

Atti del XXI Congresso Internazionale
di Linguistica e Filologia Romanza

Centro di studi filologici e linguistici siciliani
Università di Palermo 18-24 settembre 1995

a cura di Giovanni Ruffino

SEZIONE 4

Le strutture del parlato

SEZIONE 6

Storia linguistica e culturale del Mediterraneo

7
Max Niemeyer Verlag
Tübingen 1998



Peter Koch (Berlin)

Mon intervention sera centrée essentiellement sur deux points:

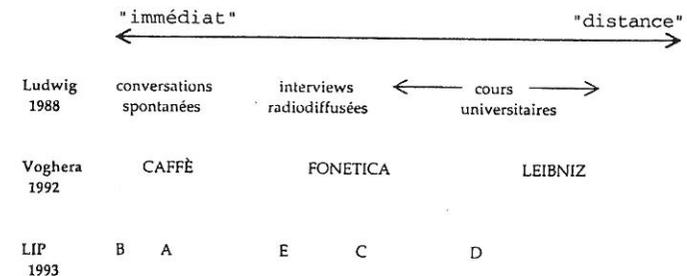
- 1) la typologie communicative de l'oral;
- 2) le travail basé sur les corpus authentiques de l'oral.

1. Comme l'a déjà démontré Wulf Oesterreicher (dans ce volume), l'oral tout court n'existe pas. Dans une perspective communicative, nous nous trouvons face à un **continuum conceptionnel**, qui va de l'**immédiat** communicatif à la **distance** communicative, c.-à-d. d'une conversation spontanée, affective, face à face, etc., à des formes de communication moins spontanée, moins affective, moins dialogique, plus détachée etc. Cela vaut, en principe, aussi bien pour la réalisation **phonique** que pour la réalisation **graphique** du langage¹.

Ceci dit, on peut donc s'attendre à une grande richesse de formes communicatives réalisées par voie phonique. Evidemment, les linguistes se sont intéressés d'abord aux formes les plus proches de l'immédiat, pour démontrer les différences par rapport au "langage écrit". Mais vu la gamme des formes communicatives réalisées par voie phonique, il y a lieu d'étudier, justement, cette variabilité conceptionnelle au niveau de la réalisation phonique elle-même.

Tout d'abord, c'est, bien entendu, un problème de **corpus**. Nous avons effectivement besoin de corpus authentiques qui reflètent la variabilité conceptionnelle de la soi-disant oralité. Pour le français, on peut signaler ici l'existence d'une excellente collection de corpus, élaborée par Ralph Ludwig et publiée en 1988. Elle contient des conversations spontanées, des interviews radiodiffusées et des cours universitaires, des cours magistraux etc. On voit tout de suite qu'il s'agit de formes communicatives très diverses qui s'échelonnent le long du continuum conceptionnel:

[1]



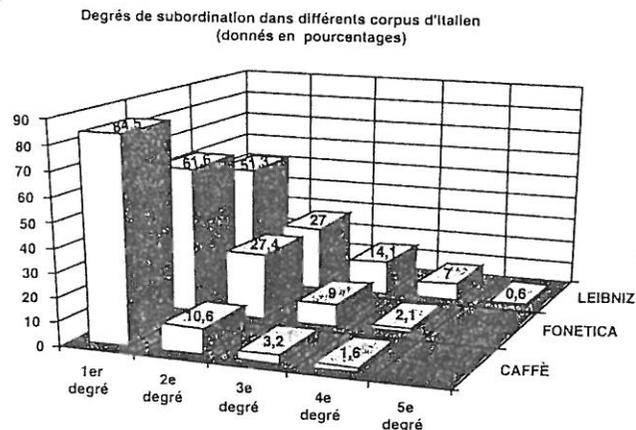
2. J'ai fait moi-même une étude qualitative, basée sur deux des corpus de Ludwig 1988, à savoir une conversation spontanée et un cours magistral (Koch 1995). Cette étude m'a permis d'examiner dans le détail les problèmes de la subordination. J'ai pu montrer que les deux corpus contiennent des subordinations même assez complexes. Or, le cours magistral exploite la complexité syntaxique d'une manière nettement plus intense, et il y a des stratégies particulières qui lui permettent d'atténuer les effets d'une complexité syntaxique trop grande².

Il s'impose de renvoyer, dans ce contexte, à Voghera (1992). Dans son étude quantitative, l'auteur avait à sa disposition cinq corpus de nature conceptionnelle assez divergente. Ses données statistiques concernant la subordination confirment très clairement le rapport qu'il y a entre distance communicative croissante et complexité syntaxique croissante. Pour simplifier les choses, je vous présente ici les chiffres qui se basent sur l'examen de trois de ces cinq corpus (quant à leurs positions relatives – tout à fait approximatives – sur le continuum conceptionnel, v. plus haut [1])³:

- CAFFÈ: conversation entre amis;
- FONETICA: séminaire de recyclage dans un contexte assez familier;
- LEIBNIZ: communication conclusive d'un congrès de philosophie.

Voici les résultats concernant les pourcentages des subordonnées à différents degrés de subordination, relevés dans ces trois corpus⁴:

[2]



On constate que le corpus CAFFÈ contient un pourcentage énorme de phrases à subordonnées du premier degré. Quand on passe aux subordonnées du second degré, la courbe descend brusquement. Cela va, malgré tout, jusqu'au quatrième degré – fait assez remarquable –, mais il n'y a plus de subordonnées du cinquième degré. Le pourcentage des subordonnées plus complexes monte, dès que l'on passe au corpus FONETICA, voire à LEIBNIZ. Ce dernier corpus contient même un – faible – pourcentage de subordonnées du cinquième degré.

3. Les corpus de Voghera sont des corpus "privés", c.-à-d. inaccessibles au public. Mais nous disposons aujourd'hui d'un matériel italien accessible à tout le monde, à savoir les corpus qui forment la base du *Lessico di frequenza dell'italiano parlato* (LIP), édité par Tullio de Mauro et al. (1993). Les transcriptions de ces textes authentiques se trouvent sur disquettes, et on les achète avec le dictionnaire. La typologie communicative de ces textes est extrêmement intéressante pour nous⁵:

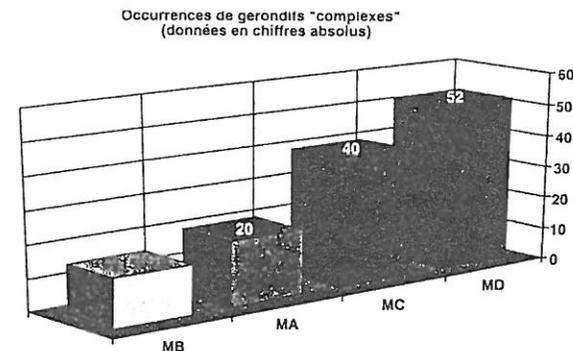
- A: échanges communicatifs bidirectionnels, face à face, avec prise de parole libre;
- B: communications téléphoniques;
- C: échanges communicatifs bidirectionnels, face à face, avec prise de parole non-libre (par ex. des discussions, des leçons etc.);
- D: échanges communicatifs unidirectionnels en présence des auditeurs;
- E: émissions radiophoniques ou télévisées.

Notons que l'échelonnement conceptionnel des corpus ne correspond pas tout à fait à l'ordre alphabétique des sigles (v. plus haut [1]). Dans le cas de E, on s'attend évidemment à un type de communication "mixte" du point de vue conceptionnel. Dans le cas de B, le caractère "plus oral" par rapport à A n'était pas prévisible (même pour les auteurs du LIP); il n'est d'ailleurs certainement pas dû au seul caractère de la forme de communication B en tant que telle, mais en partie aussi au choix des corpus du groupe A⁶.

Quoi qu'il en soit, cette typologie de corpus se prête parfaitement au genre d'analyses qui nous occupe ici.

4. Dans Koch 1994, j'ai étudié, par exemple, sur la base d'une partie de ces corpus, les occurrences des gérondifs "complexes", c.-à-d. des gérondifs accompagnés d'un complément d'objet, d'un complément adverbial etc. (par ex. *METTENDO INSIEME LE TRE COSE io poi spero che [...]*). Il s'agit là d'une construction relativement complexe que l'on évite plus ou moins dans l'immédiat communicatif; et nous voyons effectivement que le nombre d'occurrences de ces gérondifs augmente dans la mesure où l'on s'éloigne de l'immédiat communicatif⁷:

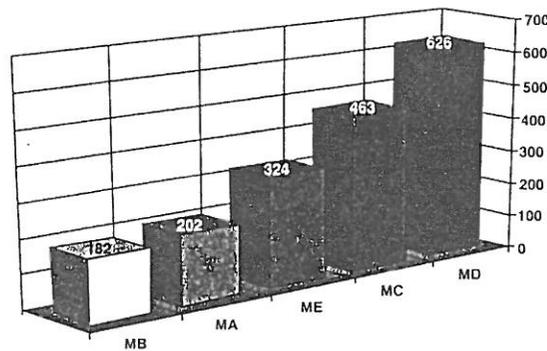
[3]



5. Dans sa thèse de maîtrise (soutenue à l'Université Libre de Berlin 1995), Simona Bellini a étudié, entre autres, la fréquence des propositions relatives en se basant justement sur les corpus milanais du LIP. Les résultats confirment que les propositions relatives constituent un type de subordonnées qui comporte, à lui seul, une complexité considérable. On ne s'étonnera donc pas de constater que le caractère conceptionnel d'un corpus et la fréquence des propositions relatives sont en corrélation directe. Plus un corpus est "scriptural" du point de vue conceptionnel, plus il contient de propositions relatives (quant aux positions des corpus MB et ME sur le continuum conceptionnel, v. plus haut [1] et 3.):

[4]

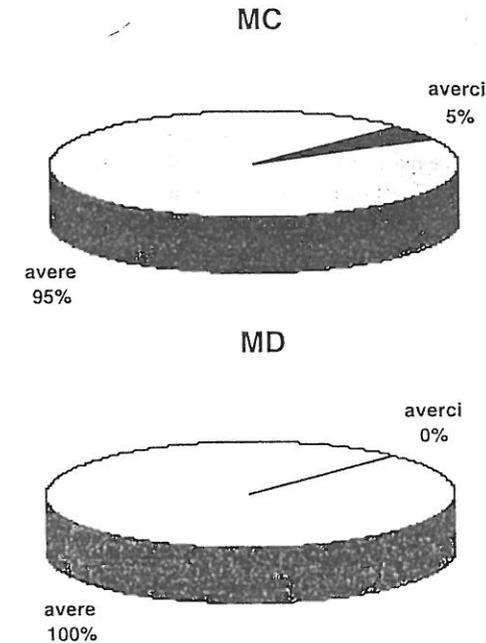
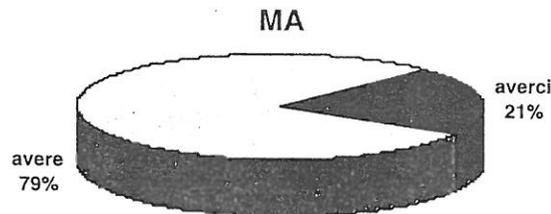
Fréquences des propositions relatives en Italien (données en chiffres absolus)



6. La subordination dont je viens de parler n'est pas un problème propre à une langue particulière, mais un problème universel de l'oralité (conceptionnelle)⁸. Evidemment, on peut aussi étudier sur la base des corpus LIP des traits spécifiques qui ne caractérisent que les variétés parlées d'une langue donnée, en l'occurrence l'italien.

Un problème propre à l'italien est l'"étoffement" du verbe *avere* sous forme de *averci*, typique de l'italien parlé (par ex. *CE L'HO a casa*). J'ai étudié moi-même trois corpus milanais du LIP qui correspondent respectivement aux constellations A, C et D (v. plus haut [1] et 3.):

[5]

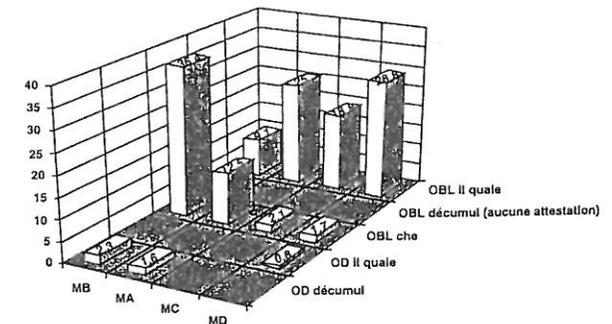


La corrélation qui existe entre le caractère conceptionnel "oral" d'un corpus et l'usage de *averci* est des plus évidentes.

7. Dans sa thèse mentionnée plus haut, Bellini a vérifié l'usage des relatifs italiens dits "non-standards" dans les corpus MA, MB, MC et MD du LIP. En voici les résultats⁹:

[6]

Occurences de relatifs "non-standards" et "scripturaux" en Italien (données en pourcentages)



Il s'avère nécessaire ici de distinguer non seulement les corpus à caractère conceptionnel différent, mais aussi les (groupes de) fonctions syntaxiques: objet direct d'une part (OD), fonctions obliques d'autre part (OBL)¹⁰.

Quant à OD, la construction non-standard que l'on pourrait appeler "décumul"¹¹ (type *il prete CHE L'hanno portato al manicomio*) est inexistante dans les corpus MC et MD – ce qui ne surprend pas – et très peu fréquente même dans les corpus MB et MA, ce qui est quand même remarquable.

Quant à OBL, il n'y a qu'une autre construction non-standard qui entre en jeu, à savoir le *che* invariable (type *questa porcheria CHE non s'è capito niente*). Cette construction est très fréquente dans MB¹² et le pourcentage descend en passant par MA vers MC et MD.

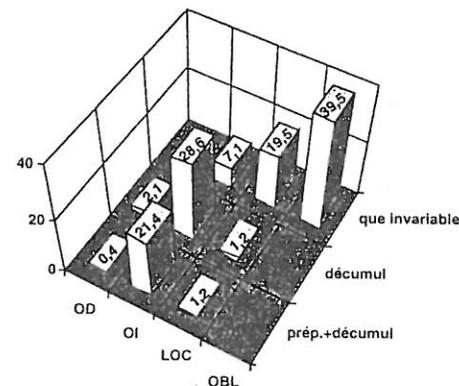
Quant à la construction tout à fait "scripturale" introduite par *il quale*, elle est extrêmement rare pour le OD, où elle n'apparaît – à un pourcentage minime – que dans le corpus MD, tandis qu'elle est assez fréquente pour les OBL, surtout dans les corpus plus "scripturaux" MC et MD, mais elle n'est pas non plus inexistante dans MB et surtout dans MA (qui, cette fois-ci, ne confirme pas une corrélation linéaire entre le caractère conceptionnel du corpus et le type de relatif).

8. Tous ces résultats nous démontrent l'importance de l'informatisation des corpus, qui nous permet de chercher certains éléments linguistiques par voie automatique et d'étudier une grande quantité de données sur la base de transcriptions authentiques.

L'espagnol dispose d'une impressionnante série de corpus authentiques du parlé élaborés dans le cadre de l' "estudio coordinado de la norma lingüística culta de las principales ciudades de Iberoamérica y de la Península Ibérica". Mais malheureusement, ces corpus n'existent pas sous forme informatisée. Voilà ce qui a amené Adela Westedt à informatiser, dans sa thèse de maîtrise (soutenue à l'Université Libre de Berlin 1995), l'un de ces corpus, à savoir *El habla de la ciudad de Madrid* (Esgueva/Cantarero 1981) pour pouvoir étudier les relatifs soi-disant "non-standards" en espagnol. Ce travail était centré beaucoup moins sur l'aspect conceptionnel que sur l'aspect des fonctions syntaxiques. Dans une perspective typologique, il était intéressant d'examiner les diverses stratégies adoptées pour exprimer les différentes fonctions syntaxiques:

[7]

Occurrences de relatifs "non-standards" en espagnol parlé selon les fonctions syntaxiques
(données en pourcentages)



Si l'on se base sur une hiérarchie du type $S > OD > OI > LOC > OBL$ ¹³, on constate que pour le OD les types non-standards sont extrêmement rares (décumul: 2,1%; préposition + décumul: 0,4%)¹⁴. Pour le OI, par contre, on trouve deux stratégies "non-standards" assez fréquentes: le décumul dans 28,6 % des cas (type *gente QUE LE gusta mucho*) et la préposition + décumul dans 21,4% des cas (type *teatro al QUE se LE ha quitado el polvo*). La stratégie du *que* invariable (type *cualquier cosa QUE les pudiese hablar*), que l'on trouve aussi pour le OI (7,1%), augmente de fréquence dans la mesure où l'on passe au LOC (19,5%), voire à l'OBL (39,5%).

Conclusion

Ce que j'ai voulu montrer, c'est tout d'abord que nous avons besoin de corpus authentiques qui reflètent toute une gamme de types conceptionnels existant au sein même de la réalisation phonique. Un autre aspect que j'ai voulu souligner est celui de l'informatisation des corpus qui nous permet d'étudier des quantités énormes de données (pourvu que les phénomènes linguistiques étudiés se prêtent à une recherche automatique). En ce sens les corpus du LIP sont – malgré certaines imperfections¹⁵ – extrêmement précieux puisqu'ils nous offrent différents types conceptionnels (A, B, etc.) et qu'on les trouve sur disquettes.

NOTES

- ¹ Cf. aussi Söll 1985, 17-25; Koch/Oesterreicher 1985, 17-24; 1990, 5-12.
² Que le cours magistral ne représente pas, en soi, une forme de communication complètement uniforme du point de vue conceptionnel, je l'ai démontré dans Koch 1996 (du moins pour ce qui est de la situation en Allemagne).
³ Cf. Voghera 1992, 80-82.
⁴ D'après op.cit., 214.
⁵ Cf. De Mauro et al. 1993, 39-41.
⁶ Cf. op. cit., 157; Koch 1994, 214 sq.
⁷ Cf. op. cit., 209.
⁸ Cf. Koch/Oesterreicher 1990, 96-100; Koch 1995, 19-21.
⁹ OD *che* n'entre pas en jeu ici, puisque c'est la forme standard normale.
¹⁰ Sous l'abréviation OBL je regroupe ici les fonctions suivantes, distinctes dans le travail de Bellini: OI (objet indirect), GEN ("génitif", catégorie qui n'inclut cependant pas les cas de possessif), LOC (locatif) et OB (toutes les autres fonctions obliques).
¹¹ Terme proposé par Frei (1929, 186-191) pour décrire des phénomènes analogues en français.
¹² Quant à la nature conceptionnelle orale particulièrement prononcée de ce corpus v. plus haut [1] et 3.
¹³ Cf. aussi l'*accessibility hierarchy*, proposée dans Keenan/Comrie 1977.
¹⁴ *Que* en tant que relatif du OD n'entre pas en jeu ici, puisque c'est la forme standard normale (sauf pour les OD animés, définis etc.).
¹⁵ Cf. Koch 1994, 214-216; Bellini 1995.

BIBLIOGRAPHIE

- BELLINI, S. (1995): *I pronomi relativi nell'italiano parlato: un'indagine basata sui corpora LP*, Thèse de Maîtrise, Université Libre de Berlin.
 DE MAURO, T. et al. (1993): *Lessico di frequenza dell'italiano parlato*, Milano.
 ESGUEVA, M./CANTARERO, M. (1981): *El habla de la ciudad de Madrid. Materiales para su estudio*, Madrid (La norma lingüística culta de la lengua española hablada en Madrid, 1).
 FREI, H. (1929): *La grammaire des fautes*, Genève/Leipzig.
 KEENAN, E.L./COMRIE, B. (1977): "Noun Phrase Accessibility and Universal Grammar", in: *Linguistic Inquiry* 8, 63-99.
 KOCH, P. (1994): "Prime esperienze con i corpora LP", in: T. De Mauro (ed.), *Come parlano gli italiani*, Scandicci, 201-216.
 KOCH, P. (1995): "Subordination, intégration syntaxique et «oralité»", in: *Etudes romanes* 34, 13-42.
 KOCH, P. (1996): "Die Vorlesung – zwischen Mündlichkeit und Schriftlichkeit", in: *Fachtagung der Deutschlehrer an Studienkollegs vom 16.-19. November 1994 in Bonn. Dokumentation*, Bochum (Studienkolleg Heft 42), 27-36.
 KOCH, P./OESTERREICHER, W. (1985): "Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte", in: *Romanistisches Jahrbuch* 36, 15-43.
 KOCH, P./OESTERREICHER, W. (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*, Tübingen (Romanistische Arbeitshefte, 31).

- LUDWIG, R. (1988): *Korpus: Texte des gesprochenen Französisch. Materialien I*, Tübingen (ScriptOra, 8).
 OESTERREICHER, W. (dans ce volume): "Pragmática del discurso oral."
 SÖLL, L. (1985): *Gesprochenes und geschriebenes Französisch*, Berlin (Grundlagen der Romanistik, 6).
 VOGHERA, M. (1992): *Sintassi e intonazione nell'italiano parlato*, Bologna.
 WESTEDT, A. (1995): *Los relativos en el habla culta de Madrid*, Thèse de Maîtrise, Université Libre de Berlin.