

LA REPRÉSENTATION COGNITIVE DU TEMPS ET DE L'ESPACE

ÉTUDE PRAGMATIQUE DES DONNÉES LINGUISTIQUES EN FRANÇAIS ET DANS D'AUTRES LANGUES

Tijana ASIC

Thèse de linguistique et sciences cognitives, Université de Genève, présentée le 6 septembre 2004
Mél : tijanaasic@hotmail.com

Thesis title — *Cognitive representation of time and space: a pragmatic study of linguistic data in French and other languages.*

Título da tese — *A representação cognitiva do tempo e do espaço: estudo pragmático dos dados lingüísticos em francês e outras línguas.*

Título de la tesis — *La representación cognitiva del tiempo y del espacio : estudio pragmático de los datos lingüísticos en francés y en otros idiomas.*

Titolo della tesi — *La rappresentazione cognitiva del tempo e dello spazio. Studio pragmatico dei dati linguistici dal francese e da altre lingue.*

Le but principal de ma thèse de doctorat était de répondre aux trois questions : 1) Pourquoi l'espace est-il utilisé pour parler du temps ? 2) Comment est-il utilisé pour parler du temps ? 3) Avec quelles variations à travers les langues ? Je suis donc partie de deux hypothèses déjà connues en linguistique et en sciences cognitives : il s'agit de l'hypothèse du localisme linguistique (Lyons, 1977) affirmant que les expressions spatiales sont grammaticalement et sémantiquement fondamentales, et de l'hypothèse des relations thématiques de Jackendoff (1985, p.209) d'après laquelle « la cognition de l'espace précède celle du temps ».

Après avoir fait une revue détaillée des travaux existants sur la linguistique et la conceptualisation du temps et de l'espace, j'ai constaté qu'on a besoin de constituer une ontologie commune pour l'espace et le temps. Pour le faire, je me suis principalement basée sur les relations méréo-topologiques¹ proposées par Casati et Varzi (1999) et sur la notion d'orientation issue de l'existence des trois différents cadres de références² (Levinson, 2003).

Cette ontologie spatio-temporelle de base m'a permis de définir de façon formelle les sémantismes de base des prépositions spatiales, temporelles et spatio-temporelles en français. Il faut souligner que mes définitions de prépositions sont minimalistes :

¹ La méréologie traite les relations du type partie-tout. La topologie est l'étude de différents types de connexion entre les entités.

² Ce sont les représentations géométriques qui supportent la cognition spatiale. Les cadres de référence servent à établir les coordonnées de l'objet référentiel, à partir desquelles on situe l'objet focal (l'objet dont la position est inconnue).

une définition minimaliste d'une préposition doit satisfaire toutes les variétés de ses usages (standard et non-standard, par exemple les usages temporels des prépositions spatiales). Elle doit aussi laisser la place aux inférences contextuelles, à savoir les hypothèses que l'on tire du contexte dans lequel l'énoncé est produit. Ainsi, j'ai adopté une perspective pragmatique³ — il s'agit de la théorie de la pertinence (Sperber et Wilson, 1986).

Voici un exemple de mes définitions minimalistes :

Définition de *sur* : $xsury = df ECxy$

$xsury = df ECxy$

(x est sur y est égal par définition à x est connecté de façon externe avec y)

Manifestement, la relation désignée par *sur* est basée sur le prédicat méreo-topologique *connexion externe* dont voici la définition :

$ECxy = df Cxy \wedge \neg Oxy$

(x est connecté de façon externe à y est égal par définition à x est connecté à y et x ne recouvre pas y).

Il est clair que dans la perspective que j'ai adoptée la préposition *sur* ne désigne pas par défaut la relation *porteur-porté*. Cette relation est un des effets interprétatifs (les implications, à savoir les conclusions faites à partir de la forme logique de l'énoncé et du contexte) associés à la préposition *sur*. Je voudrais souligner que la définition minimaliste de *sur* permet d'expliquer non seulement ses usages spatiaux mais aussi ses usages temporels, qui sont peu nombreux en français mais très courants dans les autres langues que j'ai analysées (*on Monday = lundi* en anglais ; *na taj dan / sur ce jour/ ce jour là* en serbe etc.).

En somme, les résultats de mes analyses des prépositions en question en français montrent que la division classique des prépositions entre les prépositions spatiales et spatio-temporelles est artificielle surtout dans une perspective diachronique. À titre d'exemple la préposition *devant* est aujourd'hui considérée comme purement spatiale, mais elle avait avant des usages temporels comme *chanter devant le coq*. En effet la plupart des prépositions spatiales ont des usages temporels parce que leurs sémantismes reposent sur les prédicats (tels que *connexion* ou *inclusion*) possibles dans les deux domaines.

Cependant, les prépositions uniquement spatiales existent (*au fonds de*, *à gauche de*, etc.) — elles décrivent des relations qui existent dans les dimensions purement spatiales. Donc, leurs sémantismes reposent sur une ontologie spatiale trop complexe pour le temps qui peut être vu comme un pseudo-espace mono-dimensionnel. Les prépositions à usages exclusivement temporels existent aussi (*lors de*, *pendant*) — la relation de localisation qu'elles décrivent, bien qu'exprimée en termes spatiaux, n'est possible que pour des entités temporelles à cause de leur nature ontologique (elles peuvent occuper le même endroit en même temps). Je voudrais faire ici une petite observation : même lorsque ces prépositions sont employées avec des arguments non-temporels (comme dans : *pendant dix kilomètres*, *le chemin*, *la route*) il s'agit de faux usages spatiaux. Le temps écoulé y est représenté par la distance parcourue.

³ La pragmatique peut être définie comme l'étude de l'usage du langage.

Dans la partie contrastive de mon travail, j'ai analysé les usages non-standards des prépositions et des expressions déictiques en anglais, serbe (et parfois les autres langues slaves), swahili, kikuyu, japonais, arabe et louo. Mes analyses contrastives montrent qu'il n'y a pas de différences majeures dans les sous-systèmes des prépositions spatio-temporelles. Parmi les langues examinées, il n'y en a aucune qui ait des prépositions exclusivement spatiales et temporelles sans la possibilité d'avoir des prépositions qui fonctionnent dans les deux domaines.

Quant aux variations existantes, elles peuvent être expliquées par la théorie de l'optimalité de Prince et Smolensky (1993), à savoir par la compétition entre la force de fidélité (« exprimer linguistiquement les contrastes ») et la force de marquage (« éviter les structures marquées »). Parfois c'est la première, parfois c'est la deuxième force qui gagne. De plus il faudrait souligner ici que, si une certaine relation n'est pas linguistiquement exprimée, cela ne signifie pas qu'elle ne soit pas inférable et concevable. Par conséquent, les différences entre les langues ne révèlent pas des différences dans la conceptualisation du temps et de l'espace (ce qui conteste l'hypothèse Sapir-Whorf du Déterminisme linguistique qui pose que la langue que nous parlons détermine notre conceptualisation du monde).

Il existe quand même certaines différences entre les langues que l'on pourrait résumer de la manière suivante : il y a des langues qui ont un système de prépositions spatiales et temporelles riche et un système de temps verbaux riche (comme le français), des langues qui ont un système de prépositions spatiales et temporelles riche et un système de temps verbaux pauvre (comme les langues slaves modernes), des langues qui ont un système de prépositions spatiales et temporelles pauvre et un système de temps verbaux riche (comme le kikuyu) et finalement des langues qui ont un système de prépositions spatiales et temporelles pauvre et un système de temps verbaux pauvre (comme le japonais et l'hébreu).

Par ailleurs, j'ai aussi traité le phénomène de la coïncidence ou de la non-coïncidence des points S et R proposés par Reichenbach (1947) pour le temps (S = le moment ou l'endroit de la parole ; R = le point de référence) en les étendant à l'espace. Il s'agissait de vérifier si l'explicitation de l'opposition $S=R/S \neq R$ se retrouve pour le temps et pour l'espace, dans les langues mentionnées. J'ai ainsi pu déterminer si ces langues comprennent des marques linguistiques différentes pour expliciter l'opposition entre le passé composé et le passé simple, ainsi que les oppositions lexicales suivantes : *dans trois jours/trois jours après ; demain/le lendemain ; aller/venir ; devant moi/devant la voiture*.

À partir de cela et à partir d'une opposition déjà connue en linguistique (la différence entre les expressions conceptuelles, qui encodent des concepts, et procédurales, qui encodent des procédures ; voir Blakemore, 1987), j'ai pu définir quatre catégories d'expressions : les expressions SC (spatiales et conceptuelles), les expressions SP (spatiales et procédurales), les expressions TC (temporelles et conceptuelles) et les expressions TP (temporelles et procédurales). Dans toutes les langues examinées, on obtient des résultats uniformes (positifs) pour les expressions SC, TC et SP. La seule différence existe dans la catégorie TP : certaines langues ont des paires de temps verbaux qui explicitent l'opposition en question (le français, l'anglais, le serbe, le swahili, le kikuyu, le louo) et certaines n'en ont pas (le japonais, l'hébreu et l'arabe). À titre d'exemple, en français, on a l'opposition entre les deux phrases citées ci-dessous :

(1) *J'ai mangé, je n'ai plus faim.*

(2) *Ce jour-là il mangea dans le restaurant « Chez maître Paul ».*

En japonais dans les deux phrases on emploie le même temps verbal.

De plus, certaines langues explicitent l'opposition *dans/après* (le français, l'anglais, le serbe, le swahili, le kikuyu, l'hébreu) et certaines non (l'arabe, le louo, le japonais). Ainsi, on a :

(3) *Dusan part dans trois jours.*

(4) *Dusan est parti / partira trois jours après Noël.*

Il est clair que dans l'exemple 3 on réfère au moment qui aura lieu trois jours après le moment de la parole (S), alors que dans l'exemple 4 on réfère au moment qui a eu lieu/aura lieu trois jours après un autre moment (non équivalent à S). En arabe, louo et japonais on emploie dans les deux cas la même préposition.

J'aimerais ajouter ici que, outre mon travail sur les prépositions et les temps verbaux, j'ai aussi analysé dans ma thèse les systèmes déictiques dans les langues mentionnées et notamment les usages temporels des déictiques spatiaux (*ici* temporel en français, *hapa* et *hapo* temporels en swahili et *tu* temporel en serbe). Il s'agit d'un phénomène sémantique et pragmatique universel qui est la conséquence de la nature indexicale des expressions déictiques : elles servent à pointer soit à un endroit pertinent, soit à un moment saillant dans la narration.

L'apport principal de ma thèse est non seulement dans le fait que j'ai trouvé des arguments pour défendre l'hypothèse du localisme linguistique, mais aussi dans le fait que j'ai constaté qu'il n'y a pas de relations exclusivement temporelles, mais qu'il y a des relations exclusivement spatiales, tout simplement parce que l'espace est tridimensionnel (plus complexe). Mais la différence est avant tout dans la nature ontologique des entités spatiales et temporelles. De plus, j'ai montré qu'il est possible de formuler un certain nombre de primitives conceptuelles (pour l'espace et le temps) tels que l'opposition *massif – comptable*, l'opposition *statique – dynamique*, *contact* et *contenance*.

Références bibliographiques

- Blakemore D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Casati R., Varzi A. C. (1999). *Parts and places, the structure of spatial representation*. Cambridge, Massachusetts, London, England: A Bradford Book. The MIT Press.
- Jackendoff R. (1985). *Semantics and Cognition*. Cambridge, MA: The MIT Press.
- Levinson S. C. (2003). *Space in language and cognition: Explorations in linguistic*. University of Cambridge: Cambridge University Press.
- Lyons J. (1977). *Semantics*. Cambridge University Press: Cambridge.
- Prince A. S., Smolensky P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Cambridge, Mass: MIT Press.
- Reichenbach H. (1947). *Elements of Symbolic Logic*. New York: Free Press.
- Sperber D., Wilson D. (1986). *Relevance, Communication and Cognition*. Blackwell:Oxford.

L'auteur

Tijana Asic est actuellement maître de conférences en linguistique à l'Université de Belgrade (Serbie). Elle est membre du groupe de recherche sur la causalité à l'Institut des Sciences Cognitives (CNRS Lyon) ainsi que du groupe de recherche sur la pragmatique de la référence temporelle à l'Université de Genève. Elle travaille sur la sémantique et la pragmatique de l'expression du temps et de l'espace ainsi que sur les expressions indexicales. En 2003, elle a obtenu le « Prix Naylor » de Ohio State University pour le meilleur article sur les langues slaves, et en 2005, le « Prix Charles Bally » pour la meilleure thèse de linguistique à Genève.

MÉTADYNAMIQUES EN COGNITION SOCIALE QUELLE DÉFINITION DE MEILLEUR EST LA MEILLEURE ?

David CHAVALARIAS

*Thèse de sciences cognitives et sciences des systèmes complexes,
École Polytechnique, présentée le 25 Octobre 2004.
Mél : david.chavalarias@shs.polytechnique.fr ; <http://chavalarias.free.fr>*

Thesis title — *Metadynamics in social cognition: which definition of “best” is the best?*

Título da tese — *Metadinâmicas em cognição social. Qual de difinição de melhor é a melhor?*

Título de la tesis — *Metadinámicas en cognición social. ¿Cuál definición de mejor es la mejor?*

Titolo della tesi — *Metadinamiche della cognizione sociale. Qual è la definizione migliore?*

Résumé

Quels sont les principes fondateurs de l’auto-organisation des sociétés humaines ? Produits d’une évolution culturelle rapide, qui a pris le pas sur l’évolution biologique, celles-ci réclament une approche par des formalismes radicalement différents de ceux utilisés généralement dans l’étude des autres sociétés animales.

En rupture avec les approches formelles traditionnelles s’appuyant généralement sur l’optimisation d’un critère global, nous proposons une approche qui tire parti des spécificités de l’imitation humaine pour interpréter l’hétérogénéité auto-organisée des systèmes sociaux humains comme le résultat de la cognition sociale, i.e. une différenciation par un processus de co-évolution d’une multiplicité de critères possibles.

Mots clefs : systèmes complexes, auto-organisation, autonomie, cognition sociale, imitation, réflexivité, méta-cognition, individualisme méthodologique complexe, co-évolution culturelle.

Abstract

What are the principles behind the self-organization of human societies? Studying the self-organization in human societies calls for models that are fundamentally different from those models used in studying other animal societies because cultural evolution occurs faster than biological evolution and is in fact a driving force of biological evolution.

Departing from the traditional approaches generally based on a global optimization process driven by a single criterion, this leads us to interpret the heterogeneity in self-organized human social systems as the outcome of social cognition, i.e. a differentiation resulting from a co-evolutionary process with a multiplicity of criteria.

Keywords: complex systems, self-organization, autonomy, social cognition, imitation, reflexivity, metacognition, complex methodological individualism, cultural co-evolution.

1. Contexte de recherche

La grande majorité des travaux de modélisation en sciences sociales et humaines considère des agents structurés sous la forme d'une hiérarchie de règles, chacune se plaçant en position méta par rapport aux règles du niveau inférieur. Celles-ci peuvent s'interpréter selon les niveaux comme des règles de comportement, des règles de décision, des règles de transmission culturelle ou des règles d'évolution de la population. Or, la dynamique de la plupart des systèmes sociaux qu'il nous est donné d'observer est sous-déterminée par des contraintes de type biologique ou économique, alors que la plupart des études empiriques semblent indiquer une forte hétérogénéité sur les méta-règles de comportement.

Cette remarque m'a amené à proposer une approche de la modélisation des systèmes sociaux qui s'articule autour de la question suivante : Peut-on endogénéiser la distribution sur les méta-règles de comportement dans le cadre d'une modélisation des systèmes sociaux, de manière à ce qu'elle soit le produit de la dynamique collective qu'elle définit ?

Si la plupart des formalismes en sciences sociales sont à même de représenter une hétérogénéité des systèmes de règles utilisés par les individus, un problème majeur se pose quant à sa nature et à son origine. Celui-ci ne peut se poser qu'avec pour toile de fond la question de l'évolution culturelle. Théâtres de l'émergence de ce que Hayek a qualifié d'ordre « spontané », les systèmes sociaux évoluent sous l'effet de forces internes.

Mais c'est précisément là que le bât blesse, car les modalités de cette auto-organisation sont des plus difficiles à définir. Il serait même impossible, d'après le pseudo-théorème d'Ashby (1962), de définir formellement une véritable auto-organisation au sens où il est toujours nécessaire de définir une ultime méta-règle d'organisation fixe, qui serve de cadre à cette auto-organisation.

2. Imitation et cognition sociale

Parmi les principes organisateurs candidats, l'imitation a souvent été plébiscitée. Cependant, la donnée d'une définition précise de l'imitation sous la forme d'une règle d'évolution oblige le modélisateur à définir un critère d'imitation, donnant à celle-ci le statut de règle fixe, qui nous renvoie à la position d'Ashby.

La direction de recherche que j'ai proposée est une manière de dissocier le principe d'imitation des éléments qui entrent dans la description des formes d'imitation. Ceci nous est suggéré par l'analyse de travaux en psychologie et en éthologie, qui mettent en évidence l'importance de la réflexivité et de la méta-cognition dans la définition des processus cognitifs propres à l'espèce humaine. Dans l'étude des phénomènes mimétiques nous pouvons envisager grâce à deux qualités les règles d'imitation comme des processus cognitifs modifiables, ce qui permet d'effectuer cette dissociation en posant que les composants qui définissent une règle d'imitation peuvent également être sujets à imitation. Dans ce cas de figure, *les règles d'imitation peuvent être leur propre méta-règle.*

Cette approche m'a permis de définir des systèmes particuliers, les *jeux métamimétiques*, qui formalisent cette dissociation. J'ai ainsi pu proposer un amendement à l'argument d'Ashby : si l'on ne peut modéliser l'auto-organisation au sens de l'émergence *ex-nihilo* de nouvelles règles d'organisation, il peut néanmoins

y avoir auto-organisation au sens de *sélection spontanée d'une règle d'organisation* dans un ensemble de règles possibles au cours d'un processus stylisé de cognition sociale.

Cette approche des systèmes sociaux est donc en rupture avec les approches formelles traditionnelles s'appuyant généralement sur l'optimisation d'un critère global. Elle se propose d'interpréter l'hétérogénéité auto-organisée des systèmes sociaux humains comme une différenciation par un processus de co-évolution d'une multiplicité de critères possibles, de « définitions de meilleur », qui dès lors ne peuvent se concevoir que de manière contextuelle et d'un point de vue individuel. Une application de ce type d'approche est proposée autour du problème de l'émergence de la coopération. Nous montrons au passage, que cela permet d'échapper au fameux dilemme du prisonnier.

Références bibliographiques

Ashby W.R. (1962). Principles of the self-organizing system. In Heinz von Foster, Zopf H. (eds.), *Principles of self-organization*. New-York: Pergamon. 255-278

L'auteur



David Chavalarias est en post-doc à l'École Polytechnique sur le thème de la cognition sociale, en tant que domaine de convergence des sciences cognitives et des sciences des systèmes complexes. Après des études de mathématiques à l'École Normale Supérieure de Cachan, il s'est orienté vers les sciences cognitives et s'est intéressé aux processus de découverte collective et au problème de l'émergence de la coopération. Il a ensuite effectué une thèse à l'École Polytechnique sous la direction de Paul Bourguin sur le thème de la cognition sociale. Informations complémentaires : <http://chavalarias.free.fr>

REPRÉSENTATIONS VERBALES MULTISTABLES EN MÉMOIRE DE TRAVAIL

VERS UNE PERCEPTION ACTIVE DES UNITÉS DE PAROLE

Marc SATO

Thèse de Sciences Cognitives, Institut National Polytechnique de Grenoble, présentée le 30 Septembre 2004.

Mél : marc.sato@nemo.unipr.it

Thesis title — *Multistable word representations in working memory: towards an active perception of speech units*

Título da tese — *Representações verbais multiestáveis na memória de trabalho. Para uma percepção ativa das unidades da fala.*

Título de la tesis — *Representaciones verbales multiestables en memoria de trabajo. Hacia una percepción activa de las unidades del habla.*

Titolo della tesi — *Rappresentazioni verbali multistabili in memoria di lavoro. Verso una percezione attiva delle unità di parola.*

Constitutives d'un nouveau chapitre de la cognition — la cognition motrice — les études des systèmes cérébraux et cognitifs spécialisés dans l'émergence et le traitement des représentations d'action, ont permis de montrer les liens étroits unissant l'observation, la représentation mentale d'une action, enfin, son exécution. Des premières données comportementales, soulignant le couplage fonctionnel entre production et représentation du mouvement, aux travaux récents de neurophysiologie et de neuroimagerie, le concept de représentation d'action s'est étendu au-delà de nos propres actions à celui de « représentations partagées » impliquées dans la communication et l'intersubjectivité. Face à cette extension du champ de l'action produite ou observable à l'activité de représentation, et dans le cadre des travaux visant à démontrer l'existence d'une instanciation physique et neurophysiologique, donc par la substance, des principes et paramètres de la phonologie, l'objectif premier des travaux réalisés dans le cadre de cette thèse était de montrer l'existence possible de contraintes purement motrices, liées à l'appareil périphérique articulatoire, dans l'analyse et la construction des représentations mentales phonologiques.

Dans ce but, nous avons testé la sensibilité de l'effet de transformation verbale, un phénomène de perception multistable de la parole lié aux changements pouvant survenir lors de la répétition ou lors de la perception prolongée de séquences langagières à des contraintes de production de la parole. Plus spécifiquement, nous nous sommes intéressés aux variations de stabilité perceptive pouvant survenir lors de la répétition, ouverte ou interne, de séquences monosyllabiques et dépendantes de mécanismes dynamiques articulatoires. Les données expérimentales démontrent clairement l'existence d'un biais d'asymétrie des transformations observées, biais relevant de contraintes purement motrices, actionnelles, liées aux relations de phase,

et donc à la cohérence, entre gestes articulatoires des séquences répétées. La conservation des patrons d'attractivité de ces séquences, d'une condition de répétition à voix haute vers une condition de répétition mentale, souligne la perméabilité des représentations mentales phonologiques à des contraintes purement motrices et atteste des liens unissant la représentation mentale d'une action et son exécution.

Dans le cadre des recherches portant sur l'examen des interactions possibles entre les grandes tendances des systèmes phonologiques des langues du monde et les contraintes de production et de perception de la parole, nous avons alors testé, lors de la perception prolongée de séquences réversibles dissyllabiques, l'impact perceptif d'une tendance organisationnelle spécifique du langage, l'effet « LC », indiquant une alternance consonantique préférentielle de type Labiale-(Voyelle)-Coronale plutôt que Coronale-(Voyelle)-Labiale dont les fondements seraient liés à des contraintes biomécaniques articulaires. L'existence d'une asymétrie des transformations en faveur de la structure Labiale-Coronale atteste de la validité perceptive de l'effet LC durant une tâche de transformation verbale et souligne de fait une interaction entre systèmes de production et de perception de la parole lors de la construction et l'émergence des représentations phonologiques.

Lors de ces deux premières études, c'est ainsi le potentiel de mise en forme du langage par les mécanismes de perception et d'action que nous avons voulu interroger, en adoptant une position visant à étudier la perception, comme la production, des unités de parole sous l'angle des interactions sensori-motrices et de leur lien constant avec la phonologie, ce par le biais de leurs structururations réciproques, ou co-structurations, et non comme deux systèmes indépendants et séparément étudiés. Les résultats obtenus pourraient apporter de nouveaux arguments expérimentaux en faveur d'une « théorie de la perception pour le contrôle de l'action », développée par les chercheurs de l'Institut de la Communication Parlée, pour laquelle la perception des unités de parole consisterait en l'ensemble des processus perceptifs permettant au niveau segmental de récupérer et contraindre les cibles et phases des gestes vocaux, fournissant par là-même un ensemble de représentations utilisées pour le suivi et la compréhension des actions perçues et, en retour, pour le contrôle et la spécification de nos propres actions.

Dans le cadre de cette théorie, cette interaction locuteur-auditeur consisterait donc en l'élaboration progressive de représentations internes sensori-motrices, issues du traitement des entrées sensorielles auditives et contraintes par les propriétés du système articulatoire. De manière plus spécifique, l'estimation des caractéristiques spectrales et événements temporels contenus dans le signal acoustique permettrait une première définition des cibles et du *timing* des gestes articulatoires associés. Ces traitements auditifs ne permettant qu'une caractérisation partielle des procédures de régularisation ancrées dans les mécanismes de contrôle articulatoire, contribueraient alors à une mise en forme spécifiée, une structuration, des représentations construites. Enfin, cette récupération et mise en forme des gestes de parole s'appuierait sur la nature multimodale du système perceptif, la vision pouvant servir de relais à ces mécanismes perceptifs. Les représentations ainsi construites seraient donc intrinsèquement sensori-motrices, ni purs produits sensoriels, ni purs objets moteurs inférés, mais des percepts multimodaux régularisés par l'action.

Dès lors, une question importante est celle de l'identification et localisation des processus cognitifs et cérébraux mis en œuvre lors de l'intégration et du traitement de ces représentations sensori-motrices. Un candidat susceptible de fournir une

architecture adaptée aussi bien au traitement phonologique en ligne qu'au suivi attentionnel en parole, est la boucle phonologique, la composante verbale du modèle de mémoire de travail de Baddeley et Hitch. Une étude d'imagerie par résonance magnétique fonctionnelle nous a permis de tester l'implication de ce système mnésique dans l'analyse, le contrôle et l'émergence de séquences articulatoires, lors de la recherche active de transformations verbales. L'observation d'un réseau neural activé similaire à celui observé dans des tâches de mémoire de travail, démontre l'engagement de la boucle phonologique lors des processus de *parsing* syllabique, de maintenance attentionnelle de l'information et de prise de décision nécessaires à l'émergence des transformations verbales, et suggère sa possible sensibilité et dépendance non seulement à des principes d'organisation phonologique mais aussi à des contraintes spécifiques purement motrices.

Mis en relation, ces travaux s'inscrivent dès lors dans une problématique plus large, celle du rôle possible de la mémoire de travail verbale dans l'acquisition du langage, au travers des relations entre capacités perceptives et motrices. Des recherches récentes ont en effet permis de déterminer l'implication de la boucle phonologique dans l'acquisition et le développement du langage. Ces études suggèrent que la contribution de ce système dans la rétention de matériel verbal familier ne serait qu'un produit accidentel de sa fonction première, son véritable rôle étant de servir le langage au travers de l'apprentissage de formes phonologiques nouvelles et de permettre, sinon la création, du moins le renforcement des structures phonologiques sous-jacentes. Dans le cadre des théories soulignant l'importance de l'apprentissage de coordinations orofaciales et orolaryngées nouvelles dans l'acquisition du langage et face aux études récentes indiquant l'existence possible d'un système de communication/imitation fondé sur la reconnaissance gestuelle — travaux liés notamment à la découverte des fameux neurones miroirs — nous émettons l'hypothèse que ce système de mémoire de travail incorpore un processus de contrôle de la parole, fournissant une plate-forme adaptée à la comparaison et l'évaluation des unités de parole et, de là, au développement des capacités langagières.

L'auteur

Marc Sato a réalisé une thèse de sciences cognitives à l'Institut de la communication parlée (thèse présentée à l'Institut National Polytechnique de Grenoble auquel ce laboratoire est rattaché) sous la direction conjointe du Dr. Jean-Luc Schwartz, du Prof. Christian Abry et du Dr. Hélène Loevenbruck (2001-2004). Il est actuellement en stage postdoctoral au département de Neurosciences de l'Université de Parme où il travaille, en collaboration avec le Dr. Giovanni Buccino et les Pr. Vittorio Gallese et Giacomo Rizzolatti, sur les liens possibles existants entre représentations linguistiques et représentations motrices.

ANALYSE DIACHRONIQUE DE L'ÉVOLUTION TERMINOLOGIQUE

VARIATIONS ET DISTANCES

Annie TARTIER

*Thèse d'Informatique, Université de Nantes, présentée le 4 octobre 2004.
Mél : Annie.Tartier@univ-nantes.fr*

Thesis title — *A diachronic analysis of terminological evolution: variations and distances.*

Título da tese — *Análise diacrônica da evolução terminológica. Variações e distâncias.*

Título de la tesis — *Análisis diacrónico de la evolución terminológica. Variaciones y distancias.*

Titolo della tesi — *Analisi diacronica dell'evoluzione terminologica. Variazioni e distanze.*

L'analyse diachronique de l'attestation terminologique ne peut pas se faire par introspection car l'être humain n'a qu'une perception synchronique de sa langue, même s'il s'agit d'une langue de spécialité. C'est pourquoi il est nécessaire de mettre en place des outils de repérage automatique des phénomènes évolutifs dont une empreinte est gardée dans des corpus diachroniques. Cette thèse propose une réflexion conduisant à l'élaboration de méthodes destinées à mesurer l'évolution de terminologies au cours du temps. Elle est centrée sur l'analyse du problème général suivant : *comment rendre compte de manière objective, avec peu de connaissances extérieures, de l'évolution des terminologies, en explorant automatiquement des corpus diachroniques de textes techniques ou scientifiques.* Un prototype a été mis en place pour valider les méthodes proposées.

Analyser l'évolution c'est mesurer ce qui change au cours du temps et comment cela change. Le déroulement de la thèse suit donc deux axes orthogonaux, celui de la nature des formes terminologiques et de leur variation (ce qui change), et celui des distributions temporelles des occurrences d'événements (comment cela change).

Une étude approfondie des différents travaux sur la variation terminologique passe en revue la manière dont les chercheurs définissent, classifient et traitent les différents types de variation (morphologique, syntaxique, morphosyntaxique, « sémantique »). L'étude permet d'affirmer que si le phénomène de variation est reconnu par tous, les définitions divergent. La définition de chacun est motivée par le rôle que joue la variation dans la réalisation de son objectif. Les divergences portent sur l'élasticité de l'éloignement sémantique entre terme variant et terme référent, avec, en conséquence, les modifications apportées aux formes linguistiques. Cette étude s'achève sur une présentation de trois logiciels d'acquisition terminologique qui prennent en compte la variation, soit pendant la phase de reconnaissance (Daille, 1994), soit comme moyen de reconnaissance (Jacquemin, 1997), soit, *a posteriori*, pour l'exploitation des résultats (Ibekwe-Sanjuan, 1997).

Suit un panorama des recherches et des réalisations en matière de diachronie. Il balaye des domaines très divers, dont l'objet d'étude va du « linguistique » à l' « extra linguistique », en y observant systématiquement les méthodes d'études diachroniques. Sont examinés successivement des domaines liés à la linguistique (étymologie, linguistique historique, bases posées par F. De Saussure, travaux récents), les activités et outils centrés sur la néologie, quelques exemples en terminologie qui se démarquent de la néologie, les travaux d'exploitation de corpus diachroniques, littéraires ou non, et enfin les méthodes mises en place pour la veille technologique. Il ressort de cet examen que la majorité des travaux exploitent des corpus diachroniques, les segmentent *a priori* selon différents critères, et utilisent des méthodes de statistique lexicale, en calculant des indicateurs à partir d'occurrences et de co-occurrences de mots simples. Cette thèse se positionne en complément à ces travaux, d'une part en prenant en compte les termes complexes et leurs variations, d'autre part en cherchant des stratégies plus fines et plus souples dans la manière de segmenter par le temps.

La décision de prendre en compte la réalité de la variation terminologique permet d'une part de ne pas considérer comme complètement différentes deux formes terminologiques qui sont des variantes du même terme, d'autre part de suivre les variations au cours du temps en posant comme hypothèse de travail que le phénomène macroscopique de l'évolution pourrait résulter des phénomènes microscopiques de variations des formes terminologiques (Cabré, 1998). Partant du constat que pour repérer des changements de manière opératoire il faut quantifier et mesurer l'objet observé, nous avons cherché à quantifier le phénomène de la variation en définissant une distance entre deux termes complexes. Celle-ci doit refléter le glissement de sens correspondant aux différents types de variations. Une manière d'exprimer la distance entre deux entités complexes consiste à définir un coût de transformation d'une entité vers l'autre. C'est pourquoi la méthode choisie est un aménagement de la *distance d'édition* entre chaînes de caractères (Levenshtein, 1966). L'adaptation consiste à faire jouer aux constituants des termes complexes le rôle joué par les caractères dans la distance d'édition, en introduisant dans le calcul des coûts, des connaissances provenant des variations mises en jeu. Après avoir choisi des seuils, cette distance permet de qualifier deux formes terminologiques de formes *variantes*, *parentes* ou *étrangères*. Ces distances sont destinées à établir des mesures, soit au niveau microscopique (celui du terme), soit au niveau macroscopique (celui d'un ensemble de termes structuré par exemple en un graphe complet étiqueté par les distances), puis de suivre ces mesures au cours du temps.

Puis vient la dimension du temps. Après avoir justifié l'utilisation d'un modèle à temps discret, sont définis les éléments sur lesquels s'appuient les calculs. Une période est un ensemble de formes terminologiques attestées dans un même intervalle de temps. Le corpus est segmenté en périodes de largeur paramétrable et les événements (formes stables, disparitions, apparitions) sont des résultats d'opérations sur ces ensembles. C'est cependant insuffisant. En effet, tout changement ne se traduit pas par une évolution, certains changements persistent (il y a donc évolution), alors que d'autres ne « prennent » pas. C'est pourquoi, plusieurs modalités d'examen diachronique (local, consolidé sur le passé, consolidé sur le futur, permanent) sont mises en place, qui vérifient, par exemple, qu'une nouvelle attestation est consolidée à partir d'une certaine date et n'est pas simplement éphémère. La prise en compte de la variation est réalisée en appliquant les mêmes calculs, non plus sur les périodes, mais sur les périodes étendues à leurs variantes.

La construction des données est une opération délicate. Elle s'effectue en trois étapes. La première concerne la tâche sensible de la sélection des textes et la confrontation avec le paradoxe suivant : il faut sélectionner pour obtenir un corpus homogène et représentatif de faits de langues, mais la sélection est une intervention risquée puisqu'elle influe sur les données et sur les résultats qui en découlent. Le corpus d'étude que nous avons construit est constitué de notices bibliographiques (titres et résumés) d'articles de physique (1984 à 1998) provenant de la base PASCAL maintenue par l'INIST⁴. La seconde étape décrit toutes les phases de nettoyage, d'annotation grammaticale et de lemmatisation des textes. La troisième présente l'acquisition des candidats-termes à partir du corpus qui a été effectuée par le logiciel ACABIT et pose le problème de la réutilisation des outils dans le domaine du TALN.

Le prototype qui implante les méthodes présentées ci-dessus est construit autour d'un noyau composé de modules spécialisés. En amont du noyau des modules de pré-traitement, programmés en fonction du format des données initiales et des logiciels utilisés pour l'étiquetage grammatical et la lemmatisation, sont chargés de fournir, par l'intermédiaire de filtres, des listes de formes terminologiques datées. En aval du noyau, un système d'interfaces doit être mis en place pour permettre à l'utilisateur d'exprimer ses requêtes qui sont, pour l'instant, accessibles en ligne de commande. Dans l'état actuel de développement, les modules du noyau sont constitués d'une trentaine de classes Java. Le module central est celui qui implémente les distances en faisant appel à un module sommaire de calcul morphologique. Les autres modules se répartissent en *services non diachroniques* et *services diachroniques* et, dans cette catégorie en *profils temporels* d'une part et *suivi de cohortes de termes* d'autre part. Tous les résultats sont calculables pour tous les modes d'examen diachronique et sur les formes strictes ou aux variantes près.

Les résultats obtenus sont tributaires du fait qu'il est encore difficile de construire des corpus diachroniques homogènes couvrant une période significative. Les candidats-termes extraits du corpus d'étude ne permettent pas de mettre à jour une tendance globale à l'évolution. En revanche, les outils de calculs mis en place apportent les informations nécessaires au suivi de termes ou d'ensembles de termes, si bien que le système proposé est capable d'apporter une aide dans au moins deux domaines. L'un concerne directement le matériau terminologique et linguistique en se donnant comme tâche de mettre à jour des bases terminologiques. L'autre s'appuie sur la prise de conscience de l'évolution des termes pour mieux appréhender l'évolution des domaines décrits par ces termes.

L'objectif de départ était de mettre à jour des phénomènes évolutifs par la seule observation de données. L'objectif second, au service du précédent, était de quantifier la variation pour en déduire une distance. La conclusion est que les résultats obtenus sur le non diachronique (distances et variantes) sont directement utilisables et que les idées introduites sur les découpages temporels et les modalités d'examen diachronique doivent être affinées. Ce travail devra s'enrichir d'autres types de variantes, très importantes dans le cadre de l'évolution, comme les synonymes, les acronymes et abréviations. Enfin le fait que les corpus diachroniques soient en train de se construire permet d'espérer que ce prototype puisse devenir un véritable outil au service de tous les acteurs de la terminologie.

⁴ Institut de l'Information Scientifique et Technique - CNRS - Vandœuvre-les-Nancy

Références bibliographiques

Cabré M.-T. (1998). La terminologie, théorie, méthodes et applications. Paris : Armand Colin.

Daille B. (1994). Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques. Thèse de doctorat en informatique. Université Paris 7.

Ibekwe-Sanjuan F. (1997). Recherche des tendances thématiques dans les publications scientifiques : définition d'une méthodologie fondée sur la linguistique. Thèse de doctorat. Université Stendhal.

Jacquemin C. (1997). Guessing morphology from terms and corpora. *Proceedings of 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'97)*. Philadelphia, PA. 27-31

Levenshtein V. I. (1966). Binary codes capables of correctiong deletions, insertions, and reversals. *Sov. Phys. Dokl.* Vol. 10, 707-710.

L'auteur

Initialement professeur de mathématiques, puis de sciences physiques, dans le second degré, Annie Tartier est venue à l'informatique par le biais de la formation des enseignants, d'abord en tant que « formée », puis en tant que formatrice. Elle a alors repris un cursus complet d'études en informatique. Elle a été recrutée au département informatique de la faculté des sciences de Nantes et a entrepris une thèse au sein de l'équipe de TALN de l'IRIN, devenu LINA depuis début 2004. Une partie des membres de cette équipe travaille à l'exploitation de corpus à travers l'examen de la terminologie sous-jacente. C'est en complément à ces travaux qu'a été préparée cette thèse dont Béatrice Daille a assuré l'encadrement scientifique. Annie Tartier poursuit ce travail en cherchant à affiner d'une part les méthodes de calcul des distances et d'autre part l'analyse de la diachronie.