

LA TRANSPARENCE SONORE DU LEXIQUE : DE L'EXPERIMENTATION AU CALCUL D'UN INDICE

Marie-Christine Jamet (Université Cà Foscari-Venezia)

0. Introduction

Notre intervention se situe, non pas sur le versant didactique, mais sur le versant linguistique qui précède toute intervention de type didactique, pour définir l'objet de l'enseignement/apprentissage de l'intercompréhension orale. Explorant le champ de l'intercompréhension orale au sein du groupe des langues romanes et plus particulièrement du binôme français/italien, nous nous sommes demandé si les présupposés pour l'intercompréhension existaient à l'oral comme ils existent pour l'écrit. En effet, par rapport à l'écrit, le mot entendu à l'oral présente une difficulté supplémentaire liée à sa perception auditive qui n'est pas stable comme peut l'être sa perception visuelle. Nous avons donc mené une expérience auprès de locuteurs italophones – regroupés en deux ensembles : un groupe de 14 adultes et un groupe de 14 adolescents - n'ayant jamais appris le français, langue proche inconnue, et nous leur avons demandé s'ils comprenaient des mots français qu'ils entendaient pour la première fois. Ces mots, qui avaient tous une même origine que le mot italien, étaient donc congénères. Ils étaient prononcés hors contexte, inclus dans une liste, et les variations éventuelles de sens entre les deux langues ne nous intéressaient pas ici. Nous avons privilégié pour cette expérience exclusivement les signifiants acoustiques. Il est sûr qu'un mot reconnu isolément ne le sera pas nécessairement lorsqu'il est intégré dans un discours où l'ensemble des éléments à décoder peut brouiller la reconnaissance du mot singulier. Toutefois, si le mot n'est pas reconnu isolément et associé à un congénère, il sera rare qu'il le soit en contexte même s'il n'est pas exclu que sa compréhension puisse tout de même se faire par extrapolation du sens, sans que le mot ait été associé à un signifiant. Dans cette communication, nous exposerons les résultats de cette expérimentation qui nous permet de déterminer pour le couple de langues examinées (Italien LM et français LE) ce qui pose ou ne pose pas problème dans le décodage acoustique d'un congénère en LE proche inconnue. À partir de là, nous voulons tenter de calculer un indice de ressemblance

phonémique¹, c'est-à-dire un indice qui puisse prédire, pour des mots qui sont congénères leur transparence formelle sonore.

1. Notre expérience et ses résultats

Nous avons sélectionné 200 mots congénères dans des documents authentiques radiophoniques² sur la base de plusieurs hypothèses :

- la longueur des mots : F.*cas*/I.*caso*³ ; F.*consommateur*/I.*consumatore*)
- la présence de phonèmes inexistantes en LM : F.*attention*/I.*attenzione*),
- la quantité de phonèmes identiques ou différents entre les mots : F.*chanson*/I.*canzone* (aucun phonème commun) vs F.*psychiatre* /I.*psichiatra*
- le poids des différences portant sur les voyelles ou sur les consonnes : F.*façade*/I.*facciata* (deux consonnes divergentes sur trois) vs F.*communauté*/I.*comunità*.
- la variation sur la première syllabe qui empêcherait d'activer la cohorte des mots stockés en mémoire commençant par une même syllabe : F.*impact*/I.*impatto*
- l'influence de l'accentuation différente (F.*météo*/I.*meteo*, F.*psychique*/I.*psichico*).

Les adultes et adolescents testés devaient écrire le mot congénère italien qu'ils associaient au mot entendu. On sollicitait leur capacité d'association sur la base de deux stratégies :

1) la capacité à traiter les différences intra-mots comme s'il s'agissait de bruits qui brouilleraient l'audition du mot de la LM correspondant ; en LM en effet, un locuteur active des mécanismes d'inférences à partir des bribes entendues et reconstruit le mot ; donc, nous nous sommes demandé si la quantité et la position des éléments brouillés avaient des implications sur la reconnaissance du mot étranger ;

2) la capacité à entendre les éléments semblables ou similaires. Cette idée de similitude est importante. Il s'agit alors d'associer des phonèmes différents dans les deux langues mais qui possèdent quelques traits communs. On sait qu'un auditeur d'une LE inconnue entend mal et plusieurs études ont prouvé que les erreurs typiques de prononciation résultaient moins d'une incapacité articulatoire de l'apprenant que d'une incapacité à entendre des sons différents de ceux auxquels on est habitué (Tomatis 1991, Pallier & al, 1997) Or l'audition imparfaite peut

¹ Nous utilisons volontairement le terme de « phonémique », car nous raisonnons sur la prononciation standard des phonèmes, telle qu'elle est enregistrée dans un dictionnaire.

² Nous renvoyons à notre étude complète : Jamet M.-C.(2007). Le choix des mots a été opéré sur la base de nos hypothèses à l'intérieur des documents radiophoniques authentiques utilisés pour une seconde expérience. En effet, nous pouvions ainsi évaluer les différences entre audition/compréhension hors contexte et en contexte.

³ Dans les paires de mots congénères, le F majuscule indique qu'il s'agit du mot français et le I majuscule du mot italien.

paradoxalement servir l'intercompréhension. Si les italophones n'entendent pas la différence entre des phonèmes proches mais différents (par exemple /v/ à la place de /ψ/ ou bien /e/ à la place de /↔/), c'est que les frontières entre phonèmes ne sont pas identifiées puisque la reconnaissance passe par les habitus liés à la LM. Ce phénomène perceptif va rendre plus facile le décodage du mot étranger en réception. Par exemple, le mot *F.menu* entendu erronément [μevu] ne posera aucun problème même si 2 phonèmes sur 4 sont différents. Le mot *valeur* est entendu comme [βαλ□ρ] est donc facilement associable à *I.valore*, de même que *F.successeur* probablement entendu [συσεσ□ρ] est associé sans difficulté à *I.successore*, ne pose pas de difficultés (en fait il n'y a plus que le /s/ à rapprocher de l'affriquée /τΣ/).

Les résultats de l'expérience sont tout à fait favorables au principe de l'intercompréhension orale puisque nous avons obtenu des scores élevés de reconnaissance de mots congénères. Sur les 200 mots proposés, en moyenne 127 mots sont reconnus par adolescent et 141 par adulte⁴, c'est-à-dire 67% des mots congénères en moyenne par personne. Nous avons alors fait un classement des mots selon le nombre de personnes les reconnaissant, en définissant 4 classes de mots. En arrondissant les pourcentages, pour avoir un ordre de grandeur, on peut dire qu'un mot est transparent s'il est reconnu par au moins ¾ des personnes, un mot est semi-transparent quand il est reconnu par au moins la moitié des auditeurs, un mot est semi-opaque quand il est reconnu par au moins ¼ des personnes, et un mot est opaque quand il est reconnu par moins d'¼ des personnes. On obtient pour nos 200 mots congénères, 58,5% de mots transparents, 12% de mots semi transparents, 13% de mots semi-opaques et 15,5 % de mots opaques.

Beaucoup de mots sont donc transparents malgré des variations phonétiques. Nous avons pu alors confirmer nos hypothèses.

- Le nombre de phonèmes communs intervient significativement. C'est là une intuition commune vérifiée. En principe, le fait que la syllabe finale ne corresponde pas à celle du mot italien ne pose pas de problème, sauf dans le cas où le même mot pourrait avoir deux mots associés, comme *F.monstre*, l'un avec une finale masculine en -o- (*I.mostro*), et l'une avec une finale en -a- (*I.mostra*). En général, les auditeurs ont privilégié les finales en -a- après avoir intériorisé peut-être le fait qu'un phonème final consonantique en français est le signe le plus fréquent du féminin.
- Les mots longs sont de loin plus faciles à décoder, car même si certains phonèmes ou syllabes divergent, l'auditeur a plus d'éléments pour deviner ce qui manque. Ainsi, pour le groupe des adolescents, seulement 5,5% des mots de 4 syllabes et plus sont opaques, 18%

⁴ On dispose pour chaque informateur du nombre de mots qu'il a reconnus. On a simplement fait la moyenne pour les 14 informateurs ados et les 14 informateurs adultes.

des mots de 3 syllabes, 31% des mots de 2 syllabes et 36% des monosyllabes. Pour les adultes, la progression est la même : aucun mot de 4 syllabes ou plus n'est opaque, 12% des mots de 3 syllabes, 22% des mots de 2 syllabes et 31% des mots de 1 syllabe sont opaques.

- Les voyelles divergentes ne posent pas de problème si le phonème correspondant de la LE partage des traits phonétiques avec le phonème de LM. Ceci explique que les nasales ne sont pas toutes au même titre source de difficultés. Ainsi le mot F.*simple* est-il rapproché de I.*semplice* alors que le mot F.*enfance*/I.*infanzia* est opaque. Dans le premier cas, la voyelle orale sous-jacente à la nasale française est identique à la voyelle orale de l'italien, et le décodage s'opère, dans le 2^e cas, le rapprochement ne se fait pas parce que F./A/ est trop loin de I./ɪv/. De la même façon, les phonèmes français /ɔ/ ou /O/ sont intermédiaires entre /ɔ̃/-/o/ et /E/-/ε/. Dans les suffixes en particulier, l'association a été immédiate avec le suffixe -ore- correspondant italien: F.*acteur*, *erreur*, *rédacteur*, *successeur*, *valeur* sont transparents. Toutefois F.*heure*, qui aurait pu être associé à I.*ora* selon le même principe, ne l'a pas été, sans doute à cause de sa brièveté. Et le mot F.*peur*/I.*paura* a été opaque, à cause de la divergence vocalique sans aucun trait commun. Or on constate que les mots proposés comme associés contiennent pour la plupart un son proche : I.*per*, *pera*, *perla* (en passant probablement par l'anglais *pearl*), *perché*, *povero*, *potere*. Il y avait en effet peu de candidats de deux syllabes en italien commençant pas *por-*.
- On sait que les consonnes ont un poids plus important dans la reconnaissance des mots (Nespor, M., Peña, M., & Mehler, J. 2003). En tête de syllabe, la variation consonantique est problématique. Mais un certain nombre de variations consonantiques ne font pas obstacle à la reconnaissance. Ainsi, le groupe français /κτ/ est généralement assimilé à la géminée italienne : /ττ/ (F.*acteur*/I.*attore*) ou bien à /τζ/ (F.*action*/I.*azione*). Le groupe français /ks/ est associé à l'italien /τΣ/ ou /s/ (F.*successeur*/I.*successore*, F.*expression*/I.*espressione*, et /gz/ assimilé à /ζ/ (F.*exemple*/I.*esempio*). Le phonème simple /Z/ est associé à l'affriquée italienne correspondante /dZ/ : *génial*. Mais lorsqu'un /s/ français correspond à un /τΣ/, les résultats sont variables : F.*cinéma*, *société* sont transparents, mais F.*incidence*, *façade* sont semi-opaques (les mots sont aussi moins fréquents). Il semblerait que, généralement, ce qui relève des affriquées ou des géminées ne soient pas source de difficultés. De même les consonnes françaises correspondant à une consonne + semi-voyelle en italien ne créent pas d'obstacle (F.*question*/I.*questione*). Lorsqu'un /σ/ a disparu en français par rapport à l'italien, la longueur des mots jouera un rôle majeur : F.*tête* est semi-opaque et a été systématiquement associé à I.*tetto*.

- La première syllabe joue un rôle non négligeable. Les problèmes apparaissent lorsqu'il y a variation consonantique et vocalique en même temps : F.*chemin*/I.*camino* qui cumule du reste toutes les difficultés sur le mot), ou F.*jour*/I.*giorno*, mais dans un mot long, les mêmes éléments ne font plus obstacle : F.*journaliste* est transparent). Et pour les mots français commençant par un -e- épenthétique, celui-ci ne pose problème que lorsque le mot français a perdu le /σ/ : F.*école*/I.*scuola*, F.*étude*/I.*studio*. Même le mot F.*été*/I.*estate* est opaque bien que le -e- apparaisse aussi en italien. Mais dans F.*espagnol*, la présence du /σ/ rend la décodification immédiate (plus la symétrie de l'ensemble du mot).

En faisant une synthèse des différentes observations, nous avons pu constater que les mots opaques, d'un point de vue auditif, se sont révélés être :

- les monosyllabes formés d'une syllabe ouverte (F.*cas*/I.*caso* est opaque). Avec une syllabe fermée, souvent la 2^e consonne permet le rapprochement, par exemple dans F.*hier*/I.*ieri* transparent) ;
- les monosyllabes avec une voyelle nasale (F.*temps* /I.*tempo*) ;
- les mots commençant par /Σ/ correspondant à /κ/ (F.*chose*/I.*cosa*) ;
- les mots commençant par /e/ ou /Eσ/ correspondant à /σκ/, /σπ/ en italien (F.*école*/I.*scuola*) ;
- les mots qui commencent par une syllabe inexistante en italien (F.*degré*/I.*grado*) ;
- les mots où des syllabes médianes ont disparu en français et pas en italien (F.*jeune*/I.*giovane*), les mots qui ont des équivalents phonétiques plus proches que le mot italien exact correspondant (F.*été* ≠ I.*età*). Ce sont les faux amis de l'oral.

2. Les ajustements de la transcription phonétique

À la lumière des résultats sur les facteurs de facilitation ou de difficulté dans la reconnaissance des mots singuliers, nous avons essayé d'élaborer un indice, qui permette d'avoir une mesure objective du degré de ressemblance phonétique entre un mot lexical français LE et un mot italien LM ayant la même origine latine, grecque ou internationale, et qui tienne compte de l'ensemble des facteurs, car il est rare qu'un d'entre eux soit impliqué seul. Pour le calcul de cet indice de ressemblance entre deux mots congénères, nous avons opéré un ajustement de la transcription phonémique aussi bien du mot français que du mot italien. En effet, nous avons voulu annuler les différences, pour lesquelles l'expérience a montré qu'elles ne posent pas de difficultés dans le décodage, en les rendant identiques dans les deux mots transcrits. Pour opérer les ajustements de la transcription,

nous avons donc tenu compte des éléments suivants, qui résultent de nos observations sur les tests de compréhension des mots et de nos connaissances phonologiques sur les mots.

Pour les mots italiens

On notera que ces simplifications correspondent à certaines prononciations régionales de l'italien.

- Annulation des géminées qui ne comptent que pour un seul phonème
 $/\mu\mu/ \rightarrow /μ/, /v\upsilon/ \rightarrow /v/, /\beta\beta/ \rightarrow /β/, /κκ/ \rightarrow /κ/, /δδ/ \rightarrow /δ/, etc.$
- Simplification des affriquées : $/\tau\Sigma/ \rightarrow /Σ/, /δZ/ \rightarrow /Z/.$
- Annulation de la voyelle finale après pénultième accentuée.
 $I.movimento \rightarrow /μo-πi-μEv-τ/,$ sauf si le mot français est plus long que le mot italien⁵.

Pour les mots français

- Pas de transcription du schwa en syllabe finale ou médiane:
 $F.mouvement : /μ\upsilon\pi-μA)^6 .$
- Substitution du trait de nasalité par la voyelle orale correspondante suivie de $/v/$:
 $/A/ \rightarrow /av/, /o/ \rightarrow /ov/, /E/ \rightarrow /E/, /ɔ/ \rightarrow /ɔv/$
 $F.annonce : /av\sigma/ \Rightarrow /avov\sigma/.$
- Substitution de $/\leftrightarrow/ \rightarrow /ε/$ et $/\psi/$ en $/v/$ en syllabe ouverte : $F. menu /μ\leftrightarrow\nu\psi/ \rightarrow /μ\varepsilon\nu\upsilon/$
- Elimination du trait d'aperture dans le mot français pour le rendre identique à celui de la voyelle correspondante en italien. $/E/ \leftrightarrow /ε/, /\square/ \leftrightarrow /o/$
 $F.communauté : /κ\square\mu\psi\nu\sigma\tau\varepsilon/ \rightarrow /κομψ\nu\sigma\tau\varepsilon/$ car : $I.comunità /κομ\upsilon\nu\iota\alpha/)$
 $F.erreur /E\{\downarrow\}/ \rightarrow /ε\{\downarrow\}/$ et $I.errare /ε\rho\rho\varepsilon/.$
- Substitution de $/\downarrow\{\}/ \rightarrow /op/$ sauf si le mot m est un monosyllabe.
- Substitution du phonème $/\xi/ \rightarrow /p/$ (la différence de prononciation n'est pas un problème pour la perception).
- Substitution de $/\zeta/ \rightarrow /σ/$ (car les deux phonèmes du français $/σ/$ et $/\zeta/$ sont deux allophones du même phonème en italien $/σ/.$
- Réduction des groupes $/κτ/ \rightarrow /τ/, /κσ/ \rightarrow /σ/, /γζ/ \rightarrow /ζ/ .$

En outre, il faudra tenir compte dans le calcul de l'indice, sous forme de coefficient, des facteurs suivants :

⁵ Ce faisant, nous sommes consciente d'avoir éliminé le problème des paires de mots dont la finale change le sens (*mostro* vs *mostra*).

⁶ C'est effectivement la façon dont nous avons lu les mots comportant un "e" dans une syllabe intermédiaire. Ainsi le mot *médecin* a été lu $/μ\varepsilon\delta\sigma E/$.

- la longueur des mots, car les mots longs sont plus faciles à reconnaître que les mots monosyllabiques ;
- l'importance majeure de la première syllabe dans la reconnaissance des mots et donc si la première syllabe n'est pas la même dans les deux langues, la reconnaissance du mot est rendue plus difficile ;
- le nombre de phonèmes identiques se suivant ;
- la place de l'accent tonique ;
- la possibilité pour un mot français d'être plus long que le mot italien, et donc plus difficilement reconnaissable.

3. Le calcul de l'indice de transparence sonore (T_m)

Tous les calculs seront faits sur la base des transcriptions phonémiques ajustées.

La transparence d'un mot français inconnu m par rapport à son congénère italien pourrait être définie par :

$$T_m = \frac{2I_c + I_v}{3} \times Q_m$$

- où I_c est l'indice de ressemblance consonantique (nombre de consonnes communes C_c sur le nombre de consonnes du mot italien C_i).

$$I_c = \frac{C_c}{C_i}$$

/j/ et /w/ comptent comme consonnes; l'indice est affecté ensuite du coefficient 2, pour lui donner plus de poids, vu que les consonnes sont plus importantes pour la reconnaissance des mots.

- où I_v est l'indice de ressemblance vocalique (nombre de voyelles communes V_c sur le nombre de voyelles du mot italien V_i :

$$I_v = \frac{V_c}{V_i}$$

- où le coefficient Q_m se définit comme :

$$Q_m = Q_1 [(Q_2 + Q_3)/2] Q_4 Q_5$$

et les différents facteurs sont calculés selon les critères suivants :

Q_1 mot court/mot long

- = 1, si le mot français a au moins 3 syllabes ou si tous les phonèmes sont communs ou si le mot français est plus long que le mot italien.
- = 0,75, si le mot français a deux syllabes ou 1 syllabe avec 3 consonnes,
- = 0,5, si le mot français a une syllabe.

Q₂ ressemblance de la première syllabe

- = 1, si la première syllabe est identique ou si deux consonnes au moins ou 2 phonèmes consécutifs de la 1^e syllabe sont identiques,
- = 0,75, si au moins une consonne est commune,
- = 0,5, si seule la voyelle est commune,
- = 0,25, si aucun phonème commun.

Q₃ nombre de phonèmes identiques se suivant dans le même ordre à partir de la 2^e syllabe pour $m \geq 2$ syllabes

- = 1, si au moins 2 phonèmes qui se suivent sont identiques,
- = 0,75, si 2 phonèmes non contigus sont identiques dans la même syllabe,
- = 0,5, si un seul phonème est identique,
- = 0,25, si aucun phonème n'est commun.

Q₄ place de l'accent tonique

- = 1, s'il est sur la même syllabe étymologique,
- = 0,75, s'il est déplacé.

Q₅ longueur du mot français par rapport au mot italien

- = 1, si le mot français a un nombre inférieur de phonèmes et de syllabes,
- = 0,75, si le mot français a un nombre supérieur de phonèmes et/ou un nombre supérieur de syllabes.

4. Quelques exemples

Nous donnons deux exemples de calculs complets de la transparence sonore expliquant pourquoi *Fjournaliste* est transparent alors que *Fjour* ne l'est pas.

F.jour/I.giorno

L'ajustement phonétique se fait de la façon suivante :

/ʊZv{/ /ʊδZop-vo/
/ʊZvρ/ /ʊZop-v/

L'indice de ressemblance consonantique est : $Ic = 2/3$ (2 consonnes communes sur 3 consonnes italiennes)

L'indice de ressemblance vocalique est $Iv = 0$

$$Im = (2 Ic + Iv) / 3 = 0,44$$

$Q_1 = 0,5$ (une seule syllabe)

$Q_2 = 1$ (2 consonnes communes dans la première syllabe)

$Q_4 = 1$ (accent tonique au même endroit)

$Q_5 = 1$ (mot italien plus long)

$$Qm = Q_1.Q_2.Q_4.Q_5 = 0,5 \text{ (} Q_3 \text{ n'intervient pas puisqu'il s'agit d'un monosyllabe)}$$

$$Tm = Im.Qm = 0,44.0,5 = 0,22, \text{ soit } 22\%.$$

Le mot avait été reconnu par 1 adolescent et 7 adultes, soit 28%.

F.journaliste/I.giornalista

/Zv{vα:λιστ/ /δZ□pνα:λιστα/
/Zvρ-vα-λιστ/ /Z□p-να-:λιστ/

L'indice de ressemblance consonantique est : $Ic = 1$ (toutes les consonnes sont communes)

L'indice de ressemblance vocalique est $Iv = 0,66$.

$$Im = (2 Ic + Iv) / 3 = 0,89$$

$Q_1 = 1$ ((mot long)

$Q_2 = 1$ (1^{re} syllabe avec 2 consonnes communes)
 $Q_3 = 1$ (suite de phonèmes identiques dès la 2^e syllabe)
 $Q_4 = 1$ (accent au même endroit)
 $Q_5 = 1$ (mot français aussi long que le mot italien)
 $T_m = I_m. Q_m = 0,89$

Le mot avait été reconnu par tous les informateurs.

Le tableau 1 ci-après donne les calculs pour 20 autres mots, et le tableau 2 fait la synthèse pour 22 mots des indices de transparence calculés mis en relation avec les pourcentages issus de l'expérimentation.

Tableau 1

Transcription ajustée du mot français	Transcription ajustée du mot italien	Indice de Transparence sonore pour le mot m
air /:E{/ /:Eρ/	aria /:αρφα/ /:αρφ/	$I_m = 0,33$ $Q_m = 0,38$ $T_m = 0,13$ 13 %
centre /:σA)τ{/ /:σαντρ/	centro /τΣEv:τρο / /ΣEv-:τρ/	$I_m = 0,37$ $Q_m = 0,75$ $T_m = 0,38$ 38%
chanson /ΣA):σον/ /ΣAv-:σον/	canzone /κav:τσονε/ /κav-:σον/	$I_m = 0,83$ $Q_m = 0,75$ $T_m = 0,62$ 62%
communauté /κ□μψvo:τε/ /κο-μυ-vo-:τε/	comunità /κομυvi:τα/ /κο-μυ-vi-:τα/	$I_m = 0,83$ $Q_m = 1$ $T_m = 0,83$ 83%
conscience /κο):σφα)σ/ /κον-:σφανσ/	coscienza /κοΣυΣEvτσα/ /κο-υΣEv- σ/	$I_m = 0,66$ $Q_m = 0,65$ $T_m = 0,42$ 42 %
différence /διφε:{A)σ/ /δι-φε-:ρανσ/	differenza /διφε:ρEvτσα/ /δι-φε-:ρEvσ/	$I_m = 0,89$ $Q_m = 1$ $T_m = 0,89$ 89%
distribution /διστ{ιβψ:σφο)/ /δισ-τρι-βυ-:σφον/	distribuzione /διστριβυ:τσοφονε/ /δισ-τρι-βυ-:σφον/	$I_m = 1$ $Q_m = 1$ $T_m = 1$ 100 %
école /ευκ□□/ /ε-υκο□/	scuola /υσκωο□α/ /:σκωο-□/	$I_m = 0,66$ $Q_m = 0,38$ $T_m = 0,25$ 25%
est /:Eστ/	est /:Eστ/	$I_m =$ $Q_m = 0,75$ $T_m = 0,75$ 75%
expression /Eκσπ{ε:σφο)/ /Eσ-πρε-:σφον/	espressione /Eσπρε:σφονε/ /Eσ-πρε-:σφον/	$I_m = 1$ $Q_m = 1$ $T_m = 1$ 100%
gelée /Z↔υλε/ /Ze-υλε/	gelata /δZeυλατα/ /Ze-λα-τ/	$I_m = 0,38$ $Q_m = 0,56$ $T_m = 0,21$ 21%
médecin /μEδυσE®/ /μEδ-υσEn/	medico /υμEδικο/ /υμE-δι-κ/	$I_m = 0,61$ $Q_m = 0,28$ $T_m = 0,17$ 17%
moi /:μωα/	me /μEυ/	$I_m = 0,33$ $Q_m = 0,19$ $T_m = 0,09$ 9 %
moment /μο:μA)/ /μο-:μav/	momento /μο:μEvτο/ /μο-:μEv-τ/	$I_m = 0,66$ $Q_m = 0,66$ $T_m = 0,44$ 44 %
mouvement /μυνυμA)/ /μυν-υμAv/	movimento /μοπιυμεντο/ /μο-πι-:μεν-τ/	$I = 0,53$ $Q_m = 0,5$ $T_m = 0,26$ 26%
piste /:πιστ/	pista /πι:στα/ /:πι-στ/	$I_m = 1$ $Q_m = 0,75$ $T_m = 0,75$ 75%
soir /:σωα{/ /:σωαρ/	sera /σε:ρα/ /σε-:ρ/	$I_m = 0,66$ $Q_m = 0,38$ $T_m = 0,25$ 25%
soleil	sole	$I_m = 1$ $Q_m = 0,56$

/σ□:λεφ/ /σο-:λεφ/	/:σολε/ /:σο-λε/	<i>Tm</i> = 0,56	56%
temps /:τΑ)/ /:τΑν/	tempo /:τΕμπο/ /:τΕμ-π/	<i>Im</i> = 0,22	<i>Qm</i> = 0,37
valeur /πα:λ-λ{/ /πα-:λορ/	valore /παλορε/ /πα-:λορ/	<i>Tm</i> = 0,08	8%
		<i>Im</i> = 1	<i>Qm</i> = 1
		<i>Tm</i> = 1	100%

Tableau 2

	Indice de transparence sonore <i>Tm</i>	Pourcentage de personnes ayant reconnu le mot	Pourcentage adolescents	Pourcentage adultes
temps/tempo	8%	0%	0%	0%
moi/me	9%	71%	64%	79%
air/aria	13%	25%	7%	43%
médecin/medico	17%	4%	7%	0%
gelée/gelata	21%	0%	0%	0%
jour/giorno	22%	28%	7%	50%
école/scuola	25%	21%	0%	43%
soir/sera	25%	43%	21%	64%
mouvement/movimento	27%	86%	71%	100%
centre/centro	38%	46%	43%	50%
conscience/coscenza	42%	46,4%	36%	57%
moment/momento	44%	50%	50%	50%
soleil/sole	56%	75%	71%	79%
chanson/canzone	62%	68%	50%	86%
piste/pista	75%	71%	50%	93%
est/est	75%	78,6%	64%	93%
communauté/comunità	83%	68%	64%	71
différence/differenza	89%	100%	100%	100%
journaliste/giornalista	89%	100%	100%	100%
expression/espressione	100%	100%	100%	100%
valeur/valore	100%	100%	100%	100%
distribution/distribuzione	100%	100%	100%	100%

Sur fond noir, les mots opaques ; sur fond gris foncé, les mots semi-opaques ; sur fond gris clair les mots semi-transparents ; sur fond blanc, les mots transparents.

5. Conclusion

Si on compare les indices de transparence calculés reproduits dans le tableau 1 avec ceux obtenus après l'expérimentation dans le tableau 2, malgré l'amateurisme de ce calcul, il semble que l'indice recouvre assez bien les 4 classes que nous avons définies, avec des oscillations qui restent à l'intérieur de ces divisions. Lorsqu'il n'y a pas concordance, cela signifie que d'autres facteurs ont pu intervenir dont l'indice de transparence ne tient pas compte. Par exemple, la compréhension probable de *mouvement* à travers le mot anglais connu (mieux chez les jeunes du reste), ou bien la connaissance culturelle du mot *moi*. Enfin, intervient certainement un facteur de fréquence que nous n'avons pas inclus et qui pourrait faire comprendre par exemple que le mot *communauté* soit moins bien reconnu que prévu, car moins rapidement récupérable en mémoire. Les difficultés peuvent

apparaître pour les mots qui sont à la frontière de catégories (par exemple, *école* ou *soir*). On constate que sur cet échantillon, le taux de transparence calculé est généralement inférieur au pourcentage moyen de reconnaissance dans le groupe des adultes et des adolescents cumulés, mais plus proche du taux des adolescents, moins performants en général. Notre but sera maintenant d'élaborer un programme informatique qui permettrait de classer automatiquement les mots, et de vérifier à plus grande échelle l'hypothèse avancée ici sur un échantillon réduit⁷. Nous pourrions alors coupler listes de fréquence et transparence. Si la même étude était faite d'abord sur le couple français LM/italien LE, nous pourrions voir si les résultats sont symétriques ou pas. En faisant le même travail sur d'autres couples de langues, nous pourrions alors établir, pour l'oral, selon des critères expérimentaux, un degré de proximité des langues romanes à l'oral. Il existe actuellement une étude lexicométrique qui mesure le degré de ressemblance des langues à partir de la classification de Swadesh, sur la base de 200 mots écrits présents dans toutes les langues (Dyen, Kruskal & Black, 1992). Or ces résultats sont pris en considération par les économistes pour évaluer, par exemple, quelles langues peuvent devenir officielles dans l'UE en lésant le moins de monde possible et en tenant compte du fait que les langues sont plus ou moins proches (Fidrmuc, Ginsburgh & Shlomo Weber, 2006). Il nous apparaît que les classifications sur la parenté des langues sont peu précises dès qu'on entre en détails dans les langues spécifiques d'une même famille. Par conséquent, une étude sur la transparence à l'oral mais aussi à l'écrit, faite à partir du principe de l'intercompréhension expérimentée, pourrait trouver son utilité. Sur le plan didactique, si notre travail passe le cap de la vérification sur grande échelle, il pourrait conduire à établir des listes de fréquence/transparence utiles au didacticien qui souhaiterait produire une méthode d'intercompréhension de l'oral et disposer d'un réservoir de mots à utiliser sciemment pour des activités ciblées. En outre, il pourrait permettre de décider, dans le cas d'un apprentissage immédiatement consécutif à l'oral (car une approche simultanée apparaît en effet difficile), dans quel ordre présenter les langues proches inconnues à partir de la langue maternelle.

Références

- Dyen, I., Kruskal J.B., & Black P. (1992). *An Indo-European classification : a lexicostatistical experiment*, Transactions of the American Philosophical Society 82, Philadelphia: American Philosophical Society.
- Jamet M.-C. (2007), *À l'écoute du français. La compréhension de l'oral dans le cadre de l'intercompréhension des langues romanes*. Coll. Giessener Beiträge zur Fremdsprachendidaktik, Tübingen: Gunter Narr Verlag.

⁷ Cela est possible à condition d'entrer les transcriptions phonémiques de façon lisible pour la machine, c'est-à-dire avec des caractères ASCII. Une base de données comme Lexique 3 (New, Pallier) en ligne qui donne pour tous les mots du lexique français la transcription phonémique et la division syllabique ainsi que les fréquences des mots, devrait nous aider dans ce projet.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e Linguaggio*, 2, 221–247.

New, B. & Pallier, C. Lexique 3, Université René Descartes, www.lexique.org

Pallier, C, Christophe, A., & Mehler, J. (1997), Language specific listening, *Trends in cognitive science*, 1-4, 129-132.

Tomatis, A. (1991), *Nous sommes tous nés polyglottes*, Paris: Fxot.

Jan Fidrmuc, Victor Ginsburgh & Shlomo Weber (2006), *Ever closer Union or Babylonian Discord? The official-language Problem in the European Union*, p.54.

Disponible sur :

<http://www.ecare.ulb.ac.be/ecare/people/members/ginsburgh/papers/010.babylonian.pdf>

Abstract

This article addresses the oral perception of words, presented out of context, in a foreign language that is close to one's mother tongue and that lies in the domain of inter-comprehension research. We assume that the oral perception of a word is more difficult than its written perception because of its instability. From a concrete experience of recognition by Italian speakers of 200 unknown French words, all cognates of Italian words, we try to determine which elements, in the acoustic image, lead to problems in the decodification of the unknown word, and which do not. We then use the concept of archiphoneme for the phonemes of L1 that are «transparent» and which can be assimilated into mother tongue phonemes in spite of their differences. On this basis we attempt to calculate the degree of formal oral transparency of a word, which allows us to determine the potentially difficult areas of oral perception and recognition, before thinking of meaning. The potential aim is to classify the lexicon automatically and get a transparency list and then to cross index the frequency list with that of transparency in order to provide a tool for creating target activities to develop oral inter-comprehension ability.