
Gabriel G. Bès, linguiste empiriste

François Trouilleux

*Université Clermont Auvergne, Université Blaise Pascal,
EA 999, Laboratoire de Recherche sur le Langage,
BP 10448, F-63000 Clermont-Ferrand, France
francois.trouilleux@univ-bpclermont.fr*

RÉSUMÉ. Cet article présente l'œuvre de Gabriel G. Bès, qui s'étend des années 60 aux années 2000, du structuralisme aux industries de la langue. L'œuvre de Bès est traversée par la question épistémologique et la volonté de pratiquer la linguistique comme une science empirique. On y retrouve comme problématiques centrales l'explicitation des observations sur le réel, la formalisation des hypothèses, et la pratique rigoureuse du test d'adéquation des hypothèses par rapport aux observations.

ABSTRACT. This article presents the work of Gabriel G. Bès, with extends from the 60s to the 2000s, from structuralism to industrial language analysis. The work of Gabriel G. Bès is marked throughout by the epistemological question and the will to put into practice linguistics as an empirical science. The central problems are the explicitation of the observations of reality, the formalisation of hypotheses, and the rigorous testing of the adequation of hypotheses with respect to observations.

MOTS-CLÉS: épistémologie de la linguistique, empirisme, formalisation, traitement automatique des langues, histoire de la linguistique.

KEYWORDS: epistemology of linguistics, empiricism, formalisation, natural language processing, history of linguistics.

Introduction

Le présent texte se propose de présenter l'œuvre de Gabriel Gerardo Bès, linguiste franco-argentin né le 22 novembre 1935 à Buenos Aires et mort le 28 octobre 2013 à Mendoza (Argentine). Il accompagne une bibliographie sélective (section 8), dont la plupart des textes ont été déposés dans la base HAL¹. L'objectif de l'article est double. D'une part, il vise à fournir des pointeurs vers les textes écrits par Bès, ce qui conduira parfois à présenter succinctement certains travaux, d'autre part, il tente de montrer la cohérence de l'œuvre en détaillant plus longuement d'autres travaux : sa thèse de doctorat sur la comparaison des descriptions linguistiques (section 2), ses analyses des propositions de Noam Chomsky (section 3), ses travaux en TAL (section 4) et les éléments du « paradigme 5P », cadre méthodologique pour la linguistique qu'il a proposé à la fin des années 90 (sections 1.2, 5 et 6).

Dans un domaine des sciences du langage souvent cloisonné, on rencontre avec Bès un homme de science qui, en quarante ans de carrière, a à la fois côtoyé la linguistique fonctionnelle de Martinet et les nouvelles technologies, tout en restant toujours indépendant des doctrines. En la matière, la seule affiliation qu'on peut lui reconnaître est celle de la méthode : la linguistique est une science empirique, où l'on se doit d'explicitier clairement et rigoureusement les observations faites sur le réel, au même titre que les hypothèses. Ce point de vue – d'une pertinence indémodable – lui a fait porter un regard critique sur certains travaux de référence et l'a conduit à faire, avec le paradigme 5P, des propositions profondément originales. Après une première partie qui pose le cadre général de la discussion, l'article présente les différents travaux de Bès dans l'ordre chronologique, celui de la vie d'un chercheur, en tentant de montrer le cheminement ayant conduit des uns aux autres.

1. De Buenos Aires au paradigme 5P

1.1. *Repères biographiques*

Après un baccalauréat obtenu en 1954, Gabriel Bès obtient en 1961, à l'université de Buenos Aires, un diplôme de professeur ès lettres. Il enseigne le grec classique à l'université catholique de Buenos Aires, et, en 1962, la phonétique et la grammaire historique espagnoles au Profesorado del Oeste (Ciudadela, Pvcia. de Buenos Aires). En 1963, il part suivre des études post-universitaires à Paris, sous la direction d'André Martinet. Il est d'abord soutenu par une bourse du ministère des Affaires étrangères français, puis de l'université de Buenos Aires et du Consejo Nacional de Investigaciones Científicas y Técnicas. En octobre 1965, il devient attaché de recherche au CNRS, poste qu'il occupe jusqu'en septembre 1969. En 1969, Jean-Claude Pariente, alors directeur de l'institut de linguistique de l'université de Clermont-Ferrand, recherchait un linguiste susceptible d'assurer les fonctions de professeur. Pariente avait

1. https://hal.archives-ouvertes.fr/search/index/q/*/authFullName_s/Gabriel+G.+Bès/

des contacts d'une part avec André Martinet, d'autre part avec Luis J. Prieto. C'est Prieto (Argentin) – et non Martinet – qui lui recommandera Bès. Selon les termes de Pariente, « Bès commençait déjà à échapper à Martinet »².

De 1969 à 1972, Bès est donc maître de conférences associé à l'université Blaise-Pascal. Il soutient sa thèse en 1972 et repart en Argentine. De mars 1972 à 1975, il est professeur titulaire sous contrat à l'université nationale de Cuyo, coordinateur de l'unité pédagogique de langue espagnole et études linguistiques, directeur du programme Talen (traitement automatique des langues naturelles). En 1975, son contrat n'est pas reconduit. Il travaille quelques mois à l'université nationale de San Luis, mais à cette époque, sous la présidence d'Isabel Martínez de Perón, la droite du mouvement péroniste, menée par le « sorcier » José López Rega, conduit une politique hostile aux mouvements sociaux, syndicats, universités, télévisions privées... Les perspectives de carrière en Argentine s'obscurcissent, un poste a pu s'ouvrir à Clermont, Bès rejoint l'université Blaise-Pascal en octobre 1975 et y effectuera toute sa carrière jusqu'au début des années 2000.

1.2. *Le cadre méthodologique 5P*

L'œuvre de Bès est traversée par la question épistémologique et la volonté de pratiquer la linguistique comme une science empirique. Dans les dernières années de sa carrière, cette préoccupation a conduit Bès à expliciter un cadre méthodologique pour la linguistique, le « paradigme 5P »³. Nous nous proposons, avant de parcourir chronologiquement l'œuvre de Bès, de prendre pour point de départ ce cadre méthodologique, qui nous semble donner une clé pour la compréhension globale de l'œuvre.

« Les exigences méthodologiques de 5P devraient pouvoir se déduire, toutes, de l'idée simple suivante : la linguistique est une science empirique, la linguistique telle qu'elle pourrait se pratiquer, plutôt que la linguistique telle qu'on la pratique effectivement. » [36, p. 88]⁴

De cette idée directrice découle en particulier qu'il faut :

« (i) poser le principe que le problème de fond ne commence pas avec la manipulation des grammaires qui vont traiter des données non spécifiées ou à peine suggérées par quelques exemples mais avec la spécification elle-même des données dont une grammaire ou n'importe quel formalisme doit rendre compte. [...] Spécifier des données, évaluer leur intersubjectivité, les utiliser pour vérifier les conséquences des hypothèses,

2. Communication personnelle.

3. Le paradigme 5P est présenté pour la première fois en 1999, dans un long article dont le titre, *La phrase verbale noyau en français* [35], est trompeur car la description d'un fragment du français qu'il contient illustre l'utilisation d'un *nouveau formalisme* pour la description syntaxique : les « Propriétés », formalisme qui s'inscrit dans le cadre méthodologique 5P, lui-même longuement présenté et mis en perspective.

4. Les chiffres entre crochets dans le texte renvoient à la bibliographie de Bès, section 8.

c'est un problème extraordinairement grave qu'il faut mettre à la première place des préoccupations scientifiques sur la linguistique, même si ce n'est pas le seul problème important.

« (ii) poser le principe parallèle selon lequel il n'y a pas de science sans calcul sur les hypothèses et qu'aucun calcul sur les hypothèses n'est possible sans la formalisation de ces hypothèses ⁵. »

On a là un triplet < observations, hypothèses formalisées, test des hypothèses par rapport aux observations >, qu'on retrouve aussi à la base de la grille d'analyse que Bès propose en 2002 dans *La linguistique entre science et ingénierie* [37]. Dans le cadre 5P, il se matérialise ainsi (cf. l'alinéa 1 de [35] pour une présentation détaillée).

L'observation produit le P1 de *Protocoles*. Les Protocoles représentent les données auxquelles aboutit un observateur *modélisé*, c'est-à-dire qu'on lui a attribué des capacités *explicites* pour décrire les données et fournir des *prédicats d'observation* spécifiques ⁶. Le système d'hypothèses est envisagé sous la forme d'un triplet *Propriétés* (P2), *Projections* (P3) et *Principes* (P4), trois outils qui visent à décrire et expliquer les langues, à différents niveaux de généralité. Dans le projet de Bès, seules les Propriétés ont fait l'objet d'une formalisation poussée ; nous reviendrons sur ce formalisme plus loin (section 5). À ce stade, le point important est que le système d'hypothèses {P2, P3, P4} doit être testé en le confrontant aux Protocoles. « Le résultat de la confrontation exprime l'adéquation *externe* » du système d'hypothèses [35, p. 281].

Le système d'hypothèses, pour permettre la confrontation avec les observations, se doit d'être formulé de manière explicite, calculable et cohérente. Ici, le mot clé est *formalisation* ⁷.

Protocoles et système d'hypothèses formalisées peuvent suffire sur le papier pour la pratique d'une science empirique, mais le paradigme intègre un cinquième P : celui de *Processus*, parce que (i) dès qu'un système d'hypothèses devient un peu complexe, il est impossible d'en calculer les conséquences sans l'aide d'une machine, (ii) Bès intègre dans son paradigme la problématique du traitement automatique des langues

5. Cette citation est extraite d'un document de travail non publié, préliminaire au travail sur 5P, et, à ce titre, met en avant les deux points les plus fondamentaux. Les différents articles de Bès autour de ce paradigme évoquent quelques autres aspects du cadre méthodologique. Citons entre autres l'idée que « l'observation n'est pas le résultat d'une évidence mais d'une construction faite à partir de points de vue dûment explicités » [36, p. 88], l'idée, reprise de (Bunge, 1969), qu'il n'y a pas lieu d'essayer de valider isolément chaque hypothèse particulière – c'est le système d'ensemble qui doit l'être [36, p. 77, 89], ou encore l'idée que l'observateur n'a pas d'intuition sur la représentation de la syntaxe [36, p. 17 et 105].

6. Bès précise que ses Protocoles sont les « propositions d'observation » de (Bunge, 1969). L'étude épistémologique de la physique conduite par le physicien et philosophe argentin aura eu une profonde influence sur Bès. La plupart de ses articles qui abordent la question méthodologique y font référence, cf. [9, 11, 16, 31, 34, 35, 36, 37]. Un autre ouvrage sur lequel s'appuie souvent Bès est (Bocheński, 1958).

7. L'alinéa 1.1 de *La phrase verbale noyau en français* [35] fournit toutes les justifications nécessaires à la formalisation.

en tant que domaine applicatif de la linguistique et (iii) Bès veut dissocier description de la langue et processus d'analyse et génération (cf. ci-dessous, section 5).

1.3. *Articuler théorie et application, exégèse et développement*

Les caractéristiques de la science empirique que relève Bès sont pour lui l'évidence ; s'il éprouve le besoin de les expliciter, c'est qu'il en a souvent constaté l'absence en linguistique. Bon nombre de textes de Bès ont pour point de départ des concepts établis ou des textes de référence dont il fait l'analyse. Relativement au cadre méthodologique que nous avons présenté, la démarche peut se caractériser ainsi :

- au niveau des Protocoles : explicitation des observations déterminant les conditions d'adéquation et test par rapport à ces observations ;
- au niveau du système d'hypothèses : décomposition des concepts ou reformulation des textes originaux, allant dans le sens d'une formalisation.

De ces travaux se dégage, nous le verrons, une défiance à l'égard des prétendues théories, mal explicitées, au champ observationnel irréaliste, aux critères d'adéquation internes.

Un corollaire à cette défiance est la volonté chez Bès d'articuler recherche théorique et recherche appliquée :

« Sur le plan méthodologique, il n'est pas évident qu'il soit possible de développer des notions théoriques sur le langage indépendamment de leur utilisation dans un champ d'application concret. Très sommairement, on peut caractériser la linguistique depuis Saussure jusqu'à et y compris Chomsky par la recherche de son autonomie : les notions descriptives devaient se valider à l'intérieur de la discipline même et en fonction de ses propres concepts [...]. La linguistique a réussi, à l'intérieur de cette autonomie, à développer plusieurs constructions intellectuelles importantes et fécondes ; elle n'a pas réussi, en revanche, à se donner un schéma conceptuel suffisamment explicite et riche pour pouvoir discriminer entre plusieurs solutions proposées soit par des théories différentes soit par une même théorie. Écoles, sous-écoles et tendances ont donc fleuri. [...]. Il semble bien qu'il ne soit pas possible de s'en sortir à l'intérieur du monde conceptuel établi, qui ne permet pas de valider empiriquement les notions proposées. Si cela était vrai, il serait vain de continuer à imaginer *d'abord* des notions théoriques et à les valider telles quelles pour les appliquer *ensuite*. » [9, p. 24-25]

Application et théorie doivent donc se nourrir mutuellement et les avancées théoriques ou pratiques se font à petits pas. Ainsi, par exemple, lorsqu'il s'agit de définir un système d'indexation de documents, Bès et Fauchère ne se donnent pas le but de tout traiter par la machine, mais celui de trouver un équilibre opératoire entre ce que la machine peut apporter et ce que le documentaliste doit encore traiter [14, p. 133]. Lorsque

Bès présente le formalisme des Propriétés, en l'appliquant à la phrase verbale noyau en français, il note que « tout au long de la présentation de la description proposée, grand soin a été pris de toujours parler d'une description possible du domaine observé et jamais d'une théorie ou d'une explication de ce domaine » [35, p. 351]. Lorsqu'il définit une grille d'analyse des produits de la science en linguistique, « son application dans le cadre du réalisme épistémologique conduit à un certain pointillisme : on ne corrobore pas les grandes idées mais ce que l'on a effectivement fait, c'est-à-dire des résultats concrets et parcellaires » [37, p. 76].

L'œuvre de Bès alterne donc des exégèses de travaux majeurs, à travers le prisme de l'empiricité et de la formalisation, et des développements d'outils ou de descriptions linguistiques particulières qui ne prétendent jamais au statut de « théorie », mais visent à une meilleure compréhension des problèmes. Et, de façon transversale, comme on l'a dit, on retrouve dans tous ces travaux la question d'une linguistique science empirique.

2. De la comparaison des descriptions linguistiques

Bès est venu en France pour y faire une thèse sous la direction d'André Martinet. Ses premiers travaux portent ainsi sur la comparaison des langues et la phonologie.

Dans *Parenté génétique et parenté typologique* [2], Bès décompose les définitions de ces deux concepts que propose Hjelmslev (1966), et en montre les limites. L'analyse passe par la mise au jour du fait que le mot *catégorie* est employé par Hjelmslev avec deux sens différents. Dans *Trait distinctif* [3], Bès montre que ce terme est utilisé pour désigner trois concepts distincts et il utilise cette décomposition pour réfuter l'idée que l'approche binariste de Jakobson est nécessairement la plus économique. Si l'utilisation d'un formalisme se justifie, entre autres, par le fait que « la plupart des mots du langage courant sont ambigus » (*La phrase verbale noyau...* [35], p. 278), en dégageant plusieurs sens sous un même terme du domaine de la linguistique, dès ses premiers travaux, Bès adopte une approche sinon formelle, du moins formalisante.

Dans sa thèse de doctorat, *Identités et différences dans les unités de deuxième articulation* [4], Bès pose la question du choix entre descriptions linguistiques concurrentes. En raccourci, disons qu'en présence de descriptions phonologiques de langues différentes A et B, on conclura ou non à l'identité de certains phonèmes, tandis qu'en présence d'autres descriptions des mêmes langues A et B, on conclura à l'identité d'autres phonèmes. Comment choisir ? Le travail de Bès passe par la définition d'un cadre de référence organisant les concepts du domaine, avec notamment la définition du concept de *segment*, intermédiaire entre son (produit ou perçu *hic et nunc*) et phonème (objet abstrait défini par ses traits distinctifs). Il passe aussi par la mise en évidence du fait que la description linguistique, comme la comparaison, peut se limiter à une adéquation relativement à ses propres critères, ou faire l'objet d'une adéquation *externe*, mettant en jeu des observations indépendantes de l'acte de comparaison.

La distinction entre comparaison descriptive, à adéquation interne, et comparaison explicative, à adéquation externe, organise la thèse. La partie comparaison descriptive établit la nécessité d'une définition des phonèmes s'appuyant sur leur aspect phonique – une dimension qui se rattache au monde physique – avec des traits à valeur pan-linguistique, par opposition à une définition purement différentielle et particulière à chaque langue – une définition abstraite – telle que l'envisage Martinet. Bès relève par ailleurs l'étroitesse des critères glossématiques pour la comparaison des langues, l'arbitraire des critères utilisés dans les typologies de systèmes phonologiques de Troubetskoy, Hockett ou Martinet, ou encore l'absence de critères de comparaison explicites dans le diasystème de Weinreich ou de conditions d'adéquation claires pour le *overall pattern* de Trager et Smith⁸. Au sujet de ce dernier type de description, Bès observe un « repli de la construction théorique sur elle-même : sera adéquat le *overall pattern* qui est beau, élégant, symétrique » (p. 204).

Pour sortir de l'impasse de l'arbitraire théorique (p. 247), Bès propose dans la partie comparaison explicative un protocole permettant de dégager des observations externes, qui utilise la perception des sons par des locuteurs de langues différentes. Il rapporte deux expériences de perception des sons du castillan par des Argentins⁹ et propose un système de description des unités de deuxième articulation d'une langue dont les éléments sont des segments (et non des phonèmes), organisés en sous-systèmes caractérisés chacun par un contexte différent.

On retrouve ainsi dans ce travail la préoccupation empiriste de Bès et son scepticisme à l'égard des constructions théoriques à l'élégance autoproclamée.

Le système de description proposé dans sa thèse débouche également sur un travail à visée applicative : le développement d'un « code de réduction alphabétique ». L'idée générale est d'essayer de réduire la valeur d'entropie d'un système de symboles, et donc le nombre de symboles binaires nécessaires au codage d'un message dans ce système, en associant des codes non pas à des symboles isolés, mais à des symboles correspondant aux unités qui apparaissent dans les différents contextes (c'est-à-dire qui appartiennent à différents sous-systèmes ; par exemple un *p* entre voyelles, devant consonne ou après consonne appartient à trois sous-systèmes différents) [7].

Lors de ses premières années à Clermont-Ferrand, en parallèle à la fin de son travail de thèse, Bès compose avec Jean-Claude Pariente *La linguistique contemporaine* [5], une anthologie de vingt-deux documents entourés de présentations explicatives. La première partie, *Objectifs et méthodes*, présente l'école de Prague, la glossématique, le distributionnalisme et la grammaire générative. La deuxième partie, *Le champ de la*

8. Le diasystème et le *overall pattern* (ou *overall system*) sont des outils descriptifs imaginés dans les années 50 pour une utilisation en dialectologie. Confronté à plusieurs systèmes phonologiques dialectaux, on tente de définir, en se fondant sur des similitudes partielles, un système d'ordre supérieur qui embrasse tous les systèmes particuliers.

9. L'une de ces expériences est évoquée dans *Phonétique, phonologie et communication inter-dialectale* [1], paru en 1967 mais écrit avant l'entrée de Bès au CNRS, ce qui place l'ancrage empirique tôt dans la recherche exposée dans [4].

recherche, aborde la caractérisation du langage humain, la phonologie, la grammaire, la sémantique et la variation linguistique. Enfin, une dernière partie est spécifiquement dédiée à la question des rapports entre syntaxe et sémantique, vue comme « l'un des grands problèmes de la linguistique d'aujourd'hui ».

Paraît également en 1973, dans la revue argentine *Ciencia Nueva*, un article général sur la linguistique [6] dans lequel Bès, après avoir discuté des méthodes et résultats des linguistiques structurales et de la grammaire générative, pointe l'archaïsme d'un enseignement linguistique essentiellement historique en Argentine et plaide pour une politique d'enseignement intégrant la linguistique contemporaine parmi les axes prioritaires. Bès y défend une « interpénétration constante de la théorie et de la pratique ». Les problèmes rencontrés en pratique s'avérant souvent sans solution dans les théories existantes, « le rôle de la science, du scientifique et de l'enseignant sera ainsi démystifié et une transmission effective des exigences du milieu aux centres de recherche sera établie ».

3. Analyse des propositions de Noam Chomsky

Après son retour à Clermont en 1975, Bès poursuit des travaux dans trois directions. Avec ses collègues du Groupe de recherche sur la condensation de l'information en langue naturelle (CILN), il publie dans les cahiers du CILN, *Condenser*, une série d'articles sur la contraction de texte. L'empreinte de Bès se retrouve dans l'inclusion, dans un domaine qui ne la pratique pas à l'époque, de la problématique de la formalisation. Le premier article sur le sujet [9] contient en particulier un appendice intitulé *Le cadre conceptuel*, qui expose les positions épistémologiques de Bès dans ce contexte particulier. En dehors du cadre méthodologique, Bès travaille dans ce projet à distinguer différents types d'informations dans les énoncés, à partir de facteurs proprement linguistiques *vs* extra-linguistiques [12]. On notera que, dans une sorte de prolongement de [4], le problème est posé en termes d'identités et différences entre énoncés. Les différents types d'informations identifiés sont ensuite mis en rapport avec la problématique de la contraction de texte [13]. Ils permettent de définir des points précis permettant d'identifier des informations non essentielles d'un texte, et deux principes généraux pour en repérer le noyau dur.

Avec Pierre-Maurice Fauchère, Bès développe le système Vercingétorix I [10]. Il s'agit d'un système d'indexation de documents qui se propose d'améliorer l'indexation par l'introduction de symboles notant les relations syntaxiques dans les expressions documentaires et une mise en relation des expressions *via* un réseau sémantique. Le langage défini dans ce cadre cherche un compromis entre lisibilité par l'être humain et compréhension par la machine, ce qui conduit Bès à étudier les problèmes liés à la « naturalité » du langage, autour des possibilités d'ambiguïté et de paraphrase [14].

Enfin, à titre personnel, Bès entame un travail sur l'analyse de la notion de simplicité dans les travaux de grammaire générative, qui débouche sur la publication en 1981 d'un long article : *Simplicités, empirisme et interprétation psychologique dans*

la grammaire générative [11], suivi de cinq autres articles sur les travaux de Noam Chomsky [8, 16, 17, 19, 24]. Une note de [11] signale que « des présentations partielles de ce travail ont été faites, en janvier-février 1975, à l'université de Genève et à l'université Simon Fraser, et en 1977 à Paris III ». À Genève, en 1975, Piaget était professeur émérite. On ne sait pas si les deux hommes se sont rencontrés à cette occasion, mais il ne fait aucun doute que l'œuvre de l'épistémologue Piaget a influencé celle de Bès.

La notion de simplicité, dans les travaux de grammaire générative, est censée permettre de sélectionner des grammaires concurrentes ; on retrouve là la problématique du choix des hypothèses alternatives que Bès a abordée dans sa thèse. Bès reformule à partir des textes – formalise – pas moins de huit mesures différentes, classées comme *formelles* ou *empiriques*. À travers cette analyse des schémas de raisonnement qui justifient la grammaire universelle, il fait apparaître le caractère méthodologiquement discutable des mesures de simplicité formelle – des « règles du jeu » auxquelles on soumet les résultats descriptifs – et les incohérences du discours chomskyen, qui rejette en principe la simplicité formelle, mais l'utilise en pratique. Quelques mesures de simplicité empirique sont par ailleurs identifiées, mais, soit elles sont insuffisamment explicitées pour être testées, soit des contre-exemples proposés par Bès les infirment.

Cette même problématique des mesures de simplicité conduit Bès à faire le constat, dans *Conditions d'adéquation des théories linguistiques* [8], d'une « désempirisation » de la linguistique descriptive. Outre le caractère le plus souvent formel des mesures de simplicité, Bès y observe la faiblesse des données (cf. P1 de Protocoles) sur lesquelles s'appuie la notion de généralisation : les jeux de données considérés en pratique sont de taille très réduite et le cadre chomskyen interdit l'accès aux données du processus d'acquisition. Bès défend donc l'idée que les hypothèses linguistiques doivent être « empirisées » par le biais de leur intégration à des modèles sur l'acquisition, ou des modèles d'utilisation du langage.

Compétence-performance, distinction méthodologique ou affirmation empirique étudie l'idée chomskyenne selon laquelle « la compétence est utilisée dans la performance », et plus particulièrement la formulation selon laquelle la grammaire spécifierait les entrées et sorties du modèle de la performance [16]. Interprétée comme une hypothèse empirique, cette idée pose deux problèmes : (i) comment observer les entrées et sorties du modèle de la performance (c'est-à-dire spécifier les P1), (ii) la grammaire particulière est traitée comme une hypothèse de type boîte noire : tout mécanisme produisant les entrées et sorties désirées est adéquat. La boîte noire étant non interprétable psychologiquement, Bès formule une série de conditions permettant de mesurer « l'aptitude relative de différentes hypothèses à l'interprétabilité psychologique », telles que l'accroissement du champ des observables dont rend compte une hypothèse ou la réunion d'observables de domaines différents. Selon ces critères, il n'y a pas eu entre 1957 et le début des années 80 d'accroissement de l'aptitude à l'interprétabilité psychologique des hypothèses formulées dans le domaine de la compétence. Malgré son ancienneté, l'hypothèse directrice de la grammaire générative

voulant que la grammaire particulière (compétence) s'intègre au modèle d'utilisation du langage (performance) reste pure spéculation.

Grammaire générative : année 80 [17] présente une analyse extrêmement minutieuse des textes écrits par Chomsky autour de l'année 80, plus particulièrement *Filters and Controls* et *On Binding*. Nous ne pouvons entrer ici dans le détail de l'analyse, qui relève inadéquations observationnelles, absences de pouvoir explicatif et incohérences (par exemple les filtres sont sensibles aux traces et insensibles aux traces, ou des exemples sont grammaticaux un jour, agrammaticaux un autre). Signalons simplement que le nombre de pages de l'article ne rend pas compte de la somme de travail qu'il représente – en particulier, Bès fait en quelque sorte à la main ce qu'on demanderait à une machine aujourd'hui – et rapportons-en ces deux conclusions : (i) la valeur de validité des propositions descriptives des deux textes analysés ne devrait pas être accordée à la suite d'une lecture trop rapide et (ii) il faut ajouter aux propositions méthodologiques relevées chez Chomsky les trois suivantes :

– « il n'y a pas d'hypothèse valide sans couverture d'un champ d'observables bien établis [= Protocoles], suffisamment représentatifs et, en tout cas, non présélectionnés pour favoriser la validation » ;

– « une hypothèse faiblement ou mal explicitée ne peut pas être une hypothèse valide » ;

– « une hypothèse valide ne doit pas être une hypothèse qui ne peut pas être fausse ».

On reconnaît dans ces trois points le triplet < observations, hypothèses formalisées, test des hypothèses par rapport aux observations > que nous avons évoqué en introduction (section 1.2). Après ce travail, Bès a renoncé à suivre les écrits postérieurs de Chomsky.

Il faut noter que la critique de Chomsky par Bès vient d'un homme qui partage dans une très large mesure ses principes méthodologiques (cf. la section 1 de [17]). Par ailleurs, Bès, intéressé par la formalisation et le calcul, se disait admiratif du « premier Chomsky » et a dû être séduit par les propositions concrètes de ses premiers textes. Dès sa première année d'enseignement, en 1969, peu de temps après la diffusion de la grammaire générative en France, il met à son programme – dont il reste une copie dactylographiée – les grammaires formelles de Chomsky. Il y précise cependant que Chomsky n'est pas l'inventeur de l'idée qu'il faut traiter un ensemble infini d'énoncés, avec à l'appui des renvois à Hockett, Bloch, Bloomfield, Harris et Hjelmslev [4, p. 27]. Dans cette problématique déjà posée de la génération de l'ensemble des phrases d'une langue, Chomsky apporte en 1957 un système de formalisation et calcul effectif ; Bès y a adhéré. Chomsky émet aussi une hypothèse générale qui met en relation structures syntaxiques et structures cérébrales ; Bès n'y a pas cru.

L'analyse que fait Bès des textes de Chomsky se trouve aussi au cœur de *Fonctions de représentation et de communication* [24], article paru dans le volume *Psychologie de l'Encyclopédie de la Pléiade*, sous la direction de J. Piaget, J.-P. Bronckart et P. Mounoud. Cet article est indiqué « sous presse » dès [8] et [11] et a sans

doute été écrit en parallèle à [16] et [17]. Après une première partie proposant une délimitation des concepts organisant l'étude du langage humain, il contient dans sa partie centrale une analyse de la théorie de la grammaire générative, qui partage des analyses de [16] et [17] et les complète dans le cadre d'une critique de l'hypothèse chomskyenne sur l'ontogénèse du langage – hypothèse innéiste selon laquelle la « grammaire universelle » est une fonction qui va associer une certaine expérience linguistique à la « grammaire particulière » de l'âge adulte. Dans une troisième partie, l'article présente plusieurs alternatives ouvertes dans les études formalisées sur le langage, et notamment la grammaire de Montague, l'intelligence artificielle et les modèles de grammaires LFG et GPSG – tout récents à l'époque. Pour Bès, « il est licite de penser que *Lectures on Government and Binding* exprime l'épuisement d'une manière de traiter le langage, tandis que LFG ou GPSG, en renouant avec les exigences de formalisme fort du début de la grammaire générative, en exigeant l'insertion des structures linguistiques dans des procédés effectifs et computables sous-jacents à des conduites d'utilisation, en proposant des sémantiques fortes, ouvrent la voie à une nouvelle manière de pratiquer les études sur le langage » [24, p. 1827].

En 1983, Bès intègre avec quelques membres du CILN la recherche coopérative programmée *Étude linguistique de la signification appréhendée à travers l'ambiguïté et la paraphrase*, dirigée par Catherine Fuchs. Cette collaboration débouche sur *Référence, ambiguïté et paraphrase* [18], article qui étudie la question transparence-opacité. Bès y effectue dans un premier temps une relecture de *Word and Object*, de Quine. Il y identifie une série de conventions modélisantes et principes posés qui font que « l'anglais devient un autre objet [l'anglais-Quine], dont les conditions de grammaticalité sont très différentes de celles de l'anglais » [p. 177]. Il y discrimine ainsi ce qui, chez Quine, relève de l'observation et ce qui relève de la modélisation. L'article présente ensuite un vocabulaire de notions sémantiques définissant différentes configurations d'interprétation du syntagme nominal, vocabulaire qui doit servir à la notation explicite et systématique des observations sur l'interprétation des énoncés. On retrouve ici la problématique des Protocoles du paradigme 5P.

La collaboration avec C. Fuchs débouche également sur la direction d'un volume *Lexique et paraphrase* [25]. Les contributions clermontoises, de Bès et de Baschung, y articulent la notion de paraphrase avec la problématique de l'analyse syntaxique automatique, portant là la marque d'un nouveau projet.

4. Le Groupe de recherche dans les industries de la langue

4.1. Le projet ACORD : des grammaires d'unification

1984 peut être vue comme une année charnière. Bès dépose dans le cadre des actions thématiques programmées du CNRS un projet de recherche intitulé *Les théories sur le langage, leur formalisation, leur évaluation épistémologique et leur interprétation psychologique* [15]. « Diverses observations introduisent des doutes sur la validité épistémologique de la grammaire générative, mais l'analyse interne d'une théorie lin-

guistique est handicapée par le fait de ne pas disposer d'un instrument de lecture des théories sur le langage permettant de les évaluer et de les comparer. » Le projet envisage donc l'élaboration d'une terminologie en consultation avec l'épistémologie générale qui permette d'appréhender les aspects méthodologiques d'une théorie linguistique, l'établissement d'un inventaire des problèmes méthodologiques et conceptuels dans diverses approches linguistiques (grammaire générale, générative, de Montague ; paraphrase chez Harris ; etc.) et l'élaboration, à partir de là, d'un outil de lecture des théories sur le langage. Les bailleurs de fonds n'ont pas voulu de ce projet.

En revanche, un autre projet, plus appliqué, est accepté à la même époque par les instances européennes : le projet ACORD, *Construction and Interrogation of Knowledge Bases Using Natural Language and Graphics*. Il rassemble huit équipes, académiques et industrielles, de trois pays différents, et vise à produire une interface homme machine permettant de construire et d'interroger une base de connaissances commune en anglais, français et allemand. La tâche de l'équipe clermontoise est de développer la grammaire du français qui sera utilisée dans ce cadre. Avec Karine Baschung et Rachel Panckhurst, ses deux premières doctorantes, le mathématicien Pierre-François Jurie, puis Claire Gardent en post-doctorat, s'organise sous la direction de Bès une équipe qui sera baptisée plus tard, avec valeur rétroactive, *Groupe de recherche dans les industries de la langue (GRIL)*.

Le projet ACORD commence, à Clermont, par une étude de faisabilité d'une grammaire GPSG du français [20]. Le livre de référence (Gazdar *et al.*, 1985) est encore à l'état de brouillon, les Clermontois travaillent sur les documents préparatoires et en consultation avec Ewan Klein. Pour Bès, ce projet est l'occasion de mettre à l'épreuve un modèle qui est une grammaire générative, se voulant description et explication de la compétence, mais qui présente des éléments de rupture « de taille » par rapport à la tradition chomskyenne. « Les contraintes formelles incorporées aux grammaires particulières rendent celles-ci incorporables de manière explicite à des analyseurs syntaxiques automatiques, d'où leur intérêt en linguistique computationnelle : l'incorporation chomskyenne de la compétence à la performance cesse ainsi d'être un vague espoir. Rupture aussi au niveau de l'organisation de la syntaxe : plus de transformations, plus de pluralité de structures syntaxiques, plus de règles de passage entre ces structures. Rupture enfin au niveau de la sémantique : GPSG incorpore une sémantique calculable qui profite de la tradition de Montague, les traductions sémantiques étant exprimées en logique intensionnelle » [25, p. 9].

Linguistiquement, une attention particulière sera portée aux constructions interrogatives [21, ch. 2 et 3] et aux clitiques [22, 23, 26], dont on ne trouve pas l'équivalent en anglais. Les deux types de données posent des problèmes particuliers en matière d'ordre des mots [27].

La participation au projet de l'université d'Edimbourg, où est développée UCG, conduit à utiliser par la suite ce modèle pour le français, tandis que LFG est utilisé pour l'allemand. Plusieurs grammaires sont définies à Clermont. Se pose alors, comme dans *Identités et différences dans les unités de deuxième articulation* [4], la question du choix des modèles ou des grammaires alternatives. Dans la dernière partie du pro-

jet, Bès développe donc avec Pierre-François Jurie ce qu'ils appellent le « contrôle des grammaires » [28, 31]. Il s'agit de contrôler l'ensemble des phrases générées par la grammaire relativement à l'ensemble des phrases grammaticales spécifiées par un *métalangage descriptif*. Ce métalangage prétend formuler de façon compacte des observations sur la langue naturelle. Il est l'ancêtre des Propriétés (P2) du paradigme 5P (cf. ci-dessous, section 5).

Dans le prolongement du projet ACORD, Bès et ses collaborateurs organisent en 1990 à Clermont-Ferrand un colloque « mémorable »¹⁰ sur les grammaires catégorielles (GC). Y ont participé notamment Glyn Morrill, Michael Moortgat, Dirk Roorda, Elisabet Engdahl, Mark Steedman, Anna Szabolcsi... Le livre issu de ce colloque (Lecomte, 1992) contient « des articles pionniers sur des sujets qui ont été abondamment étudiés par la suite, comme le lien des GC avec la logique linéaire, le calcul de Lambek, les réseaux de preuve, les modalités structurelles, etc. ». Bès et Baschung y présentent un article qui est un jalon dans le développement des Propriétés : aux axiomes du métalangage descriptif de [31] correspondent des filtres dans les grammaires catégorielles [30]. Martine Emorine, en thèse avec Bès, présente un travail sur l'incorporation des verbes supports dans la GC. Pierre-François Jurie, associé au GRIL, signe un article, *Orbits*, dans lequel il définit une réduction des arbres d'analyses des GC à des « structures orbitales » similaires à des arbres de dépendances et propose un algorithme de vérification de la grammaticalité utilisant ces structures. Marie-Ange Légeret, en thèse avec P.-F. Jurie, y présente quant à elle « des avancées capitales sur la voie de la démonstration d'un théorème important : l'équivalence entre les grammaires de Lambek et les grammaires hors contexte, démonstration qu'elle fera finalement peu de temps après que ce résultat n'ait été établi par un chercheur russe, Mati Pentus, en 1993, mais avec des moyens complètement différents ».

Le 10 octobre 1990, *Le Monde* publie dans son supplément *Campus* un article sur le DEA « Linguistique et informatique » de Clermont-Ferrand. Un chercheur des Laboratoires de Marcoussis y déclare « l'idée d'un pôle high-tech à Clermont-Ferrand, cela m'a paru surprenant » ; mais c'était bien réel : l'équipe du GRIL était à la pointe de la recherche en linguistique computationnelle. En juin 1990, Karine Baschung soutient sous la direction de Bès une des premières thèses en France, avec celle d'Anne Abeillé en 1991, sur les grammaires d'unification (Baschung, 1990).

4.2. *Autres projets européens et industriels*

À la suite du projet ACORD, Bès a poursuivi une politique de participation à des projets européens et industriels. « La participation à des recherches finalisées est le résultat d'un choix épistémologique explicite. Il ne s'agit pas seulement pour l'équipe de

10. Le mot, de même que les deux citations qui suivent dans ce paragraphe, est d'un des relecteurs de la revue TAL. Nous le remercions vivement d'avoir attiré notre attention sur l'importance de cet événement, ainsi que sur l'influence de Bunge sur Bès (cf. note 6) et sur le caractère pionnier de la thèse de Karine Baschung.

multiplier par un facteur d'au moins vingt les crédits officiels et de préparer les points d'insertion des étudiants avancés, mais aussi de se donner des conditions d'adéquation réelles où les modèles théoriques doivent être testés. »¹¹

Le GRIL participe d'abord en 1990-1992 au projet GENELEX, avec comme contributions un état de l'art sur la sémantique lexicale, un rapport sur la modélisation des entrées lexicales et une analyse des tables du LADL. Dans le prolongement de ces travaux, signalons deux articles. *Semantic Features in a Generic Lexicon* [32], écrit avec Alain Lecomte, présente un système de description de l'information sémantique qui se veut neutre et générique, dans le sens où différentes théories doivent pouvoir accéder à l'information que chacune demande. Le formalisme permet de rendre compte d'une part du fait que des constructions syntaxiques différentes réalisent des sens différents, d'autre part des liens éventuels entre des paires construction-sens différentes. Ces concepts se retrouvent dans le lexique génératif de (Pustejovsky, 1995). Autre texte issu du travail sur GENELEX, *Les tables de Méthodes en syntaxe : introduction à un mode d'emploi* [33] relève les défauts de formalisation des tables définies par Maurice Gross (1975), propose une solution de remédiation et aborde la question de l'adéquation des tables résultantes. On retrouve à nouveau dans ces deux articles les préoccupations centrales de Bès : formalisation et notation des observations d'une façon indépendante des théories.

En 1991, le projet européen Hyperdocsy explore la possibilité d'une génération automatique de documents techniques multilingues. Le GRIL participe à l'analyse d'un corpus de documents et à une évaluation des techniques de génération existantes [29]. À partir de 1991 également, le GRIL entame une collaboration avec Aérospatiale autour des langues contrôlées (LC), qui durera pendant près d'une décennie. Les objectifs sont de produire des outils d'aide à la rédaction en LC, d'aide à l'apprentissage de la LC et de vérification de conformité des textes aux normes de la LC. Les travaux du GRIL ont porté sur la spécification et la cohérence interne des lexiques de l'anglais simplifié et la réalisation d'un vérificateur/diagnostiqueur de l'application des règles de l'anglais simplifié. Par la suite, a été développé un système générique d'évaluation des vérificateurs de langues contrôlées.

En 1993-1994, le projet européen DELIS (*Descriptive Lexical Specification*) poursuit le travail sur le lexique avec l'exploration d'une nouvelle méthodologie pour la description détaillée et fondée sur des corpus des propriétés syntaxiques des verbes et noms. La contribution du GRIL inclut notamment un dictionnaire des verbes de perception en français, par K. Baschung, et un rapport de Bès : *Methodological Foundations of Lexicon Building* [34]. Sauf erreur, c'est dans ce texte qu'on trouve la première mention par Bès du concept de modélisation de l'observateur, central dans 5P. Un lexique est une base de données. La diversité et la complexité des informations qu'on peut ranger sous l'étiquette « sémantique », et par suite vouloir mettre dans le lexique, sont telles, que l'explicitation des principes qui vont guider les descriptions des entrées lexicales est indispensable. Il ne peut pas y avoir de formalisation ou de

11. Extrait du rapport d'activité du GRIL, 10 avril 1991.

modélisation sans une description attentive et explicite des données, qui ne sont ni évidentes ni bien définies par le sens commun.

Bès remarque qu'il n'y a pas d'accord explicite chez les linguistes sur ce que sont les données linguistiques sémantiques. Pour tenter de les cerner, il explicite deux « propositions ». Dans la première, l'observateur utilise trois prédicats sur les phrases (« syntaxiquement bien formée », « constructible », « déviante »), ce qui produit un certain jeu de données. Bès montre qu'il est impossible de modéliser la sémantique lexicale dans les termes de cette proposition, en raison d'une absence d'intersubjectivité, consécutive à l'imprécision des théories sur leur traitement des phrases « déviantes », et du faible pouvoir expressif des prédicats unaires, tels que +/- humain, utilisés en sémantique lexicale. Dans la deuxième proposition, une nouvelle compétence est attribuée à l'observateur : celle de reconnaître des paraphrases. Elle permet de partitionner les phrases déviantes en deux sous-ensembles selon qu'il existe ou non une paraphrase non déviante. Si toutes les paraphrases envisageables sont déviantes, alors le problème relève de l'ontologie, s'il existe une paraphrase non déviante, le problème relève de la description de la langue. Un principe distingue ici la sémantique lexicale de la sémantique encyclopédique.

Nous ne saurions évaluer l'apport de ce texte à l'étude de la sémantique. Bien qu'inachevé, dans le sens où des parties ne sont qu'esquissées, il apporte au moins, selon nous, une illustration concrète de la forme que peut prendre la modélisation de l'observateur et son incidence sur les données.

En 1993-1994, Bès participe aussi au projet LRE *Towards a declarative theory of discourse*. Enfin, notons qu'en parallèle à tous ces projets, le GRIL développe durant cette période de recherche appliquée ses propres outils d'analyse de corpus, notamment l'analyseur morphologique SMORPH (Aït-Mokhtar, 1998), et des ressources linguistiques pour l'analyse du français, de l'espagnol, du portugais, puis plus tard du roumain.

5. Le formalisme des Propriétés

On peut relever entre 1996 et 1999 une quasi-absence de publication dans la bibliographie de Bès. Après ces années de projets en partenariat avec l'industrie, celui-ci s'est recentré sur ses centres d'intérêt fondamentaux. Le travail débouche en 1999 sur un de ses articles majeurs, *La phrase verbale noyau en français* [35], dans lequel il expose clairement ses positions méthodologiques, sous la forme du paradigme 5P, déjà évoqué en introduction, et présente un formalisme pour la description des langues naturelles : les Propriétés.

Le formalisme des Propriétés est un prolongement du métalangage descriptif de *The Control of UCG Grammars* [31]. Les « axiomes » qui composent ce métalangage se distinguent des grammaires formelles en ce qu'ils ne sont pas des règles de construction par concaténation ; ils sont « destinés, idéalement, à présenter d'une façon ordonnée, explicite et modulaire des observations sur les langues naturelles » [31,

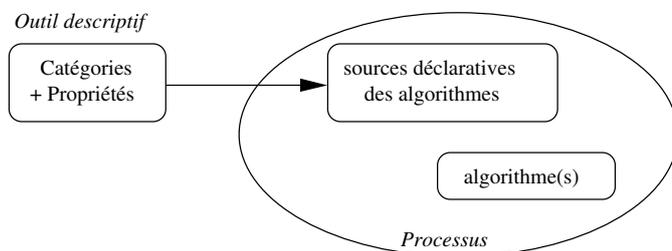


Figure 1. Place relative des Propriétés et Processus

p. 49]. De façon analogue, les Propriétés consistent en un système de formules qui visent à donner une description axiomatique des suites des langues décrites. Pour décrire un ensemble de suites, un ensemble de Propriétés dit quelles sont les catégories qui peuvent ou doivent y apparaître, une seule ou plusieurs fois, lesquelles exigent ou excluent la présence de telle(s) autre(s), quelles sont leurs relations de précédence et quelles sont les contraintes d'accord. GPSG avait introduit une séparation des informations de dominance (ID) et de linéarité (LP) ; on peut voir dans les Propriétés une décomposition maximale de l'information exprimée par un ensemble de règles ID. Les Propriétés sont interprétées conjonctivement, c'est-à-dire qu'appartiennent à la langue spécifiée les suites qui satisfont *toutes* les propriétés. Des Propriétés dites « de fléchage » permettent d'exprimer par ailleurs les liens entre les catégories apparaissant dans les suites spécifiées.

Si un ensemble de Propriétés constitue une description formalisée d'un langage, celles-ci, pour Bès, ne sont pas pour autant une grammaire. « Il n'y pas de grammaire dans 5P », écrit-il [35, p. 348].

« La notion de grammaire telle qu'elle est pratiquée dans les modèles dits calculables ou formels, recouvre une double fonction : (i) outil descriptif et (ii) source déclarative des algorithmes d'analyse ou de génération automatiques. » [35, p. 348]

« La fonction d'une grammaire "is to play the role [...] of information that fits directly into a model of processing". » [35, p. 348]

« Les fonctions qui sont censées être assurées par une grammaire calculable sont, dans 5P, analytiquement assignées aux Propriétés (P2) d'une part, aux sources déclaratives qui seront utilisées par les algorithmes d'analyse et de génération automatiques dans les Processus (P5) d'autre part. Ces sources déclaratives des algorithmes dans les Processus (P5) ne sont pas les Propriétés (P2), mais elles doivent être le résultat d'une fonction qui a comme entrée les Catégories et les Propriétés. » [35, p. 350]

À partir de ces déclarations, on peut schématiser la vision de Bès par la figure 1. La description de la langue faite par les Propriétés (P2) est distinguée des Processus (P5), c'est-à-dire les algorithmes et leurs sources déclaratives. Bès écrit par ailleurs :

« La langue (compétence) reste distincte de la parole (performance) mais la langue « n'est pas utilisée dans la parole ». Un peu plus techniquement, cela veut dire que les Propriétés (P2) ne sont pas la source déclarative utilisée dans les Processus (P5) d'analyse et de génération automatiques. La grammaire n'est donc pas utilisée dans la performance et ceci pour la simple raison qu'il n'y a pas de grammaire dans 5P. » [35, p. 348]

Bès semble donc assigner aux Propriétés la modélisation de la compétence. Si on tentait de plaquer le concept de compétence à la Chomsky sur le schéma ci-dessus, il désignerait plutôt une case fusionnant l'outil descriptif et les sources déclaratives de l'algorithme, plus, peut-être, les algorithmes eux-mêmes, autrement dit le tout. Plutôt que d'assigner aux Propriétés la modélisation de la compétence, il nous semble que Bès *décompose la compétence* en une partie « langue » et une partie « processus ». La principale raison pour faire cela, nous semble-t-il, tient à ce que les conditions d'adéquation des deux sous-parties sont très différentes ; pour la partie « langue », il s'agit d'observations sur des énoncés (est-il grammatical, ambigu, paraphrase de cet autre, etc.), pour la partie « processus », il s'agirait d'observations sur les contenus mentaux. Bès se concentre sur la langue par réalisme, parce qu'on n'a pas ou très difficilement accès aux contenus mentaux. Ainsi, dans *La linguistique entre science et ingénierie* [37, p. 31], « la paire <expr, Obs> [où *Obs* est le résultat de l'activité de l'observateur sur l'expression *expr*] [doit satisfaire] les conditions sur les données requises par les choix méthodologiques basiques [c'est-à-dire que l'objet empirique doit être accessible par des protocoles définis et renouvelables, et exister indépendamment du dispositif cognitif]. Ce n'est pas le cas des contenus mentaux, pourtant souvent invoqués ». Bès écarte ainsi délibérément toutes les justifications cognitivistes de la grammaire générative. Implicitement, il dit aussi que l'adéquation descriptive d'une grammaire particulière (par exemple relativement à un ensemble d'énoncés grammaticaux et agrammaticaux) ne devrait pas être utilisée pour corroborer l'adéquation du *modèle* de grammaire. On peut bien sûr s'intéresser aux processus mentaux et vouloir les modéliser, mais il faut se donner des critères d'adéquation externes pertinents. Inversement, le formalisme des Propriétés montre que les linguistes peuvent tout à fait ne *pas* s'intéresser aux processus mentaux ¹².

6. Derniers travaux

6.1. La sémantique en 5P et la compositionnalité

Le formalisme des Propriétés cible essentiellement la syntaxe (relations d'ordre et de cooccurrence entre catégories) mais offre un type de propriété particulier, les propriétés de fléchage, qui permet d'exprimer les liens entre les items figurant dans les suites spécifiées. Les paires de fléchage fournissent l'entrée des fonctions sémantiques, qui devraient produire en sortie les représentations sémantiques. *Empiricité en*

12. Merci à Karine Baschung-Kray de m'avoir signalé ce dernier point.

linguistique et grammaire de Montague [36] présente dans sa huitième partie un programme de recherche en sémantique pour 5P : y sont exposés d'une part le choix d'une compositionnalité qui n'est pas un homomorphisme entre structures algébriques de la langue naturelle et l'univers de dénotation, d'autre part un plaidoyer pour la définition d'un système d'observation systématique susceptible d'exprimer les observations d'ordre sémantique dans un langage *ad hoc*¹³. Ce programme de recherche est justifié par une analyse détaillée des travaux de Montague.

On sait qu'une des idées centrales de Montague est qu'il n'y a pas de différences théoriques fondamentales entre les langues naturelles et les langues artificielles des logiciens ; la syntaxe et la sémantique¹⁴ de ces différentes langues relèveraient d'une même théorie mathématique. L'étude de Bès aboutit à la conclusion qu'il n'y a rien dans la grammaire de Montague qui autorise à considérer cette idée comme une hypothèse d'une science empirique. Il s'agit plutôt d'un « commandement méthodologique *homomorphisme syntaxe-sémantique tu construiras* ».

Parmi les éléments clés qui conduisent à cette conclusion, citons les deux faits suivants :

- l'anglais traité par Montague n'est pas l'anglais réel, c'est un anglais syntaxiquement « mis au pas » (comme l'était l'anglais de Quine étudié dans [18]) et avec un vocabulaire qui en partie n'existe pas en anglais ;
- peu de restrictions sont posées sur l'algèbre de la syntaxe des langues naturelles.

Pour établir ce dernier point, Bès reformule dans le modèle des grammaires formelles le système de spécification d'un fragment de l'anglais utilisé par Montague. La grammaire résultante contient des règles de type 0. Bès expose ensuite quelques inadéquations descriptives (sur- et sous-génération de la grammaire, erreurs d'interprétations produites par les fonctions de traduction), et il montre comment les exemples non générés *peuvent l'être* en développant un peu la grammaire de PTQ. Il le fait dans le strict respect des restrictions du système, mais en utilisant ses permissivités, au premier rang desquelles les effacements et permutations des règles de type 0. L'absence de contraintes sur les opérations syntaxiques fait que rien ne peut être utilisé pour distinguer langues naturelles et langages artificiels des logiciens. L'idée qu'il n'y a pas de différence entre ces deux types de langues ne peut être falsifiée ; du point de vue de la science empirique, c'est donc une non-hypothèse.

L'étude est complétée par une tentative d'identification des différences entre les deux types de langues, qui caractérise la « naturalité des langues naturelles ».

13. Une problématique qu'on retrouve dans [18], avec son vocabulaire de notions sémantiques définissant différentes configurations d'interprétation du syntagme nominal, cf. section 3.

14. Ou seulement la *relation* syntaxe-sémantique ? Bès ne soulève pas ce problème d'ambiguïté, qui permettrait sans doute d'aborder la question sous un angle différent.

6.2. Une grille d'analyse épistémologique des sciences du langage

L'analyse des propositions de Montague se retrouve dans *La linguistique entre science et ingénierie* [37] aux côtés d'une analyse de travaux relevant de ce que Bès appelle l'approche « texto-algorithmique », en l'occurrence deux analyseurs syntaxiques de surface « robustes » fondés sur des règles. Les types de travaux abordés sont très différents, mais cette différence ne justifie pas une différence dans la méthode scientifique. Bès présente donc une grille d'analyse unique pour les appréhender. Elle porte « sur les trois éléments clés du domaine de l'empirie : le domaine des observations, des hypothèses et du test de corroboration ». Les hypothèses sont étudiées selon une sous-grille examinant (a) le type de formalisme utilisé, (b) les possibilités d'extension et de portabilité linguistiques et (c) leurs visées descriptives ou explicatives. La vocation de la grille étant de permettre l'explicitation de multiples points précis dans les approches étudiées, son application est impossible à résumer.

En conclusion, Bès relève la quasi-symétrie des avantages et inconvénients des deux approches et suggère de les considérer comme complémentaires. L'approche texto-algorithmique a révélé le problème posé par les ambiguïtés de catégorisation morphosyntaxique, de segmentation et de rattachement. La question est posée, et reste ouverte, de savoir si les formalismes de calcul existants sont capables d'être utilisés pour résoudre cette difficulté. « On ne voit pas pourquoi les deux approches ne pourraient pas collaborer pour avancer vers une solution. [...] Deux conditions sont nécessaires (mais non suffisantes) pour avancer dans cette voie : une application sincère du test de corroboration et l'accès à l'information linguistique sous-jacente aux réussites et échecs du test. » [37, p. 79]

6.3. Travaux en cours

À partir de 2002, Bès s'engage dans un travail qui se rapproche de l'approche texto-algorithmique par le type d'observation : il s'agit d'identifier dans des textes tout venant les relations entre introducteurs de subordonnées et syntagmes verbaux noyau. On trouve un aperçu de ce travail dans [39] et [40]¹⁵. Les introducteurs et les syntagmes verbaux y sont vus comme des parenthèses qui s'équilibrent : chaque introducteur – parenthèse ouvrante – est « fermé » par un syntagme verbal noyau. On peut noter que ce concept définit un genre de constituant qu'on ne retrouve pas dans l'analyse en constituants classique. On laissera à d'autres le soin de dire s'il y a là aberration ou clairvoyance ; la question aurait laissé Bès indifférent : « la non-observabilité de principe des structures syntaxiques est le choix fait dans 5P » [36, p. 17]. Le problème pour Bès est plutôt de déterminer si l'identification de ces éléments et de leurs relations est possible avec une certaine information, assez limitée, l'idée étant de faire le maximum avec le minimum d'information.

15. [38] est un document de travail qui nécessiterait une contextualisation pour être publié. [41] décrit le système de reconnaissance des syntagmes verbaux noyau.

L'approche est aussi texto-algorithmique dans la méthode implantée, mais l'objectif était de cerner précisément l'information utilisée dans les règles et les situations d'échec, autrement dit d'appliquer sincèrement le test de corroboration. Dans le même temps, du point de vue « théorique », Bès s'intéressait à l'analyse des approches statistiques (cf. [42]). Ces recherches n'ont pu être totalement abouties. En 2005, un accident vasculaire cérébral a mis un terme abrupt à la carrière de Gabriel Bès, qui est alors retourné vivre en Argentine, auprès de sa fille et ses petits-enfants.

Dans les dernières années de sa vie académique, Bès s'était aussi rapproché de l'équipe de Zulema Solana à l'Universidad Nacional de Rosario. Ensemble, ils ont créé les *Jornadas Argentinas de Lingüística Informática : Modelización e Ingeniería* (JALIMI), colloques ayant eu lieu en 2004 et 2005, puis en 2008, et la revue *Infosur*¹⁶. Bès et Solana ont développé des travaux sur le syntagme verbal noyau espagnol [43, 44], dans le prolongement de [35] et [41], et l'équipe de Rosario a continué par la suite à travailler avec les outils informatiques et conceptuels du GRIL (cf. *Infosur*).

7. Conclusion

La carrière de Gabriel Bès s'étend de la linguistique structurale au traitement automatique des langues en passant par la grammaire générative, avec comme fil directeur la préoccupation épistémologique. La question centrale est celle de la spécification des conditions d'adéquation des descriptions ou hypothèses, et derrière elle la question du choix des hypothèses concurrentes. Dans une « préface sous forme de lettre » adressée aux personnes à qui il soumettait son rapport *Indicateurs dans les chaînes* [38], Bès écrivait :

« Il y a une faute intellectuelle dans les discussions sur la comparaison des théories, bien identifiée, qui consiste dans ceci : à un moment donné, on connaît deux manières de faire, disons A et B. Les partisans de la manière de faire B démontrent que la manière A est mauvaise. Et de là ils concluent : DONC, B est la bonne. C'est encore une supercherie intellectuelle. Appelons cette supercherie la supercherie de la justification par l'exclusion. »

À la suite de quoi, il déplorait « entendre les praticiens de l'approche texto-algorithmique utiliser ce type de raisonnement à longueur de discussions ».

Pour conclure, nous voudrions montrer que le sophisme pointé par Bès se retrouve ailleurs. Au chapitre 5 de *Structures syntaxiques*, Chomsky (1957) mobilise un « faisceau de présomptions » censé montrer l'inadéquation du modèle syntagmatique. Le raisonnement passe par deux procédés. Le premier met en jeu un nouveau type d'observation – si la langue contient telle(s) phrase(s), alors elle contient telle(s) autre(s) phrase(s) – qui n'est pas à proprement parler posé comme nouveau critère d'adé-

16. <http://www.infosurrevista.com.ar>

quation ¹⁷, mais est utilisé pour arguer d'une plus grande simplicité de l'approche transformationnelle (sections 5.2 sur la coordination et 5.4 sur le passif). Le second présente une grammaire avec règle contextuelle et règle d'inversion (section 5.3 sur l'élément *Verbe* en anglais). Pour ces trois exemples, Chomsky écrit successivement :

« Ce comportement apparaît fort complexe si nous voulons intégrer ces syntagmes à une grammaire (Σ , F). » (section 5.2)

« Le lecteur peut voir aisément que pour reproduire l'effet de (28 iii) et de (29) sans sortir des limites d'un système syntagmatique (Σ , F), il faudrait faire un exposé singulièrement complexe. » (section 5.3)

« Cette répétition inélégante [ne peut] être évitée que si nous excluons délibérément le passif de la grammaire syntagmatique et le réintroduisons par une règle [transformationnelle]. » (section 5.4)

Trois énoncés qui sont autant de justifications par l'exclusion ¹⁸.

Justement attaché au problème de la justification des grammaires, Chomsky identifie au chapitre 6 trois types de relations entre la théorie générale et les grammaires qui en découlent. Il retient la troisième, la plus « modeste », les deux autres étant jugées inatteignables ; il s'agit d'une procédure d'évaluation des grammaires permettant de dire laquelle de deux grammaires est la meilleure.

Ce choix appelle deux remarques. Premièrement, on peut se demander s'il ne favorise pas l'apparition du sophisme de la justification par l'exclusion, tant il est tentant de passer de « B est meilleure que A » à « B est bonne ». Deuxièmement, il présuppose une situation idéale où l'on a deux grammaires qui satisfont toutes deux les conditions d'adéquation externes. Or s'il y a une chose que le traitement automatique des langues a montré, c'est qu'à ce jour personne n'a encore défini la moindre grammaire qui satisfasse des conditions d'adéquation externes un tant soit peu réalistes ¹⁹. Ce reproche n'est pas nouveau ; en 1970, Lyons écrivait :

« On entend parfois dire que la formulation des buts de la théorie linguistique par Chomsky en termes de comparaison de grammaires alternatives passe sur le fait qu'on n'a pas de grammaire, même partielle, pour beaucoup de langues et qu'on n'a de grammaire prête d'être complète

17. Nous voyons là un défaut dans la modélisation de l'observateur : la grammaire doit-elle rendre compte de la simple présence des deux phrases ou doit-elle rendre compte d'une *relation* entre ces deux phrases ?

18. Cf. notamment le *que* du troisième énoncé. Le procédé rhétorique est rendu manifeste par le fait que les grammaires d'unification ont apporté des réponses à ces problèmes, en associant des traits à des formes morphologiquement complexes pour 5.3 et en permettant d'associer des représentations sémantiques équivalentes à des phrases différentes pour 5.2 et 5.4.

19. Dans le sens de « qui décrivent le réel » et non de « qui se limitent à ce qu'on sait faire ». Exemple : spécifier l'ensemble des phrases grammaticales de l'anglais, cas douteux mis de côté, sans prolifération d'analyses non motivées par des différences d'interprétation et en excluant les phrases manifestement agrammaticales.

pour aucune langue. C'est un fait. Mais la conclusion selon laquelle, dans ces circonstances, il est prématuré de parler de comparer les grammaires n'en découle pas. La construction d'un ensemble de règles grammaticales pousse le linguiste à prendre des décisions pour traiter les données d'une façon plutôt qu'une autre. Même si les règles décrivent seulement une petite partie des données, il doit y avoir, explicite ou implicite, une comparaison des alternatives. C'est le but de la théorie linguistique, dit Chomsky, de rendre les alternatives explicites et de formuler des principes généraux pour choisir entre elles. » (Lyons, 1970, p. 41)

Sur cette question du choix des descriptions ou théories alternatives, la position de Bès, en tant qu'empiriste, serait qu'en présence de deux grammaires satisfaisant des conditions d'adéquation externes sommaires, le moyen de les départager est tout simplement d'étendre les conditions d'adéquation, en notant de façon explicite et contrôlée les observations dont il faut rendre compte. Il n'y a pas d'autre procédure de décision pour choisir une théorie que l'adéquation aux observations. « Spécifier des données, évaluer leur intersubjectivité, les utiliser pour vérifier les conséquences des hypothèses, c'est un problème extraordinairement grave qu'il faut mettre à la première place des préoccupations scientifiques sur la linguistique. »

À travers les multiples projets d'annotation de corpus, la communauté du traitement automatique des langues a étendu le champ d'observation de la linguistique formelle bien au-delà des jugements de grammaticalité. L'adéquation des systèmes par rapport aux observations y est couramment mesurée. L'œuvre de Bès pourra sans nul doute aider à fixer la valeur épistémologique de ces pratiques et à les améliorer.

Remerciements

Je remercie en premier lieu Karine Baschung-Kray : les échanges que j'ai eus avec elle pendant la préparation de cet article ont été déterminants. Merci également à Caroline Hagège pour ses relectures et commentaires, à Sylvain Pogodalla pour m'avoir aidé à comprendre le travail de Bès sur Montague, à Jean-Claude Pariente pour ses informations sur le début de la carrière de Bès et aux relecteurs de la revue TAL pour leurs commentaires, pour la plupart incorporés. Merci enfin à Patricia Bès et aux éditeurs pour leur autorisation de mise en ligne des articles de Bès.

8. Bibliographie partielle de Gabriel Bès

La bibliographie qui suit se limite aux articles référencés dans cet article. Une bibliographie complète, ainsi qu'une liste des thèses et HDR dirigées par Bès, est disponible en ligne : <http://lrl.univ-bpclermont.fr/spip.php?article414>.

[1] Bès G., « Phonétique, phonologie et communication interdialectale », in Schmitt L. E. (dir.), *Verhandlungen des Zweiten Internationalen Dialektologenkongresses*, vol. I, F. Steiner, Wiesbaden, p. 76-83, 1967. <hal-01100514>

- [2] Bès G., « Parenté génétique et parenté typologique », *La linguistique*, 2, P.U.F., Paris, p. 139-150, 1967. <hal-01059289>
- [3] Bès G., « Trait distinctif », *Word*, 23, p. 37-46, 1967. <hal-01059296>
- [4] Bès G., *Identités et différences dans les unités de deuxième articulation*, Thèse de doctorat, Université René Descartes - Paris V, 1972. <tel-01095229>
- [5] Pariente J.-C., Bès G., *La linguistique contemporaine, Dossiers Logos*, n° 69, P.U.F., Paris, 1973. <hal-01097664>
- [6] Bès G., « Lingüística », *Ciencia Nueva*, n° 26, p. 8-14, septembre 1973. <hal-01097665>
- [7] Bès G., « Structuration linguistique et mesure de l'information. » in Mahmoudian M. (dir.) *Linguistique fonctionnelle. Débats et perspectives*, P.U.F., Paris, p. 129-141, 1979. <hal-01100168>
- [8] Bès G., « Conditions d'adéquation des théories linguistiques », *Cahiers du Centre Interdisciplinaire des Sciences du Langage*, n° 2, p. 85-95, 1980. <hal-01064409>
- [9] Bellot-Antony M., Bès G., Hadjadj D., Pouzet R., Rousseau-Payen N., « La problématique de la contraction de texte », *Condenser*, n° 1, Adosa, Clermont-Ferrand, p. 13-44, février 1980. <hal-01071472>
- [10] Bès G., Fauchère P.-M., « Le système documentaire Vercingétorix I », *Condenser*, n° 1, Adosa, Clermont-Ferrand, p. 57-94, février 1980. <hal-01071423>
- [11] Bès G., « Simplicités, empirisme et interprétation psychologique dans la grammaire générative », *Archives et documents de la Société d'histoire et d'épistémologie des sciences du langage*, n° 1, Saint-Cloud, p. 1-84, 1981. <hal-01064415>
- [12] Bellot-Antony M., Bès G., Hadjadj D., Pouzet R., Rousseau-Payen N., « La contraction de texte. Les différents types d'information », *Condenser*, n° 3, Adosa, Clermont-Ferrand, p. 33-81, avril 1982. <hal-01071476>
- [13] Bellot-Antony M., Bès G., « Les différents types d'information et la contraction de texte », *Condenser*, n° 4, Adosa, Clermont-Ferrand, p. 5-46, février 1983. <hal-01071479>
- [14] Bès G., « Les structures syntaxiques du langage Vercingétorix I. Problèmes liés à leur naturalité », *Traitement automatique des langues naturelles et systèmes documentaires, Condenser, Supplément n° 1*, Adosa, Clermont-Ferrand, p. 109-134, 1984. <hal-01071426>
- [15] Bès G., « Les théories sur le langage, leur formalisation, leur évaluation épistémologique et leur interprétation psychologique », *Projet de recherche pour une action thématique programmée*, 1984. <hal-01143091>
- [16] Bès G., « Compétence-performance, distinction méthodologique ou affirmation empirique », in Auroux S., Glatigny M., Joly A., Nicolas A., Rosier I. (dir.) *Matériaux pour une histoire des théories linguistiques*, Université de Lille III, p. 638-647, 1984. <hal-01064441>
- [17] Bès G., « Grammaire générative : année 80 », *Lingvisticae Investigationes*, IX :1, John Benjamins B. V., Amsterdam, p. 1-54, 1985. <hal-01064446>
- [18] Bès G., « Référence, ambiguïté et paraphrase », in Fuchs C. (dir.), *Aspects de l'ambiguïté et de la paraphrase dans les langues naturelles*, Peter Lang, p. 167-206, 1985. <hal-01010045>
- [19] Bès G., « Teoría lingüística y política lingüística », *Revista Argentina de Lingüística*, vol. 1, n° 1, p. 5-16, 1985. <hal-01048572>

- [20] Bès G., Baschung K., *Feasibility of a GPSG French grammar*, ACORD ESPRIT Project 393, November 1985. <hal-01101321>
- [21] Baschung K., Bès G., Panckhurst R., Zeevat H., *Contextual phenomena in dialogue*, ACORD ESPRIT Project 393, Task 2.3, 1986. <hal-01143099>
- [22] Bès G., « Cliticos en frances y modelos linguisticos », *Revista Argentina de Lingüística*, vol. 2, n° 2, p. 246-265, 1986. <hal-01010059>
- [23] Baschung K., Bès G., Corluy A., Guillotin Th., “Auxiliaries and Clitics in French UCG Grammar”, *Conference of the European Chapter of the Association for Computational Linguistics*, p. 173-178, 1987. <halshs-00372903>
- [24] Bès G., « Fonctions de représentation et de communication », in Piaget J., Bronckart J.-P., Mounoud P. (dir.), *Psychologie, La Pléiade*, Gallimard, Paris, p. 1757-1835, 1987. <hal-01065925>
- [25] Bès G., Fuchs C. (dir.), *Lexique et paraphrase, Lexique*, vol. 6, Presses universitaires de Lille, 1988.
- [26] Bès G., « Clitiques et constructions topicalisées dans une grammaire GPSG du français », *Lexique*, vol. 6, Presses universitaires de Lille, p. 55-81, 1988. <hal-01096803>
- [27] Bès G., Gardent C., “French Order Without Order”, *Proceedings of the 4th Conference of the European Chapter of the Association for Computational Linguistics*, p. 249-255, 1989. <hal-01021826>
- [28] Bès G., Jurie P.-F., “UCG Grammars ; the control of their descriptive adequacy”, in Bès G. (dir.), *Technical Documentation*, ACORD ESPRIT Project 393, Laboratoires de Marcousis, December 8 1989. <hal-01143084>
- [29] Baschung K., Bès G., Carcagno D., Fournier C., *Text Generation*, HYPERDOCSY ESPRIT Exploratory Action 5652, Task 2, October 1991. <hal-01143504>
- [30] Bès G., Baschung K., « Filtres dans une grammaire catégorielle », in Lecomte A. (dir.), *Word Order in Categorical Grammar / L'ordre des mots dans les grammaires catégorielles*, Adosa, Clermont-Ferrand, p.93-104, 1992. <halshs-00372938>
- [31] Jurie P.-F., Bès G., “The Control of UCG Grammars”, in Bès G., Guillotin Th. (eds), *A Natural Language and Graphics Interface : Results and Perspectives from the ACORD Project*, Springer, p. 47-64, 1992. <hal-01101333>
- [32] Bès G., Lecomte A., “Semantic features in a generic lexicon”, in Saint-Dizier P., Viegas E. (dir.), *Computational Lexical Semantics*, Cambridge University Press, p. 141-162, [1993] 1995. <hal-01146172>
- [33] Bès G., « Les tables de Méthodes en syntaxe : introduction à un mode d’emploi », *Cahiers de praxématique*, vol. 22, n° 1, Université Paul-Valéry, Montpellier III, p. 55-80, 1994. <hal-00990802>
- [34] Bès G., *Methodological foundations of lexicon building*, DELIS Project LRE 61034, revised version, July 1995. <hal-01101355>
- [35] Bès G., « La phrase verbale noyau en français », *Recherches sur le français parlé*, vol. 15, Publications de l’Université de Provence, p. 237-358, 1999. <hal-01005527>
- [36] Bès G., *Empiricité en linguistique et grammaire de Montague : la sémantique en 5P et la compositionnalité*, Rapport de recherche, GRIL, Université Blaise-Pascal, Clermont-Ferrand, 2001. <hal-01101344>

- [37] Bès G., « La linguistique entre science et ingénierie », in Cori M., David S., Léon J. (dir.) *Problèmes épistémologiques du TAL*, TAL, vol. 43, n° 3, p. 57-81, 2002. <hal-01054333>
- [38] Bès G., *Indicateurs dans les chaînes*, Rapport de recherche, GRIL, Université Blaise-Pascal, 2002.
- [39] Bès G., Dahl V., Guillot D., Lamadon L., Milutinovici I., Paulo J., “A parsing system for balanced parentheses in NL texts”, *14th meeting of Computational Linguistics in the Netherlands (CLIN)*, 2003. <hal-01096733>
- [40] Bès G., Dahl V., “Balanced parentheses in NL texts : a useful cue in the syntax/semantics interface”, *Workshop on Prospects and Advances in the Syntax/Semantics Interface*, Nancy, 2003. <hal-01096737>
- [41] Bès G., Lamadon L., Trouilleux F., “Verbal chunk extraction in French using limited resources”, <http://www.arxiv.org/abs/cs.CL/0408060>, 2004. <hal-00002699>
- [42] Bès G., Solana Z., Beltrán C., « Conocimiento de la lengua y técnicas estadísticas en el análisis lingüístico », in Castel V. (comp.), *Desarrollo, implementación y utilización de modelos para el procesamiento automático de textos*, Editorial de la Facultad de Filosofía y Letras, UNCuyo, Mendoza, p. 13-22, 2005. <hal-01144065>
- [43] Solana Z., Bès G., « Extracción del sintagma verbal núcleo y resolución de ambigüedades en la asignación categorial », *Revista de Letras*, UNR / Facultad de Humanidades y Artes, p. 157-171, 2005. <hal-01145586>
- [44] Gabriel G. Bès, Zulema Solana, « Sintagma Verbal Núcleo Flexionado en Español », *Infosur*, n° 1, 2007. <hal-01145576>

9. Bibliographie

- Aït-Mokhtar S., L'analyse présyntaxique en une seule étape, PhD thesis, Université Blaise-Pascal, Clermont-Ferrand, février, 1998.
- Baschung K., Grammaires d'unification à traits et contrôle des infinitives en français, PhD thesis, Université Blaise-Pascal, Clermont-Ferrand, juin, 1990.
- Bocheński J. M., *Los métodos actuales del pensamiento*, Rialp, Madrid, 1958.
- Bunge M., *La investigación científica ; su estrategia y su filosofía*, Ariel, Barcelona, 1969.
- Gazdar G., Klein E. H., Pullum G. K., Sag I. A., *Generalized Phrase Structure Grammar*, Blackwell, Oxford, 1985.
- Gross M., *Méthodes en syntaxe*, Herman, Paris, 1975.
- Hjelmslev L., *Le langage*, Les Éditions de Minuit, Paris, 1966.
- Lecomte A. (ed.), *Word Order in Categorical Grammar / L'ordre des mots dans les grammaires catégorielles*, Adosa, Clermont-Ferrand, 1992.
- Lyons J., *Chomsky*, Fontana/Collins, 1970.
- Pustejovsky J., *The Generative Lexicon*, MIT Press, 1995.