

# Le processus de reconnaissance verbale à l'épreuve des amorces

Gabriel Bergounioux

Laboratoire Ligérien de Linguistique (LLL, UMR 72 70) – Université d'Orléans  
gabriel.bergounioux@univ-orleans.fr

**Résumé.** Cette étude se propose, à partir de l'exploitation d'un millier d'amorces dans le corpus des entretiens de l'Enquête Sociolinguistique à Orléans (ESLO2), de produire une typologie des amorces et d'en faire l'analyse. Après une présentation concernant la constitution des données, à partir des occurrences, sont examinées les différentes modalités de réalisation. Les amorces sont utilisées comme un observatoire de la conduite des auditeurs dans le cas d'informations lacunaires dans le signal. La réflexion est prolongée par un questionnement concernant le statut de la Forme Sous-Jacente (FSJ) lorsque l'unité, telle qu'elle est déclarée en morpho-phonologie, s'avère absente dans le discours, posant la question de son identification.

**Abstract. How can we know what a false start means?** Based on a French oral corpus – ESLO –, this paper discusses how we proceed to assign an interpretation when we hear a false start. It presents its special properties according to the spoken data on which this study is grounded. The discussion concludes, from the interpretation of the false starts, to the hearer's behaviour when he is confronted with that kind of riddles in the discourse.

## Introduction

L'une des opérations les plus délicates à appréhender dans la réception des messages concerne les modalités de compréhension par l'auditeur. Comment des énoncés prononcés dans des environnements bruités, avec de fortes variations sociales, inter- et intra-individuelles, parfois réalisés de façon incomplète ou fautive, peuvent-ils être néanmoins identifiés par le destinataire ? Cette question de l'*interprétation* – celle du locuteur, si le terme prend sa valeur musicale, et celle de l'auditeur en se rapprochant de l'acception psychanalytique – est le plus souvent effectuée à partir d'une approche d'inspiration sémantique. On se propose dans cette contribution d'en considérer la présence à partir de la traduction phonologique, quand le signal – les éléments acoustiques – en vient à être identifié mentalement à une (ou plusieurs) unité(s) lexicale(s), à un (ou plusieurs) signe(s).

Dans un premier temps, on présentera les données utilisées pour l'analyse, un ensemble d'amorces recueillies dans un corpus dont on exposera les caractéristiques. On indiquera ensuite de quelle façon on peut en réaliser le classement avant d'interroger la façon dont se réalise la reconnaissance du mot interrompu. Cette dernière étape a la valeur d'un test puisqu'elle met à l'épreuve la reconnaissance verbale appréhendée dans les conditions extrêmes d'une information incomplète. Autrement dit, cette étude propose d'explicitier ce qui se trouve engagé dans une telle opération en conjecturant qu'elle s'apparente à la conduite ordinaire de l'auditeur dont elle révélerait, par contraste, ce que seraient les mécanismes de son fonctionnement.

Sans rien exclure des propositions de la psychologie cognitive ou des théories de l'acquisition du langage, à l'articulation de la phonologie et de la morphologie, cette étude privilégie une approche phonologique

guidée par l'exploitation d'un corpus de français parlé. La compatibilité des résultats avec les travaux conduits dans d'autres champs d'investigation exige un décentrement de la perspective qui permettra de formuler des validations (ou des réfutations) indépendantes des propositions avancées dans cet article. Cette complémentarité est l'un des apports du repérage effectué dans les données.

## 1 Qu'est-ce qu'une amorce ?

A la différence du bégaiement pour lequel une interprétation physiologique est souvent latente et qui est de l'ordre du handicap, l'amorce, dont la phénoménologie semble proche, n'appelle pas d'explication neurologique. Même si un parallèle serait possible entre les variations de sa fréquence en discours et l'aggravation du bégaiement dans des circonstances homologues, elle n'est pas imputable à une constante du comportement expressif. On la retrouve chez tout un chacun en oral spontané, en situation de dialogue (trouverait-on des amorces dans le discours endophasique ?).

### 1.1 Description du phénomène

Le français a retenu, au moment d'assigner un nom au phénomène, son aspect positif : l'unité lexicale est présente par son (ou ses) premier(s) phonème(s). La terminologie est empruntée à la psychologie où l'« amorce » désigne le début d'un mot employé comme stimulus pour aider le sujet de l'expérience à en deviner la suite. A l'inverse, l'anglais utilise « false start » en mettant l'accent sur l'échec de la réalisation complète, un terme emprunté aux compétitions sportives, la course en particulier. Comme si le français visait la possibilité d'une compréhension par l'auditeur – on va montrer qu'elle est souvent effective – et l'anglais l'incapacité du locuteur à réaliser la production complète.

Les amorces relèvent de la catégorie des universaux de langage. Elles se retrouvent dans toutes les langues, ou plutôt dans leur actualisation en discours en sorte que le phénomène a échappé à des descriptions centrées sur le système et non sur la parole, quand bien même elles étaient amenées à se fonder sur l'examen de données orales. Les linguistes ne proposent pas d'hypothèse particulière concernant leur émergence, les catégories concernées ou la forme que devrait prendre un mot quand il se réduit à ce segment. Les études sur le sujet viennent plutôt de la psychologie ou de la théorie de la communication, éventuellement du TAL (les disfluences) voire de la linguistique clinique dans la perspective ouverte au XIX<sup>e</sup> siècle par John Hughlings Jackson.

Invisibles pour une linguistique des langues, les amorces se présentent d'ordinaire comme de simples éléments de l'échange verbal qu'il n'est pas nécessaire de restituer dans la communication pour accéder à la compréhension. Ce sont des interférences, des parasites. Elles interpolent des séquences qui ralentissent le flux discursif et le fourvoient en sorte que, lors du processus d'audition, elles seront spontanément éliminées. Pas de meilleur témoignage du tri sélectif opéré en amont de l'interprétation du message que l'expérience du transcripteur qui s'astreint à les consigner et qui éprouve les pires difficultés à en conserver la trace en mémoire immédiate.

Plusieurs explications ont été proposées pour en raisonner l'apparition, se référant :

- tantôt à quelque particularité linguistique (séquence phonotactique complexe, mot rare, construction syntaxique dérivée) si l'accent est mis sur le fonctionnement du système,
- tantôt à un dysfonctionnement des capacités cognitives (en lien avec une conception de la grammaire comme système de règles computationnelles manipulant des symboles qui correspondent à des classes d'équivalence), qu'on mette en avant
  - la difficulté de la tâche en linguistique cognitive ou
  - des défaillances neurologiques en linguistique clinique,
- tantôt à une perturbation liée aux conditions sociales de l'échange, au contexte de l'expression, résultant
  - soit d'interférences intrasubjectives – comme dans le lapsus –,

- soit d'une insécurité dans l'adresse au moment d'entrer en communication.

L'analyse emprunte ici une voie différente. Elle s'intéresse à l'audition – et non à la production – afin de déterminer de quelle façon une amorce en vient à être identifiée à un mot à partir des données extraites d'un corpus.

## 1.2 Notation : du signal à la transcription

Une fois le cas du bégaiement éliminé, l'amorce, telle qu'elle est conçue dans cette étude, doit répondre à trois propriétés : phonologique, cognitive et graphique.

Phonologiquement, elle apparaît sous la forme d'une trace phonétique, un fragment de signal qui peut être décomposé en un ou plusieurs phonèmes dont l'épellation ne coïncide pas avec un mot sans proscrire la possibilité de compléter ce qui est rétroactivement considéré comme son ou ses premiers éléments. Trois questions apparaissent alors :

(i) la nécessité de distinguer l'amorce de l'abréviation (*prof* n'est pas une amorce),

[1] vous me racontez euh depuis quand vous habitez dans l'agglo- de (2460)

« l'AgglO » est l'abréviation officielle pour la Communauté d'Agglomération d'Orléans Val de Loire.

(ii) la pertinence d'une insertion du mot conjecturé à telle place dans l'énoncé,

[2] venir les les les visit leur rendre (2411)

où « visiter » est confirmé par la reprise « rendre (visite) ».

(iii) la résolution d'une homonymie qu'accroît de façon inversement proportionnelle la réduction du nombre de phonèmes. Il y a beaucoup plus de mots commençant par *r-* que de mots commençant par *re-*, de mots commençant par *re-* que de mots commençant par *rev-* etc.

[3] avec euh y- l'aménagement- (3797)

ne laisse guère d'autre alternative que « aménagement » pour l'assignation lexicale.

Cognitivement, les sons entendus doivent être identifiés avec un mot, ce qui aboutit par exemple à l'absence, dans le corpus, de formes qui seraient notées « eu- », « heu- » et qui ont été systématiquement réinterprétées comme des « euh » par le transcripteur. Il est probable qu'il y ait aussi une sous-estimation d'amorces réduites à « m- » qui sont entendues comme une forme d'approbation et qui sont notées « mh ». Ainsi, pour que l'amorce se développe en mot, deux conditions doivent être réunies : une séquence de phonèmes présente dans un ou plusieurs mots et une ratification conditionnée par sa probabilité contextuelle, « actuelle » en [4] :

[4] Regardez la jeunesse actue- gamin (2321)

le cas le plus simple étant la répétition du mot dans sa forme complète.

[5] parce que y a y a tellement de gens qui viennent de di- régions différentes (2217)

Graphiquement, les amorces recensées sont celles consignées par la transcription, c'est-à-dire celles pour lesquelles ont été surmontées non seulement les habitudes du « sujet écrivain » dans la prise de notes par exemple, ou de « nettoyage » des corpus (en TAL notamment), mais, plus difficiles à neutraliser, ces éliminations qui sont effectuées sans intention, à un niveau subconscient. Cette réduction est le produit d'un apprentissage initial du tri qu'il faut effectuer dans le signal : on n'entend pas ce qu'on sait n'être pas nécessaire à la compréhension du message. C'est une cause d'oubli d'une importance comparable à celle qui résulte de la mauvaise qualité sonore ou de l'inattention du transcripteur.

On remarquera que le transcripteur à son tour anticipe l'interprétation par la façon dont il restitue orthographiquement les propos. Dans l'exemple suivant, la notation par « mai- » (et non « mé ») prescrit la lecture « mairie » :

[6] Avec les mai- mairies de gauche (364)

### 1.3 Les données : le corpus ESLO

Les données sont empruntées aux transcriptions du corpus de l'Enquête Sociolinguistique à Orléans entreprise depuis 2008 et sténographiée ESLO2 pour la distinguer d'ESLO1 qui correspondait à une phase précédente de collecte réalisée entre 1968 et 1971. Les deux corpus sont accessibles sur <http://eslo.huma-num.fr/> où se trouvent toutes les indications concernant le mode de collecte et les caractéristiques de la ressource.

ESLO2 est subdivisé en différents modules. Le plus important quantitativement (environ 40% du temps total) comprend une centaine d'entretiens en face à face, d'une durée moyenne d'une heure. Les propos sont recueillis à partir d'un questionnaire ouvert centré sur la trajectoire biographique et la vie des témoins dans leur environnement quotidien. Cinq éléments à souligner concernant la façon dont les analyses ultérieures sont préparées par le mode de constitution du corpus :

1. ESLO2 n'est pas conçue comme une étape préparatoire au travail linguistique mais comme une phase de la recherche de plein exercice impliquant une analyse de la constitution des données et un retour sur leur élaboration.
2. On ne s'y propose pas de définir un usage du français standard mais, comme cela avait été le cas pour ESLO1, on privilégie l'enregistrement de la variation dans toutes ses dimensions, diachronique (avec ESLO1), diastratique (l'étude est sociolinguistique), dialectale (tous les témoins ne sont pas natifs d'Orléans et une attention particulière est portée au contact de langues) et diaphasique (d'où la diversité des modules).
3. ESLO2 s'organise autour d'un programme, « le portrait sonore d'une ville », qui se focalise sur ce qu'un Orléanais capte en fait de réalisations langagières autant et plus que sur ce qu'il est à même de produire face à un micro.
4. La transcription est alignée sur le signal : il est possible de les séparer pour une analyse en phonétique instrumentale ou pour l'étude des fréquences lexicales mais le corpus est conçu comme la réunion de l'enregistrement et de sa restitution orthographique (pour les conventions adoptées, se reporter au site).
4. On intègre la variation aussi au niveau des transcriptions puisque pour chaque interview (et plus généralement pour chaque enregistrement), il existe au minimum trois versions désignées comme A, B et C :
  - A est le premier état, comprenant le formatage numérique, la segmentation du discours, une transcription « brute » et un lien aux métadonnées à des fins de catalogage et d'indexation.
  - B est une reprise de A afin d'en compléter et d'en affiner la restitution, de vérifier la pertinence ces choix opérés, de corriger des erreurs éventuelles et d'assurer l'anonymisation,
  - C est une reprise de B par un linguiste qui doit réviser l'ensemble, en particulier l'orthographe et le respect des conventions, et contrôler la qualité de l'anonymisation avant la mise en ligne.

Les trois versions sont conservées. Elles permettent d'avoir accès aux processus de compréhension à travers les alternatives d'interprétation survenant à l'audition, ce qui recoupe pour partie le type de travail présenté ici. Par ailleurs, ce processus est indéfiniment cumulatif. Tout chercheur travaillant sur ces données, par exemple pour les annoter (mais il peut s'agir aussi bien d'anthropologues, de statisticiens voire d'artistes...) a la possibilité de corriger, ou du moins de

compléter ou de modifier à son tour la transcription, selon ce qu'il aura entendu ou ce qu'il veut en faire, créant ainsi de nouvelles versions : D, E *ad. lib.*

L'enrichissement est également réalisé par l'outillage du corpus, un programme mené en partenariat avec l'Equipex Ortolang dirigé par J.-M. Pierrel et sa pérennisation relève d'un archivage effectué en lien avec la BnF, au département des documents sonores sous la supervision de P. Cordereix et A. Viault et avec COCOON (Collection de Corpus Oraux Numériques, dont les responsables sont M. Jacobson et S. Guillaume) en lien avec la TGIR Huma-Num, présidée par O. Baude.

## 2 Le traitement effectué : de la requête au tri des données

Pour recenser les amorces, une requête a été lancée par Layal Kanaan-Caillol, IR au LLL sur l'ensemble du module « entretiens » d'ESLO2, en exploitant les conventions de transcription. La demande était formulée de la façon suivante : « extraire toutes les chaînes de caractères ayant pour trois derniers éléments {lettre + trait d'union + espace} ». Il en est résulté une liste de 9504 réponses dont la présentation en colonnes, telle qu'elle a été décidée, livre les informations sous la forme d'un tableau Excel ayant les valeurs suivantes (en italiques, une explication des indications) :

Col. 1 ESLO2\_ENT\_1006\_C

*Dans le corpus « ESLO2 », il s'agit du module « entretien » avec un enregistrement codé « 1006 » dont est utilisée la version de transcription C*

Col. 2 OS6

*Identifiant anonymisé du témoin (le code est fourni par un programme aléatoire)*

Col. 3 00:50:35

*Moment de la réalisation en heure/minutes/secondes dans l'enregistrement (début)*

Col. 5 00:50:43

*Fin de la réalisation en heure/minutes/secondes dans l'enregistrement (fin)*

Col. 5 personnellement ça m'a jam- ça me touchait pas

*Séquence découpée par le transcripateur*

Col. 6 jamais

*Interprétation de l'amorce*

Col. 7 <commentaire éventuel>

*e.g. lapsus, abréviation...*

Les colonnes 1 à 5 correspondant aux résultats de la requête, les colonnes 6 et 7 ont été ajoutées pour cette étude. La référence des exemples portée à la fin entre parenthèses correspond à la numérotation des lignes proposée par Excel. On obtient, à partir de :

[7] voyez là j'ai décrété lundi je travaille pas je dis je vais aller voir les Chèvres euh du pres- du Pentagone (1872)

Soit :

Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7
ESLO2_ENT_1016_C	VB16	00:28:28	00:28:40	(exemple 1872)	Président	lapsus

A partir des résultats bruts, la première tâche à réaliser concerne l'élimination des erreurs de transcription telles que :

- l'insertion erronée d'une espace, e.g. « peut- être », ou une mauvaise interprétation d'un énoncé :

[8] j- jazz oh f- tout ça oui ça ça oui ça je j'allais (1052)

où le transcripteur n'a pas identifié le syntagme festival « jazz off », substituant une amorce en « f- ».

- les abréviations (*triso-*, *asso-*...)

- les amorces qu'une écoute attentive ne permet pas de retrouver dans l'enregistrement.

Inversement est prise en compte la multiplicité des amorces dans une même séquence :

[9] au niveau de le d'Oliv- d'Oliv- O- la S- mh la Source non je pense pas (1875)

Dans cet exemple, on compte quatre amorces – trois fois « Olivet » (une commune au sud d'Orléans) et une fois « la Source » – à l'intérieur d'une seule séquence.

On n'a pas complété le tableau en ajoutant de nouvelles amorces qui n'avaient pas été repérées par les transcripteurs et qui ont été relevées lors des vérifications sur les enregistrements. Il aurait fallu, pour donner cohérence à ce travail, reprendre l'ensemble des entretiens et créer une version D qui soit conforme aux conventions pour la mise en ligne. Il a paru préférable de conserver un état identifié de la ressource dans sa globalité.

Les résultats comprennent donc 9504 lignes pour une centaine d'heures d'enregistrement, soit en première analyse une amorce et demie par minute environ, l'élimination d'une partie d'entre elles étant partiellement compensée par la prise en compte des répétitions. L'essentiel des soustractions est conditionné par l'information fournie, ou non, dans la sixième colonne. Si les lettres suivies du tiret et de l'espace ne peuvent être assignées à un mot, on considère qu'il ne s'agit pas d'une amorce mais seulement d'un élément parasite, qui n'accède pas au statut d'objet linguistique.

Ont été également retirées de la liste :

- les tronctions morphologiques :

[10] et mais ça l'a pas empêché de se représenter et d'être re- re- réélu (1469)

- les abréviations :

[11] donc euh voilà pour organiser des des expo- voilà (1447)

- les lapsus :

[12] explicite ici pas forcément ailleurs j'ai rémarch- remarqué (1504)

On a retenu pour une analyse détaillée les lignes 1000 à 1999 sur lesquelles :

- 45 unités ont été ôtées (abréviations...),

- 106 amorces ont été ajoutées (multiplicité des amorces dans une seule séquence),

- 243 amorces ont dû être éliminées faute de correspondre à un mot identifiable (pas d'interprétation plausible dans la colonne 6).

Ainsi, en lieu et place des 1000 amorces potentielles fournies par la requête, il en demeure :

$$1000 + 106 - (45 + 243) = 818.$$

On a choisi de conserver, en les considérant comme une variété particulière d'amorces, les 22 lapsus. Chacun des exemples sera cité en référence au numéro de la ligne et le fichier sera joint à la version électronique de l'article.

### 3 Comment identifier une amorce ?

A présent, il devient possible de répondre à la question posée dès l'introduction en référence aux données recueillies : quels sont les moyens à la disposition de l'auditeur qui lui permettent d'assigner une interprétation lexicale à une séquence sonore d'un ou plusieurs phonèmes correspondant à un début de mot ?

### 3.1 La reprise

Le cas le plus simple est celui où s'effectue une reprise. Le mot interrompu est complété immédiatement :

[13] je me suis fait no- nommer ici à l'o- à l'occasion d'un changement de de grade euh (1217)

A partir de cet exemple on constate que les reprises concernent :

- soit le mot entier : « de de »,
- soit la première syllabe d'un mot : « no- nommer »,
- soit la première syllabe d'un mot compris à l'intérieur d'une locution prépositionnelle « à l'o- à l'occasion ».

Il y a une différence fonctionnelle entre la reprise d'un mot, qui reproduit une découpe grammaticale, et l'amorce. Celle-ci se conforme à un patron strictement phonologique, indépendamment de toute considération morphologique (la restitution du préfixe dans *occasion* n'est pas accessible en synchronie pour les locuteurs) et, comme le montre l'extension des deux amorces, la séquence à considérer peut être extensible, du mot au syntagme.

La taille de l'extrait à examiner pour identifier l'amorce est également à prendre en compte dès lors que la reprise peut intervenir à un ou plusieurs mots de distance :

[14] quand je suis arrivée à Orléans j'ai laissé tomber la m- j'avais laissé tom- enfin ch- (1371)

Reconstituer « tomber » à partir de « tom- » s'impose : la structure de la phrase est réitérée à l'identique et la modification concerne seulement l'aspect verbal (un plus-que-parfait au lieu d'un passé composé) et non le contenu de l'information. Le « ch- » final serait plus conjectural. Il est presque sûr, à l'audition, qu'il correspond à une réitération du « je suis » au terme d'une assimilation très fréquente à l'oral. On relèvera la permutation de l'ordre : du cas considéré jusqu'à présent, c'est-à-dire de l'amorce reformulée dans le mot entier, on passe à une succession inverse où le mot (*tomber*) précède son amorce (*tom-*), ce qui exclut une explication qui se fonderait sur des difficultés d'accès lexical.

### 3.2 La substitution

Un autre cas de figure correspond à la substitution d'un terme suffisamment proche par la forme et par le sens pour qu'il soit possible de restituer la forme enclenchée :

[15] y a une dizai- oh à peine dix ans (1927)

De « dix ans » à « une dizaine (d'années) », l'équivalence est explicite, la correction portant sur la paraphrase de l'énoncé qui, de « à peu près dix ans », passe à « un peu moins de dix ans ». La synonymie a la particularité de ménager une certaine variation dans l'interprétation.

L'assignation peut se faire sans qu'il y ait d'apparemment entre les termes, la disposition syntaxique et l'énoncé d'un terme de valeur proche dans l'environnement immédiat permettant de compléter l'amorce selon le principe de permutation paradigmatique. En voici deux exemples :

[16] je laisse couler comme j'ai dit j'aurais obli- j'aurais forcément (1828)

Il est probable que « obli- » s'explique en « obligatoirement » sans qu'il soit possible de fournir une explication aux raisons qui ont décidé d'une modification lexicale tardive. Il en va de même pour :

[17] vous avez eu en- vous vous aimez changer (1819)

où la locution « avoir envie » semble un substitut envisageable dont il serait difficile de dire en quoi elle est mieux adaptée à l'intention communicative.

Dans une approche paradigmatique de substitution se retrouvent des occurrences telles que :

[18] je crois que c'est au Quai bran- je suis pas sûre ou au musée d'Orsay je sais plus (1866)

où le contexte élargi et la référence à un second musée parisien également établi sur les quais rive gauche dans les trente dernières années incitent à compléter par « quai Branly » (en intervertissant les majuscules du transcritteur).

La restitution se rapprocherait du cas des anaphores associatives (Kleiber 2009) lorsque, dans un échange où il est fait mention du travail dans une banque, on est amené à supputer « agence » :

[19] à Touiry euh on était une agen- on était cinq hein (1704)

### 3.3 La rectification

La modification peut être rendue nécessaire par la correction d'un lapsus :

[20] de bateaux les бага- les gabares les (1744)

Un mot spécialisé, de faible fréquence, « gabares », au terme d'une permutation de consonnes, appelle « bagarres » et l'élimine avant même d'être parvenu au bout de sa réalisation.

Parfois, en lieu et place d'une inversion, on trouve une anticipation :

[21] comme ça m'av- bl- ça m'avait bien plu (1522)

où la combinaison /occlusive + liquide/ de « plu » est réalisée en avance sur « bien », ou au contraire une persévération :

[22] on lui avait attribué un lun- un logement (1582)

quand la nasale du déterminant se substitue à la première syllabe du mot suivant.

Il se rencontre aussi des cas où une construction syntaxique doit être remaniée afin d'éviter une tournure agrammaticale, comme le recours à la forme pronominale dans :

[23] donc ça complique la tâche d'adap- de s'adapter enfin moi je pensais que tu en avais (1159)

ou le respect d'un accord en genre (ou en nombre) :

[24] de par son s- sa région (1074)

voire une liaison ou, plus généralement, toutes les formes d'enchaînement, en particulier pour des monosyllabes à vocalisme en schwa qui requièrent une élision :

[25] la vieille l- l'hôtel de la Vieille Intendance (1400)

### 3.4 L'interruption

Une dernière condition permet le rétablissement du mot interrompu, c'est la réalisation d'un nombre suffisant de phonèmes pour que l'incertitude soit levée, surtout quand le terme supposé s'ajuste bien au contexte :

[26] donc là c'est pas vraiment une répartis- c'est plus une question de place (1340)

où « répartition » s'impose de lui-même, la transcription reproduisant la prononciation de la constrictive /s/ sans égard à la convention orthographique.

Si on récapitule les quatre cas de figure qui paraissent à même d'expliquer l'ensemble des occurrences pour lesquelles l'amorce peut être élucidée, on aboutit aux subdivisions suivantes :



a) Les reprises (qui s'apparentent à la répétition) comprennent :

- la reprise par un même mot
  - qu'il s'agisse de l'énoncer seul ou de rétablir l'ensemble d'un groupe syntagmatique (locution, SN, SV...),
  - que l'explicitation se fasse immédiatement ou après (voire avant) que d'autres mots se sont intercalés ;
- la substitution, que le mot soit de la même famille ou non (substitution paradigmatique), voire qu'il relève de l'anaphore associative.

En général, l'amorce précède le mot entier.

b) Les rectifications sont distinguées selon qu'elles concernent :

- un lapsus,
- le remaniement d'un enchaînement phonétique (élision, liaison...),
- la mise en conformité d'un accord grammatical.

A chacun des cas correspond une opération différente, autant lorsque l'amorce est préservée :

- en étant complétée dans le cas de la reprise,  
[27] j'ai pas fait hein maleu- enfin pas malheureusement mais (1604)
- ou corrigée pour les rectifications,  
[28] euh moi je- j'aime beaucoup le théâtre j'aime beaucoup le cinéma (1631)

que dans les cas où :

- elle est remplacée avec la sélection d'un autre élément dans le même paradigme,  
[29] et alors le conservat- enfin le directeur à l'époque s'était (1620)
- ou abandonnée quand la phrase est reprise à son point de départ, l'amorce restant inachevée.  
[30] Oui deux g- deux cousines n- (1692)

On a d'un côté un ajout (le mot est complété) ou une réduction (l'amorce reste la seule occurrence), de l'autre une modification, qu'elle soit phonologique, morphologique ou lexicale.

#### 4. Que restitue-t-on ?

On laissera le cas particulier des segments qui n'ont pas d'équivalent, sans représentant dans le lexique mental : mot inconnu du locuteur – un terme technique par exemple –, nom propre, emprunt à une langue étrangère... Dans tous ces cas, seul le signal informe l'interprétation. Dès lors que la cible est connue par l'auditeur, sa stratégie pour restituer les éléments bruités (mot incomplet, prononciation défectueuse, conditions d'écoute difficile) ne paraît pas différer sensiblement de celle exposée à propos des amorces. Le partage tiendrait à la nécessité éprouvée de reconstituer celles-là quand l'élimination de celles-ci est partie intégrante des conditions d'une bonne réception des messages. On mentionnera cependant l'exploitation qui peut en être faite dans un second temps pour interpréter ce que seraient les intentions de celui dont la langue a fourché. Depuis Freud, le lapsus est investi d'une valeur emblématique

Sans faire entrer en ligne de compte les cas d'indécision, on examinera trois cas de figure que les amorces ont permis de prendre en considération, (i) la correspondance avec la graphie, (ii) la pluralité d'identification et (iii) la détermination de la forme sous-jacente (FSJ) telle qu'elle doit être restituée.

#### 4.1 Correspondance

Comment parvient-on à déduire d'une séquence sonore un mot dont les phonèmes ne s'ajustent pas au signal phonétique ? Les conventions de transcription spécifient que les mots doivent être écrits conformément à leur orthographe sans que soient rétablis des mots absents. Par suite, une séquence phonétique entendue comme « chépa » sera reproduite sous la forme « je sais pas », le premier élément de la négation, « ne », n'apparaissant pas.

Ainsi, les conventions résolvent par avance de nombreux cas dont il ne demeure de traces sporadiques que dans des amorces où le transcripateur, faute d'une interprétation, s'est résigné à conserver l'impression phonétique. Dans l'exemple suivant, il pourrait s'agir d'un « je » suivi d'un mot commençant par une consonne sourde (tels que « je pense » ou « je crois »), jusqu'à ce qu'une construction impersonnelle conduise à l'abandon de la proposition commencée :

[31] donc euh donc ch- ça serait bien (1101)

Dans l'exemple [32], il est difficile de décider s'il s'agit d'un cas du même ordre (« chuis gênée ») ou bien d'un lapsus, la persévérance d'une réalisation palatale de la constrictive sourde en lieu et place de la dentale (\* « je chuis (gênée) ») :

[32] une fois <Nom de personne> elle dit oh je ch- je suis gênée ben je dis tu as pas (2571).

Si ces cas sont moins nombreux qu'on ne s'y attendrait, ce n'est pas qu'ils manqueraient dans les propos enregistrés, c'est qu'ils sont tranchés par avance dans la phase de traitement initial en sorte que les occurrences restantes ne soulèvent pas moins de questions d'interprétation pour l'analyse que pour la transcription.

#### 4.2 Pluralité d'identification

Autre difficulté : le choix de la forme à sélectionner dès lors que plusieurs paraissent également possibles dans le même environnement, comme dans le cas du déterminant où, à partir de « l- », l'amorce peut référer aussi bien à *le*, *la* ou *les*.

[33] je vais plutôt demander si l- si le produit sort euh (1151)

Ici, la reprise au masculin singulier ratifie rétroactivement « le » mais rien ne permet d'affirmer qu'au moment où l'énoncé a été interrompu, le locuteur avait déjà tranché. De même pour le pluriel :

[34] et et euh l- les les grandes manifestations orléanaises (1653)

Le choix opéré a posteriori laisserait supposer que la forme retenue était celle attendue mais on rencontre des cas de rectification :

[35] Par par cet- ce ce refus de rentrer dans (1081)

L'indécision s'accroît de façon inversement proportionnelle au nombre de phonèmes disponible. Une réduction à un seul phonème, la nasale labiale par exemple, sera le plus souvent interprétée comme une marque d'approbation ou d'hésitation notée « mh », ou reportée en amorce comme dans :

[36] non m- parce que c'est c'est euh (1029)

où le « m - » laisse conjecturer un « mais ».

Si on examine le cas du « s- » qui est l'amorce la plus fréquente – elle compte à elle seule pour environ 10 % des occurrences tous cas confondus –, on constate qu'elle équivaut à « ce » (1007), « ça » (1016), « c'est » (1030), « c'était » (1051), « son » (1073), « sa » (1074), « se » (1083), « si » (1111), « suis » (1142) etc.

[37] mh alors s- ce que vous avez envie de faire c'est plutôt quoi ? (1007)

- [38] donc je fais s- je je fais ça euh (1016)
- [39] c'est s- c'est un outil intéressant pour euh (1030)
- [40] dans le jardin là ça s- je trouvais très (1051)
- [41] de s- de par son origine sociale et de par euh (1073)
- [42] de par son s- sa région (1074)
- [43] euh un lieu où y s- y se se mélange beaucoup de de milieux (1083)
- [44] et servir oui parce que si s- avec votre accord donc (1111)
- [45] c'est pour ça oui alors je te disais Paris tout à l'heure mais c'est vrai que je s- je suis parti d'Orléans aussi (1142)

Dans chacun des cas, la suite de la proposition permet de lever l'incertitude pour autant que l'énoncé n'est pas remanié complètement. L'amorce apparaît comme un arrêt au moment de choisir un mot qui dépend moins de la mise à contribution du stock lexical que de l'arrangement morpho-syntaxique dont l'enchaînement présente une difficulté au moment de son énonciation. Dans les neuf exemples, il s'agit de mots-outils (pronoms, déterminants, présentatif, conjonction, auxiliaire...) en sorte que l'amorce n'est pas liée à la difficulté de le réaliser en tant que tel mais à la capacité du locuteur à décider ce qui doit suivre et qui concerne le contenu du message plus que sa phraséologie.

## 5 La « Forme Sous-Jacente »

Un dernier cas doit être considéré qui implique la forme sous-jacente ou FSJ (Brandão de Carvalho, Nguyen & Wauquier 2010). La FSJ correspond à la séquence de phonèmes (associée dans d'autres langues à des traits prosodiques) qui caractérisent structurellement un morphème ou une unité lexicale. Par sa consignation mentale, elle sert de référence pour l'identification du mot par l'auditeur, dans la superposition postulée entre l'*I-Language* et l'*E-Language*. Une équivalence globale, assez peu interrogée, postule une certaine équivalence entre la transcription en API donnée par les dictionnaires, augmentée des transformations morphologiques attendues (dans le cas d'un verbe conjugué, d'un pluriel irrégulier, d'une modification liée au genre...) et cette FSJ. Les amorces rendent possibles un réexamen de ce point de vue.

Au nombre des difficultés de circonscription de la FSJ, relevées depuis longtemps, a été pointée la pertinence de la notation de la consonne de liaison (Encrevé 1988) ou de certains schwas. Quels sont aujourd'hui les cas d'emploi réels du schwa dans un mot comme « bouleverser » en français central ? Faut-il indiquer un schwa à la fin d'un mot comme « arc » qui apparaîtra dans un syntagme comme « arc brisé » ? La prise en compte des amorces soulève une interrogation d'une autre nature.

La convention de transcription stipule que tout mot, identifiable ou non, qui n'aura pas été réalisé pleinement constitue une amorce et doit être consigné comme tel. Faisant retour des réalisations orales vers la structure de la langue, on voudrait interroger ce qu'il en est de la FSJ concernant les pronoms « il(s) » et « elle(s) » – pour celui-ci en restreignant l'observation à sa fonction de sujet, en excluant les emplois en apposition :

- [46] elle était dans quel d- domaine elle ? (1595)

ou en complément :

- [47] parce que bon i- elle est elle avait sympathisé avec elle (1655)

L'hypothèse sera formulée de la façon suivante :

Devant consonne, la forme standard de il/ils est i-, de elle/elles est e- (prononcé « è »). Devant voyelle, la forme standard est il, el et, au pluriel, iz, èz.

Un allomorphisme phonétique, d'ordre contextuel, est observable, marqué par l'apparition d'une consonne. En cas de liaison obligatoire au pluriel (« ils ont », « ils étaient ») le /l/ reste éliidé. Cela concerne une consonne qui ne figure pas dans la liste de celles qui sont impliquées, en français, dans la liaison mais surtout, de façon inverse à ce qui se produit avec les liaisons facultatives, le /l/ peut se maintenir devant une consonne, comme cela a pu être relevé pour d'autres mots tels que *but*, *fait* ou *quand*. De même, l'absence de réduction dans des fonctions autres que le sujet rejoint partiellement le comportement du schwa qui ne peut intervenir dans les positions emphatiques, soit par substitution de forme (« je » > « moi »), soit par accentuation (« dis-le » > \*« dis-l »).

Signe de l'embarras de certains transcripateurs qui ne perçoivent pas le /i/ isolé comme une amorce, la tentation de lui substituer graphiquement un « y » qui ne se justifie pas grammaticalement :

[48] Ou alors y- des fois ils font des euh des créneaux ce qu'ils appellent des réceptions en libre (1164)

Du fait que sur les mille premières phrases comprenant une amorce, aucun des {il/ils/elle/elles + consonne} en position sujet ne se présente avec sa forme complète, on devrait conclure que la FSJ ne comporte pas le /l/ comme un élément obligatoire mais plutôt comme un phonème alternant de liaison, au même titre que le /z/ du pluriel. Il est possible que le phénomène ait une certaine profondeur diachronique que la graphie a masquée.

## 6 Conclusion

Au-delà du phénomène de disfluece et de ses effets, l'amorce permet d'observer comment les locuteurs interprètent les éléments phonologiques disponibles pour interpoler les informations manquantes dans le signal. Parce que cette situation peut être généralisée à l'ensemble des cas où l'auditeur doit établir une identification lexicale (de la trace phonétique à la FSJ) en dépit d'une disponibilité lacunaire des informations contenues dans le message, l'amorce est qualifiable en tant qu'expérience et acquiert une portée supérieure à ce que le terme négatif de disfluece pourrait laisser croire.

À côté des modes compensatoires syntagmatiques (contextualisation et reprise) et paradigmatiques, qui ratifient la FSJ, on a pu noter un effet en retour de l'analyse des productions sur la FSJ. Le pronom personnel sujet de troisième personne ne se présente pas avec la forme attendue non seulement phonétiquement, comme il en va pour n'importe quel mot, mais aussi phonologiquement. La consignation d'une consonne qui a déjà disparu en cette position (de la même façon qu'elle a disparu dans le mot *outil*), au croisement de la morpho-syntaxe (seule la position sujet est concernée) et de la phonologie pose la question de la représentation abstraite de cette unité lexicale. La réduction à une prothèse vocalique (hors liaison) du pronom sujet de la 3<sup>e</sup> personne (la non personne) pourrait transformer en préfixe la marque devenue verbale, l'amuïssement de la consonne conduisant à un appui sur le mot suivant comme dans « il me dit » prononcé /im-di/ ou « il se croit » prononcé /is-kRwa/, ce qui laisse entrevoir des possibilités d'infixation (celle des pronoms réfléchis), voire l'éventualité d'un effacement complet déjà effectif dans un syntagme figé comme « (il) faut ».

## Références bibliographiques

- Brandão de Carvalho, J., Nguyen, N., Wauquier-Gravelines S. (2010) *Comprendre la phonologie*. Paris : PUF.
- Chomsky, N. (1986) *Knowledge of Language*. New York : Praeger.
- Encrevé, P. (1988) *La Liaison avec et sans enchaînement*. Paris : Seuil.
- Jackson, J. H. (1932) *Selected Writings*, édition par J. Taylor (2 vol.). Londres : Hodder & Stoughton.
- Kleiber, G. (2001) *L'Anaphore associative*. Paris : PUF.