

Représentations métalinguistiques de phrases à partir de transducteurs

Pierre-André BUVET

Abstract: The main goal of this paper is to present two types of INTEX transducers. Transducers of the first type allow to associate elementary sentences to metalinguistic representations intended for a system of translation. Those of the second type are intended for the preliminary treatment of authentic texts; their function is to report the texts' sentences in canonical elementary sentences. We also discuss linguistic theory which is underlying of the representations.

Keywords: class of objects, local grammar, machine translation, transducer.

Mots clés : classe d'objets, grammaire locale, traduction automatique, transducteur.

1. Introduction

La théorie linguistique sous-jacente à l'application (en cours de réalisation) que nous présentons ici est celle du lexique-grammaire. La phrase y est considérée en termes d'opérateur et d'arguments de telle sorte que celle-ci ne soit pas représentée d'une manière bipolaire comme c'est le cas dans les théories qui conçoivent la phrase simple sur le modèle de la proposition en logique classique¹. Ce qui différen-

¹ Cf. GROSS (M.) 1981. La représentation de la phrase simple du lexique-grammaire s'inspire de celle de Harris (1968). Sa modélisation est comparable à celle d'un atome en physique classique ; le noyau serait le prédicat, en tant que constituant central, et les électrons les divers arguments. Ce modèle a son pendant en logique contemporaine pour la proposition ; cf. FREGE 1984. Différentes théories linguis-

✉ Pierre-André BUVET, Centre Lucien Tesnière, Université de Franche-Comté
e-mail : pab@lii.univ-paris13.fr

cie fondamentalement les deux conceptions, c'est que la seconde est analogue à une fonction algébrique n'admettant en tout et pour tout qu'une seule variable (approximativement le sujet) alors que la première correspond à une fonction algébrique pouvant accepter plus d'une variable (approximativement le sujet et les éventuels compléments). La modélisation de la phrase proposée par le lexique-grammaire a fait preuve de son efficacité tant du point de vue de la description des langues naturelles que de celui du développement d'applications en TAL¹.

Avec l'éditeur de graphes du logiciel INTEX, nous avons entrepris de construire des transducteurs² qui permettent de relier des phrases à des représentations métalinguistiques (désormais RM) en vue de leur implémentation dans un système de traduction automatique basé sur le principe du transfert. Pour cela, nous avons utilisé une fonctionnalité particulière des graphes d'INTEX, dite « \$ », qui s'est avérée particulièrement opératoire pour aboutir aux résultats souhaités dans la mesure où, conjointement aux autres fonctionnalités de l'éditeur de graphes, elle permet notamment d'effectuer des effacements, des déplacements et des substitutions sur les différents éléments d'une chaîne discursive donnée représentés dans les nœuds d'un graphe représentant globalement la chaîne considérée.

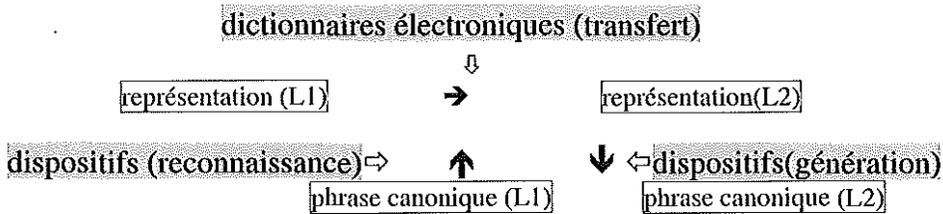
Les automates présentés ici sont destinés à un système de traduction dont l'architecture a fait l'objet de plusieurs présentations (cf. BLANCO et BUVET 1999, BLANCO et BUVET à paraître, BUVET 1999a, BUVET et GREENFIELD à paraître). Le schéma ci-dessous présente les

tiques utilisent le modèle atomique; on le retrouve par exemple chez Tesnière (1959) (toutefois, ce dernier n'a pas tiré toutes les conséquences de son modèle puisqu'il limite la nature grammaticale du noyau au seul verbe). Certaines grammaires d'unification recourent également à une telle représentation, cf. ABEILLE 1993.

¹ Cf. LECLERE 1998 pour une bibliographie complète des différents travaux du lexique-grammaire effectués tant en linguistique qu'en informatique linguistique.

² Les transducteurs sont des automates à états finis qui peuvent être représentés par des graphes où les différents états sont représentés par des nœuds et les diverses relations entre les états par des arcs. Les transducteurs permettent d'associer les informations lexicales spécifiées dans les nœuds à des informations métalinguistiques. Nous renvoyons à SILBERZTEIN (1993 et 2000) pour des explications détaillées sur le système INTEX et ses différentes fonctionnalités.

principales composantes du système en question :



En réalité, les trois modules (celui de reconnaissance, celui de transfert et celui de génération) font appel aux dictionnaires électroniques spécifiques à chacune des langues mises en correspondance puisqu'une grande partie des informations qui y sont enregistrées sont utilisées dans les dispositifs qui, dans une langue donnée, permettent le passage d'une phrase canonique à sa RM (tâche prise en charge par le module de reconnaissance) et, inversement, le passage d'une RM à l'une des phrases canoniques qui lui sont associées (tâche qui incombe au module de génération). Cependant, ces dictionnaires jouent un rôle crucial en ce qui concerne le passage d'une RM d'une phrase canonique en L1 à sa RM équivalente en L2 ; c'est pourquoi ils ont la place centrale dans le schéma ci-dessus. Comme le format des dictionnaires électroniques qui sont à la base de notre système est fondé sur les travaux du LLI¹, nous renvoyons aux travaux des principaux concepteurs de ce type de dictionnaires pour obtenir une description plus détaillée de leur structuration ; cf. GROSS (G.) 1992, MATHIEU-COLAS 1994 et 1996.

Le point important ici, pour ce qui est du traitement des données linguistiques, c'est que le module de transfert se présente sous la forme de dictionnaires monolingues coordonnés contrairement aux systèmes qui font intervenir à ce niveau des dictionnaires bilingues (cf. CLAS et BOUILLON 1999). Cela implique notamment que chaque dictionnaire décrivant une langue donnée est doté d'un champ qui présente des équivalents de traduction possibles du lemme et à partir duquel il est possible d'accéder aux dictionnaires de la langue mise en correspondance (aucune information de type métalinguistique n'inter-

¹ Laboratoire de Linguistique Informatique.

vient plus à ce niveau). Le module de transfert fonctionne donc comme un pointeur qui utilise les dictionnaires électroniques de L1 pour identifier les données lexicales majeures apparaissant dans la RM de la phrase canonique en L1 et, via le champ traduction, donner leur équivalent dans la RM correspondante en L2 de telle sorte que le module de génération puisse utiliser les diverses informations rattachées aux unités lexicales majeures de la RM en L2 afin que le dispositif qu'il comporte puisse aboutir à une phrase canonique en L2 qui sera donc la traduction de la phrase de départ.

Nous ne considérerons ici que le module de reconnaissance. Le rôle des dictionnaires électroniques, bien que non négligeable, ne sera pas pris en compte. Notre objectif est de montrer comment les transducteurs peuvent prendre en charge les différentes étapes du dispositif pour aboutir aux résultats souhaités, i.e. des représentations métalinguistiques de phrases canoniques. Bien que nos exemples portent uniquement sur le français, précisons que nous avons expérimenté les dispositifs du système de traduction sur d'autres langues (notamment l'espagnol) dans le cadre du projet DêtTAL¹.

Dans un premier temps, nous présentons les principales caractéristiques des transducteurs utilisables dans le module de reconnaissance puis nous discutons de l'intérêt des classes d'objets en tant que grammaires locales pour simplifier les procédures. Nous terminons en évoquant la nécessité d'un prétraitement des phrases (ce que nous appelons la canonisation) en faisant appel à d'autres types de transducteurs.

2. Caractéristiques des transducteurs aboutissant à des RM

Nous avons précisé que les phrases sont représentées ici en termes de prédicat et d'argument. Ainsi :

(1) *Ce garçon aime cette fille*

donne lieu à la RM suivante :

¹ Détermination et Traitement Automatique des Langues. Il s'agit d'un projet concernant divers partenaires européens et dont l'objectif principal est l'amélioration qualitative des systèmes de traduction, notamment en ce qui concerne les contraintes sur les déterminants.

P₁ → aimer (garçon , fille)

Dans le cas où l'on a affaire à des phrases complexes, les représentations sont identiques si ce n'est que l'un des arguments est lui-même une RM¹. Par exemple :

(2) *Luc dit que ce garçon aime cette fille*

aura comme RM :

P₂ → dire (Luc, aimer (garçon , fille))

Autrement dit, un discours donné, en tant qu'ensemble structuré de phrases, peut être représenté par un agencement de différentes RM qui rend compte de sa complexité phrastique. Une des conséquences majeures de ce type de modélisation est sa dimension lexicale dans le sens où une description approfondie et préalable des unités lexicales majeures de la langue prise en compte y est obligatoire. C'est pourquoi la théorie du lexique-grammaire peut-être dit lexicaliste au même titre, par exemple, que la théorie sens-texte (cf. MEL'CUK *et alii* 1995). Une autre conséquence importante, c'est que les unités fondamentales de discours sont les phrases en tant que prédicats saturés par des arguments. Les analyses et les applications que nous présentons ne portent que sur des phrases relativement élémentaires.

La transformation des RM en phrases canoniques fait appel à un dispositif qui procède obligatoirement en deux étapes (cf. GROSS (G.) 1996) : la linéarisation (i.e. la spécification du rôle syntaxique des arguments), d'une part, l'actualisation (i.e., notamment, la spécification des différentes informations relatives au temps et à la détermination des unités nominales), d'autre part. Nous avons montré (BUVET 1999a) que l'efficacité du dispositif est améliorée si l'on subdivise l'actualisation en deux sous-étapes qui tiennent compte de la possibilité d'assimiler les substantifs soit à des prédicats soit à des arguments. Inversement, après l'identification des prédicats et des arguments, la transformation des phrases canoniques en RM consiste à spécifier la nature des autres éléments de la phrase du point de vue de

¹ En réalité, les phrases complexes peuvent donner lieu à d'autres RM; cf. BUVET à paraître *α*.

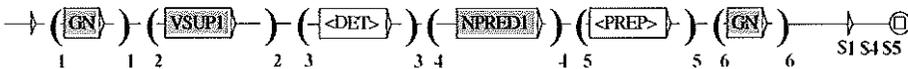
l'actualisation afin de les exclure en tant que tels et de leur associer, s'il y a lieu, un code relatif à leur fonction dans la phrase.

Ces différents points étant précisés, nous montrons à présent comment le dispositif qui construit des RM à partir de phrases canoniques peut être pris en charge par les transducteurs. Deux types de situations sont distingués, les standard et les non standard. Pour simplifier les notations, nous avons volontairement omis les paramètres afférents aux indications de temps et de nombre.

Considérons les phrases suivantes :

- (3a) *Ce garçon avait de l'estime pour son professeur*
- (3b) *Ce garçon estimait son professeur*
- (3c) *Ce garçon avait beaucoup d'estime pour son professeur*
- (3d) *Ce garçon estimait beaucoup son professeur*

La première tâche qui incombe à INTEX est de reconnaître dans cette phrase le prédicat (**estim-** sous ses formes verbale ou nominale) et ses arguments (*garçon*) et (*professeur*). La configuration de la phrase étant déterminante pour y parvenir, on lui appliquera donc un transducteur qui en tient compte. Le graphe qui correspond à un tel transducteur comporte en réalité différents sous-graphes qui eux-mêmes peuvent comporter au niveau de certains nœuds (voir tous) d'autres sous-graphes (les sous-graphes sont repérés par des nœuds grisés). Nous présentons ci-dessous uniquement les sous-graphes qui permettent un premier traitement des phrases (3a) à (3d).



Graphe 1



Graphe 2

Les graphes 1 et 2 traitent respectivement les phrases (3a) et (3b). Aucun graphe ne s'applique à (3c) ou à (3d) car ces phrases ne

nécessitent aucun traitement à ce stade. La tâche des graphes est de traiter ici de l'actualisation des prédicats. En ce qui concerne (3a), le prédicat est le substantif *estime* et les éléments qui relèvent de son actualisation sont le verbe support *avoir*, l'article partitif et la préposition *pour*. Comme ils ressortissent à l'actualisation standard du prédicat considéré, ils disparaissent du résultat final auquel aboutit l'application du transducteur (graphe 1) à cette phrase, à savoir :

(3aI) **Ce garçon estime son professeur**¹

La phrase (3c) comporte un élément actualisateur (relativement au prédicat) non standard, le déterminant adverbial *beaucoup de*, qui fonctionne aussi comme un marqueur d'intensité forte ; il ne peut donc pas être supprimé comme les éléments standard qui sont sémantiquement non marqués. Dans (3d), qui est équivalent à (3c), c'est l'adverbe *beaucoup* qui est le marqueur d'intensité forte². L'application du transducteur (graphe 2) à (3c) donne comme résultat :

(3cI) **Ce garçon beaucoup d'estime son professeur**

Comme le premier transducteur ne concerne que les phrases à prédicat nominal, (3b) et (3d) étant à prédicat verbal, elles restent inchangées après son application :

(3bI) **Ce garçon estimait son professeur**

(3dI) **Ce garçon estimait beaucoup son professeur**

Un transducteur supplémentaire permet de rendre compte uniformément de l'expression du haut degré³ :



Grappe 3

¹ L'identité de (3aI) avec (3c) est une coïncidence.

² Des verbes supports et des modifieurs peuvent également exprimer le haut degré ; cf. BUVET 1995, BUVET et GROSS 1995 et BUVET à paraître *b* sur l'expression de l'intensité.

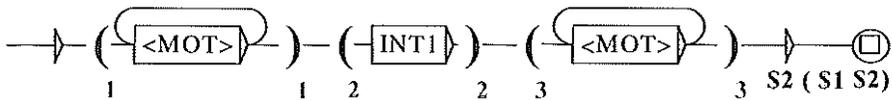
³ Sous peine de disfonctionnement, les transducteurs d'INTEX doivent nécessairement être appliqués successivement, ce qui suppose une réflexion préalable sur leur ordre d'apparition.

de telle sorte que l'on ait finalement :

(3cII) Ce garçon INT1 estime son professeur

(3dII) Ce garçon estimait INT1 son professeur

L'intensité et l'aspect s'interprètent dans la théorie du lexique-grammaire comme des prédicats de niveau supérieur dont l'unique argument est l'une des phrases comportant les marqueurs relatifs à l'une ou l'autre de ces catégories sémantiques¹. L'application d'un nouveau transducteur permet d'en faire état :



Grappe 4

afin qu'on aboutisse aux résultats suivants :

(3cIII) INT1 (Ce garçon estime son professeur)

(3dIII) INT1 (Ce garçon estimait son professeur)²

Après l'actualisation des prédicats, un transducteur doit prendre en charge l'actualisation des arguments, i.e., pour les phrases élémentaires, la détermination des noms en position sujet ou complément. Contrairement à la détermination standard des prédicats substantivaux, celle des arguments nominaux est généralement sémantiquement marquée; c'est pourquoi elles apparaissent dans les RM finales sous la forme d'un code qui renvoie à leur spécificité³. C'est un nouveau transducteur qui s'occupe de traiter ce point particulier. Par manque de place (il est beaucoup plus volumineux que les précédents), nous ne présentons pas le graphe qui lui correspond mais seulement les résultats issus de son application :

(3aIV) DDEF2 garçon estime DDEF3 professeur

¹ Sur la représentation de certains types aspectuels, cf. BUVET 1999b.

² Les résultats (3aI) et (3bII) n'étant pas concernés par l'expression du haut degré, ils restent inchangés après l'application des deuxième et troisième transducteurs.

³ Cf. BUVET à paraître c pour une typologie des différents déterminants du français.

- (3bIV) **DDEF2 garçon estimait DDEF3 professeur**
 (3cIV) **INT1 (DDEF2 garçon estime DDEF3 professeur)**
 (3dIV) **INT1 (DDEF2 garçon estimait DDEF3 professeur)**

L'étape suivante est la délinéarisation; il s'agit de réécrire les phrases de départ sous une forme qui s'inspire de la logique des prédicats à l'instar de ce que nous avons fait pour la catégorie INT1, si ce n'est que ce sont des éléments lexicaux qui doivent être spécifiés en position de prédicat et d'argument. Le graphe qui correspond au transducteur permettant ce type de traitement est également trop volumineux pour figurer dans le présent article; par conséquent, nous n'indiquons que les résultats obtenus suite à l'application de ce dernier transducteur aux résultats précédents :

- (3aV) **estime (DDEF2 garçon , DDEF3 professeur)**
 (3bV) **estimait (DDEF2 garçon , DDEF3 professeur)**
 (3cV) **INT1 (estime (DDEF2 garçon , DDEF3 professeur))**
 (3dV) **INT1 (estimait (DDEF2 garçon , DDEF3 professeur))**

Ces résultats, nonobstant l'absence d'informations sur le temps grammatical, sont conformes aux RM que l'on souhaite implémenter au niveau du module de transfert du système de traduction.

Nous montrons à présent comment la description du lexique en termes de classes d'objets peut être intégrée dans les transducteurs qui permettent de construire des RM de phrases canoniques.

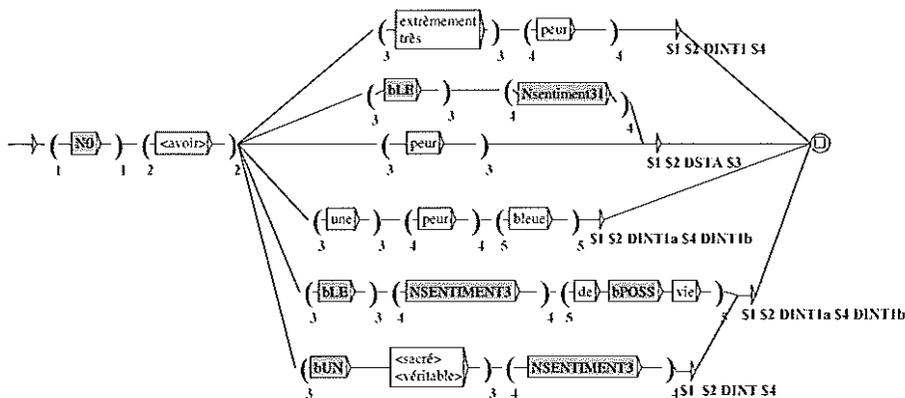
3. Intérêt de la notion de classes d'objets

Rappelons tout d'abord qu'une classe d'objets correspond à un ensemble d'items homogènes (à la fois morphologiquement et sémantiquement) qui sont caractérisés par leurs propriétés syntaxiques; les classes sont distinguées selon qu'elles correspondent à des classes d'arguments, e.g. les noms de <voie>, ou bien à des classes de prédicats, e.g. les noms de <maladies physiques>; (cf. BUVET 1998, GROSS (G.) 1992 et 1996, GUENTHNER 1998, LE PESANT et MATHIEU-COLAS 1998).

L'intérêt de la notion de classes d'objets est de deux ordres : d'une part, linguistique, d'autre part, informatique. Pour ce qui est du premier, mentionnons, entre autres, la possibilité de rendre compte des

contraintes sur les déterminants des noms prédicatifs, des valeurs non standard des déterminants, des verbes supports et des adverbes (cf. BUVET 1998, BLANCO et BUVET 1999 ainsi que GROSS (G.) 1994 pour ce qui concerne la polysémie et d'autres phénomènes linguistiques). Du point de vue informatique, les classes d'objets étant assimilables à des grammaires locales (cf. GROSS (M.) 1995), les phrases canoniques associées aux différents éléments qu'elles comportent peuvent donner lieu à la production de transducteurs les représentant et contribuant à une simplification de certains étiquetages métalinguistiques.

Nous présentons ci-dessous un des transducteurs relatifs aux noms de <sentiment> afin d'établir en quoi les classes d'objets (présentées sous la forme *ad hoc*) permettent de simplifier les procédures permettant la substitution d'une phrase canonique par sa RM.



Graphe 5

Du fait de leur homogénéité sémantique, les noms de <sentiment> autorisent (sinon totalement, du moins partiellement) des regroupements qui font état de caractéristiques syntactico-sémantiques communes. C'est ainsi que tous les noms de <sentiment> figurant dans le sous-graphe [Nsentiment31] (*bourdon, frousse, trouille*, etc.) ont un article défini comme déterminant standard lorsqu'ils se combinent avec le support *avoir*.

Signalons également que les classes d'objets présentées sous forme de grammaires locales sont également d'un grand intérêt pour le module de génération puisque après l'identification de l'équivalent

dans la langue cible du prédicat et des arguments en langue source, il est possible à l'aide des dictionnaires électroniques de spécifier de quelles classes ils relèvent et de les associer ainsi à des grammaires locales qui permettent de générer les phrases canoniques souhaitées. Autrement dit, le champ classe dans les dictionnaires utilisés dans le module de transfert est une clef à partir de laquelle le système va pouvoir accéder au module de génération.

Après ces quelques remarques sur la notion de classe d'objets, nous évoquons la nécessité d'un prétraitement des textes.

4. Les transducteurs de prétraitement de textes

On a pu constater que le point de départ des RM était jusqu'à présent des phrases canoniques. De telles phrases étant rares dans les textes, la mise en évidence des phrases canoniques d'un texte donné est donc une étape préalable et obligatoire dans le système de traduction. De ce point de vue, deux types de situations sont prises en compte :

– d'une part, celles qui font intervenir des prédicats de niveau supérieur dont les arguments sont de nature essentiellement phrastique, e.g. la comparaison :

(4) *Léa apprécie Luc autant qu'elle déteste Tom*

– d'autre part, celles qui ressortissent à des transformations de phrases simples, e.g. l'extraction dite en *c'est...qu-* :

(5) *C'est Luc qui aime Léa*¹

Le traitement de ces deux types de situations n'est pas comparable dans la mesure où, pour le premier type, le système doit aboutir à des RM qui comportent les prédicats de niveau supérieur alors que, pour le second type, il s'agit surtout d'indiquer à quelles phrases canoniques se sont appliquées les transformations².

¹ Les deux types de situations ne sont pas nécessairement indépendantes l'une de l'autre, e.g. on peut observer une extraction en *c'est...qu-* conjointement à une comparaison : *C'est Luc que Léa aime autant qu'elle déteste Tom.*

² Nous n'affirmons pas ici que les transformations sont neutres du point de vue du fonctionnement de la langue (e.g. il est clair que l'extraction a trait à la topicalisation) mais qu'il est possible de négliger leurs particularités, quelles qu'elles soient.

L'application des transducteurs évoqués ci-dessus suppose que le système INTEX ait préalablement effectué un certain nombre d'opérations, notamment celles qui portent sur l'identification des différentes phrases d'un texte donné. Par ailleurs, l'élaboration des transducteurs permettant un tel prétraitement des textes est un travail de longue haleine du fait de la variété des phrases complexes et de l'importance quantitative des transformations possibles. Toutefois, en s'appuyant sur les nombreux travaux du lexique-grammaire qui décrivent exhaustivement ces diverses classes de phénomènes, l'objectif fixé semble réalisable.

5. Perspectives

Les premières expérimentations que nous avons effectuées ont donné des résultats encourageants bien que leur champ soit pour l'instant relativement limité. L'intérêt des RM en vue d'une amélioration qualitative des systèmes de traduction apparaît indéniable comme nous l'avons déjà montré et le système INTEX s'avère suffisamment souple et puissant pour traiter informatiquement les données linguistiques dans l'optique envisagée.

D'autres expérimentations restent à faire et, une fois les premiers résultats confirmés, des programmes de vaste ampleur doivent être menés à leur terme pour élaborer un système de traduction automatique performant. Celui-ci, en termes de traitement des données linguistiques, peut s'appuyer sur la théorie du lexique-grammaire étendue à la sémantique du fait de son état d'avancement tant du point de vue de la description des faits de langue que de leur représentation formelle. Sur le plan informatique, INTEX étant un outil conçu pour implémenter le formalisme du lexique-grammaire, ses différentes fonctionnalités doivent être donc impérativement intégrées dans les différentes composantes du système.

Bibliographie

ABEILLÉ (Anne) : 1993, *Les nouvelles syntaxes : grammaire d'unification et analyse du français* (Paris : Armand Colin).

- BLANCO (Xavier) et BUVET (Pierre-André) : 1999, « La traduction automatique des déterminants de l'espagnol et du français », *Méta* 44 : 4, (Montréal : Presses de l'Université de Montréal).
- BLANCO (Xavier) et BUVET (Pierre-André) : à paraître, « Vers la mise en place d'un système de traduction automatique franco-espagnol ».
- BUVET (Pierre-André) : 1995 « Les compléments de nom intensifs », *Neophilologica*, 12 (Katowice : Université de Silésie).
- BUVET (Pierre-André) 1998, « Détermination et classes d'objets », *Langages*, 131 (Paris : Larousse).
- BUVET (Pierre-André) : 1999a, « Détermination et génération automatique », in *Actes du colloque GAT99* (Grenoble).
- BUVET (Pierre-André) : 1999b, « Vers une modélisation des quantificateurs dans les phrases simples », in *Actes du colloque Vextal* (Venise).
- BUVET (Pierre-André) : à paraître a, « Les groupes nominaux LE N MODIF avec MODIF obligatoire ».
- BUVET (Pierre-André) : à paraître b, « Les déterminants intensifs ».
- BUVET (Pierre-André) : à paraître c, « Remarques sur la détermination du français ».
- BUVET (Pierre-André) et GROSS (Gaston) : 1995, « Comparaison et expression du haut degré dans le groupe nominal », *Faits de langue*, 5 (Paris : PUF).
- BUVET (Pierre-André) et GREENFIELD (Peter) : à paraître, « Analyse et génération de phrases simples en PROLOG ».
- CLAS (André) et BOUILLON (Pierrette), eds : 1999, *TA-TAO Recherches de pointe et applications immédiates*, Association des universités partiellement ou entièrement de langue française (Montréal : URE).
- DALADIER (Anne) : 1998, « Origine adverbiale du génitif indo-européen, extractions, possessifs, anaphore associative et interprétations grammaticalisées dans les GN français », *Langue française*, 122 (Paris : Larousse), p. 101-125.
- FREGE (Gottlob) : 1971, *Écrits logiques et philosophiques* (Paris : Le Seuil).

- GROSS (Gaston) : 1992 « Forme d'un dictionnaire électronique », *L'Environnement traductionnel* (Sillery-Montréal : Presses de l'Université du Québec ; Aupelf-Uref).
- GROSS (Gaston) : 1994, « Classes d'objets et description des verbes », *Langages*, 115 (Paris : Larousse).
- GROSS (Gaston) : 1995, « Une sémantique nouvelle pour la traduction automatique. Les classes d'objets », *La Tribune des industries de la langue et de l'information électronique*, 17-18-19 (Paris).
- GROSS (Gaston) : 1996, « Prédicats nominaux et compatibilité aspectuelle », *Langages*, 121 (Paris : Larousse).
- GROSS (Maurice) : 1981, « Les bases empiriques de la notion de prédicat sémantique », *Langages*, 63 (Paris : Larousse).
- GROSS (Maurice) : 1995, « Une grammaire locale de l'expression des sentiments », *Langue française*, 105 (Paris : Larousse).
- GUENTHNER (Franz) : 1998, « Constructions, classes et domaines : concepts de base pour un dictionnaire électronique de l'allemand », *Langages*, 131 (Paris : Larousse).
- HARRIS (Zellig S.) : 1968, *Mathematical Structures of Language* (New York : John Wiley & Sons).
- LECLERE (Christian) : 1998, « Travaux récents en Lexique-Grammaire », *Travaux de Linguistique*, 37 (Louvain : Duculot).
- LE PESANT (Denis) et MATHIEU-COLAS (Michel) : 1998, « Introduction aux classes d'objets », *Langages*, 131 (Paris : Larousse).
- MATHIEU-COLAS (Michel) : 1994, *Les Mots français à trait d'union* (Paris : Didier Érudition).
- MATHIEU-COLAS (Michel) : 1996, « Représentation de la polysémie dans un dictionnaire électronique », *Actes des IVes Journées scientifiques du réseau thématique Lexicologie, Terminologie et Traduction, FMA* (Beyrouth-Montréal : Aupelf-Uref).
- MEL'CUK (Igor), CLAS (André) et POLGUERE (Alain) : 1995, *Introduction à la lexicographie explicative et combinatoire*, (Louvain-la-Neuve : Duculot).
- TESNIERE (Lucien) : 1959, *Éléments de syntaxe structurale* (Paris : Klincksieck).
- SILBERZTEIN (Max) : 1993, *Dictionnaires électroniques et analyse automatique de textes Le système INTEX* (Paris : Masson)
- SILBERZTEIN (Max) : 2000, *Intex* (Asstril).