

DONNÉES ERRONÉES : QUELLES ERREURS COMMETTENT LES TRANSCRIPTEURS ?

Parmi les nombreuses difficultés que soulève la collecte de données orales (au travers de corpus longs), je voudrais ici faire état de distorsions assez récurrentes auxquelles se livrent les transcripateurs débutants. Le problème a déjà été évoqué dans Blanche-Benveniste et Jeanjean (1987), je tenterai d'en fournir ici une présentation plus détaillée. Je me suis servi d'une vingtaine de corpus de 15 minutes transcrits à l'université de Poitiers¹. J'ai relevé de façon systématique les séquences proposées par les étudiants et les versions finales qu'après écoutes j'ai retenues². Les écarts relevés reposent sur trois procédés : l'oubli, l'ajout ou la modification d'une séquence. A travers des exemples pris dans divers domaines, je montrerai que le hasard a peu à voir avec ces erreurs - qui peuvent être plus ou moins lourdes de conséquences pour les descriptions envisagées.

1. MODES DE PRODUCTION

Les erreurs de transcription peuvent tenir à des causes diverses. La méconnaissance des modes de production de l'oral conduit fréquemment le transcripateur à occulter certains phénomènes tels que les bribes et les

1 Deux personnes travaillent conjointement, ce qui permet d'obtenir « d'emblée » des transcriptions de bonne qualité, déjà soumises à la confrontation de plusieurs écoutes. La recherche des variantes entreprise pour cet article n'en est que plus pertinente. La mention « le transcripateur » désignera donc, le plus souvent, un couple d'individus.

2 Pour ce travail, seuls ont été retenus les exemples qui ne comportent pas de multi version.

amorces. Le contexte phonétique est souvent propice à une mauvaise perception des bribes surtout lorsqu'elles portent sur des éléments monosyllabiques. Les cas d'oubli sont assez répandus. Ainsi, le débutant transcrit une occurrence là où en fait deux items sont prononcés³ :

il : écrit -> il *il* écrit (St Gelais)
 donc sous l'occupation -> donc *sous le* sous l'occupation (Notre-Dame 2)
 la même personne de la société -> la même personne *de* de la société (Vayres)
 et euh : c'est le roi -> et euh *le* c'est le roi (Aliénor)

La difficulté ne tient pas seulement au nombre réduit de syllabes en cause. Des bribes de plus grande importance peuvent aussi être oubliées :

la chute du revêtement -> la chute *du revête* - du revêtement (Notre-Dame 1)

L'erreur peut venir d'une modification des bribes. Dans cet exemple, les éléments répétés sont mal recensés⁴ :

et les les tranches -> et *des* les tranches (Notre-Dame 1)

Le nombre d'éléments répétés est souvent mal identifié :

des ang- des anguilles -> des ang- des ang- des anguilles (Anguilles)

La transcription proposée initialement tend toujours à restituer un schéma plus habituel, plus proche des formats de l'écrit. On a ici une seule tendance qui va vers la « simplification » des données. Ainsi, on a relevé un seul cas où un phénomène de bribe avait été fautivement inséré dans la transcription :

margarine *de* de : bas de gamme -> margarine de : bas de gamme (Couscous)

³ Dans les exemples on indiquera toujours en premier la version proposée par le transcripteur débutant puis, après une flèche, la version finale retenue. Le titre du corpus est fourni entre parenthèses. Les séquences divergentes seront signalées en italiques.

⁴ Les déterminants sont la source d'erreurs multiples que l'on ne traitera pas ici :
on avait des grandes rigoles -> *on avait de grandes rigoles* (Notre-Dame 1)
devant le petit côté -> *devant un petit côté* (Notre-Dame 2)
la migration des papillons -> *la migration de papillons* (Futuroscope-3)
toucher les subventions -> *toucher des subventions* (Vayres)

Les enjeux pour l'analyse semblent ici bien plus importants.

Les divers écarts relevés dans cette partie présentent finalement une importance réduite. L'oubli d'une bribe ou d'une amorce a un retentissement assez mineur sur l'organisation morphosyntaxique de la langue. Les phénomènes qui relèvent de l'approximation lexicale sont assez souvent malmenés lors de la transcription. A plusieurs reprises ont été omises des séquences telles que élément démarcatif (*là, enfin*) introductif (*alors*) ou verbe recteur faible (*je trouve que*) :

c'est par exemple euh notre remplaçant -> c'est par exemple euh notre remplaçant *là* (École)
 un peu trop chauffée à l'avance -> un peu trop chauffée à l'avance *là* (Truite)
 pas mal de de bouillon de de liquide -> pas mal de de bouillon *enfin* de de liquide (Paëlla)
 ce mur ici -> *alors* ce mur ici (Notre-Dame 2)
 on prend on prend -> on prend *je trouve qu'*on prend (Paëlla)

Dans ce cas, on rencontre à nouveau des oublis mais aucun d'ajout. Les écarts entre la version primitive et la version finale ne se font donc pas de façon anarchique. Sur les trois procédés envisagés (oubli, ajout, changement), le premier est quasi exclusivement représenté. Cette singulière représentation laisse penser que les erreurs se plient à certaines règles et que leur nature dépend du secteur auquel elles se rattachent. L'observation d'autres types d'erreurs permettra amplement de le vérifier.

2. UNITÉS LEXICALES

Les erreurs recensées ne relèvent pas uniquement des phénomènes de production de l'oral. On rencontre aussi des fautes portant sur des éléments lexicaux. On réserve pour la partie suivante celles qui ont un retentissement syntaxique.

Les transcriptions peuvent présenter, en quelques endroits, un certain flottement lexical à cause duquel certaines séquences lexicales sont imparfaitement restituées. On trouve des exemples d'oublis :

et les portails étaient -> et les portails *latéraux* étaient (Nabu)
 il faut une crème -> il faut *faire* une crème (Tarte)

Les exemples avec changement des unités lexicales sont les plus fréquents. Des variations plus ou moins plausibles sont en quelque sorte « explorées » dans le contexte. Ces cas de « reconstruction » peuvent être favorisés par une proximité phonétique des séquences en cause :



vous *avez* Adam et Eve -> vous *voyez* Adam et Eve (Nabu)
 dans une c*é*rémonie *bien peu* faste -> dans une c*é*rémonie *pleine de* faste
 (St Gelais)
 les *fondations* ont disparu -> les *fortifications* ont disparu (Notre-Dame 2)
 plein de *choses* -> plein de *jeux* (École)

Les contextes avec des éléments symétriques peuvent faciliter ce relâchement de l'attention et provoquer une permutation des éléments :

quelques *années* après allait naître une^o seconde fille donc les deux époux semblent^o réconciliés mais euh quelques *temps* après + -> quelques *temps* [...] quelques *années* (Aliénor)

La permutation est d'ailleurs souvent source de transcription erronée :

nous irons *plus près tout à l'heure* -> nous irons *tout à l'heure de plus près* (Notre-Dame 2)

Des écarts plus importants se rencontrent aussi et les variations constatées ne manquent pas alors d'être, après coup, surprenantes. On est dans un phénomène de « mirage auditif », qui ne tient pas seulement à l'homophonie ou à l'homonymie :

de *faire* une visite XXX -> d'*assurer* une visite de mariage (Vayres)
 à quoi ça ressemble là le *bidule* -> à quoi ça ressemble là le *milieu* (Nabu)
 voilà oui chacun *mange* ce qu'il veut -> voilà oui chacun *rajoute* ce qu'il veut (Paëlla)
 vous incorporez + au *feu* -> vous incorporez au *fouet* (La Périgourdine)

Un certain nombre d'erreurs proviennent aussi d'une méconnaissance culturelle. Lorsqu'elle touche le locuteur, on court le risque d'une rectification involontaire par le transcripteur. Inversement, si elle tient au transcripteur, celui-ci construit généralement une séquence à partir d'éléments phoniques perçus. Seuls ces derniers cas nous intéresseront ici :

les très riches *œuvres* du Duc de Berry -> les très riches *heures* du Duc de Berry (Notre-Dame 2)
 enlever les *sucres* qui /Ø, y/ sont collés -> enlever les *sucs* qui /Ø, y/ sont collés (Estouffade)

Des problèmes orthographiques ou de découpage des unités peuvent s'ajouter. Ils traduisent le poids de l'écrit et permettent de mesurer le jeu complexe entre forme et sens auquel est soumis le transcripteur. Identifiant une forme ou des éléments sonores, il en livre une version écrite « connue » mais inadaptée, de fait, au contexte :

vous avez eu la chance de ne pas avoir *violé* le Duc -> vous avez eu la chance de ne pas avoir *Viollet-le-Duc* (Notre-Dame 2)
 le bordelais Bordeaux est *talonneur* entre le vin /des, de/ Graves et le Lillet
 -> le bordelais Bordeaux est = *à l'honneur* entre le vin /des, de/ Graves et le Lillet (Truite)

Ces erreurs en disent long sur les diverses ambiguïtés potentielles qui ne sont pas nécessairement levées dans les échanges de la vie courante (Fuchs : 1996). Les erreurs de transcription qui touchent le lexique relèvent pour l'essentiel d'une modification des éléments produits. L'oubli, fortement représenté lorsque les difficultés provenaient des modes de production, occupe ici une place marginale. Le travail du transcritteur opère ici au niveau du syntagme. A partir d'un moule phonique, il reconstitue une séquence qui puisse s'insérer sur le contexte. C'est pourquoi il propose des « équivalents » acceptables à partir de certains indices formels ou sémantiques.

3. PHÉNOMÈNES MORPHOSYNTAXIQUES

Dans cette partie seront évoquées quelques cas d'erreurs qui peuvent avoir un retentissement dans la description morphosyntaxique du français. Nous avons, pour l'essentiel, retenu des phénomènes dont la fréquence est peu élevée à l'oral ce qui rend plus grave toute erreur sur leur non repérage lors de la transcription. On comprend mieux ainsi l'importance non seulement du travail sur corpus mais tout l'aspect délicat et minutieux qui doit s'attacher à l'établissement des données. C'est souvent, comme on le verra par la suite, une meilleure connaissance des phénomènes, grâce à leur description, qui permet de renforcer la vigilance.

Les erreurs de transcription relevées sont souvent les conséquences visibles des représentations ou des préjugés concernant l'oral. Nous traiterons des points suivants : la négation, le sujet et le complément clitiques, les relations syntaxiques.

3.1. LA NÉGATION

La plupart des locuteurs français utilisent avec parcimonie la forme « ne » à l'oral. Mais le poids de la norme écrite reste important. C'est peut-être pourquoi dans les erreurs des transcritteurs on note plus d'ajouts que d'oublis. Un seul cas d'omission erronée a été relevé :

on trouve pas -> on *ne* trouve pas (Futuroscope -2)

Les autres cas relevés portent sur le rétablissement fautif de la négation complète, plus conforme aux habitudes de l'écrit :

on *ne* va pas discuter -> on va pas discuter (St Gelais)
 ça *n'*a pas l'air -> ça a pas l'air (St Pierre)
 il *n'*y a pas eu -> il y a pas eu (St Pierre)
 la porte *n'*est pas d'origine -> la porte /*était, est peut-être/* pas d'origine (Notre-Dame 2)

Un autre phénomène intervient avec le « ne » de négation qui peut être confondu avec un autre élément clitique :

on *ne* sait pas du tout -> on *le* sait pas du tout (St Pierre)
 vous mettez pas du sel -> *ne* mettez pas du sel (Truite)

Ces fluctuations mettent en évidence à la fois la fragilité des données et l'extrême prudence qui doit accompagner les travaux de comptage, lorsqu'ils portent sur des phénomènes peu représentés. La réalisation de la négation pouvait facilement être mal évalué, si la transcription initiale avait été conservée en l'état.

3.2. LA RÉALISATION DU PRONOM SUJET

Pour des phénomènes rares à l'oral, les conséquences de telles erreurs de transcription peuvent devenir véritablement déterminantes et fausser en grande partie la pertinence des analyses. De tels problèmes sont favorisés par les éléments en cause - souvent des clitiques - difficilement perçus dans certains contextes. Mais la méconnaissance de certains pans de la grammaire du français favorise ce « relâchement » du transcripteur débutant. L'écart est d'autant plus difficile à identifier qu'aucune description détaillée ne peut servir de garde-fou. Le sujet offre deux contextes pertinents pour cette démonstration.

3.2.1. Verbes impersonnels

Dans le cas de « falloir », la présence ou l'absence du clitique « il » est bien souvent sujette à variation. Les erreurs de transcription reflètent cet usage flottant :

la aussi faudra -> là aussi *il* faudra (Nabu)

Pour cet exemple, le contexte phonétique prête à confusion. Reste que si détecter la réalisation du clitique pose quelquefois problème, c'est essentiellement dans des circonstances où il peut ne pas apparaître. Les transpositeurs proposent rarement des formes verbales (autres qu'un impersonnel) sans sujet. Le cas ci-dessous, seul de son espèce, apparaît au milieu de constructions inachevées, ce qui a pu dérouter le transpositeur. La séquence proposée « tout le monde s'appelle madame Morel » si elle frise l'absurde (dans le corpus du moins...) constitue une suite grammaticalement correcte :

il y en a une tout le monde° s'appelle madame Morel -> il y en a une tout le monde° *elle* s'appelle madame Morel (École)

Les deux versions peuvent être disposées de façon à mieux faire ressortir leur différence syntaxique :

tout le monde°	s'appelle	madame Morel	
tout le monde°			<construction inachevée>
<i>elle</i>	s'appelle	madame Morel	

3.2.2. Constructions enchaînées

Si l'on prend le cas des constructions enchaînées (juxtaposées ou coordonnées), le sujet « commun » n'est pas forcément présent devant chaque verbe à l'écrit (Blanche-Benveniste : 1995, Cappeau : 1996) :

Il fouillait, perquisitionnait, saisissait, recoupait, auditionnait. (Le Monde. 22-06-96)

Cette conformation est peu répandue à l'oral. Sur l'ensemble des corpus examinés pour ce travail, seuls 5 exemples ont pu être trouvés. On voit donc combien il importe de disposer d'une transcription fiable, d'autant plus fragile qu'elle ne repose bien souvent que sur la perception d'une forme clitique parfois réduite au phonème /i/. Les habitudes acquises au travers de l'écrit viennent ici perturber l'attention des transpositeurs et leur faire négliger cette écoute fine :

ils vont aller à Paris et vont aller être consacrés -> ils vont aller à Paris et *ils* vont aller être consacrés (Aliénor)

A nouveau, les transpositeurs « trébuchent » de façon sélective sur des réalisations clitiqes. La forme « il » n'est pas perçue aux endroits où sa non apparition est envisageable. On sait que ces écarts sont généralement

sélectifs : les formes valorisées sont plus fréquemment attribuées aux locuteurs qui ont l'air d'avoir du prestige.

Le sujet se prêterait à d'autres observations en particulier sur la concurrence des formes *l'on / on* et *ça / ce*. On retrouverait alors certaines des observations formulées pour la négation : la forme valorisée donne lieu à plus de perceptions erronées et les transcripteurs en proposent généralement un nombre trop élevé.

Par son caractère quasi obligatoire en français, les variations sur le sujet se limitent finalement à sa forme et à son « oubli » dans quelques contextes particuliers. Les erreurs relevées sont donc une nouvelle fois guidées, influencées par la structure de la langue.

3.3. LA RÉALISATION DES COMPLÉMENTS

Si pour le sujet les « oublis » étaient étroitement canalisés, pour le complément une plus grande disparité de situations existe : il suffit que la solution soit conforme à la grammaire. On trouve ainsi des exemples d'oubli d'un complément de forme clitique (monosyllabique et non accentué). La séquence incomplètement transcrite ne présente toutefois pas d'anomalie ou de bizarrerie :

on vient à la nécessité de la lumière -> on *en* vient à la nécessité de la lumière (Futuroscope -3)
 autrefois on pêchait -> autrefois on *en* pêchait (Anguilles)
 au moins il ne pleut plus -> au moins il n'y pleut plus (St Gelais)
 et mettre tremper ensuite euh dans un mélange -> et *le* mettre tremper ensuite euh dans un mélange (Anguilles)

L'ajout d'un complément, potentiellement disponible pour le verbe en cause, est aussi représenté :

écoutez ce que je vais *lui* dire -> écoutez ce que je vais dire (Vayres)
j'apprends aussi des *trucs* -> *ils nous apprennent* aussi des *choses* (Vayres)
 comment on *les* pêche -> comment on pêche (Anguilles)

Pour le complément, les erreurs présentent une variété plus grande qui correspond aussi à des schémas de construction plus nombreux. La même « liberté » se rencontrera pour les relations syntaxiques.

3. 4. LES RELATIONS SYNTAXIQUES

Les exemples recueillis sont trop peu nombreux pour permettre un examen approfondi de diverses relations syntaxiques. Mais il m'a semblé intéressant de montrer que les indications soulignées à maintes reprises dans ce travail étaient pleinement corroborées. Quelle que soit l'erreur commise, la solution proposée par l'étudiant est « grammaticalement correcte ». Les ajouts et les oublis qui, en général, changent la relation syntaxique dépendent étroitement du contexte dans lequel ils apparaissent. Ainsi, le transcripteur peut traiter comme complément ce qui apparaît comme une construction isolée :

ça dépend *de* chacun -> ça dépend *c'est* chacun (Paëlla)

Il peut aussi modifier le type de lien entre deux constructions (rection faible vs coordination) :

j'allais vous dire *que* vous ajoutez les petites noix -> j'allais vous dire *et* vous ajoutez les petites noix (Truite)

ou entre deux séquences (épithète vs relative) :

des gens disponibles -> des gens *qui sont* disponibles (Futuroscope -1)

L'indication d'un lien de rection (même faible) au travers de la présence d'une conjonction peut être proposée :

je pense *que* maintenant il est temps -> je pense maintenant il est temps (L'enjeu)

C'est encore la forme de la conjonction qui peut être atteinte :

c'est-à-dire *bien* que on met à chaque personne -> c'est-à-dire que on met à chaque personne (Estouffade)
pour qu'elles roussissent un peu ouais *pour* qu'elles brunissent on rajoute
-> pour qu'elles roussissent un peu ouais qu'elles brunissent on rajoute (Fricassée)

C'est enfin l'élément introducteur qui peut varier (verbe vs conjonction) :

un peu de moutarde *il faut* quand même que la sauce soit un peu épaisse ->
un peu de moutarde *pour* quand même que la sauce soit un peu épaisse (Paëlla)
avec un fouet *parce qu'il faut* que ça fasse un ruban -> avec un fouet *jusqu'à ce que* ça fasse un ruban (La Périgourdine)

Dans chacun de ces cas, le transcripateur retrouve ses repères. Sa version plus ou moins éloignée du résultat final, peut provenir d'oublis, d'ajouts, de changements, elle n'est jamais éloignée des structures du français. La « monstruosité » n'a pas sa place dans les erreurs que commettent les étudiants.

4. CONCLUSION

Le relevé des corrections apportées aux travaux de transcripateurs novices a permis de vérifier à quel point le locuteur est impliqué dans le travail de transcription et de quelle façon il tend à reconstruire les données. Il occulte souvent ce qu'il ignore ou maîtrise imparfaitement (les modes de production de l'oral), substitue le lexique dont il dispose, et adapte la syntaxe à certaines de ses habitudes. Par là, ce travail de correction permet de découvrir certaines zones originales de l'oral (celle où la syntaxe diverge de l'écrit par sa fréquence du moins). Les données réelles attestées ne le deviennent véritablement qu'après de multiples vérifications qui nous affranchissent de ce que nous sommes toujours prêts à entendre.

Paul CAPPEAU

Références

- BLANCHE-BENVENISTE, Claire et JEANJEAN Colette, (1987), *Le français parlé. Transcription et édition*, Paris, Didier Érudition.
- BLANCHE-BENVENISTE Claire, (1995), « Répéter ou ne pas répéter », Mélanges offerts à David Gaatone, Tel-Aviv.
- CAPPEAU Paul (1996), « Quand le sujet n'apparaît pas. Petite étude sur quelques cas », Strasbourg, *Scolia*.
- FUCHS Catherine (1996), *Les ambiguïtés du français*, Paris, Ophrys.