

Analyse des stratégies de chunking en interprétation simultanée

Myriam Piccaluga & Bernard Harmegnies

Université de Mons-Hainaut
Institut de linguistique, 22 place du Parc, 7000 Mons, Belgique
il@umh.ac.be
<http://www.umh.ac.be/linguistique.html>

ABSTRACT

In this paper, which is meant as a methodological account, we focus on a new variable (“Ecart Inter Syllabique”: EIS), intended to improve the study of speech chunks produced by subjects performing a task of simultaneous interpreting (“IS”). The variable is introduced on the basis of a discussion of the main methodological trends in the field, with the aim of improving the validity and reliability of the numerical treatments applied to the study of IS. An experimental essay is performed on a prototypical sample of 4 subjects, performing IS under several conditions. The behaviour of the variable within the design suggests its interest for future research.

1. INTRODUCTION

L’interprétation simultanée est une tâche consistant, pour le sujet, à réémettre dans une langue un message qu’il est en train de recevoir dans une autre. Cette activité de traduction orale «online» peut, comme toute activité de médiation langagière, être observée en tant que comportement spontané chez tout bilingue ; elle constitue alors un cas particulier de *traduction naturelle* [1]. Eduquée à la faveur d’une formation intensive, elle peut se développer dans l’exercice professionnel de l’*interprétation de conférence* ou de l’*interprétation de liaison*, prestations professionnelles dont l’internationalisation du monde du travail a contribué, durant les dernières décennies, à accroître la visibilité. Notre objet d’étude principal est cette tâche – nous l’appellerons ici *la tâche interprétative* : TI - quel qu’en soit le contexte d’effectuation et quelle que soit l’expertise des sujets l’exécutant.

Les différentes études cognitives, tant empiriques que spéculatives, qui ont abordé la TI [e.g., 2,3] s’accordent quant à l’idée que la *simultanéité* qui caractérise la tâche est plus complexe que la simple juxtaposition temporelle d’un processus de réception et d’un processus d’émission. Pour Paradis [4], la TI suppose l’effectuation successive de nombreuses sous-tâches permettant d’assurer la compréhension, la traduction et le monitoring (saisie en mémoire échoïque d’un segment de discours source, décodage, représentation non linguistique du message, encodage en langue cible, émission en langue cible, stockage de

la production en mémoire échoïque, représentation non linguistique du message, comparaison des deux représentations non linguistiques). Un modèle de ce genre postule évidemment la segmentation du signal d’entrée en portions suffisamment réduites pour permettre le traitement. Une chaîne de sous-tâches peut ainsi être mise en route alors qu’une autre est déjà active : plusieurs chaînes de traitement sont ainsi susceptibles de fonctionner en parallèle à condition que, en un temps t déterminé, les différentes sous-tâches actives simultanément soient de nature différente et portent sur des portions différentes du discours source. Cette idée d’un nécessaire découpage du discours d’entrée se retrouve chez d’autres auteurs, qui, avec la notion de *chunk*, réfèrent à une portion du discours source caractérisée par le fait qu’elle constitue une unité brève, mais suffisamment riche d’informations pour faire l’objet d’un traitement sémantique unitaire. Cette vision est à lier avec la présence, dans plusieurs modèles, d’un *buffer* mémoriel, qui évoque d’ailleurs, explicitement ou non, les dispositifs de stockage d’entrée que comportent les modèles plus généraux du langage [5].

Il est raisonnable de penser que l’aptitude du sujet à effectuer un chunking efficace du discours source est une condition *sine qua non* (mais non suffisante) de l’obtention d’un produit interprétatif de qualité. L’étude des capacités du sujet à ce type de découpage apparaît donc fort intéressante en vue de l’analyse des facteurs influant sur l’efficacité du sujet dans la tâche, partant, des processus cognitifs sollicités et, in fine, de la validité des modèles de la TI.

L’étude des stratégies de chunking n’est cependant pas aisée. En effet, le chunking n’est à l’origine d’aucun comportement directement observable. Une alternative intéressante à des designs complexes risquant de dénaturer la tâche peut cependant consister en la sollicitation du sujet dans le cadre d’une TI simple, avec, dans le chef du chercheur, une focalisation non pas sur l’*entrée* de la *boîte noire* mais bien sur la *sortie*. Si le traitement se déroule sans heurts, on peut en effet s’attendre à ce que le chunking d’entrée soit régulier et à ce qu’il induise un chunking de sortie lui aussi régulier. Par contre, un chunking d’entrée irrégulier ou toute perturbation, à quelque niveau de traitement que ce soit, rend probable un chunking de sortie irrégulier. L’idée qui se dégage de ces réflexions amène donc à

prendre en compte l'analyse du chunking de sortie en tant que révélateur indirect du chunking d'entrée.

Dans cet article, nous proposons une contribution à caractère exploratoire et méthodologique visant à étudier l'intérêt d'une variable (l'Ecart Inter Syllabique : EIS), escomptée apte à rendre compte des phénomènes de chunking sans nécessiter le recours à des procédures classiques dont nous montrons les limitations. A cet effet, la variable à l'étude est testée dans le cadre d'un essai expérimental exploratoire visant à en investiguer les qualités métrologiques.

2. LA QUANTIFICATION DU PHENOMENE

2.1. Des approches classiques

L'étude de la structuration temporelle du discours cible peut, en première analyse, se baser sur l'une ou l'autre de deux voies alternatives. La première se centre sur l'évaluation de l'importance relative du signal articulé par rapport au temps de la locution, soit sur base de dénombrements, soit au départ de mesures de durée des groupes syllabiques, dans la foulée des travaux fondateurs de Goldman-Eisler [e.g. 6]. La deuxième recourt à l'analyse des pauses, celles-ci pouvant être considérées comme délimiteurs de chunks.

Que l'on examine la locution avec un regard centré sur les sections de signal de parole qu'elle comporte ou qu'on la considère avec une attention plus ciblée sur ses lacunes, voire qu'on adopte simultanément les deux perspectives, on bute inévitablement sur un problème délicat, celui de l'établissement des frontières de segments.

Ainsi, il n'est ici pas toujours aisé de délimiter avec certitude l'endroit où finit une syllabe et où commence la suivante. Tel est le cas si, par exemple, la première syllabe se termine par une voyelle et la suivante débute par une voyelle ; tel est également le cas pour certaines structures consonantiques (en particulier les approximantes), qui posent également des problèmes de démarcation.

Quant à l'étude des pauses, si elle peut paraître séduisante, force est de reconnaître qu'elle se heurte cependant à plusieurs difficultés méthodologiques. En effet, l'étude des chunks basée sur l'analyse des pauses considérées comme marqueurs de frontières nécessite la détermination d'un seuil de durée : seules doivent être prises en considération les pauses significatives d'une activité cognitive de plus haut niveau que l'implémentation phonétique. Ceci nécessite que, dans l'univers des pauses, on établisse deux sous-ensembles. Le critère le plus couramment utilisé est celui de la durée qui, pour d'aucuns, doit être supérieure à 250ms [e.g. 6]. Cette valeur est cependant loin d'être la seule que propose la littérature, ce qui est probablement dû, en partie, à des différences méthodologiques entre auteurs, mais peut-être, plus profondément, à la nature-

même des pauses, dont les travaux les plus récents montrent bien la diversité [7,8]. De ce point de vue, le modèle taxinomique binaire sous-jacent à la prise de décision dichotomique sur base d'une seule valeur démarcative pose autant problème sur le plan épistémologique que sur le plan méthodologique. Enfin, il faut bien voir que l'analyse d'un phénomène (les chunks) sur base d'un non-événement (la rupture du flux phonatoire) est, par-delà ses limitations épistémologiques, à l'origine de problèmes techniques. Ainsi, si l'on tente d'automatiser un processus de délimitation de chunks, on se heurte rapidement à un autre problème de seuil concernant, cette fois-ci, l'intensité. Si les signaux ont été recueillis dans un environnement un tant soit peu écologique, les phases silencieuses risquent fort d'être contaminées par les bruits environnementaux, ce qui constitue une nouvelle source d'invalidation de la procédure.

Ces différentes approches, qui ont montré leurs limites, se caractérisent de surcroît par le fait qu'elles ne sont que difficilement automatisables, vu la quantité importante de décisions et de spéculations que requiert leur mise en œuvre.

2.2. Une approche innovante

La variable dont nous souhaitons ici tester la qualité (l'EIS) est obtenue par différenciation des moments où apparaissent les pics d'intensité des noyaux syllabiques [9]. Pour chaque pic, on calcule l'intervalle de temps le séparant du sommet syllabique précédent. Pour tout corpus de N syllabes, on obtient ainsi une liste chronologiquement ordonnée de N-1 durées inter syllabiques.

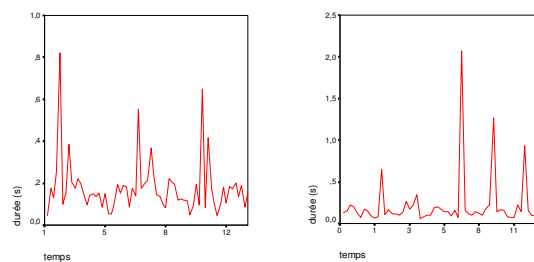


Figure 1 : évolution des valeurs d'EIS (en ordonnée) en fonction du temps (en abscisse), durant les 13 premières secondes de l'effectuation d'une TI par le même sujet, sous conditions peu perturbée (à gauche) et fortement perturbée (à droite).

Comme le montre la figure 1, les graphes d'évolution de l'EIS montrent une alternance plus ou moins régulière de pics et de vallées. Les pics correspondent aux EIS longs, signalent une baisse momentanée du débit ; les vallées, au contraire, correspondent aux périodes caractérisées par un débit plus rapide. C'est donc une forme de graphe d'évolution du débit instantané au cours du temps que nous avons ici. On peut reconnaître, dans l'alternance entre d'une part des zones assez longues caractérisées par des durées

localement faibles et peu variables, et d'autre part, des zones plus brèves caractérisées par un débit local ralenti, le séquençement du flux de parole en chunks de matériau phonique.

Lorsque le processus d'interprétation se déroule sans heurts (partie gauche de la figure 1), l'évolution de l'EIS prend la forme d'une alternance plutôt régulière de pics (les frontières de chunks) et de vallées. Par contre, lorsque le rapport *expertise/difficulté de la tâche* décroît, la régularité est perturbée : on voit apparaître des zones à EIS stablement bas (débit élevé : le sujet tente de rattraper le temps perdu) et d'autres à pics élevés et fréquents (accumulation de pauses longues : le processus interprétatif trébuche).

3. ESSAI EXPERIMENTAL

3.1. Sujets

Les sujets sont au nombre de 4, tous originaires de la région de Barcelone et y résidant depuis plusieurs décennies. Tous ont un haut niveau de maîtrise des langues française et espagnole, qu'ils exercent toutes deux fréquemment; pour chacun, l'espagnol fait cependant figure de langue dominante. Hormis ces similarités, les sujets se caractérisent par des différences en termes de leurs expertises d'une part au plan de la maîtrise de la TI et d'autre part au plan de la maîtrise linguistique (Cfr table 1).

Table 1 : ventilation des sujets en termes d'expertise linguistique (bilingues vs quasi-natifs) et d'expertise de la tâche (professionnels vs non professionnels).

	Professionnels	Non professionnels
Bilingues	<i>Int1</i>	<i>Étud</i>
Quasi natifs	<i>Int2</i>	<i>Biling</i>

3.2. Corpus sources et combinaisons linguistiques

Chaque sujet a été soumis à 6 tâches interprétatives, chacune consistant en le traitement d'une conférence originellement prononcée à la tribune du Parlement Européen. Chaque sujet a dû interpréter 3 discours dans chacune des deux combinaisons (*français vers espagnol et espagnol vers français*).

3.3. Perturbations apportées aux corpus sources

Chaque discours source a subi des perturbations introduites en laboratoire. Celles-ci consistaient d'une part en une altération locale du débit (accroissement du débit par réduction à 80%, 70% ou 60% de la durée totale de production sans modification des caractéristiques de F_0) et d'autre part en l'adjonction locale d'un bruit blanc parasitant (0 dB, 3 dB ou 6 dB re/niveau moyen du signal du discours source). Les

analyses présentées ici ne portent que sur certaines des portions de discours, caractérisées, chacune, à la fois par un traitement de parasitage et un traitement de compression temporelle.

3.4. Méthodologie statistique

Nous analysons ici les variations de l'EIS au moyen de procédures d'analyse de variance permettant de tester l'hypothèse d'un effet de nos variables indépendantes (désormais « VI ») compte tenu de leurs interactions (dont nous ne pouvons ici rendre compte, par manque de place). Nous nous interrogerons donc sur l'action des VI *parasitage*, *compression temporelle*, *combinaison linguistique* ainsi que *sujet*.

3.5. Résultats

La procédure décrite ci-dessus a abouti au recueil de 12930 valeurs, présentant une moyenne de 268ms, supérieure à la médiane (173ms), ce qui trahit une dissymétrie droite prononcée, (coefficient de 7,560), allant de pair avec une forme fortement leptocurtique (coefficient d'aplatissement de 96,462). Les percentiles 50, 90 et 95 sont d'ailleurs tous trois contenus dans l'intervalle [0-1], alors que l'étendue de la distribution est de près de 10 unités. Etant donné cet évident écart par rapport au modèle gaussien, nous avons appliqué, en vue de l'analyse de variance, une transformation argument-tangente hyperbolique visant à normaliser la distribution. Cette opération s'est révélée globalement satisfaisante, puisqu'elle a permis une diminution du coefficient de dissymétrie de 7,560 à 1,219 et du coefficient d'aplatissement de 96,462 à 3,091 (Cfr fig. 2)

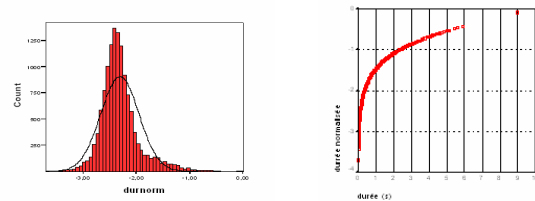


Figure 2 : distribution des durées d'EIS après normalisation ; à gauche : valeurs observées et courbe normale ajustée ; à droite, relation entre les valeurs normalisées (en ordonnée) et les valeurs d'origine.

Afin d'en faciliter la lecture, nous présentons dans les graphiques suivants non les valeurs d'EIS transformées, mais bien les valeurs dans l'unité originale (ms).

Comme le montre la figure 3, les sujets se différencient fortement en termes d'EIS ($F=70,104$, $d.l.=3$, $\alpha<.001$). On voit apparaître deux groupes distincts : d'une part, *biling* et *int2*, qui se caractérisent par des valeurs comprises entre 225 et 250ms et, d'autre part, *étud* et *int1* qui, tous deux, ont des valeurs avoisinant 310ms. Une distinction nette en termes d'expertise de la langue apparaît donc ici. Sous l'effet de la combinaison, les valeurs de l'EIS se modifient aussi. La différence, ténue mais significative ($F=216,368$, $d.l.=1$, $\alpha<.001$),

va dans le sens d'un accroissement lorsque la production est en français, la langue globalement la moins bien maîtrisée dans l'échantillon.

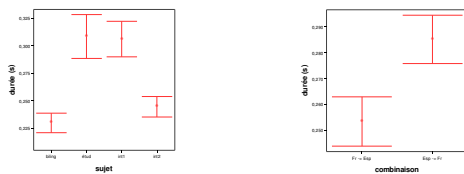


Figure 3 : évolution des valeurs d'EIS (en ordonnée) ; à gauche, en fonction du locuteur (de gauche à droite : *biling, étud, int1, int2*) ; à droite, en fonction de la combinaison (de gauche à droite : *Fr vers Esp, Esp vers Fr*).

Comme le montre la figure 4, l'accroissement du taux de parasitage induit une élévation sensible de l'EIS ($F=20,489$, $d.l.=2$, $\alpha<.001$). Celle-ci est quasiment linéaire. Tout aussi linéaire est, au contraire, la décroissance significative ($F=19,474$, $d.l.=2$, $\alpha<.001$) que produit l'accroissement du taux de compression, à l'opposé de ce que l'on constate concernant le parasitage.

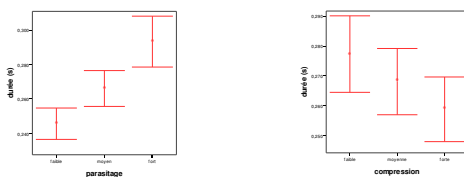


Figure 4 : évolution des valeurs d'EIS (en ordonnée) ; à gauche, en fonction du parasitage (de gauche à droite : *croissant*) ; à droite, en fonction de la compression (de gauche à droite : *croissante*)

4. CONCLUSIONS

Au plan strictement méthodologique, on peut constater que la variable à l'étude s'est montrée apte à différencier, à posteriori, des groupes de mesures correspondant à des classes à priori du dispositif expérimental (classes d'expertise linguistique, de combinaison linguistiques, de perturbations). En première analyse, il semble donc que ses intérêts théoriques (indépendance de toute logique de seuils, robustesse face aux environnements acoustiques adverses) et technique (possibilité d'automatisation aisée de la mesure) s'assortisse d'un intérêt patent en matière d'informativité.

Les observations relevées à propos des VI *sujet, combinaison et parasitage* vont dans le sens d'un EIS bas pour une situation où les contraintes sont minimales : bonne maîtrise de la langue d'émission et de la langue de réception, absence de bruitage. La variable *compression temporelle* tranche évidemment avec les autres ; il faut bien voir, cependant, qu'elle est la seule à exercer un effet direct sur le séquençage de la réception (et, en conséquence, de l'émission, dans le

contexte d'une TI), ce qui confirme en quelque sorte à contrario la sensibilité de la variable.

Notons, par ailleurs, que l'exploitation des données nous a conduits à nous centrer sur un seul aspect des mesures : leur *tendance centrale*. D'autres développements, plutôt centrés sur la *variabilité* de l'EIS, pourraient opportunément être envisagés. Par ailleurs, une étude plus approfondie des propriétés métrologiques, poussant notamment plus avant la recherche d'un caractère gaussien de l'EIS, est probablement souhaitable. Enfin, si notre essai expérimental nous a conduits à privilégier un regard macroscopique visant surtout à éprouver les qualités métrologiques d'un dispositif numérique, une étude plus clinique interrogeant microscopiquement les relations entre les variations de l'EIS et les comportements observés serait de nature à mieux éprouver sa validité ; elle pourrait participer d'une analyse investiguant les plans de la production (aspects phonétique-phonologique, lexical, morpho-syntaxique, pragmatique, etc.), de la médiation (aspects traductologiques), voire des qualités subjectives du discours cible du point de vue du récepteur final (aspects liés à l'intelligibilité).

BIBLIOGRAPHIE

- [1] A.M.B. De Groot, The cognitive study of translation and interpretation. Three approaches, in J.H. Danks, G.M. Shreve, S.B. Fountain, M.K. McBeath, *Cognitive processes in translation and interpreting*, Sage Publications, Thousand Oaks, 25-56, 1997.
- [2] D. Gerver, Empirical studies of simultaneous interpretation: a review and a model, in R.W. Brislin, (Ed.), *Translation: Applications and Research*, Gardner Press, New York, 165-207, 1976.
- [3] B. Moser, Simultaneous interpretation: a hypothetical model and its practical application, in D. Gerver and W. Sinaiko (Eds.), *Language interpretation and communication*, Plenum Press, New York and London, 353-368, 1978.
- [4] M. Paradis, Toward a neurolinguistic theory of simultaneous translation: the framework, *International Journal of Psycholinguistics*, 10, 3(29), 319-335, 1994.
- [5] A. Baddeley, *Human Memory Theory and Practice (Revised Edition)*, Allyn & Bacon, Boston, 1997
- [6] F. Goldman-Eisler, Segmentation of input in simultaneous translation, *Journal of Psycholinguistic Research*, vol. 1, n°2, 127-140, 1972.
- [7] E. Campione & J. Véronis, Pauses et hésitations en français spontané, *Actes des 25èmes Journées d'Etudes sur la Parole (JEP 2004)*, 109-112, 2004.
- [8] M. Candéa, *Contribution à l'étude des pauses silencieuses et des phénomènes dits "d'hésitation" en français oral spontané*. Thèse de doctorat, Université Paris III, 2000.
- [9] M. Piccaluga, *Approches psycholinguistiques de l'interprétation*, Thèse de doctorat sous la direction du prof. J.-L. Nespoulous, UMH, 2004.