Le dialogue oral spontané : quels objets pour quels corpora

Spontaneous spoken dialogue: which objects for which corpora

Daniel LUZZATI

LIUM, Université du Maine, France
luzzati@lium.univ-lemans.fr

Résumé. Parce qu'on le pratique en permanence et qu'il est constitutif de notre expérience du monde, on croit intuitivement savoir ce qu'est le dialogue oral spontané. Mais en fait, on connaît mal la réalité de la parole conversationnelle, et en faire un corpus est une opération complexe qui revient à transformer un processus dynamique en objet statique, comme un biologiste qui s'intéresserait au vivant en ne disposant que de matériaux morts. Des causes similaires expliquent l'échec actuel du dialogue verbal homme-machine : il faudrait considérer qu'il s'agit moins d'un programme qui doit s'exécuter que d'expériences itératives qui se produisent. On peut espérer toutefois que les corpora à venir intègrent des données audio et video, qu'ils soient open source aussi bien pour les données, les outils et les formats, qu'ils s'intègrent dans des bases évolutives et instrumentées, et que la constitution de tels corpus fasse l'objet de projets de recherches spécifiques.
Mots-clés. Dialogue oral spontané, parole conversationnelle, corpus.

Abstract. Because we permanently practice it and because it is part of our experience of the world, we believe we know what spontaneous spoken dialogue is. But in fact we know little about conversational speech, and to build a corpus is a complex work, transforming a dynamic process into a static object, much like a biologist who would study life by working with dead materials. Similar reasons explain the current failure of verbal human-machine dialog: it should be seen, less as a running computer program, but rather as iterative experiences. It is hoped, however, that future corpora will integrate audio and video, that they will be open-source both for tools, formats and data, that they will merge into evolutive databases, and that building them will become the main focus of specific research projects.
Keywords. Spontaneous spoken dialogue, conversational speech, corpus.

1 Introduction

Le dialogue oral spontané est d'abord pour chacun d'entre nous un processus, une expérience, à savoir l'expérience même du langage, des rapports sociaux et des apprentissages. Pour certains d'entre nous il est en outre l'expérience d'un objet d'étude, un objet en tant que tel pour les linguistes et les interactionnistes, un prisme

pour d'autres, à travers lequel on tente d'observer la psychologie des intervenants, le mouvement des rapports sociaux, l'émergence des acquisitions... Pour d'autres enfin, il est l'impossible mirage de ce qu'on ne parviendra jamais à faire avec une machine, domaine laboratoire autant que domaine d'application. Tous ceux qui s'y sont confrontés savent combien l'objet "dialogue oral spontané" est insaisissable et fluctuant, une image courtisée dont on se limite à avoir une idée, au demeurant rarement corroborée par les bribes de corpus qui se collationnent à grand peine

Maintenant qu'est-ce qu'un corpus en la matière ? Pour (Damourette et Pichon, 1911-1927), il s'agissait de bribes de phrases, saisies à la volée et qui leur ont permis de faire un travail de morpho-syntaxe révolutionnaire. Pour (Bally, 1929) et (Frei, 1929), c'étaient des lettres écrites aux soldats de la grande guerre par des familles "qui écrivaient comme elles parlaient", bref, c'étaient des faux, combien précieux ! Pour les protagonistes du français fondamental (Gougenheim *et al.*, 1960), c'étaient des disques de cire immédiatement "transcrits" et aussi immédiatement réutilisés, à une époque où ce microscope qu'est l'enregistrement en était à ses balbutiements. En somme, on n'accédait pas réellement au dialogue, et on se limitait à des extraits de parole spontanée et/ou conversationnelle. Pour nous, aujourd'hui, grâce aux technologies, un corpus de dialogue oral spontané peut devenir un kaléidoscope de reflets multiples et séduisants.

Mais cela revient toujours à saisir et à représenter un signal borné dont on fait un objet, quelle que soit la variété des outils à notre disposition pour y accéder. Se pose en somme une question, que nous tenterons de nourrir : comment un objet statique, fût-il multiple sous forme de corpus (données texte, audio, vidéo...), peut-il prétendre représenter un processus dont la caractéristique fondamentale est d'être dynamique, de véhiculer une expérience qui va bien au delà des signes et des référents manipulés ?

2 Quels objets ?

La première chose à définir, ce sont les mots : "interaction", "dialogue", "conversation", "langue", "parole"… On est en effet dans un univers d'ambiguïtés où l'interopérabilité des concepts et du vocabulaire jette un trouble permanent. "Interaction" tout d'abord est un concept auquel il est préférable de laisser sa neutralité. Il suppose simplement l'existence d'échanges entre des interactants qui réagissent, qu'ils soient animés ou non animés, quelle que soit la nature, l'objet et l'intensité des échanges en question. "Dialogue" présuppose ensuite que la nature des échanges soit langagière, qu'elle soit limitée à un "échange" en langue normée (composé éventuellement de multiples "interventions", "initiave" puis "réactive", voire "évaluative" (Roulet *et al.*, 1985 ; Vernant, 1997)), ou qu'elle s'étende, émaillée de superpositions, d'interruptions, de bruits, de disfluences (Adda *et al.*, 2004)... Il est enfin préférable de limiter "conversation" à un dialogue langagier dont l'objet dépasse une manipulation ontologique du monde, et qui ne pourrait se satisfaire de simples séquences de questions-réponses.

Selon que le dialogue recourt à des phrases écrites, amendables et préméditées (Coursil, 2000), ou à des énoncés oraux spontanés, dans lesquels toute correction ne peut qu'induire un supplément de message, il ne s'agit plus du tout du même objet. L'écrit normé (généralement assimilé à la "langue") est fondamentalement constitué de signes discrets, avec signifiants identifiables et signifiés exprimables (même si leur compositionnalité est pour le moins sujette à caution), associés à des référents convenus. Ce type de langage, parfois qualifié de "langue naturelle", peut suffire

lorsque le dialogue se limite à des séquences de questions-réponses, que ce soit à l'oral ou à l'écrit. Lorsqu'il s'agit de conversation, on bascule sauf exception dans l'oral spontané, dans la "parole" ou plutôt, pour être plus précis, dans la "parole conversationnelle", que nous définissons comme l'ensemble des énoncés oraux conçus et perçus dans le fil de leur énonciation. Ce type de langage, souvent mal connu, est en lui même d'une ambiguïté telle qu'il n'est compréhensible que dans un face à face, avec l'aide des gestes, des mimiques, du contexte et de la voix, où les matériaux linguistiques se doivent d'être amplement prévisibles et/ou redondants.

Il suffit pour s'en convaincre de songer un instant à la réalité d'un énoncé simple et courant, bien entendu attesté, comme : *fin euh je par exemple euh je préfère faire du social que de faire euh de faire euh je sais pas...* Tout d'abord ce qu'on vient de lire, tout parcouru de "bruitages" syntaxiques et lexicaux, n'est qu'une image *a posteriori* de l'interprétation contextuée d'une suite de sons qui ressemble à ceci (par commodité pour les non initiés, on emploiera un pseudo alphabet phonétique) : [f in eu j e p a r è g z an pl eu ch pr é fè r fè r d u s o s y a l g d e f è r eu ch è p a] avec, entre autres, *que de* qui devient [g d e], *je sais* qui se mue en [ch è] (ce qui laisse à penser combien profond est le fossé qui sépare par exemple la morphologie verbale des grammaires de celle que nous pratiquons au quotidien). Et encore, nous nous sommes dispensés de transcrire la prosodie, dont les unités, relativement peu discrètes, donnent lieu jusqu'à aujourd'hui (Avanzi *et al.*, 2007) à des interprétations et des codages parfois aussi divers que les objets d'investigation de ceux qui les proposent.

On quitte de fait l'univers de la phrase, celui de la langue préparée, celui de l'erreur qui s'efface et se rature, pour basculer du côté des énoncés non prémédités, dont l'émetteur est le premier auditeur, dans lesquels l'erreur se traduit par un allongement du message. On a coutume de présenter l'oral spontané sous la forme d'un extrait de corpus transcrit orthographiquement, assorti d'un jeu de signes conventionnels, limités à l'objet de l'étude, au nom de la nécessaire possibilité d'en faire un objet de réflexion qui ne soit pas opacifié par une surabondance de signes cabalistiques :

> *le défaut qu'ils ont / ils ont une chambre pour eux / pour payer moins cher et // ils prennent un copain ou deux et alors voilà / mais les bains qui c'est qui les paye ils payent pour un bain ils payent pas pour trois*

Exemple 1. *Extrait du corpus* café
(http://www.loria.fr/projets/asila/corpus_en_ligne.html)

Mais l'oral s'entend et ne se lit pas. L'oral spontané est destiné à provoquer une réaction et non une transcription, et une fois transcrit de la sorte, il se trouve dépouillé de son "grain" si l'on veut, de sa dimension segmentale et supra segmentale si l'on préfère. On se trouve en somme dans la situation d'un biologiste qui s'intéresserait au vivant en ne disposant que de matériaux morts.

Trois faux semblants au moins doivent en l'occurrence être présents à l'esprit lorsqu'on manipule ainsi des transcriptions. Il faut tout d'abord avoir conscience qu'on ne travaille que sur une représentation du réel. Il faut ensuite avoir une claire vision de la nature, des biais, et de l'évolution des techniques de représentations : la distance entre les disques en cire du français fondamental et les outils de représentation du signal actuels est comparable à celle qui sépare le microscope à lentilles de Swammerdam de la microscopie à champ proche. Il faut enfin éviter de croire que, parce que c'est du langage, il ne s'agit que d'un supplément à une

description constituée. Il suffit pour s'en convaincre de prendre les grammaires existantes et de les lire à l'envers (des exemples aux descriptions) : ce qui est décrit ne tient pas compte de la "parole conversationnelle" et, au cas où on en tiendrait compte, il faudrait revoir bon nombre de descriptions. A propos de l'interrogation, sans sombrer nécessairement dans la simple dichotomie interrogation directe / indirecte, interrogation partielle / totale, on pose rarement le problème tel qu'il existe, par exemple en tentant de comprendre la différence de "valeur" entre des énoncés comme :

Que fais-tu
Qu'est-ce que tu fais
Tu fais quoi
C'est quoi que tu fais
Je voudrais savoir ce que tu fais
Je voudrais savoir ce que tu peux (bien) faire
Ce que tu fais je voudrais (bien) le savoir…

En l'occurrence, on peut montrer quelque chose qui ressemble à de la "parole conversationnelle", mais il est impossible de montrer de la même manière les processus mentaux qui président à son élaboration, alors même que les énoncés en portent constamment des traces. Le caractère non prémédité du langage transparaît tout en demeurant caché, et en induisant des présuppositions à partir d'indices récurrents :[1]

- dans *je peux vous demander le téléphérique il se prend où*, a-t-on une rupture de construction après *demander* ou bien 2 "fenêtres" (Luzzati, 2004) qui correspondent à 2 étapes cognitives successives : *je peux vous demander le téléphérique* et *le téléphérique il se prend où* ?
- dans *j'aimerais savoir où est ce que je pourrais trouver des affiches de cinéma*, a-t-on un mélange entre interrogation indirecte (*j'aimerais savoir*) et directe (*où est ce que*), ou bien 2 états cognitifs, avec un second qui efface le premier, dès lors que la formulation de la requête affleure à l'esprit du locuteur ?
- dans *je voudrais savoir s(i) où est le la Caisse d'Epargne de Grenoble s'il vous plaît* la séquence initialisée par *s(i)* correspond-elle à une ébauche d'un *s'il vous plaît* à venir, ou bien à un état cognitif transitoire dans lequel l'interrogation serait de forme totale : **je voudrais savoir si vous savez où est le la Caisse d'Epargne de Grenoble*

Pour accéder à ce type d'énoncés, il faut se mettre en état de les penser de manière dynamique, en intégrant les aléas de leur production. Il est à l'inverse indispensable d'éviter une approche statique, comme s'il s'agissait de "phrases" achevées, émaillées de ruptures de construction que l'on se dipenserait de qualifier d'"anacoluthes", dans la mesure où ce seraient de vulgaires fautes, et non un effet de l'art issu de la plume avertie d'écrivains émérites. Et des énoncés de ce type, que l'on entend peut-être tous les jours, c'est précisément ce qu'on trouve à chaque "ligne" d'un corpus de "parole conversationnelle".

En l'occurrence, la nature d'un corpus a certes son importance : entre une simple transcription textuelle et un accès à des données audio ou video, on change de registre. Rendre ces données accessibles et utilisables (avec alignement et codages conventionnels, par exemple à partir de la TEI, ou de systèmes comme ToBI

[1] Les exemples qui suivent sont tirés du corpus OTG (http://www-valoria.univ-ubs.fr/antoine/parole_publique/OTG/index.html).

(Martin, 2003)) devient une entreprise uniquement accessible à des laboratoires et/ou consortiums, seuls à même d'héberger des bases de données appelées à dépasser le million de mots (pour la France, on peut citer notamment : CFP/MODYCO, CRFP/DELIC, CLAPI/ICARE, CRDO/LCP, ASILA/LORIA, BREF et ESTER/ELDA...), bases de données dont l'accessibilité, la richesse, les codages et la vitalité sont pour le moins variables (Luzzati, 2007). Il faut dire que la masse d'information à manipuler s'accroît alors considérablement, à tel point que leur représentation n'est plus lisible sur la base d'une transcription linéaire qui, compte tenu de la taille des corpus visés, doit pouvoir être assistée d'une reconnaissance de la parole (Avanzi *et al.*, 2007). L'essor de ces technologies ne fait d'ailleurs que déplacer le problème. D'une part, on a de plus en plus tendance à se focaliser sur l'un des aspects seulement de la question ; d'autre part, on est enclin à considérer que le problème réside dans la maîtrise technologique des outils de codage et de représentation. Bref, en dépit (pour ne pas dire parfois du fait) de l'accroissement des informations, on a un fâcheux penchant, qui consiste à appréhender un processus dynamique par des procédés qui le figent.

3 Quels discours ?

Il est clair que le domaine du dialogue n'est pas un domaine vierge où les recherches ne se sont jamais aventurées. Ce serait plutôt le contraire, et il donne abondamment lieu à investigations, exploitations ou surenchères. Ce qui est surtout frappant c'est le cloisonnement des approches. Telle approche donne lieu à tels "résultats", à partir de telles données, mais comme celles-ci sont par essence non partageables, les résultats et les approches en question le demeurent pour une bonne part.

Le dialogue, dans sa relation avec la parole spontanée, a en effet dès l'origine été un lieu de conflit, un lieu idéologique, un outil maïeutique pour permettre des affirmations et leurs contraires, pour acculer des interlocuteurs à convenir qu'il pleut quand il fait soleil et qu'on dialogue lorsqu'on écrit. A titre d'exemple, voici deux extraits tirés, l'un de Bally (Bally, 1929) et l'autre de Queneau (Queneau, 1950). Dans le premier, Bally nous explique qu'il existe une "langue parlée", mais qu'il s'agit d'une langue "vulgaire", la langue critiquable du peuple, qui relève d'une *forme de pensée et d'activité au-dessous de la mentalité commune.*

> *Un fait de langage qui reflète un état social supérieur ou une forme d'activité ou de pensée plus haute que celle du commun, appartient à la langue dite écrite ; mais il faut s'entendre tout de suite sur ce terme. Une expression n'a nullement besoin d'être écrite pour porter la marque de cette forme générale ; elle conserve ce caractère et même le montre mieux encore quand elle est employée dans le parler… Si un fait de langage permet de constater l'absence de toute différence sociale entre les sujets parlants, ou bien si ce fait de langage révèle un état social inférieur, ou une forme de pensée et d'activité au-dessous de la mentalité commune, dans tous les cas, on a affaire à la langue dite parlée ou langue de la conversation. Ce mode d'expression comporte à son tour de nombreuses variétés, depuis le ton simplement familier, qui en est le caractère essentiel, jusqu'à la langue dite populaire, l'expression vulgaire et l'argot le plus grossier*

Exemple 2. *Extrait des études de styles de Bally (1929)*

Dans le second, Queneau nous explique également qu'il s'agit d'une langue "vulgaire", mais comme il entend apprécier le peuple, cette langue devient la vraie,

l'authentique, celle qui est libre, la langue d'avenir. Ce qui est drôle, c'est qu'il suffit de peu de chose pour que l'un et l'autre se contredisent : Bally, quelques pages plus loin, va s'appuyer sur le syllogisme socratique, ce qui n'est pas *a priori* le prototype d'une forme de pensée inférieure : *la langue parlée dirait : "les hommes sont mortels, n'est-ce pas ? bon, ! Mais Paul n'est-il pas un homme ? Oui ? eh bien ! Alors il doit être mortel* ; Queneau, dès qu'il écrit (une fois dépassées quelques références à la phonétique ou à la disparition des discordanciels) a recours à des procédés peu oraux, qu'il s'agisse du lexique ou de morphosyntaxe : *Doukipudonktan, se demanda Gabriel excédé* (*incipit* de *Zazie dans le métro*...).

Il existe actuellement deux langues, celle qui continue à être enseignée (plus ou moins mal) dans les écoles et à être défendue (plutôt mal que bien) par les organismes officiels comme l'Académie française… l'une qui est le français qui, vers le XVème siècle, a remplacé le francien (la traduction s'impose pour presque tous les textes avant Villon), l'autre, que l'on pourrait appeler le néo-français, qui n'existe pas encore et qui ne demande qu'à naître.

Exemple 3. *Extrait de Bâtons, chiffres et lettres (Queneau, 1950)*

Parallèlement à un objet d'étude non consensuel, le domaine est en lui-même difficile à aborder de façon objective. En fait il est surtout difficile à aborder de façon exclusivement ascendante, en partant systématiquement des données. En effet, le fait même de prendre un format de données conditionne à la fois la problématique et la nature des résultats potentiels. On ne rencontre donc guère d'approches de la question qui ne soient descendantes, afin notamment de penser la parole conversationnelle de façon dynamique. Dans les deux exemples suivants, il s'agit par deux fois d'un présupposé que l'on peut tenter d'instancier, avec un succès variable selon les données, à travers des corpora :

- La théorie des grilles (Blanche Benveniste, 1979), présuppose qu'alors qu'à l'écrit on ne dispose que de l'axe syntagmatique, l'oral dialogique spontané projette des rémanences de l'axe paradigmatique ou, autrement dit, des traces des opérations de choix sur les opérations de combinaison. Cela permet des représentations telles que celle de la figure 1
- La théorie du fenêtrage syntaxique (Luzzati, 2004) présuppose que l'espace de cohérence syntaxique de l'oral dialogique spontané est un espace qui évolue de façon non linéaire dans le courant de l'énonciation et que les traces que l'on récupère à partir d'une transcription se découpent en fenêtres de cohérence syntaxique qui peuvent donner lieu à "grammaire". Celle-ci permet en outre d'isoler et de caractériser la plupart des phénomènes de bruit que l'on taxe souvent aujourd'hui de "disfluences". Cela permet des représentations telles que celle de la figure 2.

le défaut qu'ils ont
ils ont une chambre pour eux
pour payer moins cher
et ils prennent un copain ou deux
et alors voilà
mais les bains qui c'est qui les paye ils payent pour un bain
ils payent pas pour trois ah

Figure 1. *Représentation de l'exemple 1 à partir de la théorie des grilles*

D'un côté on superpose tout ce qui paraît relever du bafouillage, ce dernier pouvant ainsi être classé en différentes catégories (SV/SN/lexèmes/connecteurs, bafouillage conjoint/disjoint, enchevêtré/linéaire...). De l'autre côté, les Fenêtres de Cohérence Syntaxique (FCS) deviennent des unités caractérisées à la fois par leur empan, leur plus ou moins grande linéarité, et l'importance des disfluences qui les séparent. On peut les articuler ("le défaut qu'ils ont" + "ils ont une chambre pour eux" peut apparaître comme un mécanisme visant à remplacer un apo koïnou avec mise en commun du segment central par un bafouillement conjoint) dans la mesure où elles sont fondées sur l'appréhension de la dynamicité d'un l'oral conçu et perçu dans le fil de son énonciation. On pourrait enfin tenter de les articuler avec les problématiques actuelles des "syllabes proéminentes" (Avanzi *et al.*, 2007) (on peut faire l'hypothèse par exemple que les FCS s'organisent en amont d'un noyau vocalique avec proéminence forte).

(a)	fenêtre normale	[-----]
(b)	fenêtre interrompue	[----<
(c)	fenêtre non initiée	>----]
(d)	fenêtres de bafouillage	[---[---[---
(e)	fenêtres de recherche lexicale	---]---]---]
(f)	fenêtres avec mise en commun ou **apo koïnou** :	
	(f1) mise en commun du segment central	$[^{1a}$---$[^{1b}$----$^{1a}]$---$^{1b}]$
	(f2) mise en commun du segment gauche	$[^{1}$---$[$----$^{1a}]$---$^{1b}]$
	(f3) mise en commun du segment droit	$[^{1a}$---$[^{1b}$ ----$]$---$^{1}]$

$[^{1}$le défaut qu'ils ont$^{1}]$ / $[^{2}$ils ont une chambre [pour eux$^{2a}]$ / pour payer moins cher$^{2b}]$ et // $[^{3}$ils prennent un copain ou deux$^{3}]$ et alors voilà / mais $[^{4}$les bains qui c'est qui les paye$^{4}]$ $[^{5}$ils payent pour un bain$^{5}]$ $[^{6}$ ils payent pas pour trois$^{6}]$ / ah

Figure 2. *Représentation de l'exemple 1 à partir de la théorie du fenêtrage syntaxique*

Si on parvient toutefois à produire des représentations différentes à partir d'un même énoncé, c'est bien parce qu'on reste dans un même type d'analyse (morphosyntaxique et monologale certes, mais également descendante) et qu'on peut se satisfaire dans les deux cas de données de transcription sommaires, sans autres marques que des pauses. Il n'en irait pas de même si on avait affaire à des interventions partiellement superposées, et si on souhaitait aborder également les phénomènes intonatifs, sociolinguistiques, psycho cognitifs, ethno méthodologiques, interactifs, multi modaux… Et le problème n'est pas de parvenir à traiter un niveau, puis un autre, et encore un autre, comme si une question se réduisait à la somme de ses parties. Le problème est de parvenir à intégrer les différents niveaux d'analyse, et de ne pas cesser de traiter de l'intonation ou de la morphosyntaxe lorsque l'on parle des actes de langages, alors que c'est presque systématiquement le cas aujourd'hui.

4 Quels dialogues ?

Il en va du dialogue, ou plutôt de la conversation, comme de la parole spontanée : parce qu'on y est surexposé, qu'elle est constitutive de notre expérience du monde, on croit la connaître et chacun se convainc volontiers d'en être expert. Qui plus est, la littérature (mais également, sous une forme légèrement différente, le cinéma et la télévision), et notamment le théâtre, nous livre toute prête une fallacieuse expérience de l'écriture de la conversation, qui relève par essence de

l'illusion. En effet, la conversation écrite n'existe pas et, peut-être, n'existera jamais, sinon sous forme de représentations convenues où la parole est rarement "conversationnelle" et où l'interaction est scrupuleusement et scripturalement alternée. Seule exception sans doute : les formes d'interactions langagières médiées par la technologie (chat, SMS...), avec leurs pratiques graphiques singulières et ludiques (Luzzati *et al.*, 2007). Mais, même dans ce type de cas, sauf à considérer que le concept "envoi" n'est pas en soi la négation du dialogue oral spontané, on bascule dans une forme bien particulière d'interaction, régie par des règles spécifiques.

Il n'y a donc guère à s'étonner que le Dialogue Humain-Machine (DHM) avec entrée clavier et sortie écran soit un fiasco (peut-être non définitif, dans la mesure où la complexité croissante de nouvelles interfaces incite à permettre des interrogations en "langage naturel" susceptibles d'évoluer vers des fonctionnalités "conversationnelles"). Alors même qu'on est capable d'en concevoir, les donneurs d'ordre que sont les entreprises gestionnaires de bases de données d'intérêt général (les sociétés de transport ou de téléphonie par exemple) ont fait leur choix depuis longtemps : souris, menus déroulants, claviers tactiles... sont infiniment plus efficaces et plus sûrs. Le vocabulaire de l'ergonomie est en lui-même révélateur de l'artificialité d'un dialogue écrit : c'est un mode d'interaction trop peu "intuitif" pour espérer s'imposer. Finalement, il n'existe guère que des illusions technologiques avérées pour servir d'exemples : ELIZA (Weizenbaum 1963) (exemple médiatisé de pseudo-dialogue par mots-clés, qui simule la méthode psychanalytique rogérienne, consistant à renvoyer au patient ses propres paroles...), les avatars androgynes, comme JABBERWACKY (Carpenter et Freeman, 2005) ou ALICE (Wallace, 2003), derniers vainqueurs du "Loebner price", incontestable sommet de l'illusion technologique !

Le Dialogue Oral Humain-Machine (DOHM) n'est guère mieux loti. On rencontre certes, et depuis quelques temps déjà, des systèmes qui sont capables de détecter plusieurs mots clés dans un même énoncé de parole continue, de suivre une structure de dialogue en fonction d'une planification préétablie, et dont la coopération est bornée par le remplissage d'un formulaire nécessaire à l'exécution de la tâche. Avec entrée-sortie vocale, on trouve cela aussi bien dans l'Exemple 4, tiré du projet May I help you (AT&T (Gorin *et al.*, 2002)) que dans l'Exemple 5 tiré du projet ARISE (LIMSI-CNRS (Lamel *et al.*, 1998)). Il s'agit certes d'interaction langagière, mais sous la forme question/réponse, strictement limitée à une manipulation ontogique d'un sous-monde.

Machine*: AT&T. How may I help you?*
User: I need a credit, I got the wrong number.
Machine*: What was the number that you dialed?*
User: I dialed four one five two eight…
Machine*: Was the car billed to the phone that you're calling from now?*
User: No, to a credit card.
Machine*: What was the card number?*
Use: Four five six seven…
Machine*: I've given you credit for that call. Thank you for using AT&T.*

Exemple 4. *Dialogue effectif tenu avec May I help you (http://www.research.att.com/~jwright/hmihy/samples.html)*

Machine *: Bienvenue sur le service d'information ARISE du LIMSI-CNRS. Quel renseignement désirez-vous ?*
Utilisateur : Est-ce qu'il y a un Paris Bordeaux aujourd'hui vers 15H ?
Machine *: Dans l'après-midi vous avez un train direct de Paris à Bordeaux à 15h25 arrivant à 18h59 le 3 mars. Souhaitez-vous un autre horaire ?*
Utilisateur : et le train suivant ?
Machine *: Le train suivant part à 16h et arrive à 19h3.*
Utilisateur : OK merci c'est bon au revoir.
Machine *: au revoir. La SNCF vous souhaite un bon voyage.*

Exemple 5. *Dialogue effectif tenu avec ARISE (http://www.vecsys.fr)*

Que faudrait-il pour qu'un système de DOHM bascule vers une parole et un dialogue "conversationnels" ? Autrement dit, que faudrait-il pour qu'on ait envie de dialoguer/converser par exemple avec une machine à café, plutôt que d'appuyer sur des boutons ? Il faudrait tout d'abord que les capacités de ladite machine ne se limitent pas à la parole : reconnaître un pas, un bruit, les voix, les visages, les émotions... la rendrait certainement plus "conviviale". Il faudrait ensuite que l'utilisateur y "gagne", par exemple grâce à des fonctionnalités inédites : dosage personnalisé des ingrédients, des modes de paiement, reconnaissance et prise en compte des individus-clients, services associés de type messagerie, information... Enfin, il faudrait peut-être que la machine en question soit dotée d'une forme de subjectivité, un peu comme un serveur dans un bar.

En termes d'IA, cela suppose des capacités inédites d'adaptation, d'acquisition, voire d'apprentissage, qui présupposent l'implémentation effective de règles excessivement coûteuses comme i) un même système ne peut jamais se trouver 2 fois dans le même état ou ii) un même système mis à 2 endroits différents doit à terme diverger (Steels, 2004a, 2004b). Du point de vue langagier, cela suppose entre autres un accès maîtrisé à la variabilité du langage (capacité à produire et à interpréter de façon discrète une part importante des paraphrases possibles), un modèle de langage réversible, entre analyse et génération, ou bien la capacité à produire une génération non préméditée, c'est-à-dire qui réinterprète ses propos au fur et à mesure qu'ils sont produits, en étant capable d'apprécier leur effet, de se reprendre, de se corriger... (Nicolle 2006).

Donner un exemple de dialogue de ce type suppose par définition non pas un dialogue isolé, mais une série en cours. Les performances supposées sont nécessairement dynamiques, en mouvement d'une interaction à l'autre, condition nécessaire pour qu'appuyer sur un bouton devienne moins intéressant qu'un dialogue susceptible d'évoluer vers la conversation. Bien entendu, il ne peut s'agir aujourd'hui que de pseudo dialogues, inventés pour l'occasion, qui présupposent des capacités dont on ne dispose pas encore, et dont le coût peut parfois paraître disproportionné par rapport aux avantages qu'on en retirerait.

Machine *: bonjour Brigitte, qu'est-ce que je vous sers ?*
Utilisateur : un café svp
Machine *: sucré ?*
Utilisateur : légèrement, oui
Machine *: "légèrement", ça veut dire beaucoup ou peu de sucre ?*
Utilisateur : très peu, et avec un nuage de lait
Machine *: un nuage ?*
Utilisateur : oui
Machine *: mais "un nuage", ça veut dire quoi... ça veut dire une pluie de sucre ?*
Utilisateur : non, ça veut dire que du lait, il faut en mettre un tout petit peu
Machine *: je ne connaissais pas... pardon je... je peux me permettre une question?*
Utilisateur : oui... bien sûr
Machine *: on peut dire aussi un "nuage" de sucre ou de café ?*
Utilisateur : non... on le dira seulement avec du lait
Machine *: j'essaie... voilà... vous me direz si ça vous convient*
Utilisateur : c'est parfait, merci
Machine *: de rien, au revoir... je prends 1€ sur votre compte*

Exemple 6a. *Pseudo dialogue N°1*

Machine *: bonjour, la même chose qu'hier ?*
Utilisateur : qu'est-ce que j'avais pris ?
Machine *: un café légèrement sucré avec un "nuage" de lait*
Utilisateur : c'est parfait, oui
Machine *: je prends sur votre compte, comme d'habitude... autre chose ?*
Utilisateur : oui, est-ce que Paul est là ?
Machine *: je ne l'ai pas vu ce matin : vous voulez lui laisser un message ?*
Utilisateur : oui : dites-lui de passer me voir...
Machine *: bureau L 100, c'est ça ?*
Utilisateur : *oui... j'ai pas changé*
Machine *: entendu, je ferai ce que je pourrai... je lui dis que c'est urgent ?*
Utilisateur : pas la peine, non... merci, au revoir

Exemple 6b. *Pseudo dialogue N°2*

Machine *: la même chose Brigitte ?*
Utilisateur : oui...c'est ça... avec un "nuage"
Machine *: vous aviez l'air un peu stressée hier*
Utilisateur : ah oui, c'est vrai... au fait, Paul est passé ?
Machine *: voilà votre café... non, je ne l'ai pas vu, désolé*
Utilisateur : pas grave, d'ailleurs tu peux annuler le message
Machine *: je ne lui dis donc rien de votre part ?*
Utilisateur : non, non, c'est plus la peine
Machine *: OK, à bientôt*

Exemple 6c. *Pseudo dialogue N°3*

5 Quels corpora ?

Qu'un système de dialogue ait besoin, pour rendre ses fonctionnalités perceptibles, d'exemples itératifs est en soi révélateur (et si ce n'est pas le cas, c'est

que le système en question ne permet de produire qu'une forme de dialogue fortement réductrice). Le plus important ne réside pas dans les dialogues en eux-mêmes, mais entre ceux-ci, de l'un à l'autre, dans les transformations du "monde" qu'ils induisent. Un dialogue langagier, qui s'oriente vers la conversation, vise à devenir une expérience qui se produit et non la simple exécution d'une séquence d'instructions (Luzzati, 2006). Un corpus linéaire et alterné peut en être un effet ponctuel, mais ses résultats vont bien au delà, tout comme ses conséquences cognitives.

Il en va précisément de même avec les corpus de dialogue oral spontané. Un corpus en l'occurrence ne peut être que l'effet visible et statique d'opérations dynamiques tout aussi difficiles à percevoir qu'à visualiser. Un dialogue oral spontané est clairement une expérience qui se produit et non un corpus qui se collecte. Il induit une transformation du "monde" et une modification d'états mentaux (au moins des interactants) qu'aucun corpus ne pourra jamais reproduire, sauf à mettre en place des protocoles qui interdiraient précisément auxdits dialogues de demeurer "spontanés".

Et pourtant des corpus existent, en assez grand nombre (Cf. l'enquête réalisée pour la DGLFLF (Délégation Générale à la Langue Française et aux Langues de France) disponible sur le WEB (Cappeau et Seijido, 2005)), dont la géographie est pour le moins contrastée, tout d'abord en termes d'accessibilité. A côté de données en libre accès (base ASILA ou corpus OTG), on trouve des données à accès conditionné par une convention de prêt/échange (base CLAPI), ou des données uniquement compulsables par recherche d'occurrences (bases VALIBEL, ELICOP, CRFP ou MPI). Ces données sont ensuite extrêmement variables, par leur nature (texte nu ou enrichi à x tires (c'est-à-dire à x niveaux de transcription et/ou de commentaires visualisables à l'écran), audio avec y micros, vidéo avec z caméras...), par la richesse des codages (balisages, XMLisation, formats type TEI...) ou par la convivialité de l'interface. Aujourd'hui, l'essentiel des corpus est en effet concentré dans des bases, entretenues par des laboratoires ou des consortium, et on y accède par l'intermédiaire d'interfaces dont la convivialité dépend entre autres du nombre et de la complexité des fonctionnalités en question. Mais ce qui les différencie surtout, c'est leur vitalité : à côté de bases qui ont peu ou pas évolué depuis 2/3 ans (DELIC, VALIBEL, ELICOP, ASILA semble-t-il), certaines sont en évolution constante ou rapide (CRDO, CLAPI), et d'autres, soutenues par l'ANR, sont en émergence (VARILING, EPAC, RHAPSODIE) et devraient rapidement influer sur le paysage par des innovations spécifiques (Bazillon, 2007) (sociolinguistiques, interactionnistes, prosodiques...).

Les questions initiales demeurent néanmoins : que deviendront à l'avenir les corpora de dialogue oral spontané ? comment devraient-ils se présenter pour ne pas demeurer un matériau mort ? pouvons-nous envisager des corpora qui cessent d'être des objets statiques, constituant une image dégradée des processus dynamiques qu'ils prétendent représenter ?

Une première réponse réside dans la nature des corpus. Il est clair que de simples transcriptions textuelles sont aujourd'hui dépassées, et qu'elles sont, sinon caduques, du moins appelées à servir avant tout de témoignage. La disponibilité des données primaires, audio et/ou video, mono et/ou multisource, préserve au moins la possibilité de reprendre à volonté les transcriptions qui en découlent. La richesse des corpus en question est également en cause : signal numérisé, prosodie, identification fine des tours de parole (y compris les superpositions de locuteurs), phonétisation, annotations, insertion de balises, codages (des mimiques faciales et

gestuelle par exemple). De tels corpus sont néanmoins particulièrement lourds à mettre en oeuvre, tant juridiquement que matériellement, et la radio et/ou la télévision, avec tous les biais qu'ils véhiculent, resteront dans la pratique les sources les plus sollicitées, notamment pour des applications qui réclament d'importantes bases de données. Ce type de données, quantitativement le plus important (le corpus ESTER représente à lui tout seul bien davantage que la moitié des corpus audio + transcription existants), véhicule toutefois un biais bien particulier : les médias n'offrent qu'une image du réel, et travailler sur cette image fait courir le risque de prendre les ombres sur les parois de la caverne pour les personnes qu'elles reflètent.

Une deuxième réponse réside dans l'ouverture et la mutualisation des corpus. Travailler sur un corpus conduit presque inéluctablement à en produire un codage, une réprésentation, une interprétation. L'intégration ou l'association de ce travail au corpus source, et la perspective d'en intégrer de façon itérative, rendrait au moins le matériau ouvert, en permanence susceptible de s'enrichir et d'évoluer. La condition pour que ce type de processus se mette en place réside en grande partie dans une transparence des outils et des formats, d'emblée conçus en open source. Actuellement, les sorties XML de TRANSCRIBER (outil de transcription), de PRAAT (outil de codage, notamment prosodique) et en format TEI (consortium de normalisation) sont encore différentes, mais on peut espérer qu'elles convergent bientôt, avec balisages et DTD harmonisés. Un exemple parmi d'autres : il existe actuellement avec WINPITCHPRO un logiciel de traitement du signal de parole, de transcription, de codage et de visualisation particulièrement riche et prometteur (99 tires, interface superbe, sortie XML...) mais, à la différence de PRAAT, il demeure confidentiel car il n'est pas en open source et reste dépendant de son concepteur (Martin, 2004).

Une troisième réponse réside dans la vitalité des bases qui hébergent les corpus de dialogue oral spontané. Il faudrait en effet que ces bases, et les laboratoires qui les constituent et les entretiennent, considèrent que le plus important n'est ni dans une utilisation privilégiée de ces données, ni dans leur taille, mais dans leur vitalité. Plus un corpus circule, plus il est utilisé, cité... plus il rayonne et s'enrichit d'études qui induisent des codages et des représentations. On devrait ainsi aboutir à des corpus "multicouches" qui, à partir de données primaires audio/video, possèdent plusieurs codages qui peuvent donner lieu à des transcriptions/modélisations diverses et superposables. Les corpus cesseraient ainsi d'être un objet figé. Même en devenant d'une certaine façon "dynamiques", ils ne rendraient néanmoins pas compte en eux-mêmes de la dynamicité de la parole conversationnelle. Mais là, il faut cesser de demander aux corpus ce qui relève des modèles et de la théorie.

Une quatrième et dernière réponse réside enfin dans des projets de recherches spécifiques, comme l'EPML 50, où il s'agissait de travailler autour d'un "corpus prototypique", en envisageant toute la chaîne : recueil (avec les conditions juridiques requises), représentations (signal audio-video, transcriptions), catalogage et codage, mise à disposition (consultation, travail, intégration du travail effectué)…, sur un corpus d'observation de petite taille (10 mn de dialogue environ), avec pour objectif d'envisager l'ensemble des problèmes qui se posent, voire d'en faire émerger de nouveaux, en profitant de larges collaborations pluridisciplinaires. Il en a entre autres résulté la création de CRDO (centres de ressource pour l'oral), dans deux laboratoires (INALCO et LCP), puis le lancement deux annnées d'affilée de projets ANR "corpus" ou "bases de données", avec des projets au moins partiellement consacrés à la productions de nouvelles données (EPAC, VARILING,

RHAPSODIE), projets qui seront nécessairement confrontés à cette même problématique.

6 Conclusion

Outre une "géographie" des corpus de dialogue oral spontané (historique des corpus, inventaire de l'existant, des bases hébergeantes, des projets en cours...), doublée d'une réflexion sur la valeur de ce type de données, sur la quasi absence actuelle de systèmes de dialogues langagier, sur leur intérêt en IHM..., on aura trouvé ici un plaidoyer. Il s'agit de défendre une approche "vitaliste" de la question. Ce qui compte, nous semble-t-il, c'est tout d'abord de rendre les corpus (qui sont des objets par nature statiques) capables de refléter un mode d'expression qui relève d'un processus par essence dynamique. Cela suppose des modèles, des formats, des outils... qui soient évolutifs, et qui sont actuellement en gestation. Il faut ensuite tendre vers des bases de données dont la qualité première soit la vitalité, ce qui suppose au moins 3 caractéristiques. Tout d'abord elles doivent permettre un accès direct aux données premières (audio, video), en même temps qu'un accès aux diverses représentations alignées qui ont pu en être faites. Elles doivent ensuite être évolutives, ce qui suppose non seulement des mises à jour, mais également un travail "multicouche", c'est-à-dire une intégration des représentations effectuées par les divers utilisateurs successifs. Elles doivent enfin être ouvertes et open source, et avoir entre autres pour finalité d'être exploitées (et donc enrichies) le plus largement possible.

7 Références

Adda-Decker, M., Habert, B., Barras, C., Adda, G., Boula de Mareuil, P., Paroubek, P. (2004). Une étude des disfluences pour la transcription automatique de la parole spontanée et l'amélioration des modèles de langage. *JEP* 2004, Fès, Avril.

Avanzi, M., Goldman, J.P., Lacheret-Dujour, A., Simon, A.C., Auchlin, A. (2007). Méthodologie et algorithmes pour la détection automatique des syllabes proéminentes dans les corpus de français parlé. *Cahiers of French Language Studies,* 13/2, à paraître.

Bally, C. (1929). *Traité de stylistique française.* Klincksieck, Paris.

Bazillon, T. (2007). Le codage de la parole spontanée pour la reconnaissance automatique de la parole. *RJCP 2007*, Paris, Juillet.

Blanche Benveniste, C. (1979).Des grilles pour le français parlé. *GARS*, Vol. 2.

Coursil, J. (2000). *La fonction muette du langage.* Ibis Rouge Éditions, Presses Universitaires Créoles, Guadeloupe.

Cappeau, P., Seijido, M. (2005). *Les corpus oraux en français.* DGLFLF, accessible à : http://www.culture.gouv.fr/culture/dglf/recherche/corpus_parole/Presentation_Inventaire.pdf.

Carpenter, R., Freeman, J. (2005). Computing Machinery and the Individual: the Personal Turing Test. Accessible à : http://www.jabberwacky.com/personaltt

Damourette, J., Pichon, E. (1911-1927). *Des mots à la pensée, essai de grammaire de la langue française.* Editions d'Artrey, Paris.

Frei, H. (1929). *La grammaire des fautes*. Slatkine, Genève.

Gorin, A.L., Abella, A., Tirso, A., Riccardi, G., Wright, J.H. (2002). Automated Natural Spoken Dialog. *IEEE Computer, 35* (4), 51-56.

Gougenheim, G., Michea, R., Rivenc, P., Sauvageot, A., (1960). *L'élaboration du français fondamental*. Paris, Didier.

Lamel, L., Rosset, S., Gauvain, J.L., Prouts, B. (1998). The limsi ARISE system. In *Proceedings of IEEE 4th Workshop on Interactive Technology for Telecommunications Applications*, Torino, Italy, 209-214.

Luzzati, D., (2004). Le fenêtrage syntaxique : une méthode d'analyse et d'évaluation de l'oral spontané". *Actes de MIDL,* Paris, 29-30 Novembre.

Luzzati, D. (2006). Dialogue et apprentissage. In G. Sabah (Ed.), *Compréhension automatique des langues et interaction*, Hermès, Paris, 337-357.

Luzzati, D. (2007). Corpus d'hier et d'aujourd'hui : progrès quantitatifs ou progrès qualitatifs ? *Revue électronique Interactions & Langages*, N°1, *Grands corpus de français parlé : bilan historique et perspectives,* à paraître.

Luzzati, D., Lehuen, J., Kitliska, S. (2007). Quelques pratiques langagières dans MEPA, un dispositif de simulationglobale en ligne pour la pratique du Français-Langue Etrangère. In G. Gerbeau (Ed.), *La langue du cyberespace: de la diversité aux normes*, L'Harmattan, Paris.

Martin, P. (2003). ToBI : l'illusion scientifique ? In Aubergé, V, Lacheret-Dujour, A. (Eds.), *Actes du colloque international Journées Prosodie 2001*, Université de Grenoble, 109-113.

Martin, P. (2004). WinPichPro – a tool for text to speech alignment and prosodic analysis. In *Speech Prosody 2004*, Nara, Japan, March 23-26, 545-548.

Nicolle, A. (2006). Compréhension et interaction. In G. Sabah (Ed.), *Compréhension automatique des langues et interaction*, Hermès, Paris, 141-171.

Queneau, G. (1950). *Bâtons, chiffres et lettres*. Gallimard, Paris (réédition Folio essais N° 247).

Roulet, E., Moeschler, J., Auchlin, A., Rubbatel, C., Schelling, M. (1985). *L'articulation du discours en français contemporain*. Peter Lang, Berne.

Steels, L. (2004a). The Evolution of Communication Systems by Adaptive Agents. In Alonso E., Kudenko D., Kazakov D. (Eds.), *Adaptive Agents and Multi-Agent Systems*, Lecture Notes in AI (vol. 2636), 125-140, Springer Verlag.

Steels, L. (2004b). The Autotelic Principle. In Fumiya I., Pfeifer R., Steels L., Kunyoshi K.(Eds.), *Embodied artificial intelligence*, Springer, 231-243.

Vernant, D. (1997). *Du discours à l'action*. PUF, Paris.

Wallace, R.S. (2003). The Elements of AIML Style, ALICE A. I. Foundation.

Weizenbaum, J. (1963). Symmetric list processor. Communications of the ACM, 6, 524-544.