

L'introspection - moyen de valider un test à choix multiple de la compréhension orale du FLE

Anckar, Joanna, Universités de Jyväskylä & d'Åbo Akademi

Centre des études de la linguistique appliquée & Département de Français

Cet article présente un cas de validation d'un test à choix multiple de la compréhension orale du Français Langue Étrangère. Actuellement les docimologues recommandent une triangulation des méthodes pour analyser la qualité et la validité d'un test de compétences langagières. Nous voulons démontrer comment nous avons combiné l'analyse de contenu, l'analyse psychométrique et une nouvelle méthode qualitative : l'introspection écrite courte. L'idée de la dernière est de donner aux candidats (N=226), qui consistent en des apprenants du français avoir suivi environ 300 heures d'enseignement de la langue au lycée, un test à choix multiple de compréhension orale accompagné de la tâche de motiver en écrit, à chaque item, leur choix de réponse. Les motivations sont écrites dans les langues maternelles des candidats (finnois ou suédois.) Cette méthode permet d'entrer dans les processus de compréhension activés et dans les stratégies utilisées. La triangulation des méthodes d'analyse nous révèlent les caractéristiques des items : est-ce qu'il y a des traits dans le texte, dans les questions ou dans les options qui causent des problèmes injustes pour les candidats, menaçant la validité du test ? Des exemples de problèmes sont des mots inconnus dans les options, qui empêchent les candidats à arriver à la bonne réponse même après avoir compris le texte oral.

L'introspection – moyen de valider un test à choix multiple de la compréhension orale du FLE

Introduction

Pour pouvoir dire ce que mesure réellement un test – disons de la compréhension orale - pour garantir que les résultats obtenus par un candidat au test correspondent aux compétences que ce candidat possède réellement dans la compréhension orale – il faut en trouver des évidences. Ceci est valable surtout pour des tests « *high stakes* » c'est-à-dire des tests qui font partie des grands examens – comme les baccalauréats, les différents diplômes officiels des compétences linguistiques etc.

« La procédure d'un test ou d'une évaluation peut être considérée comme valide dans la mesure où l'on peut démontrer que ce qui est effectivement testé (le construct) est ce qui, dans le contexte donné, doit être évalué et que l'information recueillie donne une image exacte de la compétence des candidats en question » (Un Cadre européen commun de référence pour les langues : Apprendre, enseigner, évaluer. Conseil de l'Europe. Division des Langues vivantes, Strasbourg. 2001).

Dans cet article nous voulons démontrer ce que nous avons fait pour trouver des évidences pour, aussi bien que des arguments contre, la validité d'un test de compréhension orale. L'étude fait partie du projet de la thèse de doctorat.

Le contexte de notre projet est le baccalauréat finlandais, et l'examen du FLE pour les élèves ayant suivi des cours de français d'environ 300 heures de la langue au lycée (continués du collège pour certains.) L'examen dans sa totalité comprend quatre parties, dont le test de la compréhension orale est administré séparément des autres (qui consistent en la compréhension écrite, le test à trous aux structures/à la grammaire, et la production écrite.) Dans le certificat qu'obtient l'élève après avoir passé l'examen, les résultats dans ces quatre parties sont additionnés pour former une mention pour les compétences au FLE. Donc est-il question d'une évaluation aussi bien analytique que holistique.

Même si les compétences évaluées ne sont pas séparées dans le certificat final, elles sont donc tout de même mesurées dans des tests particuliers. Alors est-il important que le test de compréhension orale mesure réellement la compréhension orale aussi exactement que possible en couvrant le concept de la compréhension orale aussi bien que possible. Ces qualités essentielles du test nous mènent aux deux menaces les plus sérieuses contre la validité du test selon Messick (1989) et autres :

- 1) La situation où la compétence n'est pas évaluée assez précisément au cas où d'autres compétences non pertinentes à la compétence actuelle visée entrent en jeu excessivement (car il n'est guère possible d'isoler complètement une compétence visée dans un contexte d'évaluation) influant le résultat obtenu par un candidat
- 2) La situation où la compétence n'est pas évaluée d'une manière assez diversifiée, ce que veut dire qu'une partie seulement de ce que c'est la compréhension orale est évaluée par le test actuel. Il faut se rappeler le fait qu'un test représente toujours un échantillon des tâches possibles liées à la compétence actuelle mais que les tâches choisies devraient tout de même être représentatives de notre concept.

Pour pouvoir investiguer si l'une de ces menaces sont réelles, nous faut-il savoir un peu plus des cadres du test. Tout d'abord : qu'est-ce que nous, en tant que « consommateurs » ou constructeurs ou producteurs du test, entendons par le concept de la compréhension orale ? La source naturelle serait les spécifications des tests de baccalauréat. Néanmoins les spécifications ne nous fournissent-elles pas cette information, partiellement à cause du fait que les examens de baccalauréat sont étroitement liés au programme de base national d'enseignement. De ce côté, il est question d'une sorte d'évaluation sommative à la fin des cours du français donnés à l'école, étant pourtant au même temps une évaluation des compétences langagières générales.

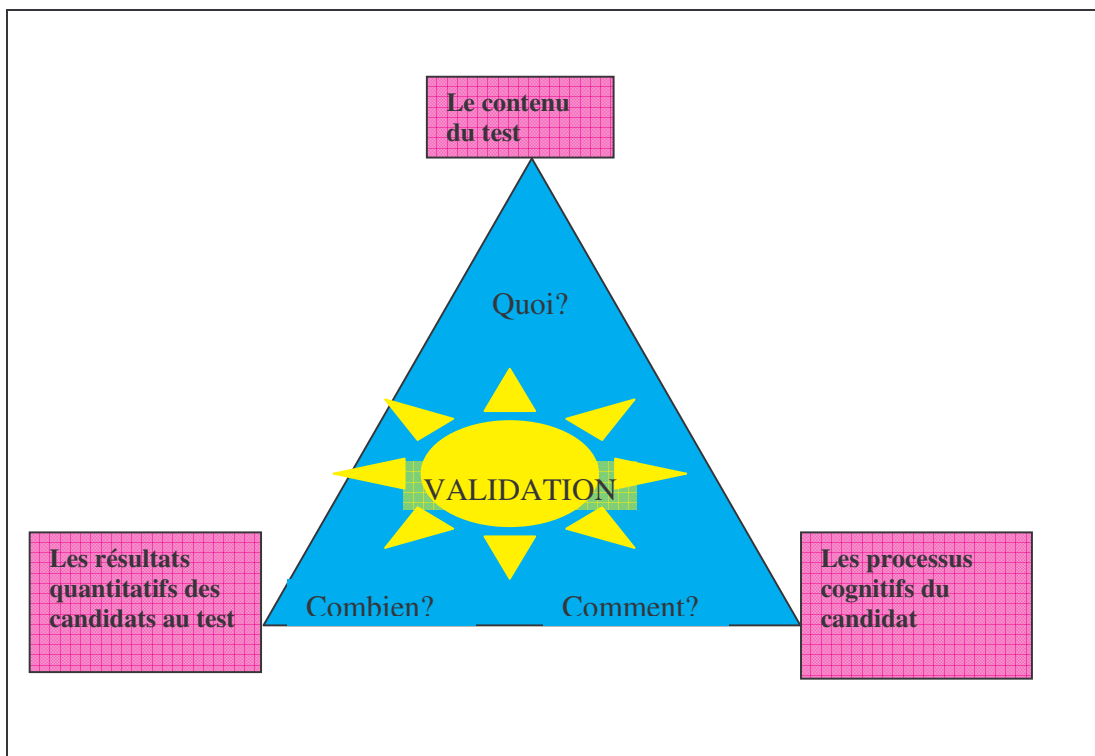
Le programme de base national d'enseignement ne nous présente pas un concept clair de la compréhension orale, mais comme principe général, on comprend qu'il est question d'une compétence communicative : le programme des langues étrangères dites courtes (d'une durée relativement courte, à cause de début tardif au collège ou au lycée) a comme but de donner aux apprenants les moyens pour pouvoir utiliser la langue dans des situations « authentiques » et dans

des activités variées liées aux intérêts et au monde des jeunes. Le programme renouvelé qui est entré en vigueur en automne 2005, suit, encore plus qu'avant, les principes présentés dans « Le Cadre européen commun de référence pour les langues. » Les grands principes de la compétence langagière communicative comprennent quatre grandes parties : la composante linguistique, la composante sociolinguistique, et la composante pragmatique – et les savoirs, habiletés et savoir-faire y liés. Ces quatre composantes devraient donc être reflétées également dans le test de compréhension orale.

Moyen de valider le test - comment savoir ce que notre test mesure effectivement ?

Il est recommandé par les théoriciens de la docimologie que la validation d'un test se fait d'autant d'angles que possible (Weir, 2005, Bachman, 2004, AERA, 1999). La « triangulation » des évidences pour et contre la validité est recherchée. Pour illustrer, voilà une figure (figure 1) montrant trois angles possibles de la validation, que nous avons appliqués dans notre étude :

Figure 1 : La triangulation de la validation



Ces trois angles de la validation sont pris en compte selon le suivant :

- 1) Le contenu du test est analysé comme une étude *a priori* en regardant du près le texte, les tâches et leur interaction dans le cadre du contexte et le concept spécifié de la compréhension orale. Grâce à cette analyse, nous obtenons une idée des thèmes, des tâches et des domaines que couvre notre test. Dans ce stade-là déjà nous pouvons former des hypothèses des points faibles du test : est-il probable qu'une des menaces potentielles contre la validité soit réalisée à l'administration du test ?
- 2) En tant qu'analyse *a posteriori*, à l'aide des résultats qu'ont obtenus les candidats au test, nous pouvons voir comment les items au test ont fonctionné. Cela implique aussi bien l'estimation de la facilité relative de chaque item (l'indice de facilité est la proportion des candidats parmi tous les participants ayant obtenu une réponse correcte) que l'estimation du pouvoir discriminant d'un item (comment l'item a-t-il su séparer les candidats forts des candidats faibles ? Un bon item devrait aboutir au succès pour les premiers, et au raté pour les derniers.
- 3) Les hypothèses établis par l'analyse du contenu et les informations quantitatives ou psychométriques fournies par l'analyse statistique peuvent être vérifiées et expliquées grâce à l'étude des processus qui entrent en jeu à une situation d'évaluation. Cette analyse qualitative des réponses introspectives des candidates face au test est pour nous essentielle en nous décidant pour ou contre la validité relative du test actuel dans son contexte déterminé.

L'introspection comme moyen de s'approcher des processus cognitifs

Étant évident que les processus cognitifs en jeu à la situation d'écoute et de test sont non seulement compliqués mais aussi invisibles, il nous faut d'autres moyens pour étudier ce qui se passe dans les cerveaux des candidats que celui d'observation directe. Même si quelques résultats ou traces de cette activité cognitive consistent en les réponses explicites aux questions posées, il nous manque l'information sur le chemin entre l'entrée (« l'input ») (le texte et les tâches/questions posées) et les réponses. Une manière de se rapprocher des processus est

d'entreprendre une étude où les candidats sont demandés de verbaliser leurs activités au cours de la situation du test. Cela peut se faire à l'oral, à l'écrit, d'une forme libre (ouverte) ou bien plus restreinte avec des questions précises posées aux candidats.

Nous avons choisi d'appliquer l'introspection écrite courte, une nouvelle méthode¹ pour mettre la main sur ce qui se passe chez les candidats dans la situation du test. Dans le suivant, nous allons présenter aussi bien notre méthode que quelques résultats préliminaires au niveau « micro » de l'analyse entreprise au contexte de la thèse de doctorat.

La justification théorique pour l'utilisation de la méthode d'introspection réside dans la supposition que la cognition humaine ressemble à un système de traitement d'information. Le processus de compréhension orale est rarement automatique, surtout pas dans une langue étrangère ou seconde au cas où une partie importante des activités devrait être contrôlée. Le processus demande une telle concentration intense sur l'entrée (*l'input*) – le texte oral – qu'une partie substantielle en reste à la mémoire. Le rapport introspectif serait donc possible par la verbalisation de ces traces de la mémoire (Yi'an, 1998.) Ericsson et Simon (1987: 25) donne un cadre pour étudier la pensée, où l'information récemment acquise par le processeur central est gardée à la mémoire à court terme (ou du travail), étant accessible au traitement continu (par exemple pour la production des rapports verbaux.)

Nous sommes bien conscientes des limitations liées à l'utilisation et à l'analyse de la méthode. Les participants passent le test sans entraînement préalable et en conséquence, la tâche qui consiste en indiquer les raisons pour leur choix des options leur est-elle toute nouvelle. Comme prévu, quelques participants la trouvent toute naturelle, facile et intéressante, tandis que d'autres ne sont pas de cet avis.

On peut dire qu'il y a des activités sur trois niveaux présentes à la méthode introspective :

- 1) Les processus cognitifs inconscients et les stratégies de compréhension ou les stratégies liées au test qui sont appliqués à l'écoute et à la tâche actuelle ;

¹ Nous n'avons pas vu appliquée cette méthode spécifique ailleurs; nous en prétendons donc être l'inventeur. Ce projet consiste donc en partie du deuxième pilotage de la méthode de l'introspection écrite courte.

- 2) La réponse, c'est-à-dire le choix d'une option ;
- 3) Les réponses introspectives.

Notre hypothèse est, donc, que les niveaux 2 et 3 nous révèlent quelque chose sur les activités au niveau 1.

Il y a évidemment des limitations aussi dans l'utilisation des résultats de l'analyse des réponses introspectives pour la tâche de validation - en ce qui concerne l'effort ajouté au test par cette activité de motivation des réponses. Ces faits limitatifs doivent, en conséquence, être considérés dans l'analyse des réponses introspectives.

Le contexte : le test d'origine et actuel

Le test analysé dans cette étude est un test de compréhension orale utilisé à l'origine dans le baccalauréat finlandais au printemps 2002. Le format du test est le choix multiple (QCM), avec 30 items au total, administré en 45 minutes.

Nous avons choisi de modifier le test et la procédure de son administration aussi bien quantitativement que qualitativement selon le suivant :

- 17 items QCM (+ 3 items à réponse courte) sont choisis du test afin de pouvoir administrer le test pendant la durée d'une leçon de 45 minutes.
- Pour la première partie (items 1-11), chaque passage oral d'une durée de moins de 60 secondes est lié à deux (ou trois dans un cas) items. Pour les six autres items QCM, ils représentent chacun des textes indépendants.
- Les participants répondent aux QCM par sélectionner une option parmi trois (ABC) après la première écoute déjà. Après la deuxième écoute, ils peuvent changer leur choix. Ils sont aussi demandés, à chaque item, de motiver leur choix de réponse, ce qui nous fournit les réponses introspectives, sur lesquelles nous nous basons pour la partie la plus importante de l'analyse qualitative du test.
- Les 226 participants sont divisés en quatre groupes selon leurs résultats aux deux parties du test, donnant 17 points au maximum. Ceci pour pouvoir comparer les manières d'agir et de répondre dans la situation de test. Les quatre groupes sont :

- Le niveau bas (0-6/ 17 points au maximum)
- Le niveau intermédiaire I (7-8 p)
- Le niveau intermédiaire II (9-11 p)
- Le niveau avancé (12-17 p)

L'exemple de l'analyse : le cas de l'item 6

Pour pouvoir montrer comment fonctionne la triangulation des évidences et surtout quelle information nous donne dans ce contexte la méthode d'introspection, nous avons choisi un item de la première partie du test à subir une analyse détaillée. Cet item est choisi présentant des traits qui nous font questionner sa validité comme un bon item mesurant la compréhension orale dans le contexte du test.

Le contenu : le texte, les questions et leur interaction

Le thème du texte visé est celui du phénomène des anciens magasins et bureaux à Paris se transformant en appartements. D'abord ce phénomène est-il introduit pour être ensuite illustré par des commentaires donnés par des gens qui ont des expériences de cette sorte d'habitation. Le texte et les questions pour les items 5 et 6 sont donnés dans la table 1 :

Table 1 : Texte oral et les items 5 et 6

<p>« Olivier et Martine ont tous les deux 35 ans. Lui est directeur commercial, elle, consultante en informatique. Ils ont acheté un vieux laboratoire de prothèses dentaires dans le 17^{ème} arrondissement. « Nous sommes originaires du Midi avec mon mari et ici, ça nous rappelle un peu la vie de province. Deux vitrines donnent sur la rue, les autres sur une cour et un jardin...mais quand j'ai essayé de dormir dans une pièce du côté de la rue, c'était trop bruyant, entre les poubelles le matin et les rondes des agents de police. L'achat a bien sûr demandé de gros investissements – et pas seulement financiers. J'ai fini par me sentir au bout de mes forces, crevée ! J'ai perdu plusieurs kilos, mais le pire, c'est que mes nerfs ont complètement craqué... Mais maintenant, après les débuts difficiles, ça a l'air d'aller bien. Le plus dur est derrière nous. »</p>	
<p>5. Que raconte Martine?</p> <p>a) Les bruits de la rue ne la dérangent pas</p> <p>b) Elle vient d'avoir des problèmes de santé</p> <p>c) Elle veut repartir en province</p>	<p>6. Que dit-elle de leur appartement?</p> <p>a) Il n'est pas trop cher</p> <p>b) Il leur évoque des souvenirs</p> <p>c) Il ne sera jamais prêt</p>

Le contenu de l'item 6 peut être décrit schématiquement selon le suivant :

- **Description du type de texte :** Monologue, genre : débat pendant émission de radio (en réalité un texte écrit, tiré d'un magazine);
- **Voix :** L'oratrice femme parle du français standard, d'une vitesse « normal »
- **Thème :** L'expérience subjective d'habiter au rez-de-chaussée à Paris
- **Tâche :** traitement du discours relativement long (contrasté aux énoncés courts)
- **Question (amorce):** Quelque support pour la localisation de l'information clé au texte
- **Options :**
 - 6a : distracteur – un fait implicite du texte nié
 - 6b : clé – paraphrase d'une phrase du texte
 - 6c : distracteur – rien qui correspond au texte

→ La négation et le modificateur 'trop' présents dans l'option 6a ajoutent un poids cognitif supplémentaire sur la lecture

→ 'Évoquer' dans 6b peut être un verbe non-familier pour les apprenants

→ 6c sans correspondance au texte peut causer de la confusion

Les conséquences de l'analyse du contenu que nous en tirons indiquent qu'item peut contenir des traits qui menacent sa validité et cela dans la formulation des options écrites. Le problème potentiel est le poids cognitif mis sur la lecture, qui diminue la capacité des candidats de centrer l'attention sur la compréhension et l'interprétation du texte oral.

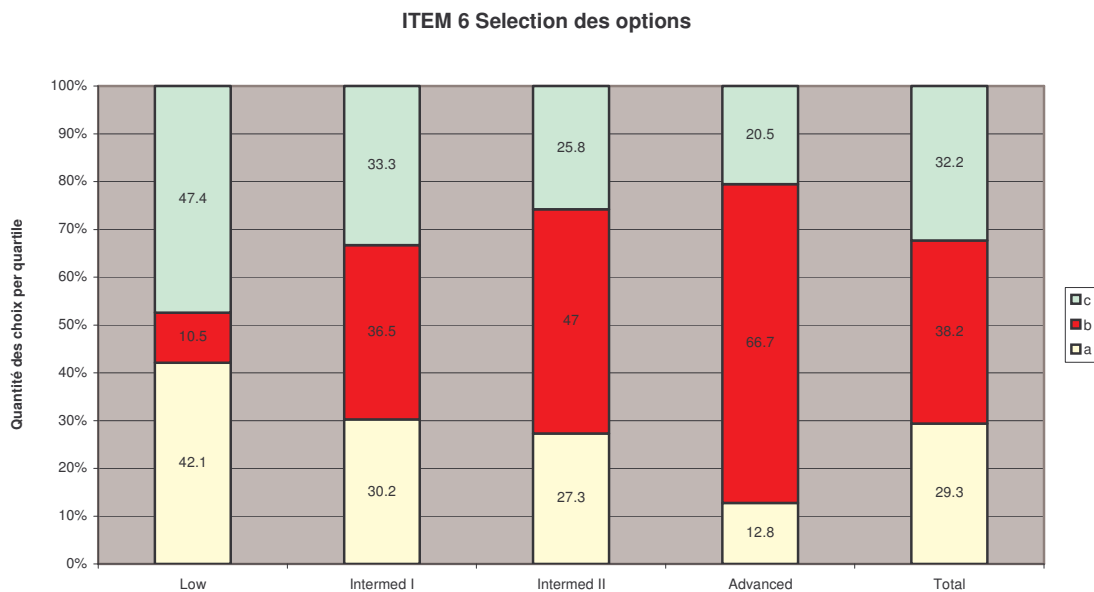
Pour pouvoir déterminer si les composantes de la compétence communicative sont bien incluses dans le test de compréhension orale, il faut évidemment voir le test complet. Le thème du texte oral fait partie de l'information socioculturelle sur le pays cible ; on peut pourtant discuter s'il est proche des intérêts des jeunes. La tâche de répondre aux questions d'un test QCM est peu authentique ; on peut néanmoins dire qu'une situation d'évaluation l'est rarement, si jamais.

De l'analyse quantitative

Nous allons passer à l'analyse quantitative des résultats obtenus par les candidats. Il y a surtout trois caractéristiques qui nous intéressent dans l'analyse : le pointage des distracteurs et de la clé (fréquence à laquelle s'effectue le choix de chaque option), l'indice de facilité et l'indice de discrimination.

Figure 2 : L'indice de facilité (moyenne) et de discrimination (corrélacion de l'item avec le totale) et la sélection des options par quartile de participants pour l'item 6

Moyenne	Corrélacion item-total
.382	.226



L'indice de facilité de .38 nous indique que l'item a été relativement difficile (comparé à la moyenne de l'indice des 11 items de la première partie du test qui est de .50, l'indice de l'item 6 étant le plus bas.) Ceci ne dit rien de la qualité de l'item en soi – on peut bien inclure dans un test des items d'un niveau de difficulté varié pour un groupe cible particulier de candidats. L'essentiel

est de savoir si les candidats plus forts ont réussi à répondre correctement à l'item, et, en conséquence, si les faibles n'ont pas réussi. Ce rapport est important pour qu'on puisse savoir si le test sait séparer efficacement les candidats aux différents niveaux de compétence les uns des autres. Pour cela nous considérons la corrélation des résultats à l'item avec les résultats des candidats au test entier. Nous pouvons constater que la corrélation pour cet item est assez basse ; selon certains psychométriciens docimologues, elle devrait atteindre .30 pour être acceptable. Néanmoins, tous les items dans ce test ont des indices de ce genre et l'indice de discrimination seul ne nous informe pas de la qualité du test. Si, par exemple, les items dans le test de compréhension orale mesurent chacun une sous-compétence de la compréhension orale qui est partiellement différente des sous-compétences mesurées par les autres items, l'indice de discrimination ne pourra pas atteindre un taux très élevé.

Par contre, dans la figure 3 (au-dessus) montrant la sélection des options – leur pointage - nous pouvons voir comment les candidats aux quatre niveaux ont choisi leurs réponses. 6b étant la clé, la proportion de son choix augmente clairement mais doucement vers les niveaux plus élevés. Ce n'est pourtant pas avant le niveau avancé que plus de la moitié des candidats sait la choisir. L'attraction du distracteur 6c persiste à travers les quatre groupes – même 32% de tous les candidats l'a choisi. 6a est également relativement attractif avec 29% des choix.

Cette analyse quantitative nous indique que l'item a été difficile avec deux distracteurs attractifs et qu'il ne sait pas particulièrement bien discriminer entre les candidats. Les raisons sont spéculatives dans l'analyse *a priori* du contenu, et doivent être vérifiés par une étude détaillée des réponses introspectives données par les candidats. Par cette analyse nous voulons trouver des explications pour les choix des options et à travers ces explications des indices supplémentaires de la qualité de l'item.

L'introspection : l'option clé 6b

La clé 6b est donc choisi par un peu plus qu'un tiers des candidats. La situation idéale du point de vue de la validité du test de compréhension orale est celle où la clé est choisie par ceux qui ont

effectivement compris le texte oral. En effet, nous n'avons que deux participants qui ont combiné le choix du 6b avec l'indication explicite de l'information clé du texte oral :

« *Muistuttaa maaseudun talosta* » (candidat Y3 :8) 'Rappelle une maison à la campagne'

« *Det påminner om landet* » (H11 :10) 'Elle rappelle la campagne'

Les quelques cas de l'épreuve d'une compréhension partielle (faits corrects selon le texte oral donnés dans la réponse introspective) ont un contenu qui n'a rien à faire avec celui de l'option :

« *Heillä on ollut vaikeaa* » (N5 :6) 'Ils ont eu un temps difficile'

« *Lägenheten var dyr. Nu går det bättre än i början* » (H6 :6) 'L'appartement était cher. Maintenant tout va mieux qu'au début'

Par contre y a-t-il même 42% des candidats qui ont indiqué avoir deviné pour arriver à leur réponse. La grande quantité de choix au hasard combinée avec la stratégie d'exclusion également donnée comme raison pour le choix ajoutées aux réponses vagues ou vides nous indiquent que l'item a été problématique. Quelques commentaires sur des difficultés ressenties face à l'item sont :

« *Mikään ei tuntunut sopivan, arvaus* » (S9 :6) 'Rien ne semblait convenir, je devine'

« *Taloudellisia ongelmia, hämää...* » (E8 :7) 'Des problèmes du genre économique, vague...'

« *En kuullut mainittavan mitään a :han tai c :hen liittyvää, joten otin b :n, jota en ymmärrä* » (N12 :11) 'Je n'ai rien entendu sur a ou c, alors j'ai pris b, que je ne comprends pas'.

L'introspection : distracteur 6a

Ce distracteur a été choisi par près d'un tiers (29%) des participants. De ces candidats, 41% ont indiqué avoir fait un choix au hasard. Il y a l'épreuve d'une grande quantité (13,5%) des malentendus ou des interprétations fausses du texte, dont des cas probablement influencés par le contenu de l'option :

« *Se maksu kai aleni tai jotain, joten arvaan A :n* » (T1 :1) 'Le prix est baissé ou quelque chose, alors je devine A'

« *Hän sanoi asunnon olevan halpa* » (R8 :5) 'Elle a dit que l'appartement était bon marché'

« *Puhui rahasta syynä asumiseen* » (Å16 :9) 'Elle parlait de l'argent comme raison d'habiter'

Les réponses varient de la stratégie d'exclusion, à la compréhension partielle et aux réponses vides. Parmi les commentaires sur l'item on trouve des cas où la compréhension des options écrites constitue un problème:

« *En saanut kovin hyvin selvää, enkä ihan ymmärtänyt täysin varsinkaan kohtaa kuusi* » (P7 :4) 'Je n'ai pas très bien distingué ce qui est dit, et je n'ai pas complètement compris l'item 6 surtout'

« *Arvasin, koska en tiennyt mitä muut vaihtoehdot tarkoittavat* » (J8 :6) 'J'ai deviné, comme je n'ai savais pas ce que voulaient dire les autres options'

L'introspection : distracteur 6c

Le dernier tiers de candidats a mis son choix sur distracteur 6c. Même 53% - plus de la moitié - de cette sélection a été faite au hasard. En général, les motivations pour choisir 6c sont vagues : la raison en peut être le fait qu'il ne semble y avoir rien au texte sur lequel on pourrait baser le choix. Les malentendus sont donc le plus souvent basés sur l'inférence des connaissances ou des expériences antérieures :

« *Eihän ne kämpät o koskaan valmiita* » (L1 :5) 'Les appartements ne sont jamais tout prêts, n'est-ce pas ?'

« *Deras hus var en stor investering, så de gör troligen remont, som inte är klar* » (H8 :7) 'Leur maison était un grand investissement, alors ils la restaurent probablement, sans avoir encore fini'

Un des commentaires sur des difficultés ressentis nous informe que le candidat a inféré à base du ton de voix de l'oratrice. Les autres problèmes qui sont mentionnés reflètent l'obscurité des options :

« *Arvauksen puolella...jotenkin toivoton sävy* » (G4 :7) 'J'ai deviné... elle avait un ton de voix sans espoir de quelque sorte '

« *Svårt det kan också vara b* » (D1 :2) 'Difficile, il peut être b aussi'

« *Asunto oli kallis, enkä ymmärtänyt b-kohtaa* » 'L'appartement était cher, et je n'ai pas compris l'option b'

« *En osaa sanoa, kohta meni jotenkin ohi* » (S11 :6) 'Je ne sais pas, j'ai manqué cet item de quelque manière'

« *prêt= ?* » (E11 :9)

Commentaires conclusifs sur l'item 6

L'analyse « triangulaire » de l'item 6 indique qu'il y a dans l'item quelque chose de problématique. Un des signes est le fait qu'il y a une grande quantité de choix au hasard : 43% des réponses au total, dont 39% sont sur l'option clé. Il y a également une proportion dominante de réponses liées au format du test, et très peu de réponses liées au texte oral. Ceci porte à croire que les problèmes résident surtout dans la formulation des options.

Il faut être conscient du fait que le traitement de l'item 5 par les candidats a une influence sur le résultat à l'item 6 (à cause du temps et de la capacité limités qui restent aux participants à l'item 6.) Ceci veut dire qu'il y a des difficultés évidentes liées à ce format avec deux items par passage du texte, comme le participant est obligé de considérer deux questions et trois plus trois options à une et même écoute.

On peut dire que la menace principale contre la validité de l'item 6 se trouve dans le format du test – la nature du format du test à choix multiple avec des traits qui empiètent sur le processus cognitif qui devrait être le plus important : la compréhension et l'interprétation du texte oral. Les traits empiétant sur la tâche des candidats consistent surtout à la présence des options obscures ou opaques et des caractéristiques de l'administration du test QCM qui empêchent le déroulement « naturel » de la compréhension orale.

La validité du test évalué serait augmenté avec l'amélioration de l'item. Il consisterait principalement dans un remaniement des options. La considération des améliorations nous donne l'occasion de toucher au côté pratique du processus de construction d'un test de compréhension orale. Comme il est toujours plus compliqué ou laborieux de remanier le texte oral qui forme la base pour la tâche ensuite donnée aux candidates, l'approche préférée est de se concentrer sur une reconstitution des options. En conséquence, s'appuyant sur l'information obtenue par l'analyse triangulaire de tous les items qui constituent le test de compréhension orale, en refaisant les options qui semblent problématique, nous, en tant que constructeurs du test, pouvons éviter au moins une de menaces contre la validité du test : la variance sans rapport avec le construct (le concept de compréhension orale).

Conclusions et spéculations

Nous nous permettons de faire les conclusions générales suivantes sur les items QCM étudiés :

1. Une stratégie typique est de deviner, plus souvent au cas des options opaques (non claires, non clairement fausses/correctes sans ambiguïtés ou contenant des expressions non familières...)
2. Les options écrites ont une influence excessivement importante sur le processus de l'interprétation du texte, et pour cela, leur qualité est essentielle pour qu'on puisse éviter la variance sans rapport avec le construct !

En ce qui concerne la méthode de l'introspection comme moyen de trianguler les évidences de la validité, nous pouvons constater le suivant :

1. L' introspection écrite courte sert de moyen complémentaire qui donne de l' information précieuse que l' analyse du contenu *a priori* ou l' analyse quantitative *a posteriori* ne peuvent pas nous fournir : des indications sur les vrais processus activés par les participants et sur le détail particulier dans l' item – le texte, les questions ou leur interaction - où retrouver les problèmes.
2. La méthode est pratique d' utiliser, comme de grands groupes de participants (aux différents niveaux micro) peuvent être testés pendant une même administration.
3. Elle peut avantageusement être utilisée au stage de pilotage dans le développement des tests, comme il indique les effets involontaires des tâches ou des options particulières.

En conséquence, le développement de la méthode de l' introspection courte comme moyen de trouver des évidences pour ou contre continue.

Bibliographie :

- Alderson, J.C. (1990): Testing Reading Comprehension Skills. Part Two. Getting students to talk about taking a reading test (pilot study). *Journal of Reading in a Foreign Language*, 7, 465-503.
- Alderson, J.C. C. Clapham & D. Wall. (1995): *Language Test Construction and Evaluation*. Cambridge Language Teaching Library. Cambridge: Cambridge University Press
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (1999): *Standards for educational and psychological testing*. Washington, DC: Author.
- Anderson, A. & Lynch, T. (1988): *Listening*. New York: Oxford University Press.
- Bachman, L.F. & A. Palmer. (1996): *Language Testing in Practice*. Oxford: Oxford University Press.
- Bachman, L.F. (2004): *Statistical Analyses for Language Assessment*. Cambridge: Cambridge University Press.
- Borsboom, D., G.J. Mellenbergh. and J. van Heerden. (2004): The Concept of Validity. *Psychological Review*. Vol 111, No.4.
- Buck, G. (2001): *Assessing Listening*. Cambridge: Cambridge University Press.
- Un Cadre européen commun de référence pour les langues : Apprendre, enseigner, évaluer. Conseil de l' Europe. Division des Langues vivantes, Strasbourg. (2001).
- Chappelle, C.A. (1998): Construct definition and validity inquiry in SLA research. In L.F. Bachman och A.D. Cohen (eds.) *Interfaces between second language acquisition and language testing research*. Cambridge: CUP
- Cornaire, C. (1998): *La compréhension orale*. CLE International.
- Ericsson, K.A., Simon, H. (1987): Verbal Reports on thinking. In: Faerch, C. Kasper, G. (Eds.). *Introspection in Second Language Research*. Multilingual Matters, Clevedon, pp. 24-53)
- Flowerdew, J. and L. Miller (2005): *Second Language Listening. Theory and practice*. New York: Cambridge University Press.
- Freedle R. & Kostin, I. (1999): Does the text matter in a multiple-choice test of comprehension? The case for the construct validity of TOEFL's minitalks. *Language Testing*, 16 (1), 2-32
- Green, E. (1997): *Verbal Protocol Analysis in language testing research: A handbook*. Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.
- Kelly, P. (1991): Lexical ignorance: the main obstacle to listening comprehension with advanced foreign language learners. *International review of Applied Linguistics*. 29 (2) 122-4
- Lhote E. (1995): *Enseigner l' oral en interaction*. Paris: Hachette.
- Malmberg, B. (1968): *Phonétique française*. Malmö: Liber.
- Messick, S. (1989): Validity. In R.L. Linn (ed.) *Educational Measurement* (3rd ed., pp. 13-103). New York: American Council on Education, Macmillan Publishing Company.

- Nagle, S. & S.Sanders. (1986): Comprehension theory and second language pedagogy. *TESOL Quarterly*, 20.1.
- Nissan S., F. De Vincenzi. & K.L Tang. (1996): *An analysis of factors affecting the difficulty of dialogue items in TOEFL listening comprehension*. (TOEFL Research Report No. 51). Princeton, NJ: Educational Testing Service.
- Noblitt, J.S. (1995): Cognitive Approaches to Listening Comprehension. *Dimension*, 1995, 1-11. Southern Conference on Language Teaching. Valdosta State University, Georgia.
- Rost, M. (1994): *Introducing Listening*. Penguin English.
- Rost, M. (2002): *Teaching and Researching Listening*. London: Longman.
- Rubin J. (1994): A review of second language listening comprehension research. *Modern Language Journal*, 78, 199-221.
- Shohamy E. & O. Inbar. (1991): Validation of Listening Comprehension Tests: The Effect of text and question Type. *Language testing* 8 : 23-40.
- Tegelberg, E. (1995): *Franskt uttal i teori och praktik*. Göteborg: Akademiförlaget.
- Thompson I. (1995): Assessment of second/foreign language listening comprehension. D. Mendelsohn & J.Rubin (red.) *A Guide for the teaching of second language listening* (pp.31-58). San Diego, CA: Dominie Press, Inc
- Tsui, A.B. & J. Fullilove. (1998) : Bottom-Up or Top-Down Processing as a Discriminator of L2 Listening Performance. *Applied Linguistics* 19/4
- Ur, P. (1984): *Teaching listening comprehension*. New York: Cambridge University Press.
- Weir, C.J. (2005): *Language Testing and Validation. An Evidence-based Approach*. Hampshire: Palgrave.
- Yi'an, Wu. (1998) :. What do tests of listening comprehension test ? – A retrospection study of test-takers performing a multiple-choice task. *Language Testing* 15 (1).