

Contraintes de contrôle articulatoire intrasyllabique dans la mémoire de travail verbale

Marc Sato, Jean-Luc Schwartz, Marie-Agnès Cathiard, Christian Abry & Hélène Loevenbruck

Institut de la Communication Parlée – INPG / Université Stendhal / CNRS UMR 5009
46, avenue Félix Viallet – 38031 Grenoble Cedex 01, France
Tél.: ++33 (0)4 76 57 48 27 - Fax: ++33 (0)4 76 82 43 35
E-mail: sato@icp.inpg.fr - <http://www.icp.inpg.fr>

ABSTRACT

Verbal transformation effect – an auditory imagery task equivalent to Necker's cube in visual imagery – recruits a specific working memory, the so-called *articulatory* or *phonological loop*. Is this mechanism sensitive to articulatory control constraints, i.e. *phase* relationships between vowel and consonant gestures? In our experiment, 56 French students repeatedly pronounced aloud non-sense syllables - all combinations of [ə] with [p] and [s] - and were asked to stop as soon as they heard a possible syllable transformation. In agreement with our in-phase predictions, the winner is syllable [psə], where all gestures can be launched in synchrony. This experiment demonstrates that verbal working memory – a primary candidate as *input memory for word learning* – is sensitive to articulatory control of syllable phasing.

1. TRANSFORMATIONS VERBALES & BOUCLE ARTICULATOIRE

Au cours d'une série de travaux sur les images mentales auditives, Reisberg et al. [Rei89] ont utilisé le paradigme de transformations verbales [War61]. Ce jeu de mot [Tre83] correspond, en parole, à la rivalité perceptive de profondeur sur le cube de Necker [Cha85] : la répétition rapide et à haute voix de certains mots produit un flux sonore qui est compatible avec plus d'une segmentation (ainsi, la répétition du mot "life" fournit un flux sonore compatible avec une segmentation de type "life" mais aussi de type "fly").

Dans leur expérience, les auteurs demandaient aux sujets de répéter rapidement le mot "stress" et de détecter une éventuelle transformation – les conditions variant d'un groupe de sujets à l'autre d'une répétition à voix haute vers une répétition imaginée. Cette expérience montre que les images auditives ne sont transformées que lorsque la subvocalisation n'est pas empêchée (on assiste alors à une transformation majoritaire du mot "stress" vers le mot "dress").

Reisberg et al. situent ces résultats dans le cadre de la boucle articulatoire, la composante de la mémoire de travail modélisée par Baddeley [Bad86]. Rappelons que cette boucle articulatoire peut être décomposée en deux sous-unités : une unité de stockage phonologique et un processus de contrôle articulatoire, appelé aussi processus de répétition subvocale. L'unité de stockage phonologique est destinée à retenir les représentations phonologiques des informations langagières. Ces traces mnésiques étant supposées s'effacer, le processus de contrôle articulatoire a pour but de rafraîchir ces traces par un mécanisme d'auto-répétition subvocale. L'exploration des capacités de cette boucle s'est faite sur le nombre d'unités mémorisables et leur composition essentiellement phonologique

[Bad90]. Des recherches récentes semblent montrer que la contribution de ce système dans la rétention de matériel verbal familier ne serait qu'un produit accidentel de la fonction première de ce système : son véritable rôle serait de servir le langage dans l'apprentissage de formes phonologiques nouvelles [Gat93]-[Gup97]-[Bad98].

Reisberg et al. font correspondre l'unité de stockage phonologique et le processus de répétition subvocale à leurs concepts de "pure imagery" (image mentale sans subvocalisation) et "enacted imagery" (image mentale avec subvocalisation). En ce qui concerne le recrutement du processus de répétition dans la tâche de transformation verbale, ce n'est pas la répétition d'un seul item qui le mobilise (qui tend vers une répétition automatique) mais bien la recherche mentale de la transformation.

2. HYPOTHÈSE DE PHASAGES ENTRE GESTES ARTICULATOIRES

Une question clé, totalement passée sous silence par Reisberg et al., est celle d'éventuelles dissymétries de la transformation verbale. Ainsi, la transformation inverse de "life ⇒ fly" en "fly ⇒ life" est beaucoup plus improbable. Ce phénomène est dû selon nous à l'existence de phasages articulatoires préférentiels. Ainsi, dans le premier cas, le locuteur peut anticiper dans la constriction labiodentale du [f], la constriction linguale du [l] - l'attaque de la syllabe [fl] a ses éléments phasés - et la position de la langue pour le départ de la diphtongue [ai]. Il peut alors lancer alors les trois gestes en synchronie. Dans le second cas, la synchronisation du [l] en attaque et du [f] en coda est impossible.

Cette hypothèse de phasages articulatoires préférentiels renvoie aux propriétés des systèmes dynamiques, mises en évidence par Haken, Kelso & Bunzt [Hak85]. L'expérience princeps consiste à mouvoir ses deux index de bas en haut et de façon anti-symétrique (produisant ainsi des mouvements en opposition de phase), puis à augmenter progressivement la vitesse de ces mouvements. On constate alors un changement brusque en un mode symétrique (mouvements phasés) impliquant l'activation simultanée des muscles homologues de flexion d'une part et d'extension d'autre part. Ce phénomène soudain et complètement involontaire de transition de phase, modélisé par les auteurs, a été interprété en termes de couplage inter-articulatoires (une interprétation perceptive plutôt que motrice de ce phénomène vient toutefois d'être proposée par Mechner et al. [Mec01]).

Tuller & Kelso [Tul90] ont tenté d'appliquer ce principe au système de production de la parole. En prenant comme test la syllabe CV [pi] et la syllabe VC [ip], les auteurs ont ainsi montré une bascule de la séquence [ip] vers la séquence [pi] par accélération du débit de répétition. L'index utilisé, en fait le

VOT (soit la coordination clusion labiale/aperture glottique), ici l'index du phasage du pic d'aperture glottique par rapport au maximum de constriction des lèvres, montre bien qu'on a changé de structure syllabique – de VC à CV et non l'inverse – par l'apparition du phénomène d'aspiration [p] → [p^h]. Mais cet index n'est pas un bon révélateur du contrôle puisque l'on passe d'une relation en phase, dans le cas de [ip], à une relation déphasée dans [pi]. L'index correspond ainsi à un contrôle du VOT, spécifique aux langues comme l'anglais, dans la coordination oro-laryngée de la consonne, et pas à la coordination (coarticulation) entre la consonne et la voyelle.

Les stimuli utilisés par Reisberg et al. ([t] vs. [d] dans "stress" vs. "dress" et [k] vs. [g] dans "kiss the sky" vs. "kiss this guy") reposent aussi essentiellement sur un contrôle orolaryngé du voisement.

La littérature subséquente n'ayant pas apporté d'index de cohésion ou de phasage ferme sur ce point [Bro00], nous proposons de tester dans le travail présenté ci-après, l'apport à cette question de l'effet de transformation verbale.

3. EXPERIENCE

L'expérience décrite ci-dessous a pour objectif de montrer la sensibilité de la boucle articuloire aux contraintes de contrôle de phases intra-syllabiques, par la mise en évidence de transformations préférentielles lors d'une tâche de transformations verbales.

Nous considérons que la répétition externe (à haute voix) d'une séquence implique sa répétition interne (en voix intérieure). En effet, il n'est pas concevable de faire en même temps et avec les mêmes articulateurs un mouvement imaginé différent d'un mouvement réel. De ce fait, l'exploration en imaginé, telle que décrite par Reisberg et al., n'est pas prioritaire pour nous puisqu'elle implique une inhibition de l'externalisation en plus de l'exécution mentale [Grè01], ce qui correspond en général à des transformations moins nombreuses. Nous nous placerons donc dans le cas plus favorable d'une répétition à haute voix.

3.1 Choix des stimuli

Pour se dégager de possibles effets lexicaux et afin de tester directement la stabilité intrasyllabique sans faire intervenir les questions de segmentation entre syllabes [Con01]–[Dum98], nous utiliserons des logatomes monosyllabiques. Nous étudierons des syllabes comprenant deux consonnes et une voyelle afin de tester les relations de phase aussi bien entre voyelle et consonnes qu'entre consonnes. Nous avons choisi la voyelle neutre [ə] et deux consonnes dont une a un point d'articulation plus avancé que l'autre, en l'espèce une labiale [p] et une coronale [s], pour qu'existe la possibilité de produire la consonne postérieure pendant la réalisation de la consonne antérieure. Le schwa n'est pas supprimable quelle que soit la combinaison. Les choix de ces consonnes permettent d'obtenir des séquences qui sont toutes phonotactiquement attestées pour un locuteur du français.

Une fois mise en boucle, chacune de ces combinaisons peut donner lieu, selon nos hypothèses, à l'une ou l'autre des séquences du même groupe :

- Groupe 1 : les séquences [psə], [səp], [əps].
- Groupe 2 : les séquences [spə], [pəs], [əsp].

Autrement dit, nous prédisons que nos sujets, à partir d'une séquence présentée oralement, ne se livreront pas à toutes les

permutations possibles, mais seront naturellement amenés à "lire mentalement" le résultat de leur répétition selon un "parsing glissant" au sein d'un groupe.

Table 1 : Classement des séquences en fonction de leur degré de phasage et transformation attendue.

Séquence	Phasage	Prédiction
psə	Complet – phasage de l'attaque et de la voyelle, phasage dans l'attaque ([s] dans [p])	psə
səp	Moyen – phasage de l'attaque et de la voyelle, déphasage du [p]	psə
əps	Faible – déphasage de la voyelle et des consonnes, phasage dans la coda	psə
spə	Moyen – phasage de l'attaque et de la voyelle, déphasage dans l'attaque	pəs/ spə
pəs	Moyen – phasage de l'attaque et de la voyelle, déphasage du [s]	pəs/ spə
əsp	Très Faible- déphasage de la voyelle et des consonnes, déphasage dans la coda	pəs/spə

La table 1 montre notre classement de ces séquences pour les deux groupes selon leur degré de phasage : c'est-à-dire le degré de synchronisation entre les différentes commandes consonantiques et vocaliques. Nous considérons ce degré de phasage, conformément à nos hypothèses, comme le facteur déterminant dans la transformation ou la non-transformation obtenue après répétition de chacune d'elles. Ainsi, plus ce degré de phasage est important et plus la transformation de la séquence devrait être difficile. Inversement, plus ce degré de phasage est faible et plus la transformation vers une séquence mieux phasée devrait être obtenue.

3.2 Méthode

56 sujets, étudiants en 1^{ère} année de D.E.U.G de Sciences du Langage, participèrent à cette expérience (51 filles et 5 garçons). Ces sujets se présentèrent individuellement au test sans avoir été au préalable renseignés sur l'expérience.

Après lecture de la consigne par le sujet, il était donné oralement l'exemple de transformation verbale : "life" ⇔ "fly" afin de clarifier la procédure à suivre en indiquant une estimation de la vitesse de répétition attendue (environ 2 répétitions/seconde). Après chaque présentation orale de chacune des 6 séquences - dont l'ordre de présentation différerait de manière aléatoire d'un sujet à l'autre - il était demandé au sujet de répéter la séquence rapidement et à haute voix. S'il trouvait une transformation, il s'arrêtait et la signalait à l'expérimentateur ; dans le cas contraire, l'examinateur l'arrêtait au bout d'une trentaine de secondes. Bien entendu, il n'était proposé aucune indication quant aux transformations envisagées. La passation de l'ensemble des sujets fut enregistrée sur bande magnétique.

3.3 Résultats

La table 2 indique les fréquences de transformations observées sur les 56 sujets et pour les 6 séquences, le maintien des séquences étant représenté en diagonale.

On constate que la tâche de parsing glissant a bien été effectuée dans 86% des cas. Les transformations observées d'un groupe vers l'autre sont extrêmement faibles : en moyenne 0.60% pour

les séquences transformées du groupe 1 vers le groupe 2 et 4.76% pour celles du groupe 2 vers le groupe 1.

Table 2 : Fréquence des transformations / non-transformations observées (en pourcentage) .

Séq.	→psə	→səp	→əps	→spə	→pəs	→əsp	→div.
psə	74.99	17.85		01.79			05.35
səp	49.99	30.35	01.78				17.85
əps	28.57		64.29				07.14
spə	03.57			39.28	46.42		10.71
pəs				42.85	41.06		16.06
əsp	07.14		03.57	05.35	19.64	55.35	08.93

La colonne " div. " représente le pourcentage de transformations non-envisagées. Quelques-unes de ces transformations étaient d'ordre sémantique (comme " ça se peut "), cependant la plupart correspondaient à une séquence syllabique de longueur supérieure (comme [psəp]). Le nombre de ces transformations, bien que non négligeable, reste cependant faible (en moyenne 11 % par séquence).

Nous avons procédé à différentes analyses statistiques concernant l'effet d'ordre de présentation des séquences, l'effet de stimulus global, l'effet groupe et enfin les effets de transformations à l'intérieur de chaque groupe. Une méthode statistique de type ANOVA n'étant pas envisageable du fait du caractère booléen des résultats de transformation (0/1), nous avons utilisé la méthode de calcul du Chi-2. Les résultats de ces analyses sont les suivants :

- Effet d'ordre global non significatif ($p < 0.6$).
- Effet de stimulus global très significatif ($p > 0.999$).
- Effet de groupe significatif ($p > 0.95$).
- Effet de stimulus significatif pour le groupe 1 ($p > 0.95$) et non significatif pour le groupe 2 ($p < 0.6$) : ces résultats indiquent donc un effet de transformation/non-transformation non-aléatoire pour les séquences du groupe 1, le résultat pour le groupe 2 s'expliquant par les résultats réciproques des stimuli [pəs] et [spə].
- Effet d'ordre de transformation préférentiel pour chaque stimulus significatif ($p > 0.95$).

Nous pouvons finalement émettre les commentaires suivants sur les transformations observées dans chaque groupe :

- Groupe 1 : La séquence [psə], supposée très stable, n'entraîne pour 75% des sujets aucune transformation, 18% la transformant en [səp] et aucun en [əps]. En revanche, [səp] est attirée pour 50% des sujets vers [psə], de même que [əps] pour 29%. Il est à noter que les transformations [səp] vers [əps] et [əps] vers [səp] sont quasi-inexistantes (2% et 0%). Le résultat inattendu est le maintien de [əps] pour 64% des sujets.
- Groupe 2 : Les séquences [pəs] et [spə], supposées moyennement stables, entraînent des transformations réciproques : 46% pour [spə] et 43% pour [pəs]. Ces séquences se maintiennent respectivement à 39% et 41%. La séquence [əsp] se transforme en [pəs] pour 20% des sujets et en [spə] pour 5%, les transformations inverses n'existant pas. Comme

pour la séquence [əps] du groupe 1, cette séquence se maintient de façon inattendue pour 55% des sujets.

En ce qui concerne les syllabes à attaque vocalique, que nous avons prédites comme très instables, l'effet majoritaire de non-transformation s'explique par la présence d'une attaque glottique (confirmée *a posteriori* par une expertise phonétique des enregistrements) qui empêche les sujets d'enchaîner rapidement les séquences et a pour conséquence de bloquer la mise en phase des articulateurs [Dej01]. Cependant, dès que les sujets abandonnent cette attaque glottique - qui peut être considérée comme une consonne et de fait transforme les syllabes VCC [əsp] et [əsp] en syllabes CVCC (cf. en français, " l'un " vs. " le un " traité comme " le hun " ou encore " onze " difficile à transformer en " zon ") - la transformation se produit en faveur de la séquence la mieux phasée de toutes, [psə] pour le groupe 1, ou la plus attractive, [pəs] pour le groupe 2.

Ainsi, globalement, le patron des transformations/non-transformations observées est compatible avec nos prédictions :

- Dans le groupe 1, nos prédictions sont validées avec [psə] > [səp] > [əps] dans l'ordre des attractions vers [psə]. Pour les stabilités, [psə] est le grand gagnant si on ajoute à son maintien ses captures. Mais [əps] n'est pas le dernier aussi longtemps que l'attaque glottale en fait une syllabe articuloire CVCC [ʔəps].
- Dans le groupe 2, même observation pour le maintien de la syllabe [əsp]. Si [pəs] et [spə] sont bien aussi moyens en termes de phasage ([pəs] = [spə] > [əsp]), nous n'avions pas prévu que des deux, [pəs] serait tout de même plus attractif.

4. DISCUSSION

Est-il possible d'expliquer nos résultats par des contraintes linguistiques du type universaux syllabiques ou fréquences lexicales en français? En fait, nous ne pouvons clairement pas tirer d'arguments de notre étude en faveur de tendances typologiques (CV préférées à CVC, elles-mêmes préférées à CCV...) puisque notre choix de syllabes était restreint aux structures CVC, CCV et VCC. Nous pouvons cependant rappeler que les syllabes VCC et CCV sont très peu fréquentes dans les inventaires phonologiques et qu'après CV, c'est la syllabe CVC qui est bien la plus fréquente (Base ULSID [Val00] : CV=54.48% - CVC=36.15% - CCV=0.5% - VCC=0.05%). Si nos résultats sont en faveur de l'évitement de syllabes à début vocalique (l'universelle CV ne passe pas à VC), ils ne sont pour autant pas en faveur de la stabilité des CVC puisque ces dernières peuvent se transformer en CCV (si la langue - comme ici le français - n'interdit pas une telle structure).

Une interprétation concurrente de nos résultats pourrait s'appuyer sur la fréquence des structures syllabiques dans le lexique de la langue des sujets (ici le français), fréquence pouvant faire intervenir un effet de transformation préférentielle du point de vue lexical. Après étude de la fréquence des séquences de type [psə] et [spə], présentes en début ou fin de mot (position initiale ou finale), relativement à l'examen de trois dictionnaires phonétisés de langue française [Aub88], il s'avère que les fréquences trouvées pour la structure de type [psə] sont, à l'exception du dictionnaire DELA pour la position finale, toujours inférieures à celles correspondant à la structure de type [spə] quels que soient le dictionnaire et la position de la séquence dans le mot.

Comment expliquer alors la faible fréquence lexicale de la séquence [psə] ? Notre interprétation est que cette séquence, pourtant mieux phasée, est cependant défavorisée du point de vue de la faible saillance perceptive du [p] devant le [s].

5. CONCLUSION

En nous posant la question du contrôle articulatoire du phasage des composantes de la syllabe, dans le cadre du paradigme de transformation verbale, nous avons pu démontrer l'influence de ces contraintes de production dans la mémoire de travail verbale. A notre connaissance, c'est la première fois que cet effet de phasage est mis en évidence, ce que n'avaient pas réussi les études sur l'observation des coordinations interarticulateurs en fonction de la vitesse de la tâche d'articulation. A notre avis, cette tâche d'imagerie auditive, "à la Necker", avec décision sur stimulus tient de l'ambiguïté maximale de ce stimulus, rendue possible par le processus de répétition. Il ne s'agit plus alors de détecter des traces de jointure entre les syllabes itérées mais de changer l'état perceptif du sujet par une perception auditive active [Gom95]-[Sch97]. Le fait qu'il existe un biais d'asymétrie (il en existe un aussi dans le biais de perspective du cube de Necker) tient pour nous à l'influence du contrôle des phases interarticulateurs, avec un régime de contrôle moteur préférentiellement en phase.

Le fait que notre résultat majeur - le succès du tout phasé [psə] - ne s'explique ni par des contraintes auditives, ni par des contraintes universelles ou développementales ou propres au français, est à nos yeux décisif pour une interprétation qui repose *exclusivement* sur le contrôle articulatoire du système de la boucle.

Remerciements : Nous remercions vivement Pascal Perrier, Anne Vilain, Laurent Dubet et les étudiants du DEUG Sciences du Langage de l'Université Stendhal pour leur contribution à la réalisation de cette étude.

REFERENCES

- [Aub88] Auberge V., Boë L.J. & Lefevre J.P. (1988), "Lexiques et groupes consonantiques", XVIIèmes Journées d'Etude sur la Parole – Nancy, pp. 55-60.
- [Bad86] Baddeley A.D. (1986), "*Working Memory*", Oxford, Clarendon Press.
- [Bad90] Baddeley A.D. (1990), "*La Mémoire Humaine, théorie et pratique*", Presses Universitaire de Grenoble.
- [Bad98] Baddeley A.D., Gathercole S.E. & Papagno C. (1998), "The phonological loop as a language learning device", *Psychological Review*, vol. 105, pp. 158-173.
- [Bro00] Browman C.P. & Goldstein L. (2000), "Competing constraints on intergestural coordination and self-organization of phonological structures", *Bulletin de la Communication Parlée*, vol. 5, pp. 25-34.
- [Cha85] Chambers D. & Reisberg D. (1985), "Can mental images be ambiguous ?", *Journal of Experimental Psychology : Human Perception and Performance*, vol. 11, pp. 317-328
- [Con01] Content A., Kearns K.R. & Frauenfelder H.U. (2001), "Boundaries versus onsets in syllabic segmentation", *Journal of Memory and Language*, vol. 45, pp. 177-199.
- [Dej01] De Jong K.J. (2001), "Rate-induced resyllabification revisited", *Language and Speech*, vol. 44, pp. 197-216.
- [Dum98] Dumay N., Banel M.H., Frauenfelder U. & Content A. (1998), "Le rôle de la syllabe: segmentation lexicale ou classification ?", XXIIèmes Journées d'Etudes sur la Parole – Martigny, pp. 33-36.
- [Gat93] Gathercole S.E. & Baddeley A.D. (1993), "*Working Memory and Language*", Hove - UK : Lawrence Erlbaum Associates Ltd.
- [Gom95] Gomez C., Argandona E.D., Solier R.G., Angulo J.C. & Vasquez M. (1995), "Timing and competition in networks representing ambiguous figures", *Brain and cognition*, vol. 29, pp. 103-114.
- [Grè01] Grèzes J. & Decety J. (2001), "Functional anatomy of execution, mental simulation, observation, and verb generation of actions : a meta-analysis", *Human Brain Mapping*, vol. 12, pp. 1-19.
- [Gup97] Gupta P. & MacWhinney B. (1997), "Vocabulary acquisition and verbal short-term memory : computational and neural bases", *Brain and Language*, vol. 59, pp. 267-333.
- [Hak85] Haken H., Kelso J.A.S. & Bunzt H. (1985), "A theoretical model of phase transitions in human hand movements", *Speech Research*, vol. 81, pp. 41-59.
- [Mec01] Mechsner F., Kerzel D., Knoblich G. & Prinz W. (2001), "Perceptual basis of bimanual coordination", *Nature*, vol. 414, pp. 69-72.
- [Rei89] Reisberg D., Smith J.D., Baxter A.D. & Sonenshine M. (1989), "Enacted auditory images are ambiguous; pure auditory images are not", *The Quarter Journal of Experimental Psychology*, vol. 41A, pp. 619-641.
- [Sch97] Schleidt M. & Kien J. (1997), "Segmentation in behavior and what it can tell us about brain function", *Human nature*, vol. 8, pp. 77-111.
- [Tre83] Treiman R. (1983), "The structure of spoken syllables : evidence from novel word games", *Cognition*, vol. 15, pp. 49-74.
- [Tul90] Tuller B. & Kelso J.A.S. (1990), "Phase transitions in speech production and their perceptual consequences", In *Motor Representation and Control, Attention and Performance XIII*, M. Jeannerod (Ed). Hove, UK : Lawrence Erlbaum Associates Ltd, pp. 429-452.
- [Val00] Vallee N., Boë L.J., Maddieson I. & Rousset I. (2000), "Des lexiques aux syllabes des langues du monde – Typologie et structures", XXIIIèmes Journées d'Etude sur la Parole – Aussois, pp. 93-96.
- [War61] Warren M.R. (1961), "Illusory changes of distinct speech upon repetition – the verbal transformation effect", *British Journal of Psychology*, vol. 52, pp. 249-258.