

Asymétries dans les transformations verbales : un corrélat perceptif de « l'effet LC »

Isabelle Rousset, Marc Sato, Jean-Luc Schwartz & Nathalie Vallée

Institut de la Communication Parlée – CNRS UMR 5009
Université Stendhal, BP 25 – 38040 Grenoble cedex 9, France
Mél : rousset, sato, schwartz, vallee@icp.inpg.fr

ABSTRACT

Since verbal transformations are mostly considered to reflect the sub-processes devoted to the perceptual organization and interpretation of speech, the present study was designed to further test this proposal by examining whether variations in the phonological structure of repeatedly presented words can lead to perceptual asymmetries in verbal transformations. For this aim, we contrasted reversible pairs of CV.CV disyllabic words, consisting in a combination of the /p/ labial and /t/ coronal consonants. The main organization of the observed transformations was always that of a pair wise coupling between the two possible forms of the reversible repeated stimuli. In addition, a perceptual preference for labial-coronal as compared to coronal-labial couple-based stimuli was observed, both in the number of transformations and in the perceptual stability duration. We interpret these findings in line with previous models of speech perception and statistical studies on universal organizational syllabic patterns.

INTRODUCTION

L'effet de transformations verbales (dorénavant TV), décrit en premier lieu par [22], apparaît lors de l'écoute d'un stimulus auditif présenté de manière répétitive et continue. Au début du test, le sujet perçoit une forme identique au stimulus d'entrée, mais au bout d'un certain temps, des mots ou des pseudo-mots illusoirement émergent, correspondant à un changement perceptif du sujet. Des études antérieures ont montré que les modifications perceptives, par comparaison avec le percept de départ, peuvent aussi bien consister en de petites déviations phonétiques qu'en de fortes distorsions sémantiques, incluant la substitution d'un phonème par un phonème proche [23-25], un regroupement auditivo-perceptif (streaming) [14-15], des transformations lexicales et sémantiques [6, 20, 23-24]. Ces TV sont généralement considérées comme le reflet de processus liés à l'organisation perceptuelle et donc à l'interprétation de la parole. Deux fonctions semblent être impliquées dans la ré-interprétation du signal auditif : celles de satiété et de changement de critère [6, 9, 25]. La présentation répétée d'un stimulus entraîne la satiété des représentations mentales correspondantes. Simultanément, le critère de catégorisation du stimulus change brusquement et une nouvelle représentation est alors élaborée.

Les TV semblent varier en fonction de différents facteurs liés au stimulus répété : densité du voisinage lexical, fréquence et statut lexical du mot. [9, 26] ont montré qu'un grand nombre de voisins lexicaux entraînait un plus grand nombre et une plus grande étendue des transformations observées. De plus, le nombre de formes perçues est moins important pour les mots que pour les pseudo-mots [13, 20] (effet du statut

lexical). [20] ont également démontré que cette différence était surtout due à la stabilité du percept original : la stabilité de la forme perçue est liée à un mécanisme de rétablissement qui apparaît être meilleur pour les mots que pour les pseudo-mots et pour les pseudo-mots phonologiquement permis que pour les pseudo-mots illégaux. Ce mécanisme de rétablissement apparaît en conformité avec la dynamique des transformations, précédemment observée par [2] qui signale l'importance d'un mécanisme de couplage par paires des TV : bien que différentes solutions alternatives émergent, les perceptions de tous les sujets sont dominées par seulement deux formes, incorporant toujours le percept d'origine.

Nous nous proposons d'explorer, pendant une tâche de TV, la validité perceptive d'un modèle organisationnel spécifique du langage, appelé "l'effet LC" [12]. Fondé à partir d'études statistiques, cet effet décrit une tendance forte dans l'organisation syllabique des lexiques des langues à préférer la structure consonantique labial-coronal plutôt que coronal-labial [11-12, 17]. Précisons que l'effet LC, bien qu'absent du babillage, est remarqué dès l'acquisition des premiers mots par les enfants [1, 11-12].

Nous avons sélectionnés trois paires réversibles de mots dissyllabiques. Chaque paire consiste en une combinaison de consonnes, labiale avec /p/ et coronale avec /t/, dans les deux positions consonantiques de la structure dissyllabique, avec l'une des trois voyelles /i a o/. Les facteurs qui pourraient potentiellement affecter la stabilité du percept ont été contrôlés de manière très précise afin d'être égalisés au mieux pour tous les stimuli : nombre et type de phonèmes, statut lexical et fréquence des mots, caractéristiques acoustiques (fréquence, durée, intensité).

Nous formulons alors l'hypothèse que l'organisation principale des transformations observées doit être un couplage par paires entre les deux formes possibles de chacun des 6 stimuli réversibles répétés. Si tel est le cas, les asymétries perceptives entre deux formes phonétiques associées, soit [pV. tV] et [tV. pV], doivent permettre de tester la persistance perceptive de l'effet LC dans une tâche de TV.

METHODE

Matériel

Trois paires de mots dissyllabiques et réversibles CV.CV ont été sélectionnés à partir de *VoCoLex*, base de données lexicales du français [3]. Pour chacune d'entre elles, le patron consonantique est /p/ devant /t/ ou /t/ devant /p/, tandis que la voyelle est fixe et choisie parmi /i a o/. Les trois paires de mots réversibles alors obtenues (c'est-à-dire /pa.ta/-/ta.pa/, /pi.ti/-/ti.pi/, /po.to/-/to.po/) ont une fréquence d'usage similaire et basse. La valeur de la densité du voisinage lexical

est de 64 pour les stimuli /pV.tV/ et de 51 pour les stimuli /tV.pV/. La fréquence lexicale (définie comme le nombre d'entrées lexicales possédant une structure CV.CV dissyllabique identique à celle du stimulus à n'importe quelle position dans le mot) est de 19 pour les stimuli /pV.tV/ et de 15 pour les stimuli /tV.pV/.

Six séquences CV, /pa/, /ta/, /po/, /to/, /pi/, /ti/, ont été enregistrées individuellement puis numérisées sur 16-bit à une fréquence d'échantillonnage de 48 kHz. Toutes ont été contrôlées pour être égalisées au niveau des caractéristiques du signal acoustique. Trois paires de formes dissyllabiques CV.CV ont ainsi pu être construites en juxtaposant dans un fichier son unique deux séquences CV appropriées avec un intervalle inter-stimuli [ISI] de 100 msec. La durée moyenne des séquences CV.CV est de 520 msec (min : 503msec ; max : 531 msec).

Procédure

Vingt-quatre étudiants naïfs, locuteurs natifs du français, sans problèmes auditifs ou articulatoires particuliers, ont participé à l'expérience. Avant le test, chaque sujet s'est vu expliquer par l'expérimentateur la tâche de TV. Il leur était précisé qu'ils allaient entendre, à travers un casque audio, un stimulus auditif répété en boucle plusieurs fois et qu'il leur était demandé d'écouter attentivement chacun des changements de l'énoncé répété. La consigne leur était donnée de signaler tout changement perçu du stimulus en prononçant clairement le nouveau percept dans un microphone placé devant eux, même si le changement correspond à une forme déjà précédemment entendue. Il était aussi indiqué aux sujets que ces changements pouvaient être subtils ou très importants et correspondre à des mots aussi bien qu'à des pseudo-mots. Pour clore cette étape préliminaire, les sujets ont été assurés qu'il n'y avait pas de bonnes ou mauvaises réponses et que s'ils n'entendaient aucune transformation, ils ne devaient rien dire. S'ensuivait un essai pratique pour s'assurer que le sujet avait bien compris la tâche, ainsi que pour vérifier les niveaux d'écoute et d'enregistrement. Dans la session même du test, les 6 stimuli étaient répétés chacun 300 fois. Des pauses ont été accordées aux sujets entre chaque écoute. La durée totale de l'expérience était de 30 minutes.

RÉSULTATS

Les formes perçues recueillies ont été extraites pour chaque sujet et chaque stimulus avec leur nombre respectif de transformations. La durée de stabilité perceptive d'une forme est calculée en additionnant le laps de temps de perception de cette forme avant le passage à une autre. Nos analyses testent les différences à la fois entre chaque paire de stimuli (ce que nous avons appelé *l'effet stimulus*, c'est-à-dire dépendant du type de voyelle /i/, /a/ ou /o/) et entre chaque forme dissyllabique possible (*effet d'amorçage*, c'est-à-dire dépendant de l'ordre des consonnes : /pV.tV/ vs. /tV.pV/).

Analyse globale

Nombre de transformations : Le nombre moyen de transformations est porté dans la table 1. Une ANOVA à double-facteurs à mesures répétées montre un effet significatif de l'effet d'amorçage [$F(1,23) = 4,61$, $MSe = 184,51$, $p < 0,05$], avec les stimuli /pV.tV/ montrant en

moyenne 2,26 fois plus de transformations que les stimuli inverses. Les conséquences de l'effet stimulus sont marginales sur le nombre de formes modifiées, avec un nombre de transformations variant entre 47,91 et 55,63 pour les trois paires de stimuli réversibles. L'interaction des deux conditions n'est pas significative.

Nombre de formes : L'analyse statistique ne montre aucun effet significatif ni de l'effet stimulus, ni de l'effet d'amorçage, ni de leur interaction.

Table 1. Nombre moyen de transformations et de formes uniques observées en fonction du stimulus (avec écart-types).

Stimulus	Transformations		Formes	
	M	ET	M	ET
/pa.ta/	26,58	16,11	3,83	2,35
/ta.pa/	22,67	14,06	4,21	1,93
/pi.ti/	24,04	12,26	4,33	1,97
/ti.pi/	23,87	13,21	4,29	2,14
/po.to/	29,17	16,45	5,04	3,62
/to.po/	26,46	12,60	4,79	2,78

Analyse des formes avec couplage par paires

Un niveau supplémentaire d'analyse statistique a été appliqué sur nos données afin de tester de possibles groupes perceptifs à travers les transformations observées. De manière à obtenir une mesure plus précise de la stabilité perceptive des stimuli, toutes les transformations ont été classées en fonction de leur adéquation au stimulus de départ (*transformation vers le percept de base*), à la forme inverse du stimulus (*transformation vers la forme associée*) ou à une autre transformation (*autre transformation*).

De manière générale, la distribution des transformations observées démontre que l'organisation principale des transformations perceptives est celle d'un couplage par paires entre la forme de base de chaque stimulus et la forme couplée : en moyenne 71 % des transformations observées. Le calcul des ANOVA double-facteurs à mesures répétées avec le nombre type de transformations comme variable dépendante (c'est-à-dire le nombre de transformations vers la forme de base, vers la forme couplée, ou vers une autre transformation) montre un effet significatif de l'effet d'amorçage pour les transformations vers le percept de base, [$F(1,23) = 9,82$, $MSe = 169,00$, $p < 0,01$], avec les stimuli /pV.tV/ qui reviennent vers leur forme de base plus souvent que les stimuli /tV.pV/ (en moyenne 2,17 fois plus). Ces conséquences de *l'effet stimulus* sont marginalement significatives [$F(2,46) = 2,60$, $MSe = 39,17$, $p < 0,09$] tandis que l'interaction entre les deux conditions ne l'est pas. Cette analyse explique ainsi en grande partie l'effet d'amorçage observé dans l'analyse globale des transformations (avec les stimuli /pV.tV/ ayant en moyenne 2,26 fois plus de transformations que les stimuli /tV.pV/). De plus, les deux analyses statistiques des transformations vers la forme couplée et vers une autre forme ne montrent pas d'effet significatif pour l'effet stimulus, ni pour l'effet d'amorçage.

Analyse des formes avec les structures syllabiques préférentielles

Afin d'évaluer la validité perceptive de l'effet LC, nous avons comparé la durée de stabilité perceptive des formes observées. Toutes les transformations vers la forme de base

ou vers la forme couplée observée sont alors classées en fonction de leur correspondance à une succession consonantique Labial-Coronal ou Coronal-Labial (i.e., de types /pV.tV/ ou /tV.pV/). Une préférence perceptive Labial-Coronal attesterait alors de l'existence de l'effet LC dans une tâche de transformation TV verbale.

Quels que soient les stimuli, le nombre de transformations /pV.tV/ est toujours plus important que celui des transformations /tV.pV/ (voir les deux colonnes "forme" de la table 2). Un test *t* unilatéral sur l'ensemble des résultats des transformations (/pV.tV/ - /tV.pV/) montre une asymétrie significative entre les deux structures phonologiques [$t(1,143) = 3,40, p < 0,001$], avec en moyenne 1,40 fois plus de transformations vers /pV.tV/ que vers /tV.pV/. Afin d'étudier davantage cette asymétrie perceptive, nous avons alors comparé la durée de stabilité perceptive des formes dominantes observées. Le nombre des transformations ainsi que la nature et la force du mécanisme de couplage par paires entre les deux formes possibles pour chaque stimulus réversible ont aussi été analysés, leur durée de stabilité perceptive représentant en moyenne 69 % de la durée totale de l'expérience (cf. table 2). La convergence de ces deux mesures distinctes sur l'effet de TV est attestée par une corrélation positive forte entre le nombre de transformations observées, liées à leur statut vers la forme de base ou vers la forme couplée, et leur durée de stabilité perceptive respective : 0,60 et 0,66.

Quels que soient les stimuli, la durée de stabilité perceptive des transformations /pV.tV/ est toujours plus importante que celle des transformations /tV.pV/. Une ANOVA à double-facteurs à mesures répétées ne montre aucun effet significatif ni de l'effet stimulus, ni de l'effet d'amorçage, ni de l'interaction entre les deux conditions. Un test *t* unilatéral montre une asymétrie significative entre les deux formes (LC vs. CL) [$t(1,143) = 3,73, p < 0,005$], avec une durée de stabilité moyenne plus longue de 13,24 s. pour les transformations /pV.tV/. Une étude plus fine de la table 2 indique que les taux de transformations (/pV.tV/ - /tV.pV/) sont tout à fait différents si on considère l'effet stimulus. En effet, la différence est importante pour /a/ (17, 94 s.) et /o/ (20,51 s.) et faible pour /i/ (1,26 s.). La significativité de ces différences dans chacun des trois contextes, en appliquant un test *t* unilatéral avec correction de *Bonferroni* mérite d'être testée. Les résultats montrent un effet significatif pour /a/ et /o/, mais aucun effet dans le cas de /i/.

DISCUSSION

Un premier résultat de cette expérience est celui de la dynamique observée des changements perceptifs. Nos données montrent que, pour tous les sujets, l'organisation principale des transformations est toujours celle d'un couplage par paires entre les deux formes possibles de chaque stimulus réversible (c'est-à-dire quand les transformations correspondent au stimulus de départ ou à sa forme syllabique inversée) : en moyenne 71 % du nombre total de transformations et 69 % de la durée totale du temps de l'expérience. Ce mécanisme de couplage par paires est pratiquement similaire à celui précédemment souligné dans [2] qui montrent que, malgré l'observation d'un certain nombre d'alternatives différentes, toutes les perceptions des sujets sont dominées par seulement deux formes, comprenant toujours la forme de base. Par conséquent, les deux études convergent vers un mécanisme de couplage par paires impliquant toujours le percept de départ. Il est intéressant de noter que ce mécanisme de couplage observé entre le stimulus de base et sa forme inversée est en accord avec plusieurs théories phonologiques de syllabation (pour un résumé, voir [4]), telles que le « Maximum Onset Principle » [5], ou le « Obligatory Contour Principle » [7]. Ce mécanisme de couplage par paires est intéressant pour valider la mesure de la durée de stabilité perceptive. En effet, nous observons une corrélation positive entre le nombre de transitions et la durée de la stabilité perceptive des formes dominantes.

Dans notre expérience l'analyse des transformations et de la durée de stabilité perceptive livrent une corrélation de l'effet LC. En effet, quel que soit le stimulus ou quel que soit le paramètre mesuré, la préférence perceptive pour les formes /pV.tV/ par rapport aux formes /tV.pV/ est nettement observée. Ce biais asymétrique ne peut pas être expliqué par des facteurs tels que la densité du voisinage lexical ou la fréquence lexicale des stimuli. Comme souligné précédemment, un grand nombre de voisins devrait augmenter le nombre de candidats lexicalement possible, ce qui entraînerait un plus grand nombre et une plus grande étendue des transformations [9, 26]. Étant donné l'importance de la valeur de la densité du voisinage lexical moyenne pour les stimuli /pV.tV/ par rapport aux stimuli /tV.pV/ (ratio de 1,16), l'effet de la densité du voisinage lexical pourrait paraître en accord avec le plus grand nombre de transformations relevées pour les stimuli /pV.tV/.

Table 2 : Nombres moyens et durées de stabilité des transformations liées au stimulus, vers la forme associée ou une autre en fonction du stimulus.

Stimulus	Type de transformations					
	Transformations liées au stimulus		Transformations vers la forme associées		Autres transformations	
	Nombre	Stabilité	Nombre	Stabilité	Nombre	Stabilité
/ba.ta/	9,96	62,88	8,04	39,95	8,58	48,04
/ta.pa/	7,42	43,39	8,42	56,34	6,83	50,64
/pi.ti/	9,79	57,20	8,17	56,65	6,08	44,31
/ti.pi/	8,04	50,87	8,83	52,85	7,00	54,11
/po.to/	11,46	67,83	8,87	46,37	8,83	43,70
/to.po/	9,25	46,41	9,75	65,97	7,46	45,52

Cependant, l'importance des transformations pour les stimuli /pV.tV/ est principalement due à la tendance pour ces stimuli de "switcher" vers leur forme de départ qui n'apparaît pas alors comme un mécanisme compétitif. De plus, l'effet de la densité du voisinage lexical ne peut expliquer la préférence perceptive pour les formes /pV.tV/ par rapport aux formes /tV.pV/. Donc, la densité du voisinage lexical ne fournit vraisemblablement pas l'explication des asymétries présentes. Pas plus que l'interprétation concurrente qui viendrait de la fréquence lexicale des stimuli CV.CV dans le lexique des sujets. En effet, cette fréquence lexicale pourrait biaiser les transformations vers des entrées lexicales plus fréquentes. Pourtant, nous observons un plus grand nombre d'entrées lexicales pour /to.po/ que pour /po.to/, un résultat qui n'est pas en accord avec les préférences perceptives observées dans les transformations de /po.to/ (ratio de 0,50).

Rappelons que « l'effet LC » est une tendance dans l'organisation phonologique et lexicale des langues à favoriser un ordre consonantique dans les structures CV.CV : Labial-Coronal plutôt que l'inverse Coronal-Labial, effet que l'on retrouve en production au stade de l'acquisition des premiers mots [1, 11-12, 17]. Ces différentes études ont supposé l'effet LC comme la conséquence de propriétés articulatoires de l'appareil de parole : les consonnes labiales résulteraient d'un mouvement basique de la mandibule, activée seule, alors qu'un mouvement additionnel de la langue est nécessaire à la recherche d'une position avant de la langue pour une consonne coronale. Comment alors réconcilier cette explication articulatoire et la validité perceptive de l'effet LC dans une tâche de transformation verbale ? Une des solutions envisageables pourrait provenir de théories sur la perception du langage qui, bien que discutées pour certaines, postulent une base de représentations articulatoires ou gestuelles [8, 10, 19]. Selon ces théories, la plus forte cohésion articulatoire des structures syllabiques LabialV-CoronalV, en comparaison aux formes syllabiques CoronalV-LabialV, peut expliquer la stabilité/attractivité perceptive plus importante des formes /pV.tV/ et ainsi valider perceptivement l'effet LC dans la tâche de transformation verbale. En d'autres mots, les contraintes de stabilité provenant du système moteur pourraient participer à la représentation mentale des formes phonologiques pendant l'écoute répétée de parole. Étant connu que les TV apparaissent aussi bien lors de la perception que lors de la production répétée d'unités de la parole [16, 18, 21], l'effet de transformation verbale pourrait alors être un élément central d'évaluation de cette proposition et donc d'investigation sur des propriétés structurales du système langagier.

REMERCIEMENTS

Nous remercions Mark Pitt pour ses commentaires utiles à propos de l'effet de transformation verbale, Barbara Davis, Peter MacNeilage et Christian Abry pour les discussions sur l'effet LC. Nous remercions aussi Alain Arnal et Christophe Savariaux pour leur aide précieuse dans l'élaboration des stimuli.

BIBLIOGRAPHIE

- [1] B.L. Davis & P.F. MacNeilage, In Proceedings of the *XVth International Congress of Phonetic Sciences*, Barcelona, 387-390, 2003.
- [2] T. Ditzinger, B. Tuller, B. & J.A.S. Kelso, *Biological Cybernetics*, 77: 23-30, 1997.
- [3] S. Dufour, R. Peereman, C. Pallier & M. Radeau, *L'Année Psychologique*, 102: 725-746, 2002.
- [4] J. Goslin & U.H. Frauenfelder, *Language and Speech*, 44(4): 409-436, 2000.
- [5] J.B. Hooper, *Language*, 48: 525-540, 1972.
- [6] Z. Kaminska, M. Pool & P. Mayer, *Brain and Language*, 71: 285-298, 2000.
- [7] M. Kenstowicz, *Phonology in Generative Grammar*. Blackwell, Oxford Press, 1994.
- [8] A.M. Liberman & I.G. Mattingly, *Cognition*, 21: 1-36, 1985.
- [9] D.G. MacKay, G. Wulf, C. Yin & L. Abrams, *Journal of Memory and Language*, 32: 624-646, 1993.
- [10] P.F. MacNeilage, *Behavioural and Brain Sciences*, 21: 499-511, 1998.
- [11] P.F. MacNeilage, B.L. Davis, C.M. Matyear & A. Kinney, *Psychological Science*, 310(5): 459-460, 2000.
- [12] P.F. MacNeilage & B.L. Davis, *Science*, 288: 527-531, 2000.
- [13] T. Natsoulas, *Journal of Psychology*, 78: 257-263, 1965.
- [14] M. Pitt & L. Shoaf, *Language and Cognitive Processes*, 16(5/6): 715-721, 2001.
- [15] M. Pitt & L. Shoaf, *Journal of Experimental Psychology: Human Perception and Performance*, 28: 150-162, 2002.
- [16] D. Reisberg, J.D. Smith, A.D. Baxter & M. Sonenshine, *Quarterly Journal of Experimental Psychology*, 41A: 619-641, 1989.
- [17] I. Rousset, In Proceedings of the *XVth International Congress of Phonetic Sciences*, Barcelona, 715-718, 2003.
- [18] M. Sato, & J-L. Schwartz, In Proceedings of the *XVth International Congress of Phonetic Sciences*, Barcelona, 435-438, 2003.
- [19] J-L. Schwartz, C. Abry, L-J. Boë & M.A. Cathiard, In Durand J. & Lacks B. (Editors.), *Phonology: from Phonetics to Cognition*, Oxford University Press, 2002.
- [20] L. Shoaf & M. Pitt, *Perception and Psychophysics*, 64(5): 795-803, 2002.
- [21] J.D. Smith, D. Reisberg & M. Wilson, *Neuropsychologia*, 11: 1433-1454, 1995.
- [22] M.R. Warren & R.L. Gregory, *American Journal of Psychology*, 71: 612-613, 1958.
- [23] M.R. Warren, *British Journal of Psychology*, 52: 249-258, 1961.
- [24] M.R. Warren, *Auditory Perception*. Pergamon Press, New York, 1982.
- [25] M.R. Warren & D.M. Meyers, *Journal of Phonetics*, 15: 169-181, 1987.
- [26] C. Yin & D.G. MacKay, *IVth Biennial Cognitive Aging Conference*, Atlanta, 1992.