C.N.R.S. I.N.a.L.F. NANCY

TRANSCRIRE, TRADUIRE, ORTHOGRAPHIER LE FRANÇAIS PARLÉ.

DE L'IMPOSSIBLE COPIE À LA FALSIFICATION DES DONNÉES ORALES

La transcription des données orales des langues standard semble présenter, depuis quelques années, un intérêt technique et théorique qu'elle avait rarement atteint auparavant. Les rencontres récentes avec des groupes de recherches français ou étrangers confirment cet engouement. Les linguistes qui travaillent sur des corpus oraux ont éprouvé le besoin d'harmoniser leurs techniques de transcription, de confronter les différentes solutions adoptées[1] en vue d'une informatisation des données à l'échelle européenne, voire internationale.

Nous savons que les transcriptions doivent se plier aux différents objectifs de recherche[2]. Ainsi des études sur les liaisons, la morpho-

1. Dans les débats pour tenter d'harmoniser la notation des pauses, un point de convergence apparaît. En effet, un consensus semble s'établir entre différents groupes de recherche pour ce qui est d'harmoniser et de réduire la notation conventionnelle des durées de pause.

Trois durées étaient initialement prévues :

pause de courte durée notée +

pause de moyenne durée notée – –

pause de longue durée notée – – –

Seules les durées extrêmes seraient maintenues. La pause moyenne, difficile à apprécier dans sa durée, serait volontiers abandonnée.

2. Une enquête menée par nos soins en septembre 1989 montre bien ces deux aspects de la transcription, à savoir : une extrême exigence de rigueur, que l'on soit lexicologue, phonéticien ou syntacticien, mais aussi une diversité dans les paramètres retenus, les appauvrissements opérés et les conventions choisies, et ce eu égard à chaque objectif. Une même portion de corpus oral proposé à différents linguistes montre bien ces écarts dans les conventions même si pour tous la transcription orthographique est retenue.

logie, les interactions entre locuteurs imposent chaque fois la prise en compte d'aspects différents. Les études du GARS sont orientées vers la recherche syntaxique; elles se font à partir de transcriptions orthographiques normées. Nous avons établi en 1977[3] les premières conventions qui répondaient à cette orientation. Depuis 1977, ces conventions ont été discutées, modifiées en partie[4].

Nous avons observé et critiqué, pendant plusieurs années, une quantité importante de transcriptions de l'oral, ce qui nous permet de mieux mesurer la portée des décisions techniques que nous devons prendre. Le statut de la transcription s'en trouve nécessairement changé.

Nous présentons ici une mise au point sur le choix de la transcription orthographique, ses limites et les problèmes particuliers que le transcripteur doit affronter. La transcription est toujours un appauvrissement des données orales et la question se pose de savoir dans quelles limites nous acceptons ce principe; nous mentionnerons ici le résultat des discussions sur les multitranscriptions, la prise en compte ou l'abandon de phénomènes comme l'intonation, les pauses, les clics, etc.

1. JUSTIFICATION ET LIMITES DE LA TRANSCRIPTION ORTHOGRAPHIQUE NORMÉE

1.1. LA TRANSCRIPTION ORTHOGRAPHIQUE NOTE LES MORPHÈMES

Notre expérience de transcripteurs nous a conduits à constater qu'en plus des problèmes de perception et d'édition des difficultés venaient d'une dimension rarement prise en compte, à savoir la confrontation de deux systèmes : celui de la langue et celui de l'écriture[5].

3. M. Raingeard et U. Lorscheider, 1977, "Édition d'un corpus de français parlé", *Recherches sur le français parlé*, I, pp. 14-29.

4. Voir : Cl. Blanche-Benveniste et C. Jeanjean 1987, *Le Français Parlé, transcription et édition*, Paris, INALF. M.-J. Savelli, Transcription et mise en grilles syntaxiques, mémoire de DEA, GARS, Aix-en-Provence. D. Giovannoni, 1988, Le carnet de bord du transcripteur, Mémoire de maîtrise, GARS, Aix-en-Provence . E. Castagnet, 1989, La holding, mémoire de maîtrise, GARS, Aix-en-Provence.

5. Cette réalité a été diversement notée par quelques linguistes dont Saussure (rééd. 1969, *C.L.G.*, Paris, Payot, p. 45) : "La langue est indépendante de l'écriture", et par Bloomfield (trad. 1961, *Langage*, Paris, Payot, p. 24) : "L'écriture n'est pas la langue mais un moyen simple d'enregistrer la langue au moyen de signes visibles".

Si, comme le fait le GARS, on adopte une notation orthographique, l'unité retenue est le morphème. On a pris le parti de ne pas aller au-delà du niveau morphémique. De ce fait, le niveau phonétique n'est jamais pris en compte dans le corps du texte. Ainsi, les *variations* de prononciations n'apparaissent jamais dans la transcription mais en notes de bas de pages comme pour les mots :

familier	prononcé	[famije]	au lieu de	[familje]
maintenant	"	[mɛ̃nã]	"	[mɛ̃tnã]
lorsque	"	[losk]	"	[lɔrsk]

Le découpage en morphèmes dépend non seulement de l'écoute mais aussi du sens reconstruit. Les bévues dans la transcription montrent bien cette saisie globale du lexème ou du syntagme. Quand on trouve par exemple dans une transcription,très attentive par ailleurs, un énoncé du type :

c'était le /dénuement, dénouement/ le plus absolu (Spilk, 3, 16)

on se rend compte que ce ne sont pas les unités phonétiques qui orientent le choix des unités. Il ne suffit pas de bien écouter pour bien transcrire et les découpages se font naturellement en conformité avec le sens local et global des unités discursives.

Les découpages ne doivent pas être absurdes mais envisagés à l'intérieur d'un faisceau de probabilités. Il ne s'agit en aucun cas, au moment du découpage en morphèmes, de perturber le sens de l'énoncé. J. Mariani[6], dans un travail sur "La synthèse de la parole", comptabilise 2 000 découpages différents pour la séquence :

/ʒemalopje/

Mais, compte tenu du contexte, des règles morpho-syntaxiques, deux seulement seront raisonnablement retenus :

1. j'ai mal au pied
2. j'ai mal aux pieds

Les découpages absurdes pour le sens sont rejetés. Par exemple si l'on prend la notation :

? j'hais malle lot pied

le sens global nous échappe, et dans :

6. J. Mariani, 1982, Un système de compréhension de la parole continue. Thèse de doctorat d'État, Paris-VI.

geai malle eau pied

le sens global n'est intelligible que dans un contexte qui accepterait cette liste.

Les découpages agrammaticaux sont bien évidemment rejetés :

* j'aima lot pied

De la même façon, tout corpus possède ses servitudes grammaticales, lexicales, culturelles qu'il nous faut respecter au mieux.

On peut proposer comme exemple limite pour illustrer ce dernier point l'énoncé suivant :

on trouve de moins en moins de /Mongols fiers, mongolfières/

Si l'on s'en tient à cette portion d'énoncé, les deux éléments proposés sont plausibles localement. C'est le contexte qui va nous éclairer et nous permettre d'orienter notre choix sur la première graphie :

on trouve de moins en moins de Mongols fiers d'être communistes
(TF1, 90)

D'autres problèmes peuvent surgir par ailleurs au moment du découpage syntaxique. Dans :

il m'a dit lundi je viendrai

comme la ponctuation n'est pas prise en compte, deux interprétations sont ici possibles entre :

il m'a dit lundi / je viendrai

et

il m'a dit / lundi je viendrai

Ce dernier cas mériterait d'être pris en compte et réglé peut-être par la notation des effets rythmiques.

Enfin, il faut rappeler que lors de la saisie globale en syntagme, seules quelques probabilités sont prédictibles :

on la prend / on l'apprend
c'est intolérable / c'était tolérable
il en porte / il emporte / il l'emporte

Ce phénomène n'est pas spécifique au français. En anglais, par exemple, on peut noter les couples bien connus :

a sister / assist her
in sight / incite

1.2. L'INFORMATION PERDUE

Un inconvénient cependant : nos morphèmes écrits ne donnent pas toujours une représentation réaliste de ce qui est dit. Ainsi, par exemple :

la différence entre *plus* positif et *plus* négatif
[plys] [ply]

n'est pas marquée diacritiquement.

Par ailleurs, une part importante des caractérisations socioculturelles peut disparaître.

Enfin, on est sans cesse amené à interpréter une forme parce que, par exemple, la différence entre conditionnel et futur :

chanterais et chanterai

entre possessif et démonstratif :

ses et ces

n'est pas perceptible à l'oral en français.

Prolonger notre travail de transcription au niveau de la morphologie et de la phonologie apporterait sans aucun doute un complément d'informations. Des recherches pourraient être entamées dans ce sens.

1.3. OÙ S'ARRÊTER DANS LES TRUCAGES ORTHOGRAPHIQUES ?

À côté de ces inconvénients négligeables puisque l'information pertinente au niveau syntaxique peut être éventuellement "récupérée" grâce à un jeu de notes, les avantages de la transcription en orthographe normée sont appréciables. En effet, la lisibilité de tels corpus est optimale car elle se rapproche au mieux de nos habitudes de lecture. Contrairement à la transcription A.P.I., elle est largement connue de tout individu alphabétisé. Par exemple, le médecin non phonéticien qui s'intéresse aux troubles du langage peut travailler plus facilement, semble-t-il, sur un corpus transcrit orthographiquement. D'autre part, le français parlé ressort moins dévalorisé par ce traitement que si l'on se livre à des mélanges de genres (mêler orthographe et A.P.I.) ou si l'on a recours à du bricolage orthographique. Ainsi donc, il apparaît plus pertinent de conserver au mieux l'image graphique du mot au lieu de se livrer à des excentricités dont on maîtrise mal les limites.

Le mot *peut-être* prononcé /ptɛtr/ et noté *p'têtre* pose plusieurs problèmes : d'abord celui du repérage — à coup sûr — de ce *e* latent, ensuite de celui du signifié. L'absence de signe diacritique ambiguïse le sens de l'énoncé. Si l'on s'amuse à pousser plus loin encore la caricature, il ne serait pas exclu de devoir trouver la représentation graphique d'une double altération du type *p'êt'*.

Comme ces phénomènes sont pour certains locuteurs extrêmement fréquents, la lisibilité de corpus dans lesquels toutes les altérations sont systématiquement notées serait amoindrie. On pourra apprécier cet argument au travers de l'exemple qui suit. En notation orthographique normée on pourra transcrire l'énoncé :

[ptɛtkʃverɛpysdptizɛt]

ainsi :

peut-être que je verrai plus de petits êtres (Col. 89)

ou bien, en tenant compte des altérations :

p'êt' que j'verrai plus de p'tits êt'.

Encore, une fois, il nous semble plus raisonnable de donner en notes ces faits de prononciations. C'est en bas de page que l'on fait état des effacements de sons ou des hyperliaisons, très fréquentes à l'oral, mais qui posent un problème de notation orthographique. Comment noter en effet des productions telles que :

entre quatre Z yeux
donne-moi Z en
il devra T avoir
les femmes ont eu T un livre (TF1, 90)

Les représentations graphiques de ces liaisons ne correspondent pas à des morphèmes répertoriés. On ne peut pas non plus les assimiler à des amorces de morphèmes. Nous préférons donc, pour ces mêmes raisons, la transcription de l'énoncé :

c'était des maisons tout à fait ordinaires (Masa, 37, 9)

et compléter en note de bas de page :

1. énoncé prononcé [tutafɛzɔrdinɛr]

1.4. QUELQUES PROBLÈMES D'INTERPRÉTATION

Que décider en l'absence de norme, comme c'est le cas dans certains domaines scientifiques, pour des énoncés du type :

des bactéries qui n'ont pas d'exigences nutritives (Microbio, 89)

Là où le linguiste transcripteur aurait tendance à noter une variation possible du nombre :

des bactéries qui n'ont pas d'exigence(s) nutritive(s),

le biologiste est catégorique, ne retient qu'une seule graphie et avance un argument scientifique pour justifier l'orthographe exclusive : une bactérie se nourrit obligatoirement de plusieurs substances nutritives. Dans un tel cas, une variante orthographique malenconreuse peut engendrer des conséquences importantes sur le contenu scientifique d'un énoncé en le rendant absurde.

Un autre exemple extrait du corpus *Microbio 89* nous interpelle. En effet la locutrice parle de

bactéries qui sont gram plus.

On est dans ce cas précis confronté à un problème de notation orthographique car plusieurs transcriptions entrent en concurrence : doit-on noter

plus	ou	+	signe mathématique ?
gram	ou	grams	avec l'accord ?
gram-plus	ou	gram plus	avec tiret ?
GRAM	ou	gram	en majuscules ou en minuscules ?

Difficile d'affirmer ici si, pour l'énoncé *gram plus*, il s'agit d'un morphème, d'une amorce ou d'une abréviation.

2. CONSÉQUENCES THÉORIQUES DE CERTAINS CHOIX DE TRANSCRIPTION

2.1. LES RÉDUCTIONS PHONIQUES

Certains choix de notation peuvent orienter ultérieurement l'analyse syntaxique et révèlent un phénomène d'écriture intéressant. Dans un

premier temps en effet peut surgir un éventail de formes graphiques possibles pour une même séquence orale mais on ne retiendra, dans un deuxième temps, qu'une seule forme.

Ainsi, la séquence *il y a* en français pourra être produite différemment par un locuteur. Nous posons la convention suivante :

La graphie *il y a* représente plusieurs possibilités de prononciation, en particulier les trois types de syllabation :

/ilija/
/ilja/
/ja/

Pourquoi préférer représenter la prononciation /ja/, équivalente à une syllabe, par *il y a* et non par *y a* ? Cette dernière représentation ne serait-elle pas plus réaliste et ne ferait-elle pas apparaître l'absence du pronom *il*, comme pour *faut* et *paraît* dans :

faut qu'il vienne
paraît qu'il a disparu.

Nous proposons un élément de réponse :

Dans la séquence phonique /ilja/, on a un /l/ entre voyelles et /j/ qui disparaît souvent en débit rapide, comme c'est le cas pour :

soulier	/sulje/	/suje/
milieu	/miljø/	/mijø/
million	/miljɔ̃/	/mijɔ̃/
collier	/kolje/	/koje/

et un deuxième temps où /i/ et /j/ se réduisent à /ja/ comme dans :

riions	/riijɔ̃/	/rjɔ̃/
chantilly	/ʃãtiji/	/ʃãtji/

La réduction phonique ne peut être assimilée sans autre argument à une absence de pronom sujet. Tout choix graphique doit être intégré dans la transcription qui permet :

1. de segmenter la chaîne sonore et de régler de ce fait les problèmes de découpages en morphèmes.
2. de sélectionner des morphèmes dans l'éventail des possibilités, et ce en conservant toutes les traces des variantes.

On peut saisir, au travers des exemples précités, un cas typique de problème posé par la transcription orthographique française où la fréquence des homonymies est remarquable, ce qui entraîne les cas de multitranscriptions que connaît tout transcripteur prudent.

2.2. LES MULTITRANSCRIPTIONS

Les multitranscriptions s'opèrent sur des zones bien précises appelées points fragiles[7]. Les variabilités sont prédictibles au niveau de certaines alternances auditives. Aussi, le linguiste averti est attentif quand il rencontre certaines structures. Ainsi, dans les variabilités prévisibles, on a retenu :

— Les noms propres :

Seliah, Sheila

— Les termes techniques :

P.H.B., P.H.V.

— Les confusions phonologiques :

– perception des consonnes dans une suite grammaticale :

ton / son
peut / veut
vont / font
le / de
les / des
ce sont / les, des / pauvres gens

– perception des consonnes liquides :

qui / qu'il
ce / qui, qu'il / a de pas mal c'est la compétition (Fayek)
des enfants pour / qui, qui il / fallait autre chose (Leggio, 3, 2)

– perception des semi-consonnes et des liquides :

y / lui
on / y, lui / a dit de se taire (Col. 90)

– perception des voyelles géminées :

ça / / ça
a été / était

Dans ce type d'exemples, plusieurs éléments entrent également en concurrence, sans qu'il soit possible d'en sélectionner un seul, même si l'on prend en compte le contexte immédiat ou périphérique. De la même manière, l'alternance :

parce que / puisque

7. D. Giovannoni, 1988, Le carnet de bord du transcripteur.

gagnerait à être étudiée plus particulièrement, compte tenu des problèmes de syntaxe (valence ou associé) et de sémantique (argumentation) qu'ils entraînent.

Ici, la perception est difficile et l'on est souvent amené à utiliser, par prudence, la multitranscription quand il est impossible de les discriminer. *Parce que* repose sur la qualité du vocalisme et en débit rapide dans *puisque* le *k* est particulièrement tonique et favorise les confusions.

En plus de ces problèmes dus à l'alternance auditive, on se trouve de nouveau confronté à celui de la restitution orthographique, ce qui peut être illustré avec les exemples suivants :

— qui concernent l'alternance temporelle :

– futur / conditionnel :

elle me dit moi je /retaperai, retaperais/ une maison mais même ceux qui me l'ont vendue je leur /permettrai, permettrais/ de venir autant que bon leur semble et leur /ferai, ferais/ plaisir et ça leur permettrait d'honorer le culte dans une maison (Masa, 55, 7)

– indicatif présent / subjonctif présent :

courir	je cours / je coure
voir	je vois / je voie

— qui concernent certains accords :

À l'oral en effet, les découpages de certains accords sont difficiles à percevoir et à transcrire, surtout si l'on a des liaisons. La difficulté vient des consonnes finales et entraîne des conséquences dans la mesure où l'évaluation des accords est perturbée, et par là même complique les recherches sur ce point.

À titre d'exemple nous donnerons :

– les accords de participes passés :

elle s'est /fait, faite/ opérer (Col. 89)
ils les ont /pris, prises/ et torturé(e)s (Ent. 26)

– les accords d'adjectifs :

boisson /spécial, spéciales/ apéro (FR3, 89)
bon il y en a qui en feraient des œuvres d'art /moderne, modernes/ (France-Info, interview, 89)

Chaque fois qu'on identifie, pour une séquence orale donnée, des morphèmes différents, on note la multitranscription.

En cas de phénomènes homonymiques, on se rend compte que, localement, pour une portion réduite d'énoncé, ça marche bien, le sens est intelligible, non altéré :

> alors avant toute chose dans un labo -/- euh pratiquement on est obligé d'isoler -/- pourquoi -/- si on veut étudier les bactéries présentes dans un produit les étudier ça voudra dire trouver des caractères précis -/- pour /eux, euh/ savoir que /c'est telle, c'était/ (l')espèce et donc trouver des caractères de bactéries -/- il faut avoir de très nombreuses bactéries -/- mais évidemment toutes identiques (Microbio, 89)

Avec ce seul exemple, on se rend compte combien un découpage univoque peut être arbitraire et peut en conséquence engendrer des répercussions sur une analyse syntaxique ultérieure. Ici l'incidence sur l'alternance qui concerne la morphologie verbale :

c'est / c'était

est évidente et incite à la prudence.

De la même manière pour la suite phonique :

/ setɛlɛspɛs /

On n'a pas retenu la transcription orthographique :

c'est telle l'espèce

peu probable dans le contexte du corpus *Microbio 89*.

On n'a pas davantage retenu l'éventualité d'une bribe dans l'optique d'une approximation lexicale du type :

/ ses, ces / telle(s) espèce(s)

Seules seront retenus en contexte les deux découpages suivants :

/ setɛlɛspɛs /
[1] [2]
c'était l'espèce

et

/ setɛlɛspɛs /
[1] [2] [3]
c'est telle espèce

et un troisième cas limite :

c'est elle l' espèce
[1] [2] [3] [4]

On peut citer ici Jespersen :

> Les mots sont des unités linguistiques mais ce ne sont pas des unités phonétiques; l'analyse exclusivement phonétique d'une corde sonore ne peut préciser ni le nombre de mots qui la composent ni les frontières qui séparent ces mots. (*Philosophie de la grammaire*, p. 315)

La multitranscription nous permet de jouer la carte de la prudence et de restituer le plus fidèlement possible les productions orales. Il faut dire cependant qu'elle alourdit sensiblement la lecture lorsque plus de deux possibilités se présentent. Ainsi, par exemple, pour la suite phonique :

/ lɔmkilamɛn /

on a la possibilité de multiples interprétations. Elles respectent toutes le découpage cohérent en morphèmes et la syntaxe :

l'homme qui l'amène
l'homme qui la mène
l'homme qu'il amène.

Localement, et si l'on s'en tient à cette portion d'énoncé, il faudrait retenir et noter ces trois variantes.

Ou encore dans cet exemple à cinq variantes :

mais peut-être parce / que, que eux, que euh, qu'euh euh, que euh /

Le contexte sémantique ou grammatical plus ou moins large va nous aider à sélectionner certaines formes et à en rejeter d'autres :

* voici le cheval et l'homme qui la mène

L'importance du contexte pourra encore être apprécié dans l'énoncé suivant :

c'était une personne qui avait un code et vie

Dans un contexte plus large on est amené à reconsidérer ces deux éléments pour leur préférer le mot CODEVI dans :

> c'était une personne qui avait un codevi[8] un plan d'épargne logement enfin qui avait pas mal de livrets avec au total euh: vingt millions à peu près (Laubreton, 9, 14)

Il serait, semble-t-il, intéressant d'établir une typologie des problèmes liés à la multitranscription en regroupant certains phénomènes comme :

8. Un *codevi* est en fait le sigle de "COmpte pour le DEVeloppement Industriel". Ici encore se pose le problème de la notation normative. Au lieu de noter CO.DEV.I., on adopte l'usage.

– la présence ou l'absence d'un élément :

/ je, je ne / sais pas
il y a / une, Ø / foule (Col., 89)

en particulier lorsque les points d'articulations sont proches comme dans :

il voit pas, il ne voit pas il / Ø, ne / voit pas
il ne le voit pas, il le voit pas il / ne, Ø / le voit pas

– la syntaxe et le lexique :

/ c'est tout, ces toux / ça m'arrive (Giova., 8, 9)
les pigments donc / eux, euh / colorent (Microbio, 89)
/ ça paye, sa paye / c'était lui (France-Inter, 90)
un traitement / afin, enfin / d'être soigné (Giova., 29, 14)
c'est de / cela, ceux-la / qu'elle se souvient (Col. 90)

on met par-dessus une lamelle ⊣ il est très fortement conseillé de bien faire tenir /la lamelle, la lame elle/ sur la lame ⊣ c'est ce qu'on appelle
L2 qu'est-ce que c'est la différence lamelle lame
L1 la lamelle est beaucoup plus petite ⊣ et elle est beaucoup plus fine (Microbio, 89)

– la morphologie verbale :

/ c'est, c'était / un homme extraordinaire (Col., 89)

à la fin il / la quittait, l'a quittée / que la pauvre elle a dû aller à l'asile (Giova., 23, 14)

/ je dis, j'ai dit / cela (Col., 89)
tu / as, vas / trouve(r) cela génial (Col., 89)

– le lexique seul :

est-ce / la Rome, l'arome/ d'aujourd'hui (FR3, 90)
c'est un / des astres, désastre / (FR3, 90)
il parlait de / la moralité, l'amoralité / (Col., 89)
dans le /filtre là, filtrat / (Microbio, 89)
il a / subi, suivi / un cours pendant des années (France-Info, 89)
c'était le / dénuement, dénouement / le plus absolu (Spilk, 3, 16)

Une remarque s'impose ici à propos des noms de nombre qui posent aussi des problèmes.

Dans l'énoncé :

un grossissement environ / de, deux / mille au Gram- (Microbio, 89)

doit-on retenir le nombre 1000 ou le nombre 2000 ? Il semble que, d'un point de vue scientifique, le contenu de l'information ne peut souffrir ce genre de variation. De même pour :

en / quatre vingt, quatre vingt hein, quatre vingt un /
80: 80 hein 81

/ soixante dix, soixante quinze /
70: 75

ou

/ soit cent dix, soit cent quinze / /
110: 115

/ soixante et dix huit, soixante dix huit /
60 et 18 78
en liste un seul nombre ?

Comment comprendre l'énoncé :

/ ãmilnœfsãswasãtdøspeleolɔgzesjãtifiksʃdesãdydãsɛtkrvas /

Que faut-il transcrire ?

en 1960 deux spéléologues et scientifiques sont descendus dans cette crevasse

en 1962 spéléologues et scientifiques sont descendus dans cette crevasse

Il n'est pas de transcription purement locale et on ne saurait jamais trop insister sur l'importance du contexte.

Pour le linguiste praticien averti de la transcription, le rêve du copiste qui saisirait la matière vivante telle qu'elle est, objectivement, et la restituerait *ipso facto*, est illusoire. "Il n'est pas de texte sans interprète, donc sans déformation"[9]. Le réel transcrit n'est jamais le réel émis mais un réel perçu et traduit. Au cours de ce processus, on assiste à une série de filtrages, d'appauvrissements, d'interprétations.

Ceci est inévitable : la transcription est toujours l'interprétation d'une forme mentale à une autre forme de représentation, et la traduction d'un système de communication à un système d'expression.

Par ailleurs, le français parlé ne saurait être, on le rappelle, appréhendé par le GARS comme une variante appauvrie du français écrit. La connaissance des faits de l'oralité nous amène à prendre en compte des créations lexicales et syntaxiques, des phénomènes dont la présence est souvent gommée à l'écrit en particulier :

9. O. Jespersen, trad. 1971, *Philosophie de la grammaire*, Paris, Éd. de Minuit, p. 315.

- les *euh* de remplissage
- les répétitions
- les corrections (approximations lexicales, hypercorrections ...)
- les bribes
- les hésitations
- les ruptures de constructions
- les allongements vocaliques
- les variantes non normatives
- les pauses.

Il s'agit d'une saisie positive de tous les faits de langue qui va nous permettre, au moment de l'analyse, de découvrir des microsystèmes étrangers à notre conscience métalinguistique. Cependant, l'écriture orthographique normée n'est pas toujours adaptée pour traduire toutes les réalisations de l'oral.

3. DISCUSSION SUR LA PRISE EN COMPTE DE CERTAINS PHÉNOMÈNES

3.1. PHÉNOMÈNES RETENUS

Comment noter les amorces, si spécifiques de l'oral, sans tomber dans la monstruosité orthographique ?

> euh: aussi peut-être T je peux te dire deux mots tout de suite sur les instruments (Microbio, 89)
>
> alors je vais: euh faire une petite: diver on va diverger là pour expliquer ça (Microbio, 89)

Il est gênant, du point de vue orthographique, de maintenir ces éléments sous ces formes atypiques isolées :

* t
* diver

On se refuse cependant à les éliminer des corpus. En les notant avec un système phonétique, on alourdit inutilement la transcription et on a du mal à les reconsidérer dans une analyse syntaxique ultérieure :

> elle a changé de [kul]

En les notant par le truchement du système orthographique, on se risque à créer des aberrations orthographiques comme dans l'énoncé :

? elle a changé de coul

Ainsi noté, ce dernier élément est difficilement acceptable au regard de la norme. Il est préférable alors d'aménager une convention qui réduise la difficulté, et le signe – (comme dans *coul–*) signale mieux, semble-t-il, qu'il s'agit d'un élément en attente. Cette convention a l'avantage de conserver la lisibilité du corpus, de noter ces faits de langue et d'être aisément traitées par l'informatique.

De même dans le corpus *Microbio 89*, la locutrice utilise successivement :

1. antibiogramme
2. antibiotique
3. antibio

Il est hasardeux ici de prévoir la troisième forme à partir des précédentes. On ne possède aucun argument pour dire s'il s'agit ici d'une abréviation ou d'une bribe, et on se trouve encore ici confronté au problème délicat de la notation conventionnelle.

Déjà quand on a affaire certainement à une abréviation, ce n'est pas simple. En effet, plusieurs notations seraient possibles pour l'élément *manip* dans :

tu veux que je te parle de quelques manip

Il est possible de noter :

manip. (selon la norme)
manips
manip

et pour nous :

manip-

Un autre cas difficile se pose pour certains mots techniques ou savants qu'on ignore. Ainsi dans :

c'est pas le bilan du tout et on dira un coccobacille (Microbio, 89)

Dans la mesure où le transcripteur ignore l'existence des coccobacilles, il peut interpréter un cas de bribe et noter :

un co- cobacille

La prudence s'impose donc et il faut se méfier des interprétations hâtives, définitives, surtout si l'on sait qu'à l'oral les retouches, les recherches

lexicales sont fréquentes et que le choix d'une graphie peut changer radicalement le sens d'un énoncé, comme dans :

> mais c'est plus haut que Volpajola — c'est environ à — trois kilo deux kilomètres et demi trois kilomètres de Volpajola — par la route

Ici, la graphie *kilos* introduirait l'idée qu'il s'agit d'une unité de poids. Même si les arguments pour interpréter *kilo-* comme l'amorce de *kilomètre* sont intuitifs, on retiendra cette interprétation, seule plausible dans le contexte.

Les bafouillages intralexémiques sont également difficiles à repérer et à représenter en orthographe. La notation *tech- nec- nec- nique* (Microbio, 89) semble lourde mais restitue l'oralité. La notation de cette production par la restitution normative *technique* respecte l'orthographe mais masque l'oralité. Ce problème, longuement débattu, n'a pas encore trouvé de solution.

Il faut cependant ajouter que nous devons tenir compte, bien entendu, des objectifs respectifs de recherche, et de la spécificité de chaque langue. Ainsi, on ne dispose pas en orthographe française de certaines abréviations de prépositions comme en portugais où, par exemple, *para* est conventionnellement noté *pra.*

3.2. PHÉNOMÈNES ÉCARTÉS

Au GARS, la notation de l'intonation est remise à plus tard par insuffisance de moyens scientifiques nous permettant de la faire figurer avec fiabilité dans nos corpus. Ne sont pas non plus pris en compte d'autres éléments supra-segmentaux, non verbaux, ceux qui ne passent pas nécessairement par le langage articulé tels que les mimiques, la gestuelle, les rires, les bruits, l'accentuation, le débit ...

On peut citer, à titre d'exemple, la notation du clic apical, très fréquent dans la communication orale des Français où il semble tenir la place d'un élément ou de plusieurs éléments syntaxiques : un clic peut remplacer le mot *non*, quelques clics peuvent, dans un certain contexte, signifier "du calme s.v.p.".

Les bruits, les silences qui font du sens sont des paramètres très utiles en analyse conversationnelle où agit l'interaction que l'on étudie, précisément, mais dans l'état actuel de la recherche, le syntacticien néglige la notation de certains chuchotements, rires ou rythmes discursifs alors que

pour les psychanalystes, les psychologues ou les psycholinguistes, ils sont nécessairement chargés de sens et intégrés à leurs analyses.

Nous sommes cependant conscients que l'absence de notation des éléments intonatifs si particuliers constitue une perte d'information considérable pour l'analyse syntaxique.

4. EN CONCLUSION : LE STATUT DE LA TRANSCRIPTION

L'activité de transcription peut présenter, selon les méthodes adoptées, des statuts très différents. Nous nous rangeons du côté de ceux pour qui la transcription est une activité cognitive complexe, un savoir-faire qui suppose des savoirs implicites, des pré-requis en matière de langue dont le choix conditionne l'établissement des données. Lors du passage de l'oral au système graphique, les "connaissances" du sujet transcripteur, pour l'essentiel très peu contrôlées, ne seront pas pour autant neutres. Ainsi, aborder la transcription en appliquant un découpage qui dépendrait directement de la représentation de l'écrit avec des réflexes et/ou des contraintes normatives aura des conséquences importantes d'autant que, en français, on doit gérer l'inadéquation partielle entre la graphie et la phonie.

Transcrire, c'est orthographier en français les séquences saisies globalement, c'est donc traduire. Pour transcrire, il ne suffit pas de maîtriser l'orthographe ordinaire. Il nous a été nécessaire parfois de recourir à des ouvrages spécialisés pour vérifier la graphie de certains mots techniques ou de recourir au locuteur quand c'était possible. On peut citer là, à titre d'exemples, certains lexèmes :

danse de Saint-Guy
quadrant
Institut Mérieux
une fois que c'est luté (Microbio, 89)

Le travail du transcripteur spécialisé dans les faits de l'oralité exige aussi des connaissances approfondies de l'orthographe. L'orthographe française, notant des morphèmes, implique une interprétation déjà très poussée. Ceci nous oblige à aborder des questions qui sont souvent posées aux psychologues ou à d'autres chercheurs qui travaillent sur corpus.

Mettre en mots, orthographier, c'est déjà toute une opération d'analyse. C'est préparer tous les morphèmes et cette opération est fondamentale.

Pris entre les deux objectifs de l'exactitude optimale (même si on en connaît les limites) et de la lisibilité du document, le transcripteur est sommé de faire un travail interdisciplinaire. Sa tâche, souvent minimisée par rapport aux autres activités d'analyse linguistique, nous paraît être de première importance; elle nous a souvent obligées à déplacer le champ de nos observations et à modifier notre problématique.

Dominique-Catherine GIOVANNONI
Marie-Josée SAVELLI