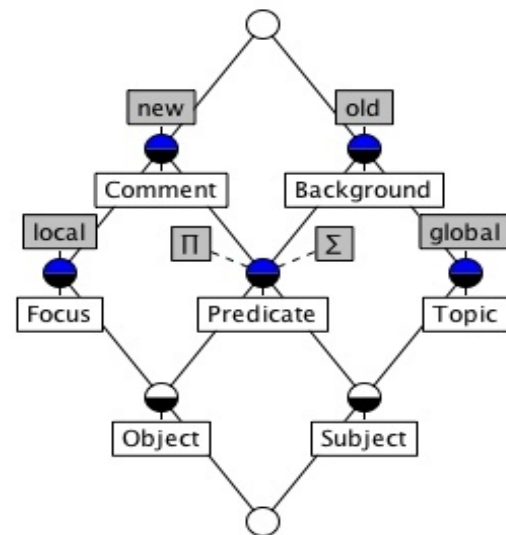
La 3^e conférence du cycle:

Les Polonais dans la culture et la civilisation française

Le Langage au carrefour des sciences

André WŁODARCZYK

wlodarczyk.andre@gmail.com



Paris, le 21 mars 2014, Bibliothèque Polonaise de Paris

Faisceaux de connexions intracérébrales (représentation automatisée)



Denis Le Bihan, *Le cerveau de cristal : Ce que nous révèle la neuro-imagerie*, Odile Jacob Sciences, Paris 2012

Plan

1. Comment définir le langage humain ?
2. Des paradigmes linguistiques à l'interdisciplinarité
3. Les Origines du langage et des langues
4. Digression sur le « langage » des peintures préhistoriques
5. Bilinguisme et multilinguisme
6. Grammaires formelles
7. Traduction automatique
8. Théorie du Centrage Méta-informatif (MIC)
9. Conclusions

Grande complexité

Les langues font partie des systèmes complexes. On suppose aussi que les systèmes complexes comportent des régularités locales. Dans ce cas, les langues seraient non-linéaires (ne pas confondre la linéarité mathématique avec la succession des unités linguistiques).

Influence des paradigmes des sciences sur l'évolution de la linguistique

1	législation	linguistique normative
2	biologie	linguistique comparée
3	chimie	linguistique structurale
4	mathématiques	linguistique générative
5	sciences de la nature	linguistique cognitive
6	informatique	linguistique informatique

WINOGRAD Terry (1983) *Language as a cognitive process* (vol. 1: Syntax), Addison-Wesley Publishing Co. (en selon la théorie de Thomas KUHN concernant les révolutions scientifiques)

Vers la linguistique interdisciplinaire

Face à la complexité de la cognition et des facultés linguistiques, il ne suffit pas d'attendre l'arrivée d'un paradigme suivant mais il est nécessaire de réunir les méthodes de toutes les sciences qui ont trait au langage pour en faire des modèles.

Sciences concernées par les langues naturelles

SCIENCES "MOLLES"

Linguistique
Stylistique & Rhétorique
Psychologie & Sociologie
Histoire

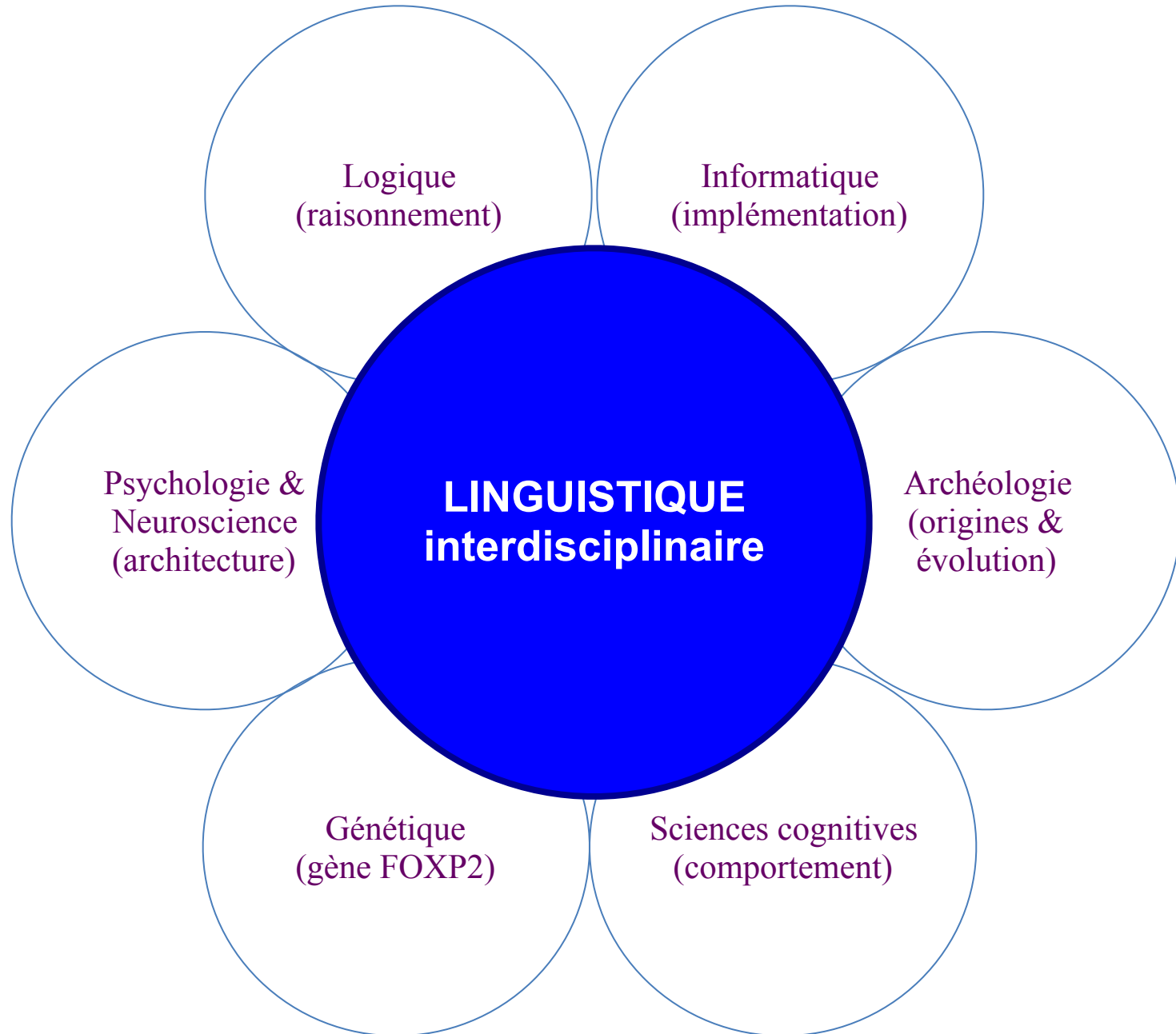
SCIENCES "INTERMEDIAIRES"

Sciences cognitives
Anthropologie et Ethnologie
Archéologie

SCIENCES "DURES"

Mathématique & Statistique
Informatique & Logique
Neurologie et Neurosciences
(du cerveau humain)
Génétique

Interdisciplinarité



Langage et langues

L'exigence de première importance pour toute science veut que les objets qui appartiennent à son domaine d'intérêt présentent des caractéristiques communes.

C'est en cela que la Science se distingue de la Culture car les objets de cette dernière peuvent (même doivent) manifester des caractéristiques originales.

Force est de constater que les langues sont des phénomènes naturels propres à toute l'espèce humaine. C'est pourquoi il paraît raisonnable de parler du "langage" dont toutes les "langues" ne seraient que des manifestations culturelles (F. de Saussure, N. Chomsky).

L'Universalité du langage et la Relativité des langues

Linguistique et Sciences du langage:

Les sens des expressions varient considérablement (ce qui est dû à la partialité des messages linguistiques) d'une langue à une autre. Malgré cela il y a beaucoup de chances que les contenus de la « mémoire de travail » (à la conception et à la compréhension) présentent beaucoup de points communs.

La dynamique de conception/compréhension va du précis au concis et vice versa. Si nous comparons les conceptions/compréhensions des individus porteurs des langues différentes avec une forte précision (interprétation à granularité fine), nous devons constater des similarités notoires.

Génétique et Neurosciences

Génétique:

Le gène nommé FOXP2 (forkhead box P2) découvert en 2001 par des chercheurs britanniques d'Oxford serait spécifique du langage humain. Mis en évidence dans la famille KE dont plusieurs membres souffraient d'altérations spécifiques du langage, le gène FOXP2 est la première trace découverte du langage dans le patrimoine génétique.

Neurosciences:

Les fonctions cérébrales qui semblent concerner la production et la compréhension des messages linguistiques sont concentrées dans deux régions du cerveau humain: l'aire de Paul Broca (1878) et celle de Carl Wernicke (1874).

Langues naturelles et Langage artificiels

Langues naturelles (moyens de communication entre les hommes)

Langages artificiels

- Langages de programmation (moyens de communication entre les hommes et les machines)
- Langages de spécification des connaissances
- Langages outils - notations de la logique (moyen de raisonner)

*? Langage naturel

Les codes génétiques ne sont pas des langages.

Origines des langues et du langage

Âge du langage

- Archéologie:

100.000 ans ou 200.000 ans ou 400.000 ans vu les résultats de recherche sur l'ancien ADN.

1.000.000 ans sur la base des recherches sur les ossements des premiers hommes

- Linguistique:

Abandon de toute recherche sur la question par la Société de Linguistique de Paris (*Statuts de 1866*)

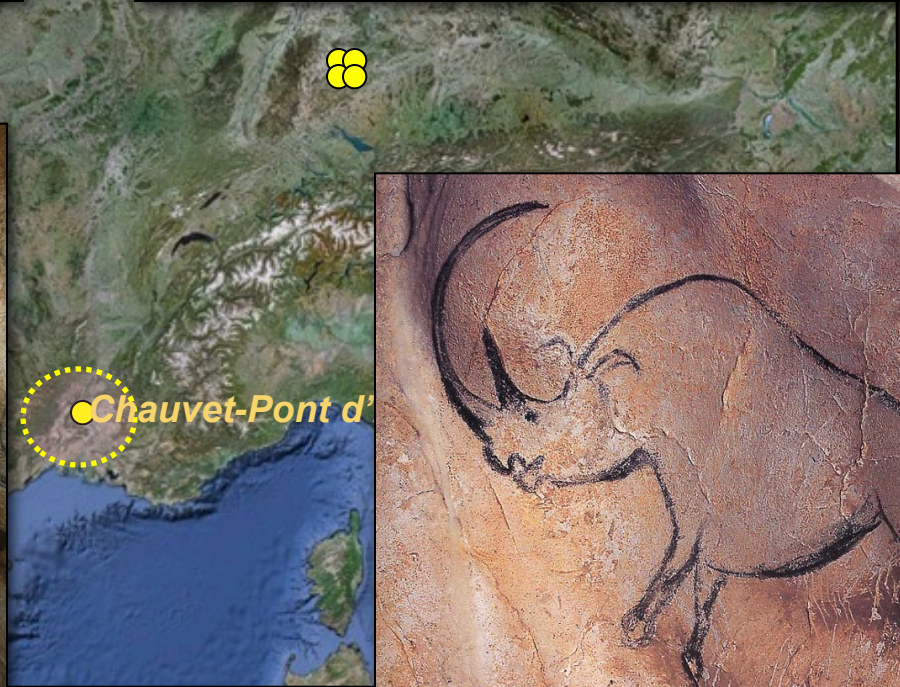
« Saussure et l'origine du langage. Un interdit à dépasser par la philosophie linguistique », par Thomas Robert, RIFL (2010) 3: 12-10 (Saussure filosofo del linguaggio, 10)

Peintures préhistoriques

Les peintures pariétales préhistoriques obéissent à des règles de composition qui trahissent des contenus sémantiques (mythes ?), cela constitue la preuve que les artistes possédaient un langage.

[http://perso.numericable.com/andre.wlodarczyk/
PubliListeArtPaleo.html](http://perso.numericable.com/andre.wlodarczyk/PubliListeArtPaleo.html)

Art pariétal de la grotte Chauvet (32-30000 BP)



Très peu de scènes et très peu d'humains



Lascaux : « scène du puits »

Des motifs non figuratifs nombreux et structurés

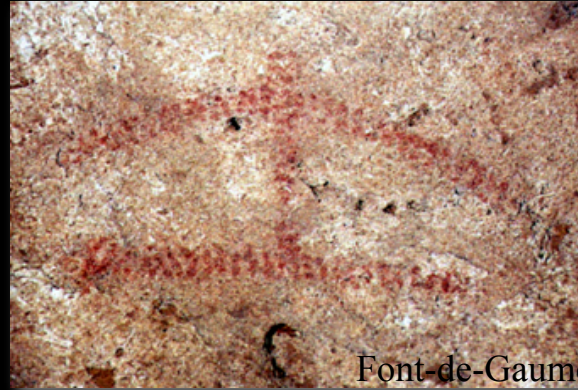


El Castillo (Espagne)

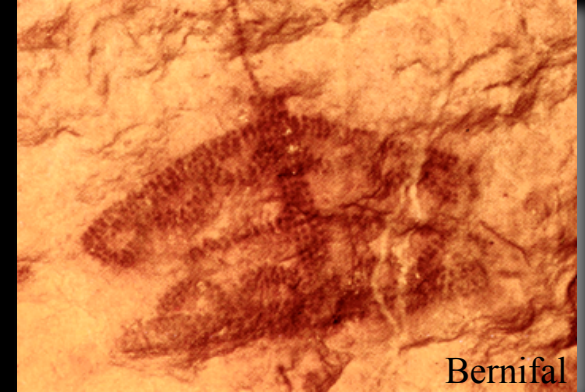
Des motifs non figuratifs nombreux et structurés



Rouffignac



Font-de-Gaume



Bernifal

« Signes tectiformes »



Fontanet



El Pindal



La Cullalvera

« Signes claviformes »

Âge du langage

- Linguistique générative:

Noam Chomsky 100.000 ans (innéisme)

AW:

1.000.000 ans (vu que la cuisson a commencé à cette époque)

Âge des langues

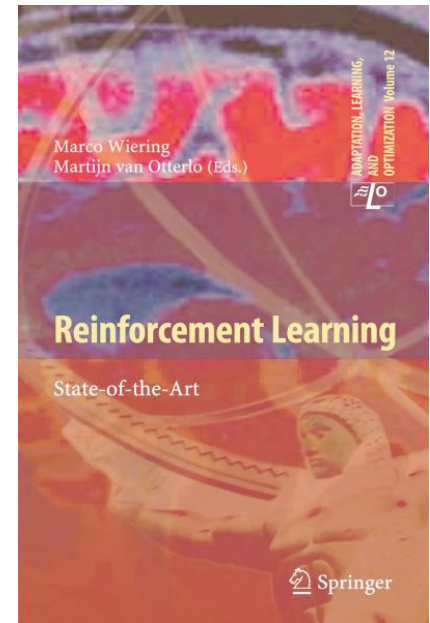
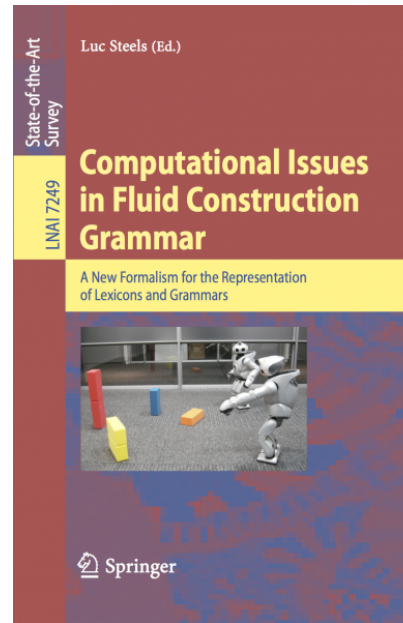
- Linguistique quantitative (Zipf et Swadesh):

Calcul de la différenciation linguistique en étudiant les écarts de la distribution et de la fréquence des mots entre les langues.

Cependant, l'efficacité de ce calcul est limitée à environ 10.000 ans.

Émergence du langage

- Informatique (Luc Steels & al.): Simulation des mécanismes conduisant à l'apparition spontanée du langage.



Cf. aussi Projet CNRS – Origines de l'Homme, du Langage et des Langues

Sciences neuronales

Imagerie



Systeme IRM 3 T de NeuroSpin.
Les techniques d'imagerie par resonance magnetique

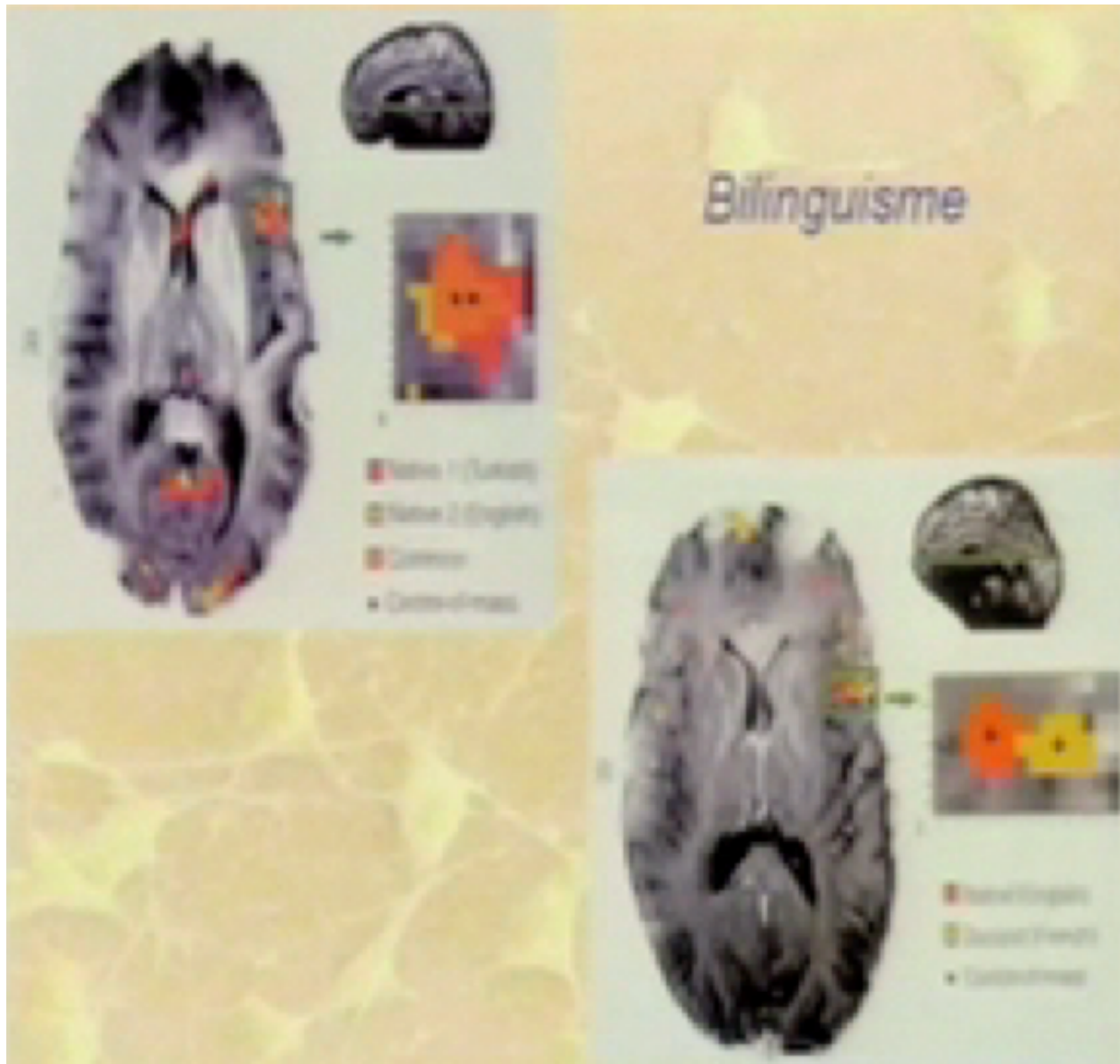
Une ou deux aires de Broca

Pour les personnes réellement bilingues, quand on leur demande de penser à des mots dans une langue, puis dans l'autre, la même localisation dans la région de Broca s'active quelle que soit la langue.

Pour la plupart d'entre nous, qui avons appris une langue étrangère plus tard, on s'aperçoit qu'il existe 2 aires de Broca, très proches, mais néanmoins différentes pour chaque langue.

Selon *Le cerveau de cristal* par Denis le Bihan – 2012

Une ou deux zones de Broca



Diapositive
(copie d'écran)

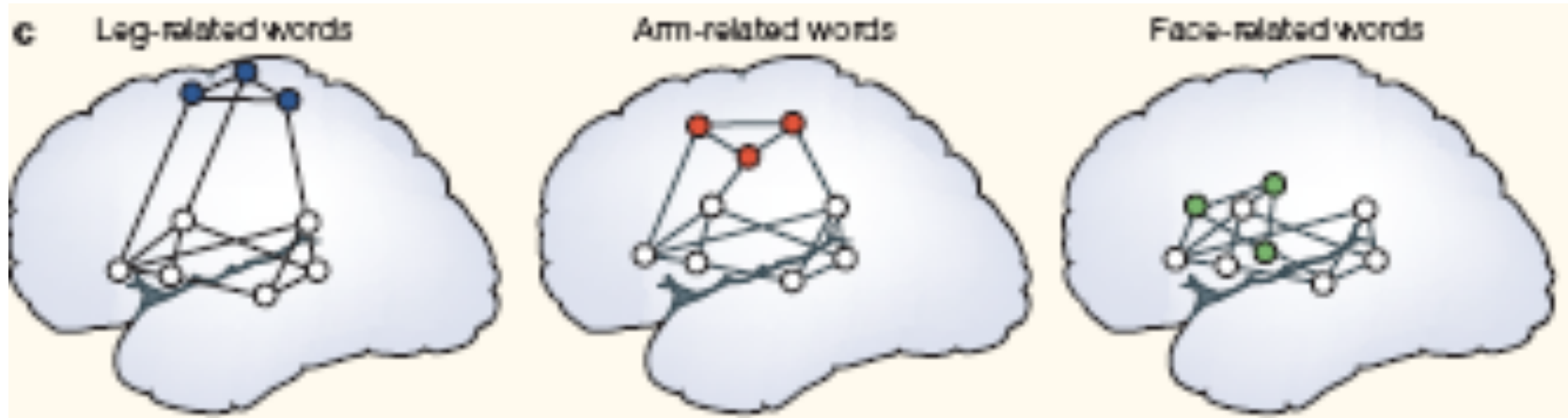
*Le cerveau de
cristal*

par

Denis Le Bihan

2012

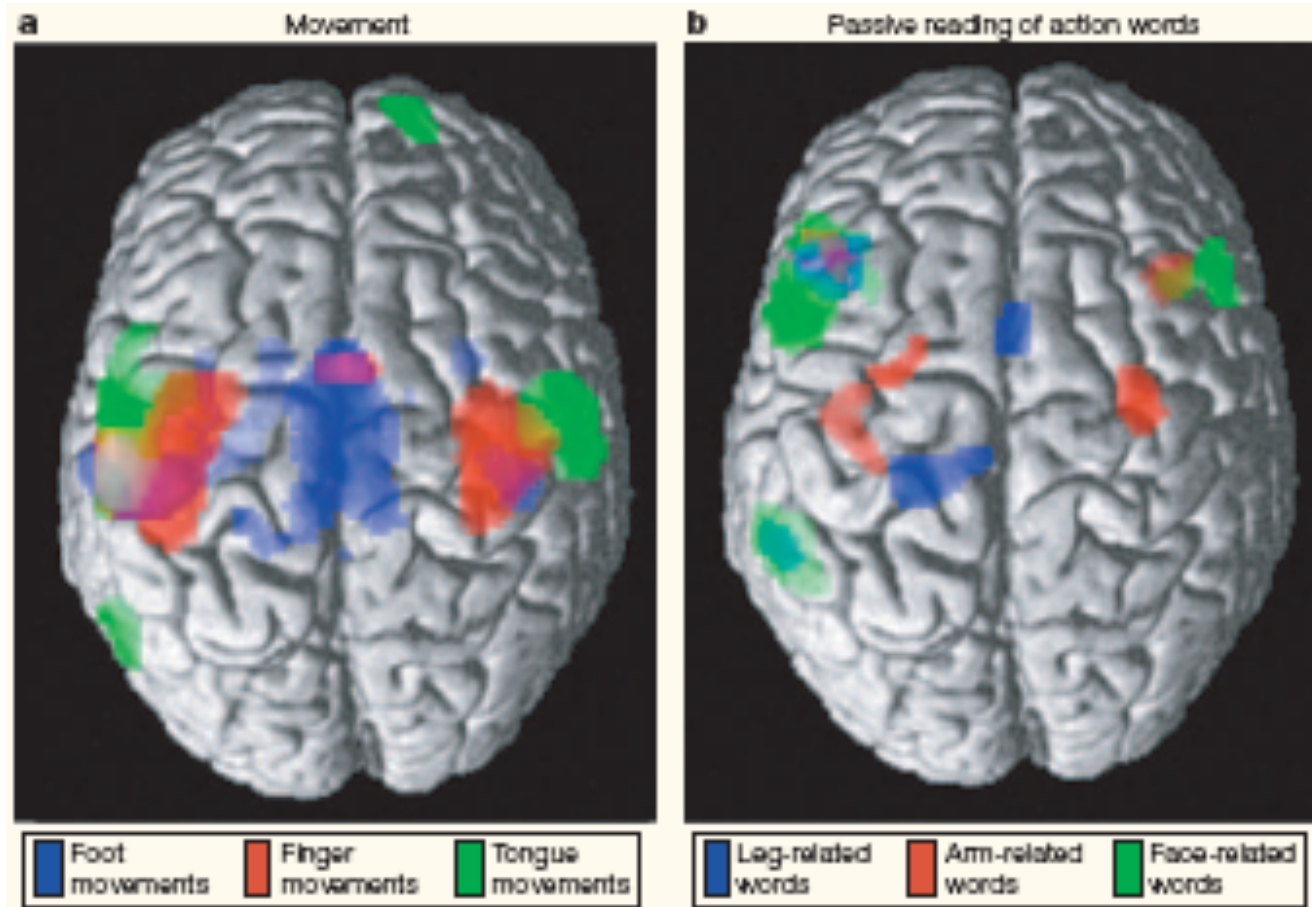
Somatotopie



Selon ce qu'on dit (ou lit seulement) sur telle ou telle partie de son corps, on observe que différentes parties du cerveau sont activées.

Friedemann Pulvermüller: Semantic somatotopy of action words

Somatotopie



De plus, des réseaux neuronaux se créent aux mêmes endroits quand on met en mouvement les mêmes parties du corps.

Friedemann Pulvermüller: Semantic somatotopy of action words

Linguistique

Grammaires formelles des langues naturelles

1. Grammaires transformationnelles
2. Grammaires catégorielles et d'unification
3. Grammaires logiques

Théories transformationnelles

CHOMSKY:

Théorie Standard (1957, 1965, 1970)

Théorie Standard Etendue (1971)

Théorie du Gouvernement et Liage (1981)

Théorie des Principes et Paramètres (1986, 1993)

Théorie du Programme Minimaliste (1995)

Grammaires catégorielles

Grammaire de Lambek (1958)

Grammaire de Montague (1970)

Grammaires d'unification

LFG : Lexico-Functional Grammar (Bresnan 1982)

FUG : Functional Unification Grammars (Key, 1983)

Head Grammars : (Pollard, 1984) and HPSG : Head-driven
Phrase Structure Grammar (Pollard, 1985)

GPSG : Generalized Phrase Structure Grammar (Gazdar et
al., 1985)

UCG : Unification Categorical Grammar (Uszkoreit, 1986)

TAG : Tree-Adjoining Grammar (Joshi, 1987)

Grammaires logiques

System Q (Colmerauer, 1970)

Metamorphosis Grammar (Colmerauer, 1975)

DCG : Definite Clause Grammars (Pereira & Warren, 1980)

MSG : Modifier Structure Grammars (Dahl & McCord, 1983)

DG : Discontinuous Grammars (Dahl, 1981)

DCTG : Definite Clause Translation Grammars (Abramson, 1984)

Grammaires Attributives

Linguistique structurale

Reconnaissance des formes

Traitement automatique des langues

Représentation des connaissances

Grammaires Attributives

- 1963 - Harman G. : Indices added to non-terminal symbols
- 1963 - Katz J.J. & Fodor J.A. : Binary semantic attributes (+ and -)
- 1965 - Chomsky N. : (in "Transformational Grammar")
- 1968 - Knuth : Semantic Attributes (in "Context-free Grammars")
- 1970 - Woods (in "Augmented Transition Network Grammar")
- 1972 - Winograd T. (in "Understanding Natural Language")
- 1973 - Fu K.S. (in "Attributed Grammars")
- 1973 - Kaplan R. M. (in "General Syntactic Processor")
- 1976 - Heidorn (in "Augmented Phrase Structure Grammar")
- 1978 - Pyster A. & Buttelmann H.W. (in Semantic-Syntax)
- 1982 - Robinson J. (in "Annotated Phrase Structure Grammar")
- 1982 - Bresnan J. (in "Lexical Functional Grammar")
- 1983 - Shieber S. M. and al. (in PATR-II)
- 1984 - Pollard C. (in "Head-driven Phrase Structure Grammar")
- 1985 - Kay M. (in "Functional Unification Grammar")
- 1985 - Gazdar G. & al. (in "Generalized Phrase Structure Grammar")
- 1987 - Gunji T. (in "Japanese Phrase Structure Grammar »)

TRADUCTOLOGIE
et
TRADUCTIQUE

Vieux RÊVE des linguistes



La traduction est-elle POSSIBLE ?

Réponse:

La traduction est possible au même titre que l'expression elle-même l'est, c'est-à-dire dans la mesure où nous sommes capables de

“dire ce que nous voulons”.

QUESTIONS DE FOND

Q1: Quelle sorte d'opération la traduction est-elle ?

Q2: Quels faits linguistiques concernant deux langues sont caractéristiques de chacune d'elles et quelles paires de phénomènes sont comparables ?

Q3: Jusqu'à quel degré les traductions sont-elles semblables à l'original ?

COMMUNICATION

Dialogue

Le sens d'une expression se construit en 2 temps :

En encodage: Conception(s)= s' , Production(s')= e ,

En décodage: Interprétation(e)= s' , (2) Compréhension(s')= s''

encodage(s')= e

décodage(e)= s''

Locuteur



Allocutée



Expression e



TRADUCTION

Truchement($s1'$)= $s2$

décodage($eL1$)= $s1'$

Expression
en langue de départ $eL1$

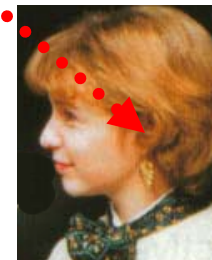


encodage($s1$)= $eL1$



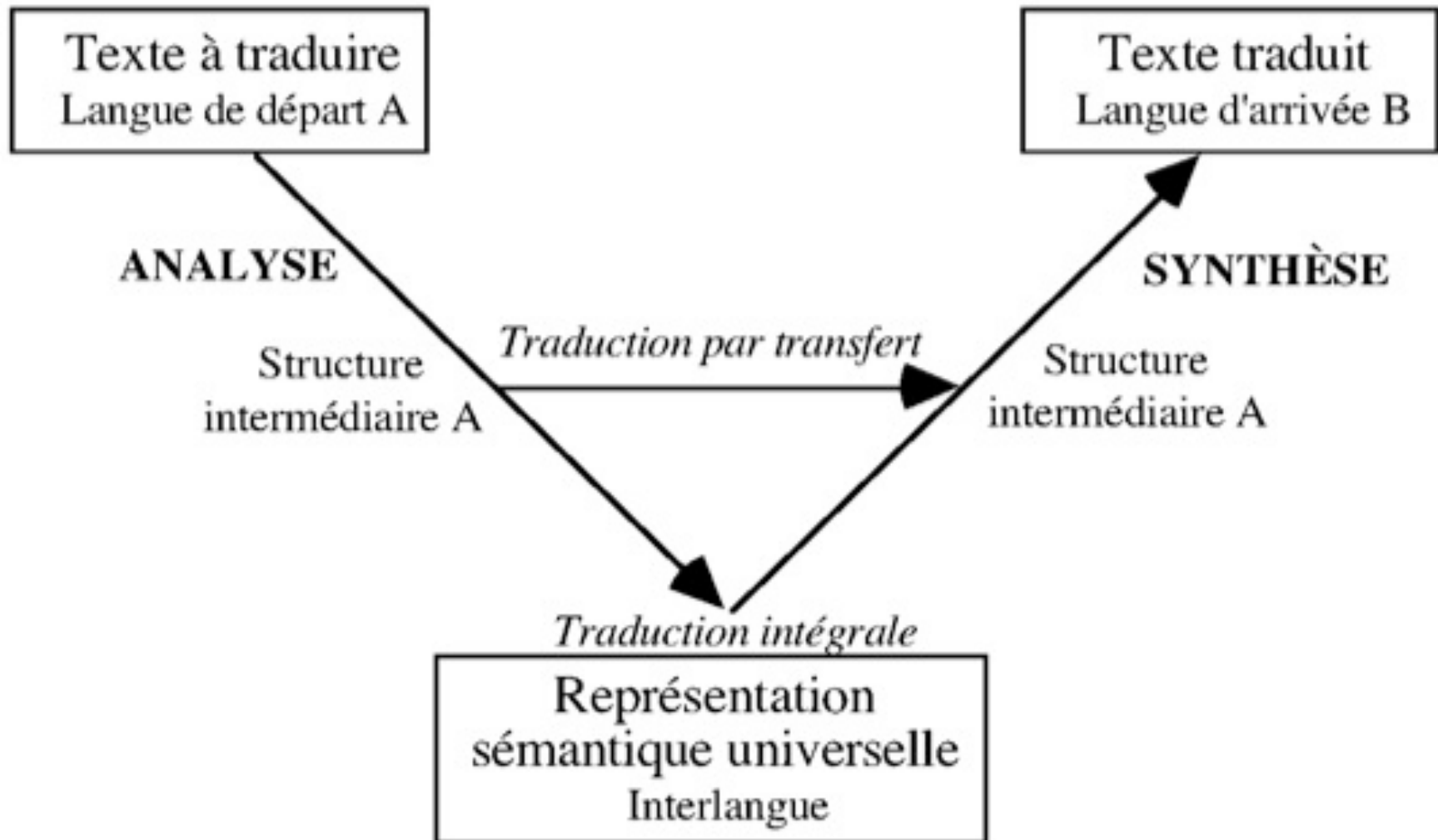
encodage($s2$)= $eL2$

Expression
en langue cible $eL2$



décodage($eL2$)= $s2'$

Représentation schématique des premières approches en traduction automatique

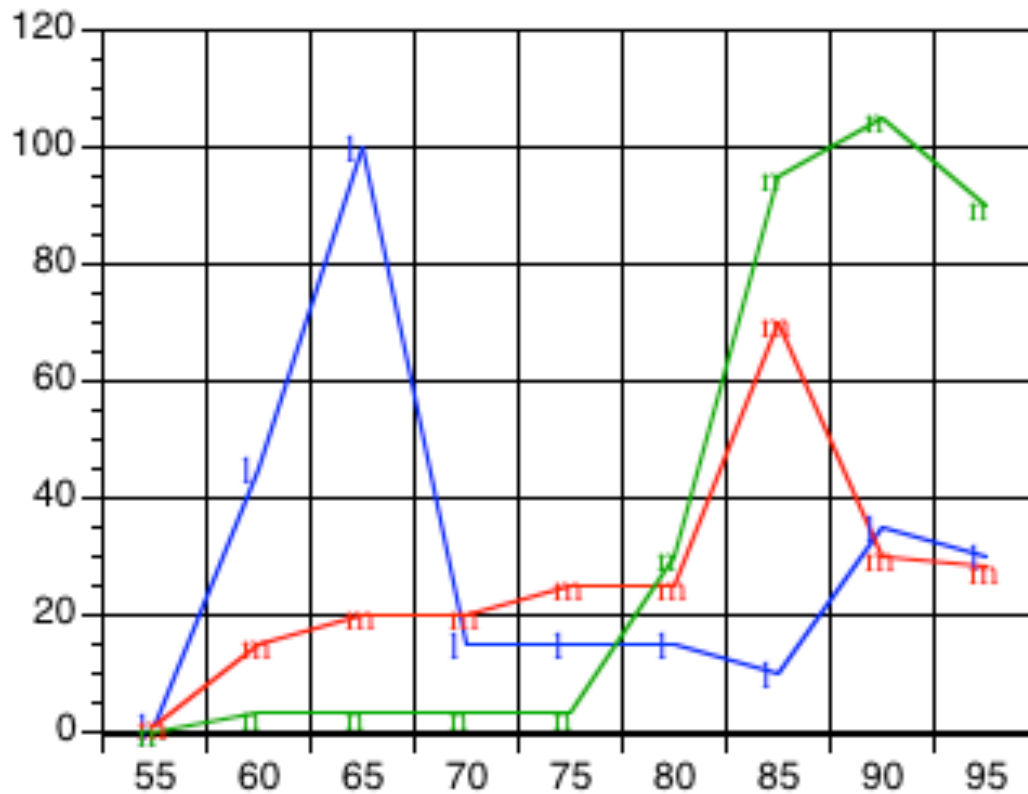


Fidélité

Étant donné la complexité du sens, la traduction ne peut être qu'approximative (jamais exacte).

Pour cette raison, on parle de la « fidélité » (non des « équivalences ») des traductions.

Dépenses pour la recherche en Traduction Automatique (en MF)



—+— Etats-Unis —m— Union Européenne —n— Japon

RÉALITÉ: nécessité de continuer les recherches



Théorie du Centrage
Métainformatif
MIC

Forme et Matière

	OBJET	APPROCHE	THÉORIE
FORME (structures)	<i>Type</i> <i>universel</i> <i>homogène</i> <i>statique</i>	<i>Déduction</i> <i>Synthèse par des</i> <i>règles</i>	$L = (V, G)$ <i>Langage = ensemble de phrases/</i> <i>textes <u>produit(e)s</u> par des règles</i> <i>de grammaire</i> <i>Sources des explications:</i> <i>uniquement à partir des règles</i>
MATIÈRE (données)	<i>Instance</i> <i>spécifique</i> <i>hétérogène</i> <i>dynamique</i>	<i>Induction</i> <i>Analyse par des</i> <i>analogies</i>	$L = (V, L)$ <i>Langage = ensemble de phrases/</i> <i>textes <u>reproductibles</u> à partir des</i> <i>mots</i> <i>Sources des prédictions:</i> <i>à partir des contextes (emplois)</i>

Structural and Computational Linguistics

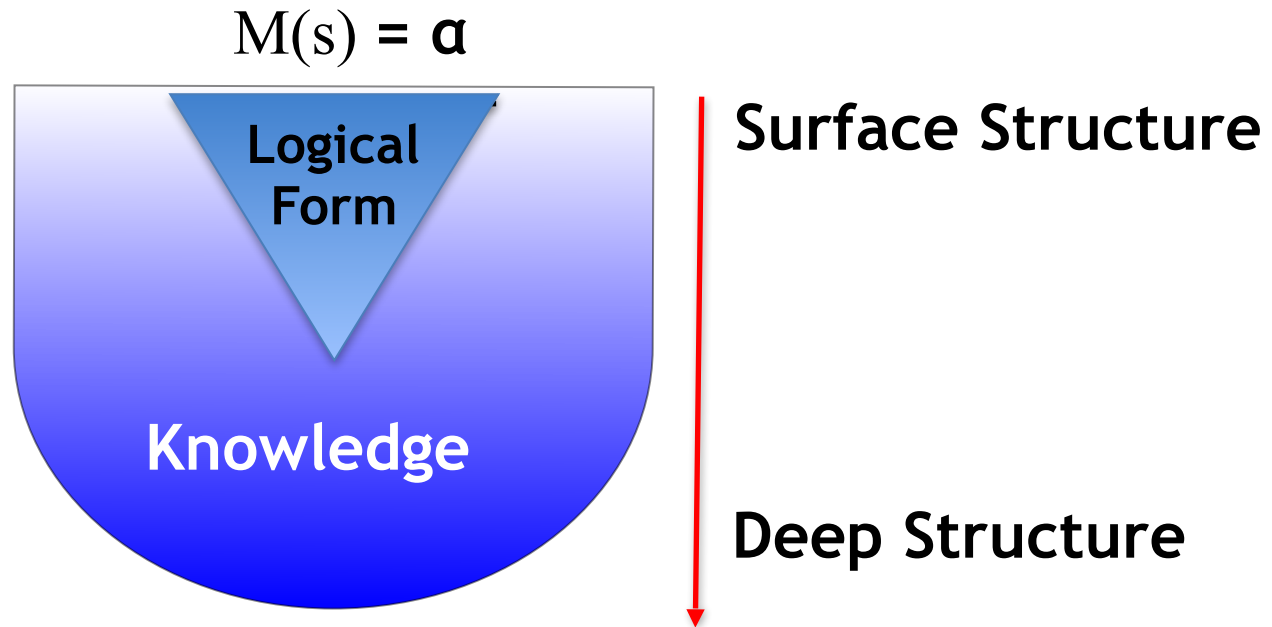
	Structural Linguistics	Computational Linguistics
FORM (Structures)	THEORY-oriented Linguistics (Formal Generative Linguistics)	NATURAL LANGUAGE PROCESSING (NLP) Lexicon-Functional Grammars Unification Grammars Logic Grammars, etc.
MATTER (Data)	DATA-oriented Linguistics (Linguistic Typology)	HUMAN LANGUAGE PROCESSING (HLP) Corpus Linguistics Interactive Linguistics

OBJECTS – APPROACHES - TASKS

Objects	Text Data	Symbolic Data
Approaches	Corpus Linguistics Text Document Exploration (Text Mining)	Interactive Linguistics Linguistic Knowledge Extraction (Data Mining)
Tasks	1. Selection 2a. Preprocessing 2b. Filtering	3. Transformation 4. Analysis 5. Evaluation

The Meaning of a Sentence in a mono-dimensional development hypothesis

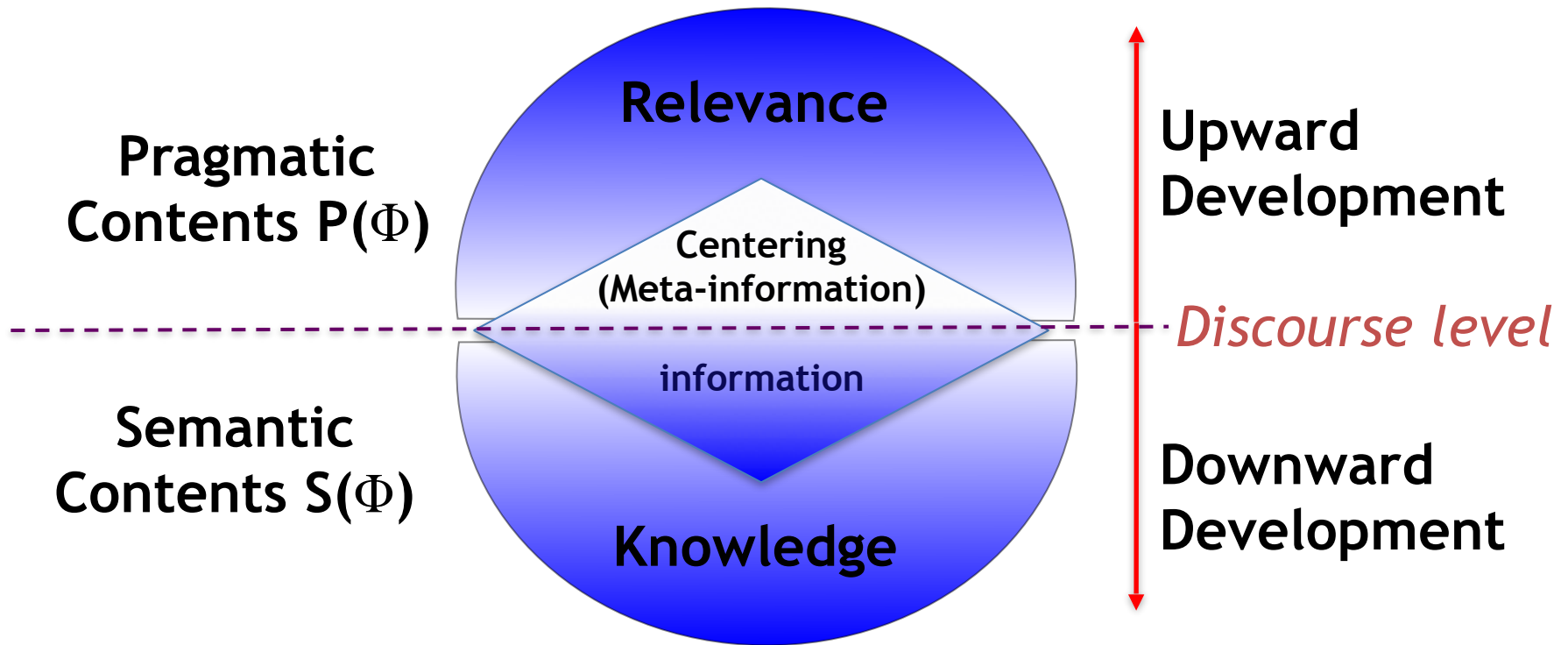
In **Generative Semantics**
 α stands for the *meaning of a sentence s as transformed into a Logical Form.*



The “Surface Structure” of a sentence s is *transformed* into the “*deep tree structure representation*” α

The Contents of Discourse

bi-dimensional development hypothesis



$S(\Phi)$ stands for *a set of information-related formulae*
 $P(\Phi)$ stands for *a set of centering-related formulae*

Les deux « axes » de la sémantique et de la pragmatique

	Sémantique	Pragmatique
<i>Deux « axes du discours »</i>	<i>Information</i>	<i>Méta-information</i>
Sélection (sur l'axe paradigmatique)	Comparaison des Propriétés $x \in X$ (espace d'attributs)	Centrage distinguer x
Combinaison (sur l'axe syntagmatique)	Création de scénarios $r(x)$ (<i>relation sur x</i>)	Prédication dire p à propos de x

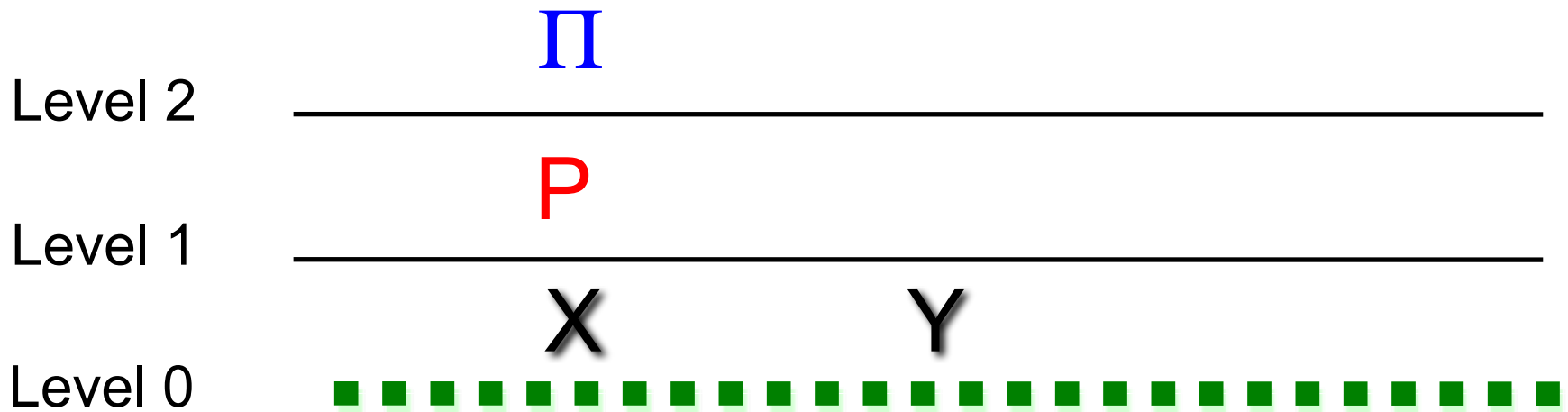
PROPRIÉTÉS de l'ATTENTION

Sélection

Orientation

Contrôle

Information and meta-information



Def. 1: X is a situation iff $\exists P$ such that $\text{represent}(P, X)$ – cf. mental situation

Def. 2: P is an information on X iff $\exists X, Y$ $\text{represent}(P, X) \wedge \text{not-represent}(P, Y)$

Def. 3: P is an absurd information iff $\forall P$ $\text{not-represent}(P, X)$

Def. 4: P is a trivial information iff $\forall P$ $\text{represent}(P, X)$

Def. 5: Π is a meta-information iff P is an information on X and Π is an information on P .

GRENIIEWSKI Henryk (1968) "Język nauki" (The Language of Science), *Zagadnienia Naukoznawstwa* (Problems of Epistemology), quarterly, vol. IV, tome 1(13), Polish Academy of Science, Warsaw (in Polish).

Centres d'Attention et Méta-information

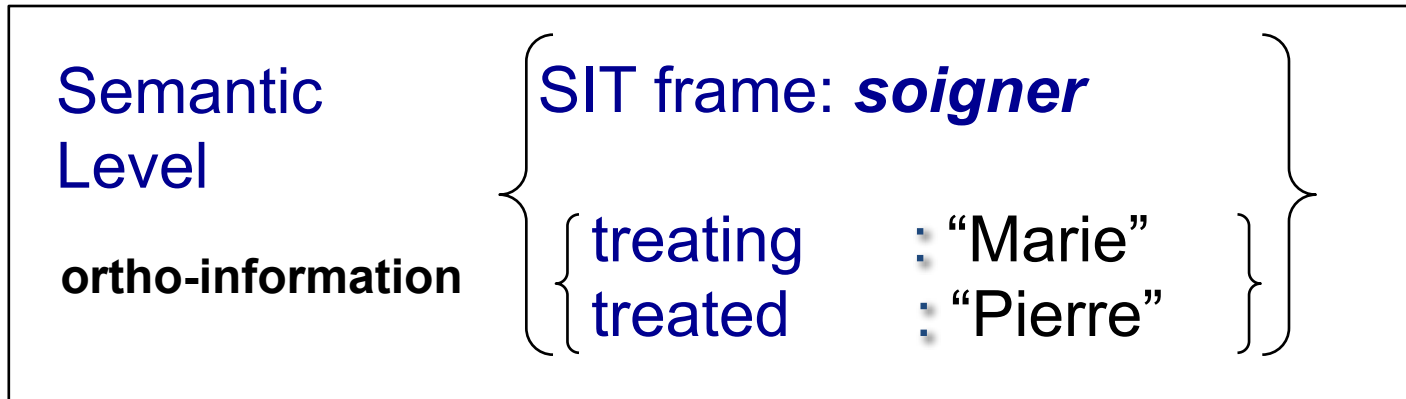
L'*information* contenue dans les énoncés est enrichie par des *pointeurs méta-informatifs* qui y distinguent les *centres d'attention* du locuteur.

La méta-information appartient à la **pragmatique** mais occupe une place centrale dans la grammaire de chaque langue puisqu'elle concerne la présentation linéaire du contenu sémantique dans l'énoncé.

On ne peut donc analyser l'énoncé sans tenir compte de sa structure méta-informative qui n'est pas une simple mise en ordre ou *présentation* s'ajoutant à la structure syntaxique.

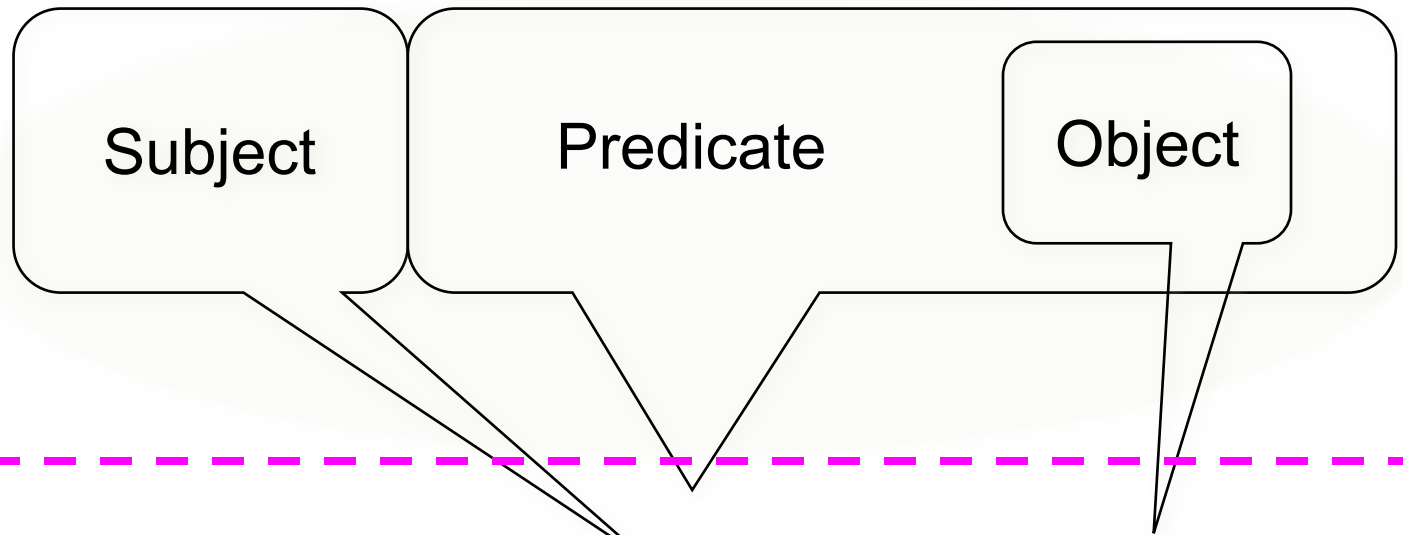
Semantic Content

Utterance: *Marie soigne Pierre.*



Syntactic Constituency as Meta-Information

Utterance: *Marie soigne Pierre*



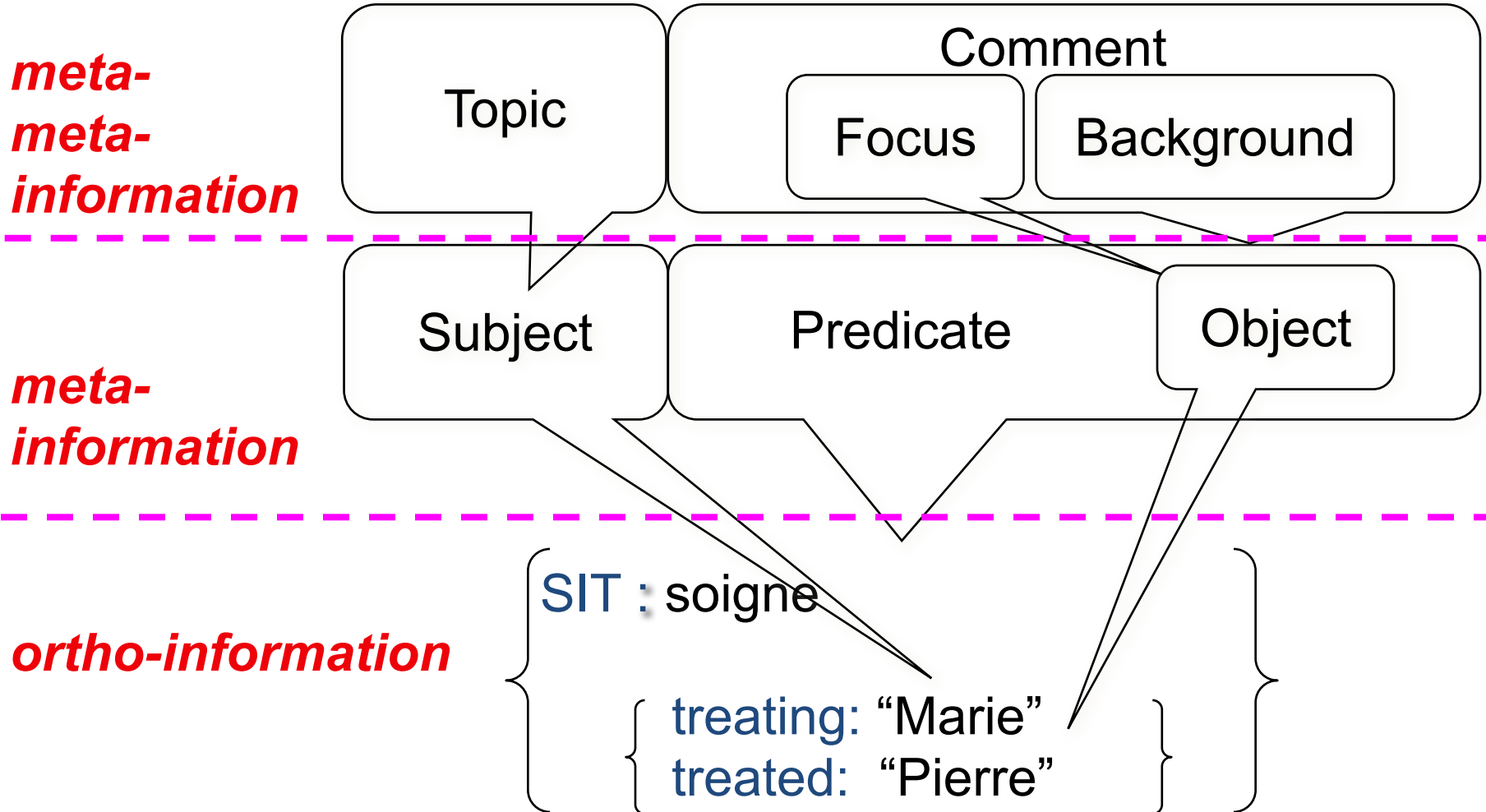
meta-information

ortho-information

SIT : soigne { treating: "Marie", treated: "Pierre" }

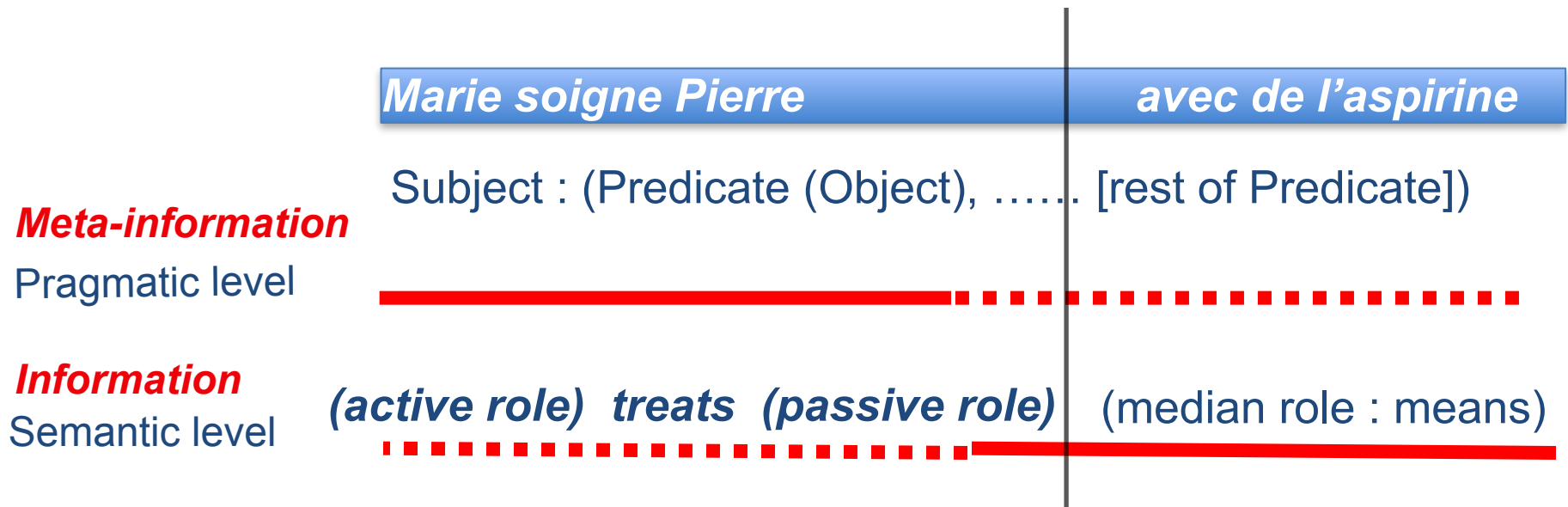
Predication and its Extensions

Utterance: *Quant à Marie, c'est Pierre qu'elle soigne.*



Semantic and pragmatic levels

Utterance: *“Marie soigne Pierre avec de l’aspirine.”*



Seul le rôle sémantique du moyen est exprimé (“avec”).

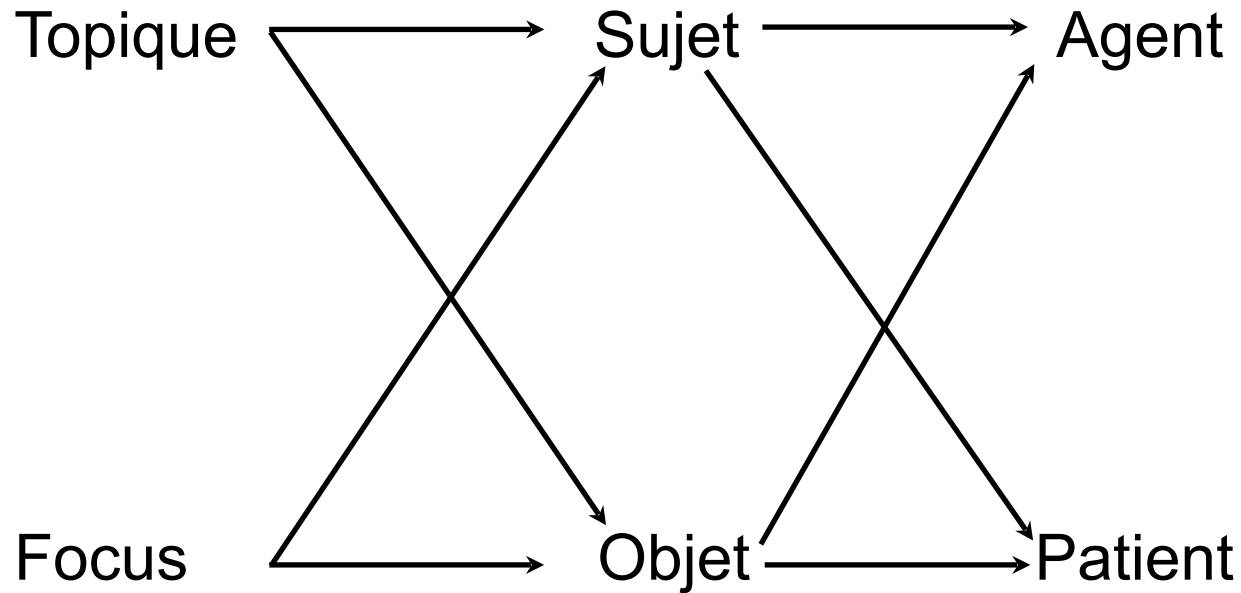
Homogeneous and Heterogeneous meta-informative status

Base Utterance (Schemas)	Base Utterance (Examples)
(New) Subject : (New) Predicate	#1 <i>Un satellite vient d'être lancé.</i>
(Old) Subject : (Old) Predicate	#2 <i>Les satellites tournent autour de la Terre.</i>
Extended Utterance (Schemas)	Extended Utterance (Examples)
(Old) Topic : (New) Comment	#3 <i>Quant au satellite X03, il a été détruit.</i>
(New) Focus : (Old) Background	#4 <i>C'est le satellite X03 qui a été détruit.</i>

Pivots du discours par niveau méta-informatif

Niveaux pragmatiques	Unités centrées par l'attention	
	Global	Local
Niveau 1 : (Énoncé de base)	Sujet	Objet
Niveau 2 : (Énoncé étendu)	Topique	Focus
Niveau 3 : (Dialogue/Texte)	Thème général	Thème particulier

Combinatoire des centres d'attention et rôles sémantiques



Combinaison

des centres d'attention et des rôles sémantiques

1a. Marie soigne Pierre.

(Actif + [Sujet || R actif] + [Objet || R passif])

1b. Pierre est soigné par Marie.

(Passif + [Sujet || R passif] + [Objet || R actif])

2a. Quant à Marie, elle soigne Pierre.

(Actif + [Topique || Sujet || R actif] + [Objet || R passif])

2b. Quant à Pierre, il est soigné par Marie.

(Passif + [Topique || Sujet || R passif] + [Objet || R actif])

3a. Quant à Marie, c'est Pierre qu'elle soigne .

(Actif + [Topique || Sujet || R actif] + [Focus || Objet || R passif])

3b. Quant à Pierre, c'est Marie qui le soigne.

(Actif + [Topique || Objet || R passif] + [Focus || Sujet || R actif])

4a. Quant à Pierre, c'est par Marie qu'il est soigné .

(Passif + [Topique || Sujet || R passif] + [Focus || Objet || R actif])

4b. ?? Quant à Marie, c'est par elle que Pierre est soigné .

(Passif + [Topique || Objet || R actif] + [Focus || Sujet || R passif])

.....

Équivalence sémantique et distinction pragmatique

Les énoncés A et B sont sémantiquement équivalents. Seuls les calculs « profonds » du sens les distinguent, pas le résultat: il s'agit de *l'interprétation des variables* dans A et de *l'application générative* dans B.

$$A. \text{ loves}(\text{John}, \text{Mary}) = B. \text{ loves}(\text{Mary})(\text{John})$$

Cependant, ces énoncés diffèrent au point de vue pragmatique. Les sujets de B sont plus « forts » (mis en exergue) par comparaison avec les sujets de A. Cela est dû au fait que le calcul du sens de B se fait en deux temps: 1° « localement » par rapport à l'agent passif et 2° « globalement » par rapport à l'agent actif.

English:

A. *John loves Mary.* ≠ B. *John is in love with Mary.*

French:

A. *Jean aime Marie.* ≠ B. *Jean est amoureux de Marie.*

Japanese:

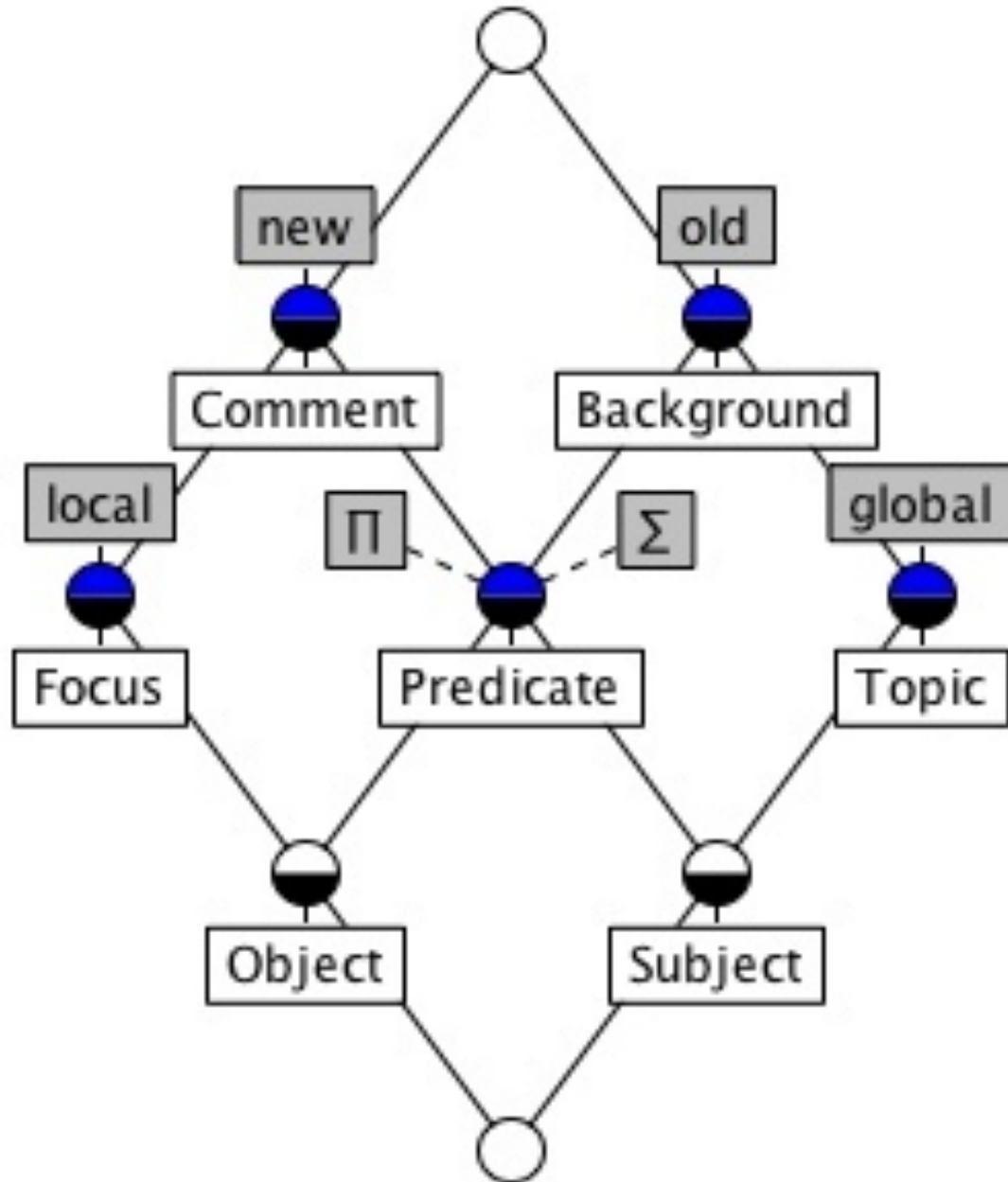
A. 太郎は花子をすきだ。 ≠ B. 太郎は花子がすきだ。

A. *Tarô wa Hanako o suki da.* ≠ *Tarô wa Hanako ga suki da.*

Polish:

A. *Janek kocha Marysię.* ≠ B. *Janek jest zakochany w Marysi.*

Structure du Centrage Méta-informatif



Le Centrage et l'Ordre des mots

Ordre des mots dans les langues du monde selon:

“The World Atlas of Language Structures Online” - WALS

Sujet-Objet-Verbe (SOV)	– 565	(41,03 %)
Sujet-Verb-Objet (SVO)	– 488	(35,43 %)
Verbe-Sujet-Objet (VSO)	– 95	(6,90 %)
Verbe-Objet-Sujet (VOS)	– 25	(1,81 %)
Objet-Verbe-Sujet (OVS)	– 11	(0,79 %)
Objet-Sujet-Verbe (OSV)	– 4	(0,29 %)
Manquant d'ordre dominant	– 189	(9,55 %)
<i>total: 1377 languages</i>		

Il est plus que probable que les langues VSO et OSV appartiennent au groupe des langues ergatives où le *rapport par défaut* concerne le sujet et l'agent passif qui dans le WALS est assimilé par erreur d'analyse à l'objet.

En ce qui concerne les ordres VOS et OVS,

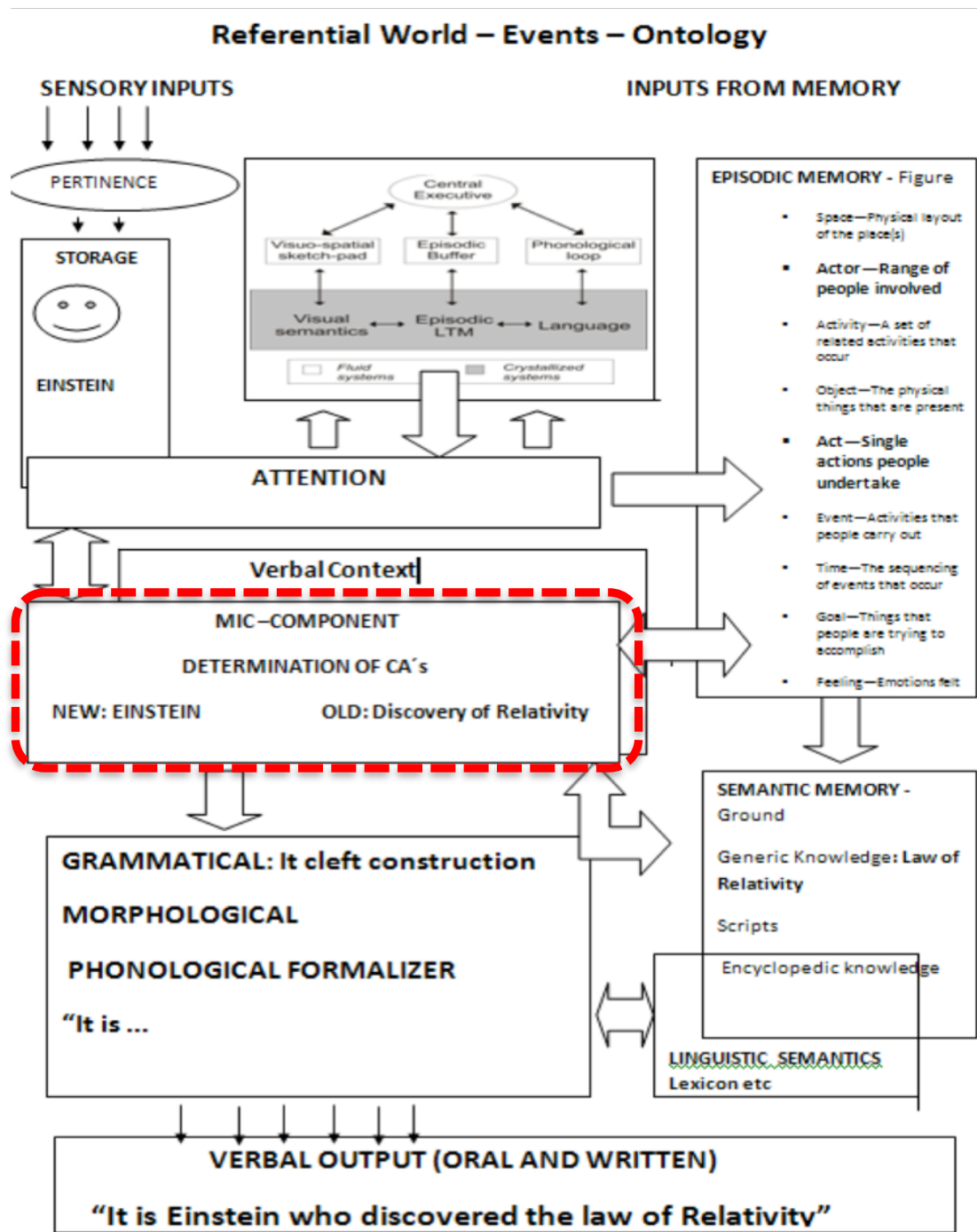
MIC and Neuroscience

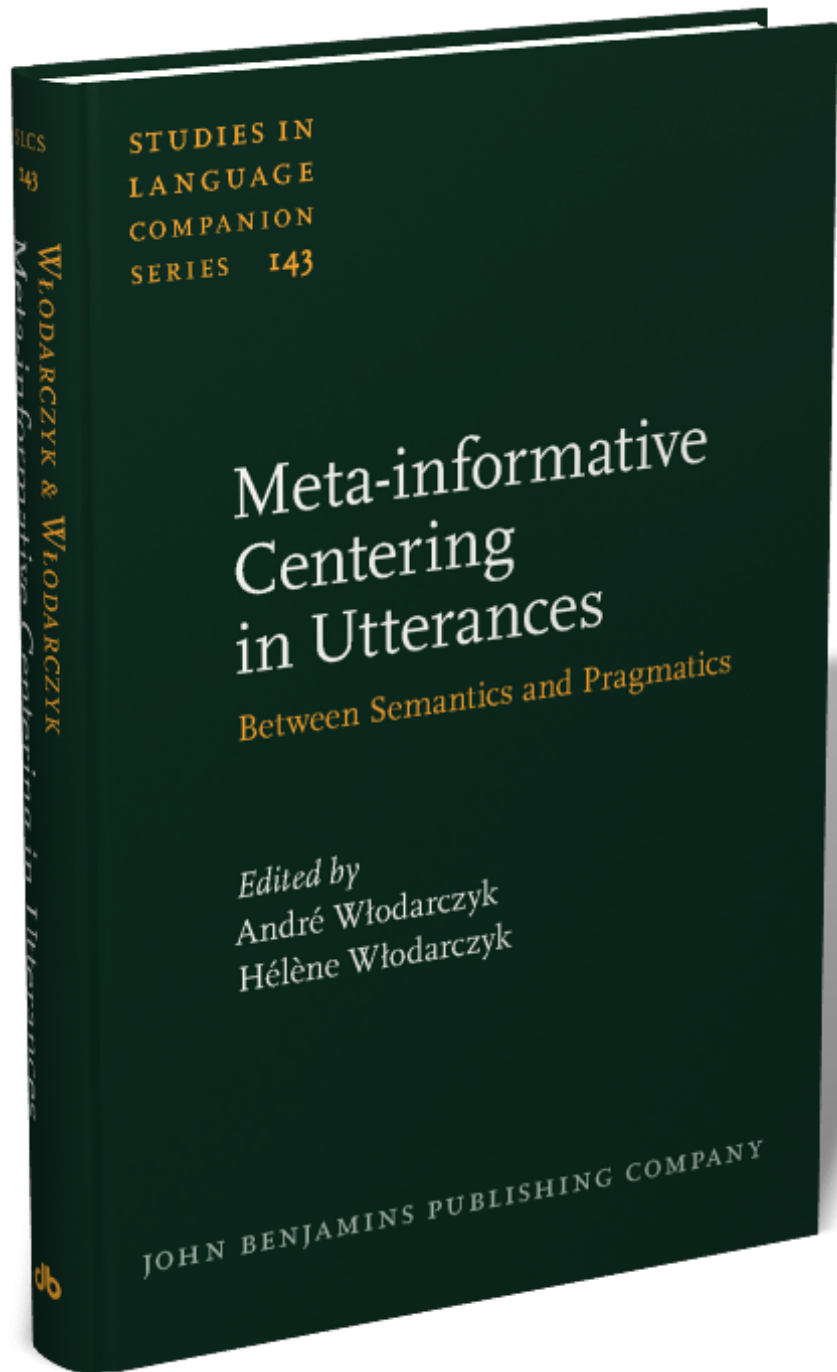
Meta-Informative Centering in Utterances - Between Semantics and Pragmatics, edited by André Włodarczyk & Hélène Włodarczyk, John Benjamins, Expected December 2013, xvii, 301 pp. + index, (page 115)

Chapter:

“Semantic and episodic memory by reference to the ontological grounding of the old and new meta-informative status”, by Franz J. Stachowiak
Charles Sturt University Albury, NSW, Australia

Embedding of MIC-Component in production of utterance





The MIC Book

The notion of information has nowadays become crucial both in our daily life and in many branches of science and technology. In language studies, this notion was used as a technical term for the first time about at least fifty years ago. It is argued, however, that "Old" and "New", used traditionally for characterising information, refer in fact to the meta-informative status of communicated chunks of information. They provide information about other information. Since subjects and objects, as attention-driven phrases, are also related to aboutness, the presented Meta-Informative Centering (MIC) framework includes predication theory.

By applying the MIC theory to their analyses of English, German, French, Polish, Russian, Greek, Latin, and Japanese, the authors provide comprehensive explanations of the most puzzling aspects of the pragmatic use of basic universal linguistic categories. It seems clear now that canonical syntactic patterns, their permutations, and diverse transformations do indeed reflect very truly the meta-informative encapsulation of utterances.

As a consequence, this book presents new and coherent theoretical solutions as well as their very efficient applications.

<https://benjamins.com/#catalog/books/slcs.143/>

Conclusion

Le Langage naturel est un système complexe. Pour bien comprendre ce système, il ne suffit pas de l'aborder d'un seul point de vue. Une recherche interdisciplinaire s'impose.

Cela permettra d'améliorer:

1. l'enseignement des langues
2. la traduction automatique
3. la compréhension des langues par les machines
4. tous les outils de traitement de textes et d'aide à la rédaction et
5. de soigner les pathologies du langage.

Merci de votre attention

wlodarczyk.andre@gmail.com