



**HAL**  
open science

# Étude de l'impact du contexte sonore environnemental sur la compréhension de dialogues pédagogiques : une démarche expérimentale en didactique du français langue étrangère

Marie-Mandarine Colle Quesada

► **To cite this version:**

Marie-Mandarine Colle Quesada. Étude de l'impact du contexte sonore environnemental sur la compréhension de dialogues pédagogiques : une démarche expérimentale en didactique du français langue étrangère. Linguistique. Université Toulouse le Mirail - Toulouse II, 2014. Français. NNT : 2014TOU20098 . tel-01275436

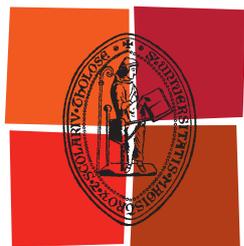
**HAL Id: tel-01275436**

**<https://theses.hal.science/tel-01275436>**

Submitted on 17 Feb 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université  
de Toulouse

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 2 Le Mirail (UT2 Le Mirail)

---

**Présentée et soutenue par :**

**Marie-Mandarine Colle Quesada**

Le 21 novembre 2014

**Titre :**

Étude de l'impact du contexte sonore environnemental sur la compréhension de dialogues pédagogiques : une démarche expérimentale en Didactique du Français Langue Étrangère

---

ED CLESCO : Acoustique, Science du langage

**Unité de recherche :**

U.R.I. Octogone EA 4156

**Directeur(s) de Thèse :**

Nathalie Spanghero-Gaillard, Professeure des Universités, Université Toulouse 2 - Jean Jaurès  
Pascal Gaillard, Maître de Conférences, Université Toulouse 2 - Jean Jaurès

**Rapporteurs :**

Danièle Dubois, Directrice de recherche CNRS émérite, Université Pierre et Marie Curie, Paris 6

Jean-Rémi Lapaire, Professeur des Universités, Université Bordeaux Montaigne

**Autre(s) membre(s) du jury :**

Jérémi Sauvage, Maître de Conférences - HDR, Université Paul Valéry, Montpellier 3  
André Tricot, Professeur des Universités, Université Toulouse 2, Jean Jaurès







# Remerciements

Je tiens tout d'abord à remercier chaleureusement mes directeurs de thèse *Nathalie Spanghero-Gaillard* et *Pascal Gaillard* pour leur soutien, leurs encouragements, leurs conseils et leur patience. Votre accompagnement tout au long de ces années m'a été très précieux et sans vous ce travail n'aurait pas abouti. À travers nos discussions et votre écoute, j'ai grandi. Vous m'avez donné le goût de la recherche et du questionnement. Vous m'avez accompagnée sur ce chemin, vous m'avez guidée et permis d'évoluer. Merci pour votre confiance. J'espère pouvoir continuer à profiter de vos conseils et de votre soutien pour les prochaines étapes de mon parcours. Vous avez contribué à la fois au développement de mes compétences et à mon épanouissement... Je vous en serai toujours reconnaissante.

Je tiens également à remercier l'ensemble des membres du jury qui ont accepté de lire mon travail, de l'évaluer et d'en discuter : *Danièle Dubois* et *Jean-Rémi Lapaire* qui ont accepté le rôle de rapporteurs, et *André Tricot* et *Jérémi Sauvage* qui les rejoignent pour cette soutenance. Je vous remercie du temps que vous accordez à mon travail, de votre attention et de vos remarques. Une soutenance de thèse est à la fois un aboutissement et un départ. Je vous remercie d'y participer.

Ce travail de thèse n'aurait pas pu aboutir sans le concours des différents participants aux études menées. Mes pensées vont vers les professionnels de Radio France Internationale qui m'ont reçue et m'ont accordé un temps précieux, et dont les conseils et remarques m'ont accompagnée tout au long de cette *aventure sonore*. Merci à *Manou* de m'avoir ouvert ces portes.

Je remercie également *Cécile* pour son excellent travail de dessin. Tu as été réactive, à l'écoute et patiente. Ça a été un vrai plaisir de travailler avec toi.

Je pense aussi aux enseignants de Français Langue Étrangère de *l'Alliance Française de Toulouse* et du *Département d'Enseignement du Français Langue Étrangère de l'UT2J*, qui se sont prêtés au jeu de l'évaluation des dialogues.

Enfin, je tiens à remercier tous les doctorants et les étudiants de l'Université de Toulouse et de l'Alliance Française qui ont participé aux expérimentations, m'ont accordé du temps et se sont intéressés à ma démarche. Vos remarques ont grandement contribué à ma réflexion.

Je n'oublie évidemment pas les étudiants auprès de qui j'ai eu la chance d'enseigner, avec qui j'ai eu l'occasion de partager mon intérêt pour la didactique et l'enseignement du français. Vous avez vous aussi participé à ce travail en me rappelant toujours l'intérêt et l'enrichissement que l'on trouve à enseigner. Je pense ainsi aux étudiants de la licence et du master Apprentissage/didactique du Français Langue Étrangère et Seconde de *l'UT2J*, ainsi qu'aux apprenants du Département de l'Enseignement du Français à l'International de l'Université de Lille 3.

Il est difficile ou impossible parler des étudiants sans se souvenir également de l'accueil et de la bienveillance des équipes pédagogiques avec lesquelles j'ai pu collaborer, à Toulouse et à Lille. Certains d'entre vous ont d'abord été mes enseignants, mais vous m'avez tous beaucoup appris. Je pense notamment, et de manière non exhaustive, à *Christiane Soum, Barbara Köpke, Corine Astesano, Olga Théophanous, Michel Billières, Cecilia Gunnarsson, Mélanie Jucla, Ines Saddour, Christiane M'Bayo, Sylvie Dubart, Jean-Noël Jego, Martine Eisenbeis, Brigitte Lepez, Isabelle Boneu, Marie-Pascale Hamez, Vannessa Obin, et Anne-Sophie Thuillier*. Travailler à vos côtés m'a enrichie, ouverte à de nouveaux horizons et surtout rappelé le goût de l'enseignement...

Je remercie également Évelyne pour sa disponibilité, sa réactivité et sa gestion des aspects pragmatiques comme émotionnels.

Le cheminement de la thèse est un parcours solitaire mais que je n'aurais pu parcourir sans le soutien et la complicité de mes amis et collègues. Travailler avec vous tous a été une vraie chance. Nous avons ri et craqué, fait des tours de parking (quand il y avait encore un parking pour abriter nos conversations)...

J'adresse mes remerciements à *Émilie* et *Kamran* pour leur soutien indéfectible et très significatif pour les statistiques. Vous avez fait preuve d'une patience infinie!

*Maman, Lucie, Charlotte, et Hélène*, je vous remercie également, pour vos relectures et conseils, pour votre disponibilité et votre patience.

Mes pensées vont vers tous les doctorants et chercheurs des laboratoires Octogone, CLLÉ et CPST *Émilie L, Kamran, Quentin, Sophie, Lionel, Charlotte A, Marie, Clémentine, Antoine, Sören, Aurore, Karine, Cynthia, Julien, Luis, Xueting, Émilie M, Claire, Charlotte V, Ajhar, Lucie, Caroline, Cécile, Guillaume* et tous ceux que je suis désolée d'oublier... Nous nous sommes suivis et croisés, nous avons échangé et partagé. Vous avez jalonné ce parcours de sourires et de fous rires et fait de la thèse un moment que j'ai aimé partager avec vous. Nous avons vécu tant de choses dans ces boîtes grises !

Je tiens enfin à remercier, pour leur présence, leur écoute, leur complicité et leur soutien ma famille et mes proches : mes parents, qui m'ont offert un soutien sans faille tout au long de ces années (et même bien avant) et mes sœurs, *Manou, Cécile*, et leurs familles, qui m'ont entourée et encouragée.

Je pense également à remercier la famille un peu plus élargie, (je ne peux pas vous citer tous ici, mais je ne vous oublie pas) et à mes amis *Émilie, Lucie, Charlotte, Clémentine, Hélène, Cédric, Julien et David*. Vous m'avez aidée à avancer et à me dépasser. Vous m'avez écoutée, encouragée et soutenue et avez su m'apporter des moments de joie et de complicité quand j'en ai eu besoin.

Enfin, je ne pourrais terminer ces remerciements sans une pensée à *Émilie F*. J'ai mis mes pas dans les tiens...

...merci



# Sommaire

<b>REMERCIEMENTS</b>	<b>3</b>
<b>SOMMAIRE</b>	<b>7</b>
<b>INTRODUCTION GENERALE</b>	<b>11</b>
<b>CONTEXTE DIDACTIQUE : L'ORAL EN FRANÇAIS LANGUE ETRANGERE</b>	<b>17</b>
1. Qu'est-ce que l'oral	17
2. La place de l'oral dans trois grandes méthodologies d'enseignement/apprentissage des langues étrangères	27
3. Les supports de CO en FLE	34
<b>PARTIE 1 : CADRE THEORIQUE</b>	<b>49</b>
<b>PREMIER CHAPITRE : DES ETUDES SUR LA PERCEPTION...</b>	<b>49</b>
1. Qu'est-ce que percevoir ?	50
2. L'analyse des scènes auditives (ASA)	59
3. Le sémantisme des sons de l'environnement	72
Ce qu'il faut retenir...	79
<b>DEUXIEME CHAPITRE : ... AUX ETUDES SUR LA COMPREHENSION</b>	<b>81</b>
1. Éléments définitoires : Qu'est-ce que comprendre ?	82
2. Des études sur la compréhension orale : de la perception du signal à la (re-) construction de sens	86
3. Des études sur la compréhension de textes narratifs	95
Ce qu'il faut retenir	104
<b>TROISIEME CHAPITRE : HYPOTHESES</b>	<b>105</b>

1. Qu'est-ce qu'être débutant en langue ?	107
2. Le rôle du contexte sonore environnemental dans la compréhension d'un dialogue	111
<b>PARTIE 2 : ELABORATION DU MATERIEL EXPERIMENTAL</b>	<b>119</b>
<b>PREMIER CHAPITRE : ÉLABORATION DU MATERIEL SONORE ENVIRONNEMENTAL</b>	<b>121</b>
1. Une étude sur les attentes auditives	121
2. Construction des scènes sonores	132
3. Identification des scènes sonores	138
4. Sélection des scènes sonores	146
<b>DEUXIEME CHAPITRE : ÉLABORATION DU MATERIEL LINGUISTIQUE</b>	<b>151</b>
1. Méthode	151
2. Résultats	158
<b>PARTIE 3 : EXPERIMENTATION PRINCIPALE</b>	<b>165</b>
1. Méthode	165
2. Résultats	187
<b>PARTIE 4 : DISCUSSION</b>	<b>211</b>
1. Discussion des résultats du Classe-Images	212
2. Discussion méthodologique	229
3. Discussion didactique et théorique	245
<b>CONCLUSION GENERALE</b>	<b>261</b>
<b>BIBLIOGRAPHIE</b>	<b>269</b>
<b>WEBOGRAPHIE</b>	<b>287</b>
<b>TABLE DES ILLUSTRATIONS</b>	<b>289</b>
<b>TABLE DES FIGURES</b>	<b>289</b>
<b>TABLE DES TABLEAUX</b>	<b>290</b>
<b>TABLE DES DESSINS</b>	<b>293</b>
<b>TABLE DES IMAGES</b>	<b>294</b>
<b>LISTE DES ABREVIATIONS</b>	<b>295</b>
<b>PRESENTATION DES ANNEXES</b>	<b>297</b>
<b>TABLE DES MATIERES</b>	<b>299</b>





# Introduction générale

« La didactique cognitive des langues s'assigne pour objectif de caractériser les processus et les stratégies sur le plan cognitif qui sous-tendent aussi bien l'acquisition/apprentissage des langues que les pratiques enseignantes. Elle vise à favoriser le développement d'un regard croisé entre, d'une part les recherches et méthodes utilisées en sciences cognitives – principalement en psycholinguistique et en psychologie cognitive – et d'autre part, les méthodes, besoins et impératifs liés à l'apprentissage des langues, et ce du côté de l'apprenant (...) comme du formateur » (Spanghero-Gaillard & Billières, 2007: avant propos)

Ce travail de recherche trouve son point de départ dans une pratique enseignante, celle du Français Langue Étrangère. Dans ce cadre, il apparaît que la compréhension orale est une compétence de premier ordre bien que trop souvent laissée de côté par la recherche en didactique des langues. En effet, l'oral correspond à la première modalité de la langue, tant chronologiquement (phylogénèse et ontogénèse) qu'au niveau de sa fréquence d'utilisation. Toutefois, cette modalité n'est pas toujours considérée comme un objet d'enseignement (Fasel Lauzon, Pekarek Doehler, & Pochon-Berger, 2009; Lafontaine & Pharand, 2011), et ce malgré des positionnements en ce sens (Dolz & Schneuwly, 1998; Grandaty, 2004; Lafontaine, 2000, 2005; N. Rey, 2004) et un intérêt grandissant dans les méthodologies dominantes dans l'enseignement des langues étrangères. C'est donc dans ce contexte que nous nous sommes intéressée à l'enseignement de la compréhension de l'oral et à son apprentissage.

Dans notre pratique enseignante, nous avons pu constater que les dialogues pédagogiques (destinés à des fins d'enseignement) étaient très fréquemment employés pour développer la compréhension orale des adultes apprenants, notamment débutants. Nous avons également remarqué que ces dialogues étaient souvent accompagnés d'un contexte sonore environnemental, sans toutefois qu'aucune préconisation pédagogique ne les accompagne en vue d'une utilisation optimale en classe de langue. Nous exposons ces deux aspects (emploi fréquent des dialogues pédagogiques et présence des contextes sonores environnementaux) à partir d'une enquête que nous avons réalisée, présentée dans le *Contexte didactique* (voir p.17). Comme leur utilisation est bien souvent laissée à la discrétion de l'enseignant, nous nous sommes interrogée sur : *Comment enseigner la compréhension orale ? Les supports de compréhension orale disponibles sur le marché sont-ils adaptés aux enjeux de cet enseignement/apprentissage ? Quelles sont les difficultés spécifiques des apprenants débutants en tâche de compréhension orale ? Comment l'évaluer ? Quelle(s) fonction(s) assument les contextes sonores environnementaux qui peuvent accompagner les dialogues pédagogiques ?* Ces interrogations correspondent donc à des problématiques de terrain, auxquelles nous avons tenté d'apporter des éléments de réponse à l'aide d'une méthodologie adaptée, en lien avec la démarche de la didactique cognitive.

Deux grands axes d'hypothèses ont été élaborés (voir chapitre *Hypothèses* p.105). Le premier s'organise autour des spécificités du public apprenant débutant : *qu'est-ce qu'être apprenant en langue ? Qu'est-ce qu'être débutant en langue implique en compréhension orale ?* En effet, si nous considérons que le rôle de l'enseignant est de permettre aux apprenants de développer de nouvelles compétences (Conseil de l'Europe, 2001), il est important de s'appuyer sur leurs connaissances et compétences déjà acquises. Dès lors, connaître leur niveau pour comprendre leurs besoins devient une nécessité. Notre deuxième axe d'hypothèses s'organise autour du rôle du contexte sonore environnemental dans la compréhension de dialogues oraux. *Ces contextes sonores représentent-ils une aide à la compréhension ? À quel niveau cette aide est-elle susceptible d'intervenir ?*

Afin de répondre à ces questions nous avons convoqué un cadre théorique principalement issu de la psychologie cognitive et de la psycholinguistique. Le premier chapitre de cette synthèse vise la mise en lumière des processus cognitifs qui régissent

les activités de perception en général, et la perception des sons environnementaux en particulier.

Dans ce chapitre, nous verrons que percevoir c'est percevoir un « objet » (Barabas, 1994; Shepard, 1981) et que dans cette activité, l'individu et ses connaissances (Barsalou, 2008; Berkeley, 1970; Dubois, 2003; Guillaume, 1979) jouent un rôle actif, et ce d'autant plus que nous nous situons dans la perspective d'une perception descendante, c'est-à-dire trouvant son point de départ chez l'individu, *via* la formulation d'hypothèses (Gaillard, 2000). L'action du sujet-percevant implique un traitement des stimuli qui lui parviennent à travers, notamment l'Analyse des Scènes Auditives (ASA; Bregman, 1990), elle-même dirigée par les processus attentionnels (Camus, 1996; Shamma & Micheyl, 2010; Sussman & Steinschneider, 2009; Thompson, Carlyon, & Cusack, 2011). Nous verrons également dans ce chapitre que la perception des sons environnementaux donne lieu à leur compréhension (Dubois, Guastavino, & Raimbault, 2006; Shamma & Micheyl, 2010). Ce chapitre nous permettra de conforter notre hypothèse sur l'aide que peut représenter un contexte sonore environnemental et de faire apparaître que perception et compréhension correspondent à deux processus co-dépendants.

Le deuxième chapitre du cadre théorique porte sur les mécanismes de la compréhension (voir p.81). Comprendre correspond à l'élaboration d'une signification nouvelle de la part du compreneur (Blanc & Brouillet, 2003; Chanquoy, Tricot, & Sweller, 2007; Le Ny, 2005). Ce processus d'élaboration fait intervenir à la fois les connaissances antérieures de l'individu et les éléments du perçu et c'est de leur interaction que naît le compris. Selon ce modèle, le perçu entraîne l'activation de représentations déjà élaborées et stockées en mémoire et dont la propagation conduit à l'activation d'un réseau de connaissances (Anderson, 1983). Dans ce chapitre, nous nous pencherons aussi sur la compréhension de la parole, c'est-à-dire sur les spécificités que le langage oral entraîne sur ce processus cognitif général. En effet, comprendre la parole implique une étape de reconnaissance des unités de la chaîne parlée (W. Marslen-Wilson & Warren, 1994; Norris, Cutler, McQueen, & Butterfield, 2006) avant d'élaborer une nouvelle signification à partir de ces unités (Culicover & Jackendoff, 2006; W. D. Marslen-Wilson, 1989). Nous nous interrogerons également sur les conditions de cette compréhension en salle de classe et à partir de supports artificiels (dans le sens de « construits pour l'enseignement et l'apprentissage ») dans lesquels le langage est à la fois produit (G.A. Miller & Chomsky, 1963; Tanenhaus & Brown-Schmidt, 2008) et action (Austin, 1970; Grice, 1957; Tanenhaus & Brown-Schmidt, 2008). Enfin, nous verrons en quoi la compréhension de ces supports particuliers est à rapprocher de la compréhension de textes narratifs (Charaudeau, 1992) et dans quelle mesure les

modèles qui découlent de ce courant de recherche peuvent être convoqués pour considérer notre objet de recherche —principalement le modèle de situation (Kintsch, 1998; Kintsch, 1988; Van Dijk & Kintsch, 1983) ; et le modèle *Landscape* (Magliano, Zwaan, & Graesser, 1999; Rolf A. Zwaan, Magliano, & Graesser, 1995)

Ces rappels théoriques nous permettent donc de mieux cerner nos deux principales questions de recherche : Qu'est-ce qu'être apprenant débutant en langue ? Le contexte sonore environnemental qui peut accompagner les dialogues pédagogiques représente-t-ils une aide à la compréhension ?

Nous avons mis en place un protocole expérimental nouveau afin de répondre à ces questions (partie *Expérimentation principale* p.165). Nous avons en effet élaboré un test qui vise l'évaluation de la compréhension orale, que nous avons appelé le Classe-Images. Cette évaluation passe par une tâche de sélection de dessins pour rendre compte de dialogues entendus (et compris). Différents groupes de population participent à ce test : un groupe de participants francophones qui écoutent les dialogues présentés seuls, un groupe de participants francophones qui écoutent les mêmes dialogues accompagnés d'un contexte sonore environnemental ; un groupe de participants apprenants débutants en FLE soumis aux dialogues seuls, et un groupe de participants apprenants débutants en FLE face aux dialogues accompagnés d'un contexte sonore environnemental. À partir de la comparaison des résultats obtenus par chaque groupe, nous obtenons des éléments de réponse sur nos deux axes d'interrogations. La comparaison des résultats des francophones et des apprenants débutants en FLE nous permettra de définir plus précisément les implications d'un faible niveau d'expertise linguistique sur la compréhension de dialogues courts. D'autre part, la comparaison des résultats des participants ayant écouté les dialogues seuls ou accompagnés d'un contexte sonore environnemental nous permettra de voir si la présence de ce contexte a donné lieu à de meilleures performances dans la tâche.

Afin de contrôler un maximum de paramètres de ce test, et de développer des compétences dans l'élaboration de matériel pédagogique, nous avons fait le choix de construire entièrement les supports de ce test. Trois dialogues ont été élaborés, ainsi que les scènes sonores environnementales qui les accompagnent. Ces élaborations ont, à leur tour, fait l'objet d'une démarche expérimentale que nous présenterons dans la partie *Élaboration du matériel expérimental* (p.119).

Les principaux résultats obtenus nous permettent, comme nous le verrons dans le chapitre *Expérimentation principale* (p.165) et dans la *Discussion* (p.211), de

considérer qu'un déficit de connaissances en langue entraîne une augmentation de la charge cognitive dont les principales conséquences sont visibles principalement sur l'appréhension d'une part de la situation d'énonciation (à laquelle participe le cadre spatio-temporel) et d'autre part, du contenu propositionnel des dialogues. Nous constatons par ailleurs que la présence d'un contexte sonore environnemental implique le plus souvent de meilleures performances sur les deux aspects cités ci-avant. Cette amélioration est notamment notable au niveau des « profils d'erreurs », c'est-à-dire des difficultés observées dans chaque groupe de participants.

Ces observations nous amèneront, enfin, à proposer des pistes didactiques pour un enseignement stratégique de la compréhension orale en classe de FLE transposables en classes de langues étrangères.



# **Contexte Didactique : L'oral en Français Langue Étrangère**

Notre travail de recherche porte sur la compréhension orale auprès d'apprenants débutants et notre démarche s'inscrit dans le cadre de la didactique cognitive des langues. Notre objectif est de questionner et de mieux connaître les mécanismes en jeu lorsque ces apprenants sont confrontés à des dialogues pédagogiques. Dans cette première partie, nous tenterons dans un premier temps de définir l'oral (sans pour autant l'opposer nécessairement à l'écrit) et verrons dans quelle mesure il est, pour nous, une modalité première de la langue. Dans un second temps nous observerons l'évolution de la place de l'oral en classe de langue, à travers l'étude de trois grandes méthodologies d'enseignement/apprentissage des langues étrangères qui constituent la didactique du français langue étrangère depuis les années 1960 en France (la méthodologie Structuro-Globale Audio-Visuelle, les approches communicatives et la perspective actionnelle). Enfin, dans un troisième temps, nous présenterons les différents supports de compréhension orale utilisés en salle de classe à l'heure actuelle, notamment avec les apprenants débutants.

## **1. Qu'est-ce que l'oral**

L'oral peut être considéré comme la première modalité de la langue tant du point de vue de sa fréquence d'utilisation qu'au niveau de son apparition (tant phylogénétique qu'ontogénétique). Toutefois, en didactique des langues, les activités pédagogiques portant sur l'oral restent trop souvent mal définies. Or, malgré l'intérêt des chercheurs pour ce domaine et les difficultés qui l'accompagnent, l'oral est une modalité primordiale

dans l'enseignement/apprentissage des langues et doit être considéré comme un objet d'enseignement à part entière, tant en production qu'en compréhension. Parallèlement, les recherches en Didactique des Langues Étrangères (désormais DLE) semblent moins nombreuses que celles en didactique des langues maternelles sur ce sujet. Nous nous appuyons donc ici tant sur les travaux portant sur les langues maternelles que ceux portant sur les langues étrangères.

## **1.1. L'oral comme modalité première de la langue**

### ***1.1.1. L'antériorité de l'oral sur l'écrit (phylogénèse et ontogénèse)***

Du point de vue chronologique de la phylogénèse (Rondal, Esperet, Gombert, Thibaut, & Comblain, 1999), l'oral est la première modalité de la langue. En effet, au niveau du développement de l'espèce, « *on considère généralement que les hommes utilisent le langage oral, dans la communication linguistique depuis au moins 30 000 ans* » (Gineste & Le Ny, 2002:1-2). L'oral est donc antérieur à l'écrit dont on situe l'origine à environ 3000 ans avant notre ère chez les Sumériens (Auroux, 2000). Ainsi, « *l'expression écrite est donc une invention humaine toute récente* » par rapport à l'oral (Gineste & Le Ny, 2002:4). De plus, du point de vue de l'étendue des pratiques, l'oral est également dominant : à l'échelle de l'espèce humaine, il y a plus de locuteurs que de scripteurs. Deux facteurs explicatifs peuvent être avancés. D'une part, un grand nombre de langues ne disposent pas de système d'écriture alors que, comme nous le rappelle Durand (2000), on ne connaît —du moins aujourd'hui— aucune société ne disposant pas de système linguistique oral. D'autre part, le taux d'analphabétisme (incapacité ou difficulté à lire, écrire et compter faute d'apprentissage) dans le monde reste important. En 2008, le taux d'alphabétisation des adultes était de 83,6%<sup>1</sup> et la prévision pour 2015 est de 87%. Ainsi, en 2014, entre 13 et 16% de la population mondiale ne dispose que de l'oral, soit environ 774 millions d'adultes. L'oral est donc une modalité primordiale dans la pratique des langues.

Au niveau du développement ontogénétique (à l'échelle de l'individu), il apparaît que l'oral est également premier. En effet, l'enfant acquiert généralement le langage oral entre 9 mois et 3 ans. Cet apprentissage est long, en raison notamment de la complexité du système à apprendre par l'enfant (Jisa, 2003). En revanche, l'apprentissage de

---

<sup>1</sup> Site de l'ONU présentant l'actualité de l'alphabétisation : <http://www.un.org/apps/newsFr/storyF.asp?NewsID=17467&Cr=alphab%C3%A9tisation&Cr1=UNESCO#.Uf51woWiE7A>

l'écriture se fait plus tardivement (à partir des 6 ans de l'individu généralement) et le plus souvent par le biais de la scolarisation. Il s'agit d'un apprentissage dirigé, explicite et programmé, le plus souvent assumé par une personne extérieure au giron familial de l'enfant (Durand, 2000) et dont la durée varie sensiblement d'une langue à l'autre, notamment en fonction de la transparence du système orthographique de la langue (Forst, Katz, & Bentin, 1987; Forst, 1998; Vialettes, 2012).

Ainsi, « (...) *la parole est ontogénétiquement et phylogénétiquement première par rapport à l'écriture, dont l'une des fonctions primordiales est précisément d'en constituer une certaine forme de matérialisation.* » (Detey, 2007:23). On peut donc parler de primauté de l'oral (Leroy, 1985) dans les usages. Cependant, il apparaît que l'oral reste bien souvent défini **par rapport** à l'écrit (Durand, 2000) et sa position d'objet d'apprentissage reste à préciser tant en langue maternelle qu'en langue étrangère. En effet, « *si la compréhension orale est première dans l'acquisition de la langue maternelle et est considérée comme une compétence de base dans la conception occidentale de l'enseignement apprentissage des langues, elle n'en est pas moins souvent considérée comme la compétence la plus difficile à acquérir par les apprenants. Pourtant, celle-ci n'est pas l'objet d'un apprentissage spécifique et elle est peu présente dans les manuels et dans les cours en comparaison de la compétence de lecture.* » (Berdal-Masuy & Briet, 2010:16). Il apparaît donc que l'oral en général, et la compréhension orale en particulier, restent les parents pauvres de l'enseignement des langues tant maternelles qu'étrangères.

### ***1.1.2. L'oral, un objet mal défini***

Malgré sa primauté (Gueunier, 2005) l'oral reste bien souvent un objet mal défini par la didactique des langues. Trop souvent laissé de côté dans l'enseignement et l'apprentissage, il n'a été intégré que tardivement à l'enseignement du Français Langue Maternelle (désormais FLM) (Bergeron, 2000). C'est sans doute une des raisons de la frilosité à l'égard de l'oral que nombre de chercheurs ont *diagnostiquée* (Gueunier, 2005; Parpette & Peutot, 2006; Parpette, 2008 par exemple). Aussi, on peut qualifier l'oral d'« *OVMi : objet verbal mal identifié. Longtemps « non objet », ni didactique, ni pédagogique, l'oral est aujourd'hui une espèce d'objet verbal mal identifié (...)* » (Halté, 2005:12).

Jusqu'au début du XX<sup>ème</sup> siècle et l'arrivée de la linguistique saussurienne en France (et des travaux de Bloomfield aux États-Unis), l'étude de la langue était avant tout normative et uniquement fondée sur les textes littéraires canoniques (Durand, 2000). En d'autres termes, l'oral n'était quasiment pas considéré comme un objet d'étude du langage (en raison notamment de l'absence d'outils techniques permettant de

l'enregistrer en vue de son étude). Il est encore moins conçu comme un objet d'apprentissage, et absolument pas comme objet d'enseignement. Sa participation à la didactique des langues étrangères est toute récente avec une introduction timide par la méthodologie directe de la fin du XIX<sup>ème</sup> siècle en Europe (Germain, 1993). Toutefois, l'oral est aujourd'hui un réel objet d'études tant dans les sciences du langage en général (voir (Blanche-Benveniste, 1985, 1990, 2010; Blanche-Benveniste, Bergounioux, Chevalier, Dumont, Encrevé, Perdue, & Simoni-Arembou., 1992; C Germain, 1993; Kerbrat-Orecchioni & Traverso, 2004 par exemple), qu'en Didactique des Langues (DDL) en particulier.

L'introduction d'activités orales comme objets d'enseignement s'est principalement faite dans les années 1970 avec le recours aux documents authentiques en classe de langue (Cuq & Gruca, 2002). Mais cette introduction n'a pas stabilisé l'oral comme objet d'enseignement ou de recherche en DDL puisque, comme nous le rappelle (Nonnon, 1999), entre 1980 et 1990, les recherches sur l'oral sont passées de 25% à 10% des recherches sur le français. Un deuxième embarras idéologique et technique que l'on peut identifier à propos de l'oral tient à sa nature *immatérielle*. Insaisissable par sa nature et son statut social, l'oral est aussi difficile à maîtriser en situation de classe : que ce soit en compréhension ou en production, l'oral est toujours « *en direct* » et échappe à la prévisibilité (Billières, 2009). L'oral est ainsi difficilement contrôlable, empreint d'émotion et dans une temporalité imposée. Par ailleurs, l'oral reste également problématique dans la formation initiale des enseignants (cf. par exemple Lafontaine, 2005 pour le cas du FLM au Québec). Ces derniers se trouvent donc parfois en difficulté lors d'activités portant sur l'oral. Ainsi, nombreuses sont les idées reçues chez les enseignants à propos de l'oral, encore aujourd'hui.

Jusqu'à récemment, les travaux sur l'oral en DDL étaient encore peu nombreux et les supports d'enseignement proposés par les manuels relevaient plus de présentations parsemées que d'une réelle approche méthodologique. Encore aujourd'hui, nombreuses sont les approches de l'oral qui passent par l'écrit en postulant un transfert de compétences (Bergeron, 2000), nombreux aussi sont les dialogues pédagogiques (sensés être des représentations d'interactions orales spontanées entre des natifs de la langue cible) qui sont plus calqués sur l'écrit que sur les interactions orales attestées (Onursal, 2008; Parpette, 1997, 2008). Mais, il est aussi fréquent de voir l'inverse : que l'oral soit au service de l'écrit. « *Ainsi, de prime abord, l'oral occupe une place primordiale non seulement pour faire apprendre à communiquer oralement mais, également, pour faire apprendre à lire (les rapports son-graphie) et pour faire apprendre à écrire (les structures écrites n'étant qu'une transposition, en début d'apprentissage d'une L2, des structures*

*appries tout d'abord à l'oral*» (Germain & Netten, 2005:8). Oral et écrit sont donc souvent entremêlés dans la DDL et peu de place est accordée à l'oral en tant que compétence propre (et encore moins à la **compréhension** orale cf. Clerc, 1999:49). « *Il est plus facile d'utiliser l'oral pour aller vers la lecture et l'écriture que de l'enseigner* » (Lafontaine, 2005). Et, comme le souligne Sylvain Detey (2007:24) : « (...) *même s'il existe de nombreuses similitudes entre la perception de la parole orale et la perception de l'écrit, chacune possède ses caractéristiques propres, ce qui conduit, d'un point de vue didactique à devoir entraîner les étudiants aux deux, si l'on souhaite développer chez eux les deux types de compétences* ».

Dans cet entremêlement, deux écueils semblent être à éviter : **ne pas distinguer l'oral et l'écrit** (estimer que ces deux formes n'impliquent pas de différences dans l'enseignement/apprentissage) et **considérer que l'oral et l'écrit sont tout à fait différents et qu'il n'y a pas de transfert possible**. « *Il apparaît en effet que la langue n'est ni orale, ni écrite : elle est cette connaissance partagée par ses locuteurs linguistiquement compétents (au sens chomskyen) (...). Cependant, cette langue s'incarne à travers les manifestations audibles et visibles que sont la parole et l'écriture (et nous devrions également en mentionner les autres manifestations corporelles, en particulier mimo-gestuelles)* » (Detey, 2007:22). Ainsi, l'enseignant de langue doit tout à la fois enseigner la langue cible dans sa globalité et recourir à des approches différenciées selon qu'il traite de l'écrit (et de toutes les formes d'écrit) ou de l'oral (et de toutes les formes d'oral).

La question de la didactique de l'oral reste entière. Reconnu comme objet d'apprentissage, le statut d'objet d'enseignement ne lui est pas encore toujours accordé. Cinq questions résument ainsi la problématique de l'enseignement de l'oral : faut-il l'enseigner ?; Comment le définir ? Quel oral enseigner ? Peut-on l'enseigner ? Comment ? (Roulet, 1991).

Comme nous le verrons, ces interrogations bien que formulées il y a plus de 20 ans, restent d'actualité, puisque, comme le dit Boulton (2009:6) « (...) *la compréhension orale reste une aptitude sous-entraînée – même si on peut espérer que cet état des choses se fait plus rare qu'il y a trente ans* ».

## 1.2. L'oral en classe de langue

Bien qu'il soit aujourd'hui généralement admis que l'oral doit être enseigné, la question du « comment » reste posée. Ainsi, comme nous le font remarquer Germain & Netten (2005:8), nous pouvons nous interroger sur l'enseignement de l'oral : « *si l'on accepte qu'une langue est d'abord une habileté servant à communiquer, se pose alors la question du mode d'acquisition de l'habileté à communiquer oralement en L2. Une habileté ne peut s'acquérir que par son utilisation* ».

### 1.2.1. L'oral, objet d'apprentissage... mais pas toujours d'enseignement

L'enseignement de l'oral n'a pas réellement fait l'objet d'un travail de systématisation. Il est souvent laissé à la discrétion de l'enseignant (Lafontaine, 2005:97). Reconnu en tant qu'objet d'apprentissage, il n'est pas encore toujours considéré comme objet d'enseignement (Fasel Lauzon, Pekarek Doehler, Pochon-Berger, & Steinbech Kohler, 2009; Lafontaine & Pharand, 2011), malgré de nombreuses prises de position dans ce sens (par exemple Dolz & Schneuwly, 1998; Grandaty, 2004; Lafontaine, 2000, 2005; Rey, 2004). Toutefois, il apparaît que l'ensemble de ces travaux renvoient majoritairement à la production orale (désormais PO), les travaux sur la compréhension orale (désormais CO) relevant plus de la didactique des langues étrangères que de l'enseignement des langues maternelles. En effet, la compréhension orale, en langue maternelle semble considérée comme relevant de l'intuitif, ou du moins, des apprentissages et acquisitions préscolaires. C'est sans doute l'une des explications de la rareté des travaux sur la CO en didactique des langues étrangères aussi. Or, la distinction compréhension/production orales semble primordiale (Carton, 1995; Gremmo, 1986/7). En effet, différencier oral et écrit s'avère nécessaire mais insuffisant pour une approche méthodologique de l'enseignement des langues étrangères : il faut également discriminer les activités de compréhension et de production, entraîner les apprenants à chacune des compétences fondamentales (Cuq & Gruca, 2002) qui placent l'apprenant tantôt en locuteur, tantôt en compreneur, tantôt en interlocuteur.

Parpette (2008) et Porcher (1995) nous rappellent que l'écoute<sup>2</sup>, reste l'habileté à la fois la plus indispensable, eu égard à la fréquence de son utilisation dans des

---

<sup>2</sup> D'un point de vue terminologique, il importe ici de distinguer l'écoute (prise en compte volontaire des indices acoustiques et implique donc une certaine intervention de l'attention) de la réception (identification des unités de sens du discours cible (voir chapitre *des études sur la perception* p.47)) et de la compréhension (les informations traitées intègrent le système de connaissances et implique donc une certaine intervention de la mémoire (voir chapitre *... aux études sur la compréhension* p.79)) de l'oral.

situations naturelles, et la plus difficile à acquérir, notamment du fait de son exigence dans ses conditions de réussite (développé plus bas). Malheureusement, il semble que la majorité des travaux en didactique de l'oral porte sur la production, et non sur la compréhension. On peut même noter que parfois, la CO est envisagée comme une étape du développement de la PO (Bergeron, 2000; Fasel Lauzon, Pekarek Doehler, Pochon-Berger, *et al.*, 2009), mais encore une fois, peu d'intérêt semble porté à la CO en tant que telle. En effet, Parpette (2008) nous fait remarquer que de nombreux exercices et activités entraînent notamment l'apprenant à la mobilisation rapide de connaissances afin de développer une réactivité langagière et des compétences liées à la production d'un discours oral à la fois fluide et adapté à la situation. Mais rien de semblable n'est proposé pour la CO. Cependant, « (...) *C'est en effet en traitant par construction et vérifications d'hypothèses, ce qu'il aura saisi des comportements langagiers auxquels il est exposé que l'apprenant fera progresser sa connaissance de la langue à acquérir (il s'agit bien de connaissances **de** langue et non **sur** la langue)* » (Holec, 1990:67).

Un autre écueil du traitement de la CO par la didactique des langues étrangères est la non prise en compte du caractère fugace de l'oral (Parpette, 2008). En effet, si de nombreux exercices sont proposés dans les manuels sur les dimensions phonétique et sémantique de l'oral, notamment *via* des dialogues pédagogiques, peu de supports permettent à l'enseignant de travailler la réactivité immédiate que suppose la CO. L'oral est en effet un flux continu (voir chapitre ... *aux études sur la compréhension* p.81) sur lequel l'auditeur ne peut que rarement intervenir, par exemple, les répétitions ne sont possibles que dans les cas d'interaction directe, et ne peuvent pas intervenir pendant une conférence, au cinéma, au théâtre ou lors de l'écoute de la radio « en direct ». Ainsi, la construction de sens par les apprenants passent souvent pas des écoutes successives et séquencées des supports oraux (É. Nonnon, 2004; Parpette, 2008). Or, recourir à la segmentation et à la répétition laisse peu de place à l'élaboration, par les apprenants, de stratégies d'écoute permettant d'appréhender la fugacité de l'oral. **Selon nous, une attention portée à la mise en place d'un contexte dans lequel prend place une parole (un contexte sonore environnemental par exemple) pourrait contribuer à enseigner certaines stratégies pertinentes pour la compréhension linguistique orale.** En effet, si des stratégies de compréhension en direct sont développées dans le plus jeune âge, lors de l'acquisition de la langue maternelle (désormais LM), il semblerait hasardeux de les considérer comme définitivement acquises et disponibles pour l'apprentissage d'une langue étrangère. En outre un adulte apprenant en situation d'immersion n'est pas dans la même situation qu'un jeune enfant natif (les besoins

langagiers étant sensiblement différents) et il revient donc à l'enseignant de l'entraîner sur l'ensemble de ces aspects et d'étayer son apprentissage (Vallat, 2012).

Il nous apparaît donc que l'oral doit être un objet d'enseignement et que la compréhension orale gagnerait à faire l'objet d'une attention particulière. Dit autrement, nous nous positionnons en faveur d'une élaboration méthodologique permettant de travailler l'oral en général, et la CO en particulier. Ces apports méthodologiques peuvent prendre la forme d'études sur la compréhension orale (comme celle que nous nous proposons de présenter dans ce travail (voir partie *Expérimentation principale* p.165), et d'outils pédagogiques accompagnés de pistes quant à leur exploitation (voir *Discussion didactique et théorique* p.245).

La question de l'enseignement de l'oral soulève celle de son évaluation, quelle soit *diagnostique* (qui permet aux enseignants d'évaluer le niveau et les besoins des apprenants), *formative* (qui permet aux enseignants comme aux apprenants de prendre conscience des acquis en construction et éventuellement de proposer une remédiation) ou *sommative* (qui permet aux enseignants et aux apprenants de dresser le bilan des enseignements et des apprentissages). Nous allons donc maintenant nous pencher sur l'évaluation de l'oral.

### ***1.2.2. L'évaluation de l'oral: une difficile distinction entre compréhension et production***

Les compétences orales sont difficiles à mesurer. Ce constat peut en partie s'expliquer par les mêmes raisons qui les rendent difficiles à enseigner. Il n'en reste pas moins qu'il est nécessaire d'en effectuer l'évaluation, c'est-à-dire de prendre conscience, pour l'enseignant comme pour l'apprenant, de ce qui est acquis, ce qui est en cours d'acquisition et ce qui reste à acquérir. Mise en place régulièrement, l'évaluation permet aux acteurs de la situation pédagogique (enseignant(s) et apprenant(s)) d'adapter la progression et la méthodologie de travail.

Que se soit en Français Langue Maternelle (désormais FLM) ou Langue Etrangère (désormais FLE), en production ou en compréhension, l'oral reste toujours délicat à évaluer tant il est personnel et soumis à la variation. Claudine Garcia-Debanc, (1999:194-199) a listé dix principales sources de difficulté pour l'évaluation de la **production** orale en FLM. Parmi ces dix items, cinq peuvent s'appliquer à la compréhension orale en FLE :

« - La pratique de l'oral est transversale à toutes les disciplines et à toutes les situations

- L'oral est difficile à observer et complexe à analyser

- L'évaluation de l'oral nécessite beaucoup de temps

- L'évaluation de l'oral suppose un détour par l'écrit par le biais des retranscriptions

- L'oral est souvent mal connu ».

Le premier point soulevé par l'auteure tient à la transdisciplinarité. Or la CO est tout aussi transversale que la PO : au niveau de la situation de classe, les cours de FLE sont le plus souvent (dans les situations d'immersion du moins) donnés en langue cible et à l'oral. Dans ce cadre, la CO est une compétence sans cesse sollicitée, et si l'apprenant suit d'autres programmes de formation (non-linguistique), la CO en langue cible sera de premier ordre. De plus, en dehors de la salle de classe, en situation d'immersion, la CO est requise très fréquemment, aussi bien dans la vie de tous les jours, que dans les situations professionnelles. Le deuxième point avancé par Claudine Garcia-Debanc est la difficulté à observer l'oral. La CO, par le fait qu'elle engendre le plus souvent des manifestations non-verbales est tout aussi difficile à observer. En outre Lafontaine et Pharand (2011:45) nous rappellent que le *programme de formation de l'école québécoise* (MEQ-2001) propose comme critère d'évaluation de la CO d'observer les « *réactions témoignant d'une écoute efficace* » (Ministère de l'Éducation du Québec, 2006:82), sans toutefois clairement définir ces réactions. La durée liée aux évaluations de l'oral et à sa préparation est un autre point soulevé par les auteures. Évaluer la CO implique la définition des éléments à évaluer, la construction d'un support audio, mais aussi la définition de la forme de l'évaluation permettant d'isoler les connaissances interrogées. Un QCM par exemple, fait intervenir non seulement la compréhension orale, mais aussi la compréhension écrite. Dès lors, comment s'assurer que les éventuelles difficultés relèvent de la CO et non de la lecture ? Il en va de même pour les questions ouvertes pour lesquelles un discours, oral ou écrit, même court, est attendu en réponse. Ce point est à rapprocher de la 4<sup>ème</sup> source de difficulté mise en avant par Garcia Debanc (1999). Enfin, la difficulté suivante correspond à la méconnaissance de l'oral. Ce point semble commun à la PO et à la CO, et est peut-être même plus fort au niveau de la CO du fait de la moindre proportion des recherches qui lui ont trait.

Malgré ces difficultés, il semble indispensable d'évaluer l'oral et Claudine Garcia-Debanc, (1999 : 199-200) nous en donne six bonnes raisons, dont quatre seront retenues ici.

*« - Qu'il soit ou non enseigné, l'oral est évalué, notamment dans les examens ou les entretiens d'embauche, de sortes que les enjeux sociaux de la maîtrise de l'oral sont immenses*

*- Une évaluation objective est nécessaire aux enseignants pour procéder à un étayage efficace*

*- Une évaluation objective est nécessaire aux enseignants pour programmer efficacement un enseignement de l'oral*

*- Une évaluation objective est nécessaire aux élèves pour savoir comment progresser ».*

Ces différents arguments en faveur de l'évaluation de l'oral, bien que proposés pour la PO en langue maternelle correspondent tout à fait à nos préoccupations. Concernant l'évaluation de l'oral en dehors de la salle de classe (premier point soulevé par l'auteure), il apparaît que dès l'instant où l'apprenant de FLE se trouve dans un pays francophone, sa compréhension orale sera sollicitée à tout moment et lui sera indispensable pour son intégration. La compréhension orale est donc primordiale, non seulement en raison de sa fréquence, mais aussi pour son caractère social et interpersonnel et pour sa dimension cognitive (É. Nonnon, 2004). Les deux points suivants concernent l'étayage adapté, potentiellement fourni par l'évaluation. En effet, évaluer les performances des apprenants permet à l'enseignant de connaître leurs difficultés, et c'est à partir de cette connaissance qu'il peut mettre en place différents procédés d'enseignement/apprentissage, une progression répondant réellement aux besoins des apprenants. Enfin, le dernier point porte sur l'apport de l'évaluation à l'apprenant. Si l'enseignant a besoin de l'évaluation pour élaborer la progression, l'apprenant a besoin, lui, de l'évaluation pour se positionner vis à vis de ses savoirs acquis ou en cours. Ceci est en outre particulièrement important puisque, en tant qu'enseignante nous attendons des apprenants une attitude volontaire d'implication dans l'apprentissage. Il faut donc leur en donner les moyens.

## 2. La place de l'oral dans trois grandes méthodologies d'enseignement/apprentissage des langues étrangères

Jusqu'à la fin du XIX<sup>ème</sup> siècle, la méthodologie dominante dans l'enseignement des langues étrangères était la **méthode traditionnelle**. Principalement basée sur les textes littéraires considérés comme le canon, cette méthodologie a laissé l'oral de côté. Les enseignements, reposant sur l'exercice du thème et de la version, étaient tous dispensés en langue maternelle. À la fin du XIX<sup>ème</sup> siècle, la **méthode traditionnelle** se voit remise en question par la **méthodologie naturelle**. Celle-ci s'oppose radicalement à la précédente en considérant que l'apprentissage d'une langue étrangère (LE) doit reproduire les conditions d'apprentissage de la langue maternelle. C'est ce changement radical entre ces deux méthodologies qui donnera naissance à la didactique des langues étrangères et à la **méthodologie directe**. L'objectif de cette dernière est « *la maîtrise effective de la langue en tant qu'instrument de communication* » (Vallat, 2012:103). Mais cette méthodologie est mal acceptée par les enseignants (Puren, 1994) et de ce refus naît la **méthodologie active** qui apporte quelques modifications à la **méthodologie directe**.

C'est dans ce contexte qu'apparaît au cours de la Seconde Guerre Mondiale, la **méthodologie audio-orale (MAO)**. Ce courant méthodologique est issu de l'*army method* conçu par l'armée américaine pour former ses soldats en LE (Cuq & Gruca, 2002). Il proposait la compréhension globale et la mémorisation de dialogues pédagogiques avant l'analyse grammaticale. Dans les années 1940, les nouveaux outils techniques, l'intégration des apports de la psychologie (du courant béhavioriste principalement) d'une part et de la linguistique de Bloomfield, d'autre part, viennent enrichir la méthode de l'armée en la convertissant définitivement en MAO. L'enseignement de l'oral prend alors une réelle place dans l'enseignement des langues étrangères.

### 2.1. L'oral dans la méthodologie structuro-globale audio-visuelle

À la suite des MAO, les méthodologies audio-visuelles (MAV) ont émergé, et plus particulièrement la méthodologie structuro-globale audio-visuelle (SGAV) développée autour de Petar Guberina (à Zagreb), Paul Rivenc (à St Cloud) et Raymond Renard (à Mons). La MAV « *sera définie ici comme cette méthodologie dominante en France dans les années 1960-1970, et dont la cohérence est construite autour de l'utilisation conjointe de*

*l'image et du son.* » (Puren, 1988:284). Tout comme pour le fondement des MAO aux États-Unis, c'est l'introduction de la linguistique dans l'enseignement des langues qui marque un réel tournant. En 1951, le Centre d'Étude du Français Élémentaire est créé et chargé de mener une étude sur le lexique et la grammaire du français permettant de favoriser son apprentissage et par là même, sa diffusion. Georges Gougenheim et Paul Rivenc sont chargés de diriger ces recherches qui aboutiront à la rédaction du Français Fondamental (désormais FF) Premier degré (1475 mots) et Deuxième degré en (1609 mots) en 1954 (Gougenheim, Michéa, Rivenc, & Sauvageot, 1956 ; Ministère de l'éducation nationale, 1970, 1972). C'est la première fois qu'une telle étude est menée sur le français, et sa base méthodologique (appliquer la statistique à un corpus de conversations spontanées) est à l'époque révolutionnaire. À partir du FF, les objectifs d'enseignement en FLE répondront à des objectifs d'utilité et d'efficacité. « (...) *il s'agit pour l'enseignant de mettre l'apprenant en contact avec la culture quotidienne des natifs de la langue cible, et non plus de lui présenter seulement les représentants de lettres et de philosophie de cette culture* » (Spanghero-Gaillard & Garnier, 2013:115).

En 1959, le CRÉDIF (Centre de Recherche et d'Études pour la Diffusion du Français) est créé avec à sa tête les deux principaux auteurs du FF : G. Gougenheim et P. Rivenc. La même année naît le BEL (Bureau d'Étude et de Liaison) qui deviendra, en 1966 le BELC (Bureau d'Enseignement pour la Langue et la Civilisation encore aujourd'hui très actif dans la formation des enseignants de FLE). Ainsi, une nouvelle dynamique est donnée à la didactique des langues étrangères et la didactique du FLE devient un véritable objet de recherche.

Dans la méthodologie SGAV, le noyau dur de toute leçon est le dialogue pédagogique qui met en scènes des interactions du quotidien et dont les répliques sont élaborées en référence au FF (les mots les plus fréquents sont les premiers introduits). Ce dialogue est accompagné d'un film d'images fixes qui sert à l'illustrer notamment à travers les dimensions spatio-temporelle, sociale et psychologique. Les apprenants sont donc plongés dans une situation de perception *globale* de la scène. Cette globalité renvoie aux fondements de la théorie psychologique de la Gestalt (théorie de la forme (Guillaume, 1979)) qui propose que le tout ne saurait être égal à la somme des parties (et les parties ne sauraient être réductibles à des fractions de ce tout). Nous percevons globalement une scène et c'est à partir de cette appréhension globale que nous construisons une interprétation. Ainsi, la méthodologie SGAV incite l'enseignant à exploiter les éléments prosodiques, mimo-gestuels et contextuels présents dans les dialogues accompagnés d'illustrations. Dans ce cadre, l'étude de la langue se veut onomasiologique (du sens vers la forme) et l'apprentissage de la grammaire par exemple est inductif et se base sur des énoncés apparus en situation dialoguée. Ainsi, dans la

préface de *Voix et Images de Frances* (VIF) (Rivenc & Guberina, 1962) la grammaire est définie comme « *résolument fonctionnelle* » puisque « *tout en se conformant scrupuleusement à l'usage, elle vise un enseignement pratique* » (cité par Avram, 2006:57). Bien qu'inductif, l'enseignement/apprentissage de la grammaire est soutenu par des exercices structuraux qui visent l'automatisation des réflexes langagiers, tout en considérant le langage comme un comportement (Coste, 1970:13).

La méthodologie SGAV se caractérise également par sa progression scrupuleuse, tant au niveau global (d'une séance à l'autre, considéré ici comme le niveau mezzo) qu'au niveau de l'organisation de chaque classe (niveau micro)<sup>3</sup>. La leçon SGAV s'articule en effet en cinq phases successives, cinq moments de la classe : la présentation, l'explication, la répétition, l'exploitation (ou le réemploi) et la phase de transposition (Avram, 2006; Cuq & Gruca, 2002; Ivan & Avram, 2006). Enfin, une des caractéristiques les plus innovantes de la méthodologie SGAV est le décalage entre l'introduction de l'oral et celle de l'écrit. En effet, les formes écrites sont proposées aux apprenants après plusieurs leçons. Grâce à la méthodologie SGAV depuis les années 1960, la priorité est accordée à la Compréhension Orale auprès des apprenants de niveaux débutants.

L'oral, et la CO en particulier sont donc mis sur le devant de la scène et le dialogue pédagogique (enregistré et illustré) est le principal support de cours (Puren, 1988). Pour ces différentes raisons, ce courant méthodologique est à la base de notre réflexion.

## **2.2. L'oral dans les approches communicatives**

Les Approches Communicatives (AC) s'inscrivent dans la continuité de la méthodologie SGAV, mais au début des années 70 la DDL s'émancipe de la linguistique structurale et considère comme disciplines de références la linguistique de l'énonciation, l'analyse du discours et la pragmatique. La publication du *Niveau Seuil* (Coste, Courtillon, Ferenczi, Martins-Baltar, Papo et Roulet, 1976) constitue le point de départ de cette nouvelle approche. « *À la suite des méthodologies audio-visuelles, ces approches ont intégré des enregistrements sonores, des reportages vidéo (...)* » (Spanghero-Gaillard & Garnier, 2013:115). Dans l'élaboration des supports audio, il s'agit de procéder aux enseignements à partir des actes de parole adaptés aux situations. De plus, la notion de besoin devient primordiale et les AC placent comme objectif principal la compétence de

---

<sup>3</sup> Nous considérons le niveau macro comme l'ancrage des savoirs langagiers dans les connaissances générales de l'individu. Nous empruntons cette terminologie à Andrée Tiberghien (2013)

communication en langue étrangère : savoir communiquer ne se réduit pas à connaître la langue, il faut aussi connaître les règles d'emploi de cette langue (Onursal, 2008).

La compétence de communication se base sur quatre composantes (Cuq & Gruca, 2002) : la composante linguistique (structures de la langue), la composante sociolinguistique (règles socio-culturelles d'emploi des structures linguistiques, correspondance avec la situation d'énonciation), la composante discursive (cohérence et cohésion des discours) et la composante stratégique (stratégies verbales ou non verbales pour compenser les défaillances liées à l'apprentissage de la langue cible)

Au niveau de la composante linguistique, la grammaire redevient déductive et est abordée dans une approche notionnelle. Toutefois, les exercices structuraux qui visent l'automatisation des processus langagiers perdurent, bien que moins présents que dans le SGAV.

Par ailleurs, dans ce nouveau courant, une approche différenciée est encouragée et les matériaux d'apprentissage sont plus variés afin de s'adapter aux besoins de chaque apprenant. Le dialogue pédagogique n'est donc plus au centre du dispositif pédagogique ; les documents authentiques sont introduits comme supports d'enseignement/apprentissage (voir p.35 pour plus de détails). La progression est beaucoup plus souple qu'avec les MAV : il revient à l'enseignant de s'adapter au public apprenant. L'oral garde une place de choix dans ces approches si les attentes des apprenants le justifient, tant en compréhension qu'en production. L'écrit, quant à lui, est réhabilité en début d'apprentissage. « *Les dialogues de l'AC sont organisés à partir de projets de communication et d'actes de langage servant à réaliser le projet défini ; c'est l'aspect pragmatique du langage et la conformité de la production au contexte socioculturel qui y sont prioritaires* » (Onursal, 2008:69).

La compréhension orale s'oriente vers la compréhension globale et la production orale prend une nouvelle dimension avec l'introduction des activités de simulation (déjà en germe dans la SGAV avec les jeux de rôles). Pour ces dernières, comme pour l'ensemble des composants des leçons des AC, les situations de la vie quotidienne sont privilégiées, quel que soit le niveau linguistique des apprenants.

Enfin, dans ces approches, il revient à l'enseignant d'adapter sa méthodologie aux apprenants. C'est le point d'entrée dans l'éclectisme, dénoncé par Puren en 1994 et qui préfigurera la perspective actionnelle constituée à partir du Cadre Européen Commun de Référence pour les Langues (Conseil de l'Europe, 2001).

### 2.3. L'oral dans la perspective actionnelle

La perspective actionnelle (désormais PA) élaborée à partir du Cadre Européen Commun de Référence pour les Langues (désormais CECRL) prend aujourd'hui le relai dans les méthodologies en langues étrangères.

Le CECRL, dans une approche résolument européenne, vise le développement d'une compétence plurilingue dont le sujet idéal *« ne classe pas ces langues et ces cultures dans des compartiments séparés mais construit plutôt une compétence communicative à laquelle contribuent toute connaissance et toute expérience des langues et dans laquelle les langues sont en corrélation et interagissent. Dans des situations différentes, un locuteur peut faire appel avec souplesse aux différentes parties de cette compétence pour entrer efficacement en communication avec un interlocuteur donné »* (Conseil de l'Europe, 2001:11). C'est donc une approche plurilingue qui est retenue, visant la capacité d'agir dans une (LE) plus ou moins bien maîtrisée. Dit autrement, il ne s'agit plus de former les individus à communiquer mais de les amener à travailler ensemble (Puren, 2006a). L'apprenant devient un usager de la langue impliqué dans un agir social. Ainsi, *« la perspective privilégiée ici est, très généralement aussi, de type actionnel en ce qu'elle considère avant tout l'usager et l'apprenant d'une langue comme des acteurs sociaux ayant à accomplir des tâches (qui ne sont pas seulement langagières) dans des circonstances et un environnement donnés, à l'intérieur d'un domaine d'action particulier. Si les actes de parole se réalisent dans des activités langagières, celles-ci s'inscrivent elles-mêmes à l'intérieur d'actions en contexte social qui seules leur donnent leur pleine signification »* (Conseil de l'Europe, 2001:15). Dans cette perspective on dépasse donc la langue, il s'agit en effet de former les apprenants à l'action, dont la langue devient alors un médium. L'apprenant n'est plus seulement envisagé comme un individu-locuteur à qui proposer les moyens linguistiques de communiquer, il est considéré comme un individu-acteur, à qui il faut proposer les moyens d'agir, de *jouer* à l'échelle européenne. Plus que des locuteurs, il s'agit de former, à travers (entre autres) l'enseignement des langues étrangères (européennes), des citoyens européens, redonnant ainsi à la didactique des langues sa dimension éminemment politique.

D'un point de vue méthodologique, l'approche par compétences des AC est reprise dans la perspective actionnelle ; celles-ci se déclinent en : une compétence linguistique, une compétence sociolinguistique, et une compétence pragmatique, chacune d'elles participant à la compétence à communiquer langagièrement (Conseil de l'Europe, 2001:17-18). Nous voyons donc que, malgré quelques ajustements, la base de la perspective actionnelle reste la même que celle des AC, ce qui nous amène à ne pas

considérer les recommandations du CECRL comme révolutionnaires par rapport aux approches précédentes mais bien comme des avancées méthodologiques de l'enseignement/apprentissage des langues étrangères en phase avec une évolution sociétale.

Les trois grandes nouveautés introduites par le CECRL sont globalement : une prise de distance vis à vis des méthodologies « dirigistes », l'introduction de la « capacité à interagir » en tant que cinquième compétence et la définition de niveaux communs. Du point de vue de la distance vis à vis des méthodologies, on laisse les praticiens de terrain faire leurs choix. Le CECRL se différencie donc assez nettement du *Niveau Seuil*. « (...) comme exposé (...) tout au long du présent document, le cadre de référence n'a pas pour vocation de promouvoir une méthode d'enseignement particulière mais bien de présenter des choix » (Conseil de l'Europe, 2001:110). Ainsi, plus qu'un ouvrage dogmatique, le CECRL se présente comme un outil de cadrage épistémologique et social. Cependant, comme le souligne (Puren, 2006b), on peut regretter que « *les auteurs du CECR se soient contentés de justifier leur ouverture méthodologique par des arguments négatifs (l'absence de certitudes scientifiques et les dangers du dogmatisme) ou empiriques (la reconnaissance de pratiques éclectiques généralisées chez les apprenants et enseignants)* » et non pas avec des arguments positifs.

L'autre grand changement apporté par le CECRL, est la définition de niveaux linguistiques communs de référence allant de A1 (premier niveau) à C2 (niveau le plus avancé). Nous nous concentrerons, dans leur présentation, sur les compétences de Compréhension Orale chez les apprenants de niveaux A. En effet, c'est spécifiquement sur ce public et sur cette habileté que porte notre travail.

Ainsi, l'échelle globale des niveaux linguistiques communs de compétences (Conseil de l'Europe, 2001:25) décrit les apprenants de niveau A comme des « *utilisateurs élémentaires* ». Ainsi, l'apprenant de niveau A1 « *peut comprendre et utiliser des expressions familières et quotidiennes ainsi que des énoncés très simples qui visent à satisfaire des besoins concrets* » parallèlement, l'apprenant de niveau A2 « *peut comprendre des phrases isolées et des expressions fréquemment utilisées en relation avec des domaines immédiats de priorité (par exemple, informations personnelles et familiales simples, environnement proche, travail)* ». La grille d'autoévaluation des compétences des niveaux linguistiques communs se veut, elle, plus détaillée (Conseil de l'Europe, 2001:26). Ainsi, un apprenant s'auto-évalue au niveau A1 s'il « *peu(t) comprendre des mots familiers et des expressions très courantes au sujet de (lui)-même, de (sa) famille et de l'environnement concret et immédiat, si les gens parlent lentement et distinctement* » et au niveau A2 s'il « *peu(t) comprendre des expressions et un vocabulaire très fréquent relatifs*

à ce qui (le) concerne de très près (par exemple (lui)-même, (sa) famille, les achats, l'environnement proche, le travail). (Il) peut saisir l'essentiel d'annonces et de messages simples et clairs ». En plus de ces descriptions générales sur les compétences langagières de l'apprenant en fonction du niveau, le CECRL propose des descriptions plus détaillées sur les quatre compétences fondamentales : la Compréhension Orale, la Compréhension Écrite, la Production Orale et la Production Écrite.

Ainsi, si nous nous penchons sur les compétences en CO aux niveaux A1 et A2, alors, nous pouvons constater que :

- en *compréhension générale de l'oral* (Conseil de l'Europe, 2001:55), l'apprenant de niveau A1 « *peut comprendre une intervention si elle est lente et soigneusement articulée et comprend de longues pauses qui permettent d'en assimiler le sens* » ; parallèlement, l'apprenant de niveau A2 « *peut comprendre des expressions et des mots porteurs de sens relatifs à des domaines de priorité immédiate (par exemple, information personnelle et familiale de base, achats, géographie locale, emploi) ; « peut comprendre assez pour pouvoir répondre à des besoins concrets à condition que la diction soit claire et le débit lent* ».

- Pour la compétence à *comprendre une interaction entre locuteurs natifs* (Conseil de l'Europe, 2001:55), il apparaît que le Cadre ne propose pas de descripteur pour le niveau A1. Pour le niveau A2, le Cadre considère que l'apprenant « *peut identifier le sujet d'une discussion se déroulant en sa présence si l'échange est mené lentement et si l'on articule clairement* ». C'est spécifiquement ce type de situations que les dialogues pédagogiques mettent en scène.

- Pour la compétence à *comprendre des annonces et (des) instructions orales*, le Cadre définit l'apprenant A1 comme pouvant « *comprendre des instructions qui lui sont adressées lentement et avec soin et suivre des directives courtes et simples* » ; et l'apprenant A2 comme pouvant « *saisir le point essentiel d'une annonce ou d'un message brefs, simples et clairs. (il peut aussi) comprendre des indications simples relatives à la façon d'aller d'un point à un autre à pied ou avec les transports en commun* » (Conseil de l'Europe, 2001:56).

Enfin, concernant la *compréhension des émissions de radio et des enregistrements*, le Cadre ne propose pas de descripteur pour le niveau A1. En revanche, pour le niveau A2, l'apprenant « *peut comprendre et extraire l'information essentielle de courts passages enregistrés ayant trait à un sujet courant prévisible, si le débit est lent et la langue clairement articulée* » (Conseil de l'Europe, 2001:56).

Les compétences en CO sont donc claires et il ne subsiste aucun doute quant à leur importance. Toutefois, on peut observer (et regretter) que les gestes pédagogiques pour atteindre ces objectifs ne soient pas aiguillés.

Ainsi, comme nous l'avons vu, la place de l'oral a beaucoup évolué depuis la naissance de la didactique des langues en tant que discipline. Inscrite dans la perspective actionnelle, notre pratique enseignante nous a amenée à nous interroger sur l'enseignement/apprentissage de la CO qui reste relativement peu documenté et aiguillé. Notre questionnement s'est principalement porté sur les supports de CO, leur élaboration, leur utilisation en classe et leur compréhension par les apprenants débutants. Dans ce cadre, nous allons maintenant nous concentrer sur ces supports dont nous présenterons une rapide typologie.

### 3. Les supports de CO en FLE

Comme nous l'avons vu, les différents courants méthodologiques dominants depuis les années 1960 ont tous accordé une place de choix à la CO, sans pour autant en proposer une méthodologie systématique pour son développement. En effet, le choix des supports d'enseignement et d'apprentissage de la CO, ainsi que leur exploitation en classe restent bien souvent à la discrétion de l'enseignant. Adami (2009:168) propose une nomenclature des types de supports utilisés concourant à aborder la compréhension orale en classe de langue étrangère :

- « - *les documents bruts : ce sont les moins artificiels ; ils sont extraits directement de la réalité interactionnelle et utilisés tels quels : ce sont les documents dits « authentiques » (...)*
- *les documents réformés : ce sont des documents bruts qui, pour des raisons de confidentialité, de droits ou d'éthique, ont été toilettés : on a changé les noms, les signes, les logos, éliminé toute référence personnelle, etc. tout en conservant l'intégralité du contenu (...)*
- *les documents pastiches : ils reproduisent fidèlement la réalité ; ce sont par exemple des interactions orales reproduites et réenregistrées par des acteurs, professionnels ou d'occasion, sur la base de véritables interactions*
- *les documents réalistes : ils sont entièrement fabriqués mais en s'inspirant très fortement de la réalité ; les tours de paroles sont écrits et lus ou improvisés en tentant de rester au plus près de la réalité ; ils peuvent ainsi reconstruire la réalité en ne retenant d'une interaction réelle que les éléments intéressants pour atteindre un objectif à un moment donné*
- *Les documents fabriqués : ils sont conçus autour d'un objectif didactique en tentant*

*de ne pas trop s'éloigner de la réalité d'une interaction verbale ; ils s'inscrivent dans une démarche de progression du simple au complexe et s'articulent autour de la notion de niveau*

- *Les documents artificiels : ils ne tiennent pas compte de la réalité d'une interaction verbale et servent uniquement un objectif didactique ; ce type de document est de plus en plus rare et correspond à une époque révolue. »*

Nous retiendrons ici une distinction entre les documents authentiques (*bruts, réformés* ou *pastiches*) d'une part et l'ensemble des documents pédagogiques (*réalistes, fabriqués* ou *artificiels*) d'autre part. Ce distinguo repose sur les objectifs de leur création : à des fins de communication ou d'enseignement. En effet, le débat qui oppose les documents authentiques aux supports pédagogiques (et plus particulièrement les documents *réalistes*) n'est pas clos (Gallien, 2001) et ces deux types de supports sont largement utilisés en classe de langue. Dans cette partie, nous allons donc considérer, dans un premier temps quels sont les supports les plus fréquemment utilisés en classe de FLE avec les niveaux débutants. Pour ce premier aspect, nous avons choisi de nous concentrer sur l'opposition documents authentiques/dialogues pédagogiques. Ce point s'articulera en deux étapes : une étude bibliographique sur chacun de ces deux types de documents et une enquête réalisée auprès d'enseignants de FLE pour connaître leurs habitudes d'utilisation de ces supports. Dans un deuxième temps, nous présenterons une étude sur les supports de CO disponibles dans les manuels de FLE (destiné à des apprenants de niveau débutant) afin d'en dégager une typologie.

### **3.1. Documents authentiques ou dialogues fabriqués**

De manière empirique nous avons pu constater que les documents authentiques sont plus souvent utilisés avec des apprenants de niveaux avancés que de niveau débutant. Sans doute cette distinction tient-elle à la difficulté de trouver des documents authentiques adaptés aux premiers niveaux. Nous avons réalisé une enquête pour vérifier cette hypothèse.

#### ***3.1.1. Documents authentiques ou pédagogiques ?***

Les documents authentiques sont des documents non prévus à des fins d'apprentissage où la langue apparaît dans son contexte d'utilisation. « *En termes positifs, c'est un énoncé produit dans une situation réelle de communication : en sont exclus les textes écrits et oraux fabriqués ou modifiés en vue d'enseigner une langue, le discours*

*d'un enseignant en langue ou d'un « native speaker », si ce discours a pour fonction d'enseigner une langue et les productions en langue étrangère des apprenants (échanges entre apprenants et enseignants ou apprenants) si ces productions ont pour but d'apprendre la langue étrangère. » (Abe, Carton, Cembalo, & Regent, 1979:2-3). Nathalie Spanghero-Gaillard et Emmanuelle Garnier (2013) définissent à leur tour les documents authentiques comme « (...) des énoncés produits dans des situations réelles de communication d'une communauté donnée ; destinés aux membres de cette communauté, ils sont utilisés en vue de l'apprentissage de la langue et de la culture de cette communauté linguistique. Ils appartiennent ainsi à un ensemble très étendu de situations de communication, ils se présentent sous la forme de messages verbaux, pictoraux, chorégraphiques, empruntant tant un canal visuel qu'auditif ». L'authenticité des documents est donc définie au niveau de leur encodage : s'ils sont destinés à la communication en dehors de la salle de classe, ils sont authentiques (Adami, 2009). La question de leur authenticité est alors posée. En effet, dès lors qu'ils sont exploités en salle de classe, ces documents sont extraits de leur contexte, et donc, dans une certaine mesure, moins authentiques, (voir à ce propos Abe *et al.* 1979 ; Adami, 2009). Ils n'en restent pas moins des productions attestées, le plus souvent émises par des natifs, pour des natifs et en ce sens représentant de bons exemples langagiers. De plus, dans la mesure où ils sont issus de la vie extérieure à la classe, les documents authentiques permettent une procédure d'expansion (« *procédure méthodologique qui permet de faire des savoirs ou des activités didactiques des savoir-faire non didactiques* » (Cuq & Gruca, 2002:55) plus aisée puisque les liens avec les situations non-didactiques sont plus évidents.*

Par ailleurs, ces supports répondent à plusieurs objectifs d'enseignement/apprentissage. En situation d'apprentissage du FLE dans le pays d'origine, les documents authentiques permettent de faire découvrir les pays francophones et leur culture à travers des éléments de communication attestés. « (...) *Le document authentique est une sorte de photographie d'un discours donné à un moment donné dans un endroit donné. Et comme un cliché, il a une existence propre. Il n'est pas le réel, mais une attestation de réel avec son cadre, ses limites et son caractère relatif et instantané* » (Abe *et al.*, 1979:4). Quand l'apprentissage du FLE se fait dans un pays francophone (situation de migration principalement) le document authentique permet de travailler « *l'adaptation des apprenants à la réalité. Dans ce cadre, le document authentique n'est plus un moyen mais un objectif en soi : il s'agit, en situation d'apprentissage guidé, d'aider les apprenants à interagir dans des situations d'oral spontané, à lire des documents de tous ordre et à écrire dans des situations de communication que n'importe quel natif est quotidiennement contraint de gérer.* » (Adami,

2009:164). Dans ce cadre, travailler à partir de documents authentiques permet de répondre aux besoins directs et spécifiques des apprenants.

Enfin, les documents authentiques fournissent une exposition de la langue dans son contexte, l'apprenant découvre ainsi non seulement la langue mais aussi « *les règles discursives qui déterminent qui a le droit de dire **quoi**, à **qui**, **comment**, **où** et **quand*** » (Holec, 1990:67).

Dans le cadre des activités de CO, on peut constater que les documents authentiques sont très souvent médiatiques, audiovisuels (télévision, cinéma ou maintenant internet) ou radiophoniques (notamment avec les *podcasts*). Ces documents peuvent être de différentes natures et peuvent présenter différents degrés de difficulté pour les apprenants. Ainsi, Kamber & Skupien (2009) ont réalisé une enquête auprès d'apprenants de niveaux B1 à C2 et ont pu observer que « *que tous les documents considérés comme faciles proviennent d'émissions de type magazines, reportages ou débats, alors que ceux qui sont considérés comme difficiles proviennent systématiquement de revues de presse (nationale ou internationale). En fait, les documents considérés comme faciles, situés en dessous de la moyenne de difficulté (...), sont des documents oraux spontanés, alors que ceux qui sont perçus comme difficiles sont des documents écrits oralisés* » (p.180). D'après les auteurs, les apprenants désignent trois sources de difficultés majeures : le rythme, le vocabulaire et l'accent. Le rythme est la difficulté la plus citée par les apprenants interrogés, il apparaît donc que cet aspect est particulièrement important. Le vocabulaire, semble être identifié comme source de difficulté par les participants pour l'ensemble des documents soumis au test. Nous pouvons donc supposer que le vocabulaire fait partie des difficultés, lorsqu'il est inconnu des apprenants, et ce quel que soit le type de documents (et donc également des dialogues pédagogiques). Enfin, d'après l'étude présentée, l'accent ne semble pas être considéré fréquemment comme une difficulté.

En outre, les auteurs (*ibidem*) arrivent à la conclusion que « (leurs) *intuitions premières, à savoir que les documents spontanés, présentant un bruit de fond ou un brouillage importants, un registre de langue populaire, de fréquentes ruptures syntaxiques, des interactions entre interlocuteurs, des superpositions, de l'humour et de l'ironie, pouvaient s'avérer particulièrement ardues pour des apprenants allophones, sont en fait très largement contredites par les résultats de cette enquête. La longueur de l'extrait non plus ne joue aucun rôle dans l'appréciation des étudiants* » (p.179-180). Il apparaît donc que la présence d'un contexte sonore environnemental ne perturbe pas la compréhension des dialogues pédagogiques et pourraient même la faciliter. Cette remarque nous amène à avancer l'hypothèse principale de ce travail de thèse : **un**

**contexte sonore environnemental, s'il respecte certaines conditions d'élaboration et de présentation vis à vis du dialogue favoriserait la compréhension de ce dernier par des apprenants débutants en langue cible** (voir chapitre *Hypothèses*, p.105).

Les documents authentiques sont donc de bons supports d'enseignement, notamment pour la Compréhension Orale. Toutefois, comme il est apparu dans l'étude que nous avons réalisée auprès d'enseignants de FLE (voir p. 39), nous pouvons voir que les documents authentiques sont majoritairement exploités auprès des apprenants intermédiaires et avancés. Les publics débutants sont invités à travailler, pour la CO avec des dialogues pédagogiques. Ainsi, Abe, Carton, Cembalo, & Regent (1979:8) s'interrogent sur le recours aux documents authentiques pour les publics que nous visons (A1-A2) : « *qu'est-ce qu'un document de langue simple, facile à appréhender pour un débutant ? Les textes construits présentent l'avantage de faire varier à volonté la complexité morphosyntaxique et d'établir une progression allant, de ce point de vue, du simple au complexe.* ». Les documents fabriqués à des fins d'enseignement (par les concepteurs de manuels, ou par les enseignants), présentent en effet l'avantage d'être conçus en fonction d'un certain nombre d'objectifs pédagogiques (lexicaux, grammaticaux, thématiques etc.) ce qui, auprès des niveaux débutants nous semble indispensable. En effet, les premiers niveaux d'apprentissage et de maîtrise d'une langue impliquent la nécessité d'une progression rigoureuse : les besoins des apprenants sont facilement identifiables par les enseignants à qui il revient d'y répondre. Ainsi, on peut voir, à partir d'une rapide analyse des supports d'enseignement destinés aux apprenants débutants, pour la CO, les manuels de FLE présentent majoritairement des dialogues pédagogiques (voir p.42 et dans le CD en annexe). Globalement, les dialogues pédagogiques peuvent être de deux types : ils peuvent présenter de la parole seule, ou de la parole accompagnée d'un contexte sonore environnemental (CSE). C'est sur ce type de supports, parole + CSE, que porte la suite de nos travaux.

Toutefois, il apparaît que les dialogues pédagogiques, bien que représentant de nombreux avantages dans l'enseignement/apprentissage des LE (et plus spécifiquement du FLE), entraînent également un certain nombre de travers. En effet, bien que censés reproduire des productions de natifs (des interactions spontanées dans le cadre des dialogues pédagogiques), ils sont fréquemment très proches de l'écrit et présentent, le plus souvent une trop grande unité thématique en leur sein (peu de glissements) et sont peu empreints de marques de l'oral (Arditty & Vasseur, 1999; Parpette, 1997, 2008).

### **3.1.2. Une étude sur les supports utilisés par des enseignants de FLE**

Douze enseignants de FLE ont été interrogés sur les supports de CO auxquels ils ont recours en classe de FLE. Il s'est agi pour nous de déterminer si certains supports (documents authentiques ou dialogues fabriqués) sont privilégiés par rapport à d'autres, notamment en fonction du niveau des apprenants, et si dans le cas du travail à partir de dialogues pédagogiques, le Contexte Sonore Environnemental (désormais CSE) était exploité (Colle-Quesada & Spanghero-Gaillard, sous presse). Ces enseignants exercent en France, au sein de l'Alliance Française de Toulouse (où le choix du manuel est décidé collégalement par l'équipe enseignante) ou du Département d'Étude du FLE de l'Université de Toulouse 2 Le Mirail (DEFLE, où les enseignants sont généralement libres de choisir leurs supports). L'ensemble de ces enseignants a en moyenne 10,5 ans d'expérience (8 de médiane)<sup>4</sup>.

Les participants ont été interrogés sur trois points : 1) la fréquence de leur utilisation des dialogues pédagogiques fabriqués/documents authentiques pour la CO, 2) les niveaux (du CECRL) avec lesquels les documents authentiques sont les plus utilisés (toujours pour la CO) et 3) dans l'exploitation de dialogues pédagogiques fabriqués pour la CO, l'attention particulière éventuellement portée au Contexte Sonore Environnemental (CSE). Un questionnaire anonyme auto-administré leur a été soumis afin de recueillir leur réponses via une l'échelle de Likert (jamais / absolument pas < exceptionnellement / faiblement < rarement / modérément < occasionnellement / sensiblement < souvent / considérablement < toujours / énormément)<sup>5</sup>. Nos hypothèses à ce niveau sont que ces deux types de supports seront fréquemment utilisés mais de manière différente en fonction du niveau des apprenants. En effet, nous considérons que les besoins spécifiques de chaque niveau implique une démarche enseignante différenciée : les documents pédagogiques étant plus facilement adaptables aux premiers niveaux (A1 et A2) et les documents authentiques requérant des connaissances linguistiques et para linguistiques plus avancés pour être appréhendés.

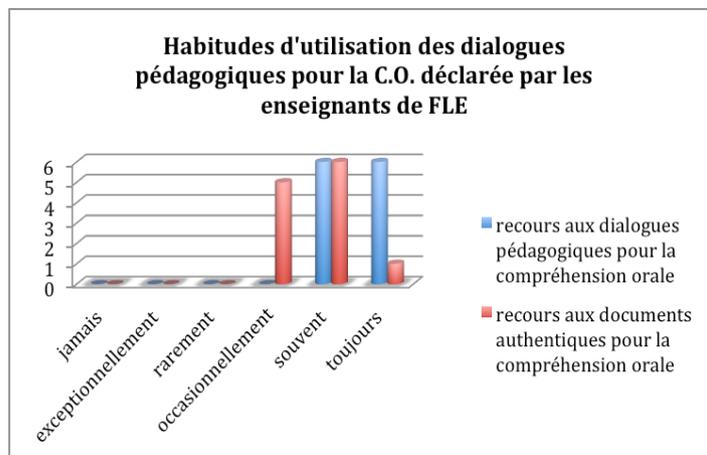
Concernant le recours aux dialogues pédagogiques fabriqués et aux documents authentiques, nous pouvons voir que ces deux types de supports sont fréquemment utilisés. Les réponses à la première question, qui porte sur la fréquence d'utilisation des dialogues pédagogiques fabriqués, font apparaître une utilisation fréquente : 6 *souvent* et 6 *toujours*. De même les documents authentiques s'avèrent très fréquemment utilisés

---

<sup>4</sup> Pour plus de précisions, se reporter au CD en annexe

<sup>5</sup> Pour plus de précision, voir la fiche questionnaire sur le CD en annexe

puisque 5 enseignants déclarent y avoir recours *occasionnellement*, 6 *souvent* et 1 *toujours*.



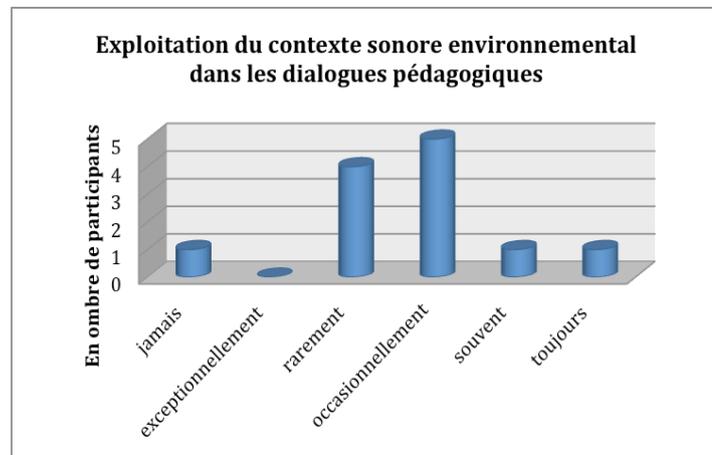
**Figure 1: Fréquences de recours aux supports de Compréhension Orale déclarées par les enseignants de FLE**

Toutefois, ces résultats sont à nuancer à la lumière des réponses portant sur l'utilisation des documents authentiques en fonction des niveaux linguistiques des apprenants. En effet, il apparaît que les enseignants interrogés exploitent les supports d'enseignement de la compréhension orale différemment selon le niveau des apprenants car pour les **débutants (A1-A2)**, les documents authentiques ne sont pas fréquemment sollicités. Avec les niveaux intermédiaires et avancés, leur utilisation est en revanche plus fréquente (voir résultats détaillés en annexe sur le CD).

Conformément à l'hypothèse avancée au départ, le recours aux documents authentiques augmente avec le niveau linguistique des apprenants. Autrement dit, auprès des niveaux débutants, les dialogues enregistrés à des fins d'enseignement sont privilégiés. Toutefois, il apparaît, après observation de différents manuels de FLE destinés à un niveau débutant, que les dialogues pédagogiques (sur lesquels seuls porte notre travail) ne représentent pas la totalité des supports de compréhension orale disponibles. En effet, nous pouvons également noter la présence d'oraux visant d'autres compétences que la compréhension orale (modèles de prononciation par exemple) ou des supports de CO différents des dialogues pédagogiques (monologues suivis ou « conversations par répondants interposés » par exemple).

Par ailleurs, et comme dit plus haut, ces dialogues peuvent, ou non, être accompagnés d'un contexte sonore environnemental, sans que celui-ci fasse l'objet d'une attention particulière dans les recommandations aux enseignants.

Les participants ont donc également été interrogés sur l'exploitation potentielle du contexte sonore environnemental dans les dialogues pédagogiques. Parmi les 12 enseignants interrogés, un enseignant ne fait jamais appel à ce composant, 4 y ont rarement recours, 5 s'en servent occasionnellement, 1 l'utilise souvent et 1 l'emploie toujours.



**Figure 2 : Exploitation du contexte sonore environnemental dans les dialogues pédagogiques**

Globalement, ces résultats nous indiquent que la majorité des enseignants interrogés a régulièrement recours au contexte sonore dans l'exploitation des dialogues pédagogiques en salle de classe.

Il apparaît donc que les dialogues pédagogiques sont des supports de premier ordre et que le contexte sonore qui les accompagne parfois est considéré par la moitié des enseignants interrogés comme très important pour une exploitation plus efficace du support. Ce constat nous amène à chercher à évaluer **en quoi le contexte sonore environnemental modifie la compréhension du dialogue (partie langue)**. Cette dimension sera l'objet principal de notre partie expérimentale où nous tenterons à la fois de voir dans quelle mesure la présence d'un CSE a eu un impact sur la compréhension du dialogue, mais aussi d'identifier les paramètres de la compréhension affectés par cette présence.

## 3.2. Une étude sur les dialogues pédagogiques de niveau 1 en FLE

Nous avons donc vu qu'avec les niveaux débutants (A1 et A2), les dialogues pédagogiques sont privilégiés. Ceux-ci peuvent être de deux sortes : ils peuvent présenter de la parole seule ou de la parole accompagnée d'un contexte sonore environnemental. Nous allons nous intéresser à ces types de supports. Quelles peuvent être les fonctions assumées par ces contextes sonores environnementaux ? Quel panorama nous offre aujourd'hui les manuels d'enseignement/apprentissage du FLE aux niveaux débutants ?

### 3.2.1. Les fonctions des sons environnementaux : un parallèle avec les images de la méthodologie SGAV

Dans la méthodologie SGAV (voir p.27), les dialogues pédagogiques étaient le plus souvent accompagnés d'images de films fixes) visant à reconstruire visuellement certaines parties du discours des locuteurs. Ces images faisaient l'objet d'une attention particulière en classe afin de participer au processus de compréhension. « *La perception culturelle est elle aussi un fait culturel et social, mais elle varie moins radicalement d'une culture à une autre, que ne le font les symbolismes linguistiques (...). Les « langages de l'image » quels qu'ils soient (cinéma, télévision, etc.), ont tous ceci en commun de prendre au départ un large appui sur la perception visuelle : celle-ci (...) ne rend pas compte de l'intellection de toutes les données visuelles, il s'en faut de beaucoup ; mais elle assure du moins une première couche d'intelligibilité qui n'a aucun équivalent dans les langues et qui, dans une large mesure, n'a pas à être enseignée* » (Metz, 1970:163). Henri Besse (1974) distingue deux types d'images : les images situationnelles et les images de transcodage : « *Les éléments des icônes des images qui renvoient aux composantes de la situation d'énonciation de discours, seront appelés icônes situationnelles, et les éléments qui renvoient au contenu sémantique des énoncés étrangers seront appelés icônes de transcodage* » (p.30). Ainsi, images situationnelles et de transcodage se complètent et impliquent différentes tâches de compréhension chez l'apprenant compreneur et leur « décodage » doit faire partie des séquences du cours, s'inscrivant ainsi réellement dans la pédagogie.

**Notre hypothèse est que les sons environnementaux peuvent assumer les mêmes fonctions.** En effet, les sons environnementaux peuvent rendre compte de la situation d'énonciation dans laquelle se déroule le dialogue (par exemple, pour une discussion à une terrasse de café, il est possible de rendre compte du lieu), et/ou du contenu propositionnel du dialogue (par exemple, pour une interaction portant sur une

recette de cuisine, il est possible de rendre compte des actions par le son (découpe d'un légume, mijotage, etc.). Certaines études sur la multimodalité (par exemple Spanghero-Gaillard (2008); Spanghero-Gaillard & Gaillard (2007); Tricot (1998)), permettent d'avancer l'hypothèse d'un environnement sonore offrant des éléments de redondance complémentaire de celle des éléments linguistiques. Comme le dit Clerc (1999:56), ces éléments peuvent faciliter la compréhension : « *s'il agit d'un document audio, il est indispensable de le contextualiser, ou bien, s'il s'agit d'un document vidéo, celui-ci doit proposer des images de micro-gestualité ou de situationnalisation ou de redondance par rapport au message audio. Ces images devraient faciliter la compréhension du message* »

Toutefois, il apparaît que le son et l'image n'offrent pas les mêmes possibilités de décodage, puisqu'ils ne relèvent pas de la même modalité. Si l'image fixe permet une durée de vue illimitée (en dehors des contraintes extérieures), les enregistrements audio ne permettent pas de telles expositions. Les montages de supports faisant intervenir à la fois la dimension linguistique (le dialogue) et une dimension sonore environnementale, n'offrent donc pas les mêmes possibilités d'exploitation. En outre, accompagner un dialogue audio d'images implique une perception multimodale, alors que les scènes sonores en position contextuelle ne renvoient qu'à la seule modalité auditive. Cette distinction aura un impact certain sur l'appréhension du support par l'individu.

Ainsi, plutôt que de reprendre la terminologie établie par Henri Besse pour les images, nous avons établi une nouvelle nomenclature des éléments sonores environnementaux qui peuvent accompagner les dialogues. En effet, nous avons pu observer empiriquement (voir *l'Élaboration du matériel sonore environnemental* p.121) que les sons environnementaux ne permettent pas, dans leur création les mêmes manipulations que les images fixes en raison, notamment des contraintes temporelles ou de réalisme. En outre de nombreux objets (ou actions) sont difficiles à mettre en ondes, comme la mimo-gestuelle par exemple.

Dans ce cadre, les éléments de distinction des différents types de contextes sonores environnementaux se situent au niveau de l'encodage (élaboration, degré de redondance entre le dialogue et la scène sonore), et au niveau du décodage (perception, qualité narrative). Ainsi, les scènes sonores qui rendent compte de la situation d'énonciation dans laquelle se déroule le dialogue seront appelées *images sonores*, parallèlement, les scènes sonores qui rendent compte des actions des actants seront appelées *histoire sonore*. Toutefois, et bien que dans leur nature ces scènes sonores soient fondamentalement différentes des images de films fixes de la méthodologie SGAV, leurs fonctions semblent relativement parallèles : généralement les *images sonores* assument une fonction *situationnelle* alors que les *histoires sonores* assument le plus

souvent une fonction de transcodage. Ainsi, nous considérons comme *situationnelles* les scènes sonores qui mettent en ondes des composants de la situation d'énonciation ; et comme transcodantes les scènes sonores qui donnent à entendre des aspects du contenu propositionnel du dialogue.

### **3.2.2. Des dialogues avec et sans contexte sonore**

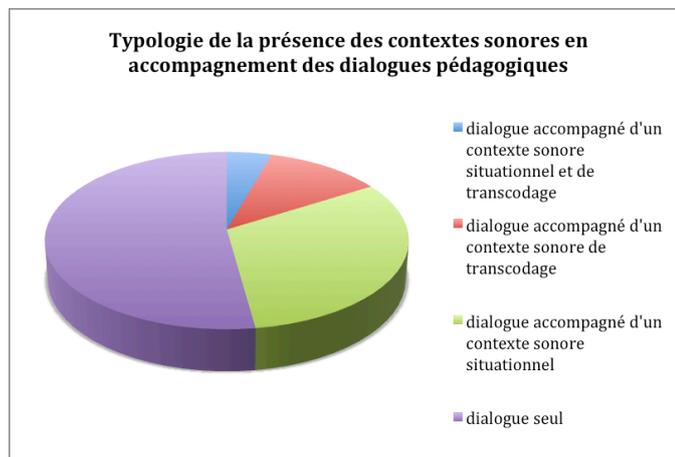
Afin d'obtenir un panorama général des dialogues utilisés en classe de FLE, nous avons mené une étude sur ces supports. Son objectif était de voir dans quelle mesure les dialogues pédagogiques proposés par des manuels récents sont accompagnés d'un contexte sonore environnemental et, lorsqu'ils le sont, quelles fonctions peuvent assumer ces contextes. Dans ce cadre, 12 manuels de FLE de niveau débutant, actuellement sur le marché et quotidiennement utilisés par les enseignants ont été analysés (livre de l'élève (Baglieto, Girardeau, & Mistichelli, 2011; Berthet, Hugot, Kizirian, Sampsonis, & Waendendries, 2006; Denyer, Garmendia, & Lions-Olivieri, 2009; Di Giura & Beacco, 2007; Flumian, Labascoule, Lause, & Royer, 2011; Grand-Clément, 2010; Labascoule, Lause, & Royer, 2009, 2010; Miquel, 2007; Miquel, 2005; Poisson-Quinton, Siréjols, & Bruley, 2011)) Ce matériel représente un total de 348 dialogues (pour une présentation détaillée des données de cette étude, se reporter au CD annexe). Rappelons que nous nous sommes concentrée sur les seuls dialogues pédagogiques, en laissant de côté les autres types de supports audio tels que les supports du manuel *Ici 2* (Abry *et al.*, 2007) qui donnent à entendre des documents authentiques (lectures d'œuvres littéraires, chansons...). Une typologie a ainsi pu être mise au jour : parmi ces dialogues, 171 présentent un contexte sonore environnemental et 177 n'en présentent pas. Dit autrement, malgré l'intérêt intuitif que représente l'ajout d'un contexte sonore environnemental au dialogue pédagogique, et bien que l'écart soit faible, la majorité des dialogues que nous avons pu observer (50,8%) ne présente que la dimension verbale des dialogues<sup>6</sup>.

Le deuxième objectif de cette étude était de voir dans quelle mesure les fonctions (situationnelle ou de transcodage) des illustrations mises en lumière par Besse (1974) peuvent apparaître avec les contextes sonores environnementaux. Dans ce cadre, nous avons pu voir que certaines scènes sonores présentent la seule fonction *situationnelle* ; d'autres, la seule fonction de *transcodage* et enfin, d'autres assument les deux fonctions à la fois. Parmi les 171 dialogues qui présentent un contexte sonore environnemental, 113 présentent un contexte sonore uniquement situationnel, 42 scènes sonores assument

---

<sup>6</sup> Cet écart, il n'est pas statistiquement significatif (test de comparaison de fréquences,  $p > 0,05$ )

une fonction de transcodage, et 16 présentent à la fois une dimension situationnelle et de transcodage.



**Graphique 1: Typologie des dialogues pédagogiques en fonction de la présence d'un contexte sonore environnemental, sur 361 dialogues, ou 12 manuels**

En outre, nous pouvons constater que les fonctions situationnelles sont largement plus mises en œuvre avec les contextes sonores environnementaux. En effet, l'écart entre contextes sonores situationnel et de transcodage est statistiquement significatif (test de comparaison de proportion,  $p=6,217 \times 10^{-15}$ ,  $z=7,712$ ).

Ainsi, ces différentes observations nous amènent à considérer que bien que la présence d'un contexte sonore reste légèrement minoritaire **les éléments sonores environnementaux en accompagnement des dialogues sont aujourd'hui fortement présents dans la salle de classe**, et peuvent assumer les différentes fonctions des illustrations dégagées par Besse (1974). Toutefois, il apparaît également que la fonction situationnelle (*i.e.*, lorsque les sons environnementaux donnent à entendre certains composants de la situation d'énonciation sans présenter d'éléments de redondance avec le contenu propositionnel du dialogue, saillants dans la scènes sonore) soit plus fréquente. Parallèlement, nous pouvons également constater que les contextes sonores assumant les deux fonctions (rendant compte à la fois de la situation d'énonciation et du contenu propositionnel du dialogue) sont très largement minoritaires. En outre, il est important de mettre en lumière le fait que certains manuels (par exemple *Taxi* et *Le nouveau Taxi* (Capelle & Menand, 2009; Menand, 2006)) présentent les dialogues dans une configuration qui permet à l'enseignant de choisir délibérément de faire écouter le dialogue avec ou sans contexte sonore. La présence sur le marché de tels supports nous

amène à penser que l'édition de manuels de FLE à destination des apprenants débutants s'est déjà saisie (au moins en partie) de l'enjeu que peut représenter la dimension sonore environnementale à la compréhension d'un dialogue pédagogique.

En conclusion, il semble que le sujet de cette thèse renvoie à une problématique réellement existante dans le domaine de la didactique du FLE. En effet, les dialogues pédagogiques accompagnés d'un contexte sonore environnemental sont largement présents dans les manuels de français langue étrangère et il semble tout à fait pertinent, dans ce cadre, de s'interroger sur leur rôle et leur participation au processus de compréhension que les apprenants suivent lorsqu'ils sont confrontés à ce type de matériel.





# Partie 1 : Cadre théorique

## Premier Chapitre : Des études sur la perception...

L'objectif de notre travail est l'évaluation de la compréhension de courts dialogues audio en français, présentés avec ou sans contexte sonore environnemental, à des locuteurs francophones natifs et des apprenants débutants de Français Langue Étrangère. Dans ce cadre, nous devons proposer quelques éclaircissements sur ce que nous savons de la perception des *soundscales*. En effet, nous pouvons considérer la compréhension comme une activité cognitive unique et intégrée, qui prend en compte l'ensemble des « flux auditifs » (parole ou non) dans un seul flux complexe et non-discriminable en unités plus petites. Cette vision n'a toutefois pas reçu l'attention nécessaire ces dernières années. C'est en effet la perception de flux différents qui a fait l'objet des études les plus avancées (cf. Bregman et beaucoup d'autres). Si nous considérons donc la perception des *soundscales* et la perception de la parole comme deux « activités cognitives » concomitantes résultant de deux flux différents et dont la compréhension finale dépend de la cohérence des résultats de chacune d'elles, alors il convient de décrire la perception auditive des scènes sonores en soi. Ici, nous nous interrogerons donc sur la perception et les mécanismes cognitifs qu'elle implique, la perception de multiples flux auditifs et les éventuelles spécificités des différents plans qui constituent les scènes sonores. Dans ce chapitre, nous verrons en quoi la perception des sons amène à l'identification des sources et à l'activation de représentations

cognitives associées. En ce sens, nous défendrons une approche psychologique de la perception auditive.

## 1. Qu'est-ce que percevoir ?

L'être humain, comme la totalité des espèces vivantes, évolue au sein d'un environnement complexe qu'il doit impérativement appréhender et comprendre pour s'y adapter. Dans ce but, il interagit avec son milieu, notamment au niveau de la perception, à travers ses organes sensoriels. Ces derniers sont considérés comme « *l'interface privilégié avec le monde extérieur* » (Houix, 2003:7), puisque c'est à travers eux que nous accédons au milieu qui nous entoure. Barabas (2009:44) définit la perception comme « *l'acte par lequel un sujet se rapporte à une réalité* ». En d'autres termes, elle est le processus par lequel l'individu accède à ce qui lui est extérieur. Selon l'auteur, la perception « *est en effet ce qui nous donne accès à quelque chose, à ce qu'il y a* » (p.7), à ce qui est, à ce qui environne l'individu. Les autres éléments de définition que propose Barabas (2009) se caractérisent par leur dimension *négative* puisqu'il ne définit pas directement ce qu'est (ou pourrait être) la perception, mais explicite ce dont elle se distingue : de la *pensée* (puisque la perception repose sur la présence concrète d'un objet), du *sentiment*, de l'*imagination*, et de la *mémoire* (puisque pour qu'il y ait perception, l'objet doit être présent physiquement). En ce sens, la perception peut être définie comme un processus cognitif émergeant de l'interaction entre l'individu et son environnement. Selon Al-Saleh, (2011) « *Les notions sensorielles ont trait à tout ce qui concerne la captation d'une information ou d'un complexe d'informations par un organe sensoriel* ». L'auteur précise que, pour lui, ces informations correspondent, au départ, à des paramètres physiques tels que, dans le cadre de la perception auditive, la fréquence ou l'intensité du son par exemple. Toutefois, comme nous le verrons plus loin, nous pouvons aussi considérer que les informations extraites de l'environnement sonore ont une valeur sémantique, plus ou moins dépendante de ces paramètres physiques, que l'individu est à même d'interpréter. Guastavino (2003:19) définit la perception comme l'attribution d'une « *signification à la source responsable de la sensation, en fonction des connaissances préalables du sujet, de son activité, ses intentions et du contexte (...)* Ainsi le sujet percevant participe à configurer le monde sensible ». Toutefois, nous ne considérons pas que l'information soit comprise dans l'input mais bien le fruit de l'interaction entre ce dernier et l'individu percevant. Dit autrement, dans ce cadre, la perception correspond à l'interprétation de signaux physiques *via* l'identification des sources à l'origine de ces signaux et l'élaboration d'une représentation (cognitive) de l'objet perçu et/ou de la situation.

Perception et compréhension sont donc intimement liées, voir co-dépendantes. Percevoir c'est avant tout percevoir quelque chose et, au niveau auditif, l'individu percevant cherche à attribuer les sons qui lui parviennent à des objets du monde présents dans son entourage. La perception (en générale et auditive en particulier) est donc le processus par lequel un sujet (l'individu percevant) accède aux sens des « objets sensoriels » qui l'entourent.

Roger Shepard (1981) précise que la notion d'objet<sup>7</sup> implique une double entité : à la fois physique et perceptive, les deux étant, en ce sens, les deux côtés d'une même pièce. Selon le Grand Robert de la langue française (A. Rey, 2008), un objet est « *ce qui, en affectant les sens d'un sujet, peut être distingué de ce qui l'entoure* ». Dans ce cadre, un objet n'existe que lorsqu'il est perçu. En outre, l'auteur reprend ici les principes de la théorie de la forme (Guillaume, 1979): ce qui est perçu n'est pas nécessairement l'objet, mais la forme. « *La psychologie de la forme apparaît donc bien comme une récusation du concept empiriste de sensation : nous ne percevons que des formes et, par conséquent, le perçu n'est pas composé d'atomes sensibles. Si la sensation conserve un sens, ce n'est plus en tant que fondement de la perception mais comme résultat d'un processus de désintégration et de dissociation d'une forme, processus qui est encore, notons-le, un mode de mise en forme* » (Barabas, 2009:23-24). Le courant de psychologie de la Théorie de la forme (Gestalt) considère en effet que les phénomènes perceptifs sont avant tout une expérience globale et que le tout n'est pas équivalent à la somme des parties (Guillaume, 1979). Ainsi, Barabas, (2009:14) suggère même que « *l'objet n'est donc bien qu'une collection de sensations* », celles-ci étant définies comme une « *réalité subjective* ». Parallèlement, l'auteur fait état des divergences entre Locke (1972, cité par Barabas 2009), pour qui la réalité extérieure serait à l'origine de nos sensations, et Berkeley (1970, cité par Barabas 2009) qui affirme qu'envisager la réalité extérieure en dehors des perceptions sensibles de l'individu n'a pas de sens. Ainsi, Barabas (2009:16) rappelle que la « *réalité que nous percevons comme une réalité n'est rien d'autre que la perception que nous en avons* ». En synthèse, nous pouvons considérer que l'objet est à la fois un item physique, qui existe en dehors de la perception de l'individu, et ce qui est perçu (c'est-à-dire distingué) par l'individu. Le perçu ne réside pas seulement dans la dimension physique de l'objet, mais cette dimension ne peut être considérée qu'à partir (ou à travers) la perception qu'un individu en a, ou en fait. En ce sens, l'objet perçu existe à travers sa dimension phénoménologique et ne saurait en être totalement dissocié.

---

<sup>7</sup> Pour plus de précision sur la notion d'objet, se référer par exemple aux travaux de Griffiths & Warren (2004), Kubovy & Schutz (2010), Matthen (2010) Mervis & Rosch (1981), Nudds, (2010), Sheppard, Duncan, Shapiro, & Hillstrom (2002).

C'est notamment dans ce cadre que nous considérons les objets sonores de l'environnement dans leur perspective perceptuelle.

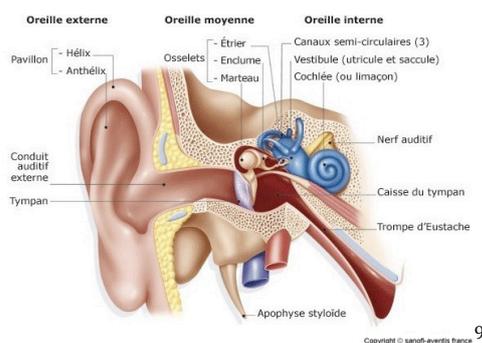
### **1.1. Comment perçoit-on auditivement ?**

Dans le cadre de notre étude, la perception des objets du monde (et des objets sonores en particulier) correspond à la première étape d'interaction entre l'individu-auditeur et son environnement<sup>8</sup>. Nous considérerons la perception comme une partie du processus qui permet à l'individu d'utiliser les qualités intrinsèques et extrinsèques d'un stimulus pour en faire une ou plusieurs unité(s) signifiante(s).

Les éléments définitoires de la perception présentés jusqu'ici, sont très larges, et laissent entendre qu'il s'agit d'un processus global. En effet, la perception, en tant que processus, présente des éléments communs à toutes les sensorialités potentiellement engagées. En outre, il apparaît que la perception est, le plus souvent, multimodale. Dit autrement, l'individu, dans son environnement, reçoit des informations de toute sorte, qui se présentent à ses différents organes sensoriels, donnant lieu à une perception multimodale (ou multi sensorielle : l'ouïe, la vue, l'odorat, le goût et le toucher). Au niveau de l'ouïe, le type d'objet perçu est le son. Celui-ci présente, comme nous le verrons tout au long de ce chapitre, différentes propriétés physiques ainsi qu'une dimension phénoménologique (Al-Saleh, 2011). Nous pouvons définir la perception auditive comme l'ouverture de l'individu à la dimension sonore de son environnement. Cette ouverture consiste à appréhender, à travers différents processus perceptifs et cognitifs, les objets qui l'entourent. Le système auditif humain est complexe et fait intervenir l'oreille et le cortex auditif. Nous allons en présenter une rapide description :

---

<sup>8</sup> Dans le cadre de nos expérimentations (voir *Élaboration du matériel sonore environnemental* p.117) l'environnement est reproduit et ne correspond pas à l'environnement immédiat de l'individu participant au test mais à celui des locuteurs mis en scène dans le dialogue.



**Image 1 : Représentation schématique de l'oreille humaine (externe, moyenne, interne)**

Composée du pavillon et du conduit auditif externe, l'oreille externe capte et rassemble les ondes sonores (variations de pression) et les transmet jusqu'au tympan. Celui-ci participe à ce que l'on appelle l'oreille moyenne qui est également composée des osselets. À ce niveau, les ondes sont transformées en vibrations mécaniques par la membrane tympanique et qui sont alors transmises à l'oreille interne. Celle-ci est constituée de la cochlée dont les cellules ciliées convertissent, à l'aide d'un codage tonotopique, les vibrations mécaniques en un influx nerveux transmis aux aires auditives primaires par le nerf auditif. À chaque étape, des neurones activent des réponses complexes. L'oreille humaine détecte les sons entre 20 et 20 000 Hz, mais sa sensibilité optimale, particulièrement bien adaptée à la parole, se situe entre 2 000 et 4 000 Hz. Certaines données d'IRMf (Imagerie par Résonnance Magnétique fonctionnelle), suggèrent la possibilité d'une contribution des aires corticales au-delà du cortex auditif (Shamma & Micheyl, 2010). Une étude des auteurs a montré une augmentation significative de la scissure interpariétale quand les auditeurs sont soumis à des séquences de sons ambigus, laissant supposer que le traitement du son dépasse largement le système auditif périphérique. La perception auditive ne peut donc pas se résumer au simple traitement des propriétés physiques du signal.

En conditions expérimentales, comme celles que nous présenterons dans la deuxième partie de cette étude, l'objet de la perception auditive devient généralement « stimuli », c'est-à-dire un « fragment de réalité (...) reconstruit en laboratoire » (Cance,

<sup>9</sup>[http://www.google.fr/imgres?client=firefox-a&hs=Wmf&sa=X&rls=org.mozilla:fr:official&biw=1387&bih=707&tbnid=yT32qSDP\\_XLjKM:&imgrefurl=http://coccinelledemoiselle.wordpress.com/2010/02/2le-systeme-auditif/&docid=t19urmfzfoU6qM&imgurl=http://coccinelledemoiselle.files.wordpress.com/2010/02/anatomie\\_oreille.jpg&w=582&h=410&ei=RDkrUqqxAcjI0wXF14HACQ&zoom=1&iact=rc&dur=486&page=1&tbnh=133&tbnw=190&start=0&ndsp=27&ved=1t:429,r:11,s:0,i:115&tx=103&ty=81](http://www.google.fr/imgres?client=firefox-a&hs=Wmf&sa=X&rls=org.mozilla:fr:official&biw=1387&bih=707&tbnid=yT32qSDP_XLjKM:&imgrefurl=http://coccinelledemoiselle.wordpress.com/2010/02/2le-systeme-auditif/&docid=t19urmfzfoU6qM&imgurl=http://coccinelledemoiselle.files.wordpress.com/2010/02/anatomie_oreille.jpg&w=582&h=410&ei=RDkrUqqxAcjI0wXF14HACQ&zoom=1&iact=rc&dur=486&page=1&tbnh=133&tbnw=190&start=0&ndsp=27&ved=1t:429,r:11,s:0,i:115&tx=103&ty=81)

2009:112). Dans la littérature scientifique, ces stimuli ont souvent été des sons *de synthèse* ce qui n'est pas le cas de nos travaux expérimentaux (voir *Élaboration du matériel sonore environnemental* p.121). Toutefois, que les stimuli correspondent à des sons *de synthèse*, à des *enregistrements authentiques* (prise de son brute, sans manipulation en dehors de l'enregistrement) ou *semi authentiques* (prise de son brut mais dont les enregistrements ont été manipulés)<sup>10</sup>, l'écoute en conditions expérimentales est nécessairement différente de l'écoute en condition naturelle. En effet, les objectifs varient et font varier la « conscientisation » de la tâche à effectuer, notamment par l'intermédiaire des processus attentionnels (Botteldooren, *et al.*, 2011 ; voir le sous-chapitre sur l'Analyse des Scènes Auditives p.59). Selon les conditions et les objectifs de l'écoute, le processus de perception auditif variera. Pierre Schaeffer (1977) propose une terminologie particulière. Dans ce cadre, l'*Ouïr* renvoie au premier niveau de perception auditive, caractérisé notamment par la passivité du sujet percevant. À un niveau un peu plus élaboré interviendrait l'*entendre* qui correspond, pour l'auteur, à une écoute plus active de la part du sujet percevant, qui, lorsqu'il *entend* cherche à qualifier le signal sonore qui lui parvient. *Écouter* renverrait alors à la recherche des sources sonores à partir des phénomènes acoustiques perçus (voir p.63). Enfin, *comprendre*, reviendrait, dans ce contexte à saisir le sens du son. Celui-ci est alors considéré comme un signe. Dans le cadre des situations expérimentales que nous avons mises en place, les auditeurs-participants, conscients et informés à la fois de la dimension expérimentale de la tâche, des conditions (récupération des résultats pour analyse, observation des comportements...) et de la tâche elle-même (consigne et objectifs) ont *a priori*, selon la terminologie de Schaeffer, à la fois *entendu*, *écouté*, et *compris* les sons (verbaux et expérimentaux) auxquels nous les avons soumis.

## 1.2. Les modèles de perception

Quelques soient les objectifs de l'écoute et les conditions dans lesquelles elle intervient, il apparaît que la première étape des traitements cognitifs que l'individu percevant appliquera aux *stimuli* correspond à ce que nous avons défini comme participant à la perception (elle même définie dans une acception très large). En effet, nous considérons ici la perception comme le premier processus permettant l'interaction entre l'individu et son environnement.

Traditionnellement, deux modèles de la perception coexistent. Le premier modèle considère que la perception est ascendante, c'est-à-dire qu'elle trouve son point

---

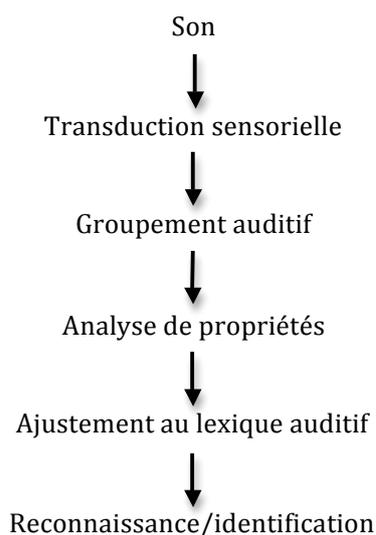
<sup>10</sup> ce qui est le cas de nos supports audio

d'entrée dans le signal. Il s'agit d'un processus qui exploite les informations qui proviennent de l'environnement. Le deuxième modèle considère que la perception est descendante, c'est-à-dire qu'elle trouve son point d'entrée dans la cognition de l'individu percevant. Ce processus intègre, en premier lieu dans la perception, les connaissances de l'individu.

Le modèle ascendant, également appelé *Bottom-up*, envisage la perception comme un processus dont le point de départ se trouve dans le signal sonore. Selon Carlson, (2010:1011) « *Le processus Bottom-up est le traitement de l'information qui parvient aux récepteurs sensoriels depuis le monde (...) Nous ne commençons à la percevoir que lorsque l'information stimule nos récepteurs (...) Le traitement Bottom-up de l'information disponible lors de la stimulation des récepteurs sensoriels constitue le début de la perception* »<sup>11</sup>. En d'autres termes, dans cette perspective, c'est à partir du son, et de sa transduction sensorielle que le processus perceptif démarre (Houix, 2003; McAdams, 1994). La transduction sensorielle correspond à une « *représentation brute du signal acoustique dans le système nerveux auditif périphérique* » (Houix, 2003:35). À la suite de cette étape de transduction viendrait une étape de groupements auditifs. Cette étape correspond à la séparation des sources lorsque plusieurs d'entre elles composent notre environnement (voir l'analyse des scènes auditives p.59). La troisième étape correspond à l'analyse des propriétés de chaque forme constitués dans l'étape de groupement. Dans ce modèle, aux propriétés prises en compte à ce niveau correspondraient certains paramètres physiques tels que la fréquence du son, sa durée ou son intensité par exemple. Enfin la dernière étape de ce modèle correspondrait au lexique auditif. Dans cette étape, les éléments du signal, extraits lors de l'analyse de ses propriétés, seraient associés à des catégories d'événements (sonores), permettant l'identification des sources. Houix, (2003) propose un schéma pour illustrer ce modèle (inspiré des travaux de Stephen McAdams (1994)) que nous reprenons et adaptons à notre tour.

---

<sup>11</sup> (Carlson, 2009:1011) « *Bottom-up processing is the processing of information from the world that falls on your sensory receptors (...) You only begin to perceive it when the information stimulates your receptors (...) Bottom-up processing of available information as it stimulates your sensory receptors constitutes the beginning of perception* »

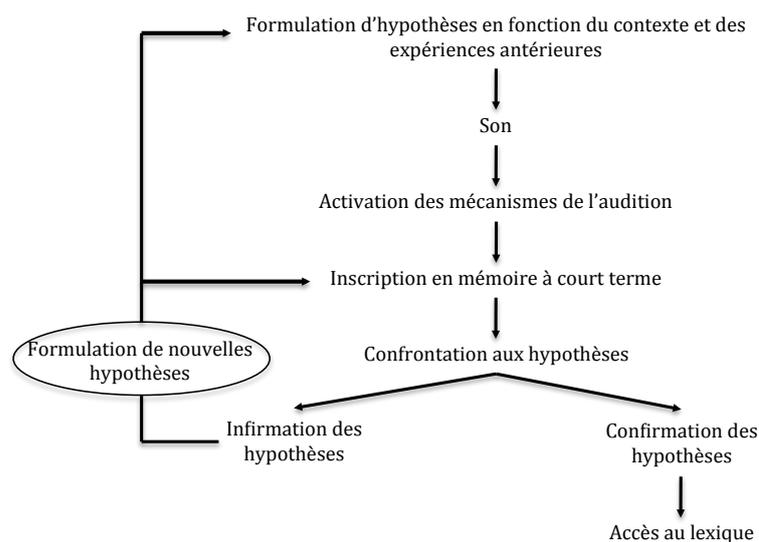


**Figure 3: Schéma du modèle *bottom-up* de la perception auditive selon Houix (2003) et McAdams (1994)**

Dans cette perspective, il apparaît également que l'information est considérée comme contenue dans le signal, voire même qu'information et signal ne font qu'un. Or, comme nous l'avons déjà dit, nous considérons que l'information résulte du traitement cognitif que l'individu applique au signal.

Parallèlement au modèle descendant, certains chercheurs se positionnent en faveur d'un modèle ascendant également appelé *Top-Down*. Celui-ci considère que la perception trouve son point d'entrée dans le système cognitif de l'individu. Celui-ci construirait des attentes (directement dépendantes du contexte, des objectifs d'écoute et de ses connaissances antérieures) et « irait chercher » dans le signal des indices lui permettant de confirmer ou d'infirmer ses hypothèses.

Dans ce contexte d'hypothèses formulées en amont de la confrontation au signal acoustique, les mécanismes de l'audition présentés plus haut (transduction sensorielle, groupement auditif, analyse des propriétés) s'activent. Après inscription en mémoire à court terme, notre système cognitif expérimenterait les hypothèses formulées en les comparant au lexique auditif activé et serait alors en mesure de confirmer ou d'infirmer les hypothèses. Dans ce cadre, si les hypothèses ne sont pas confirmées, d'autres sont formulées. La « boucle de traitement se rapporterait alors à l'inscription en mémoire à court terme (ou au son lui-même si celui-ci est répété). Pascal Gaillard (2000) a en outre proposé une représentation schématique de ce modèle, que nous avons reprise et adaptée :



**Figure 4: Schéma du modèle de perception Top-down, selon Gaillard (2000)**

Le point de départ de ce modèle correspond donc à la formulation d'hypothèses. En d'autres termes, ce modèle met en avant la formulation, par l'individu, d'hypothèses basées sur ces connaissances antérieures activées par le contexte et les objectifs de l'écoute, avant la présence du son. Ce sont ces hypothèses qui pilotent ensuite sa perception du monde (théorie du *knowledge driven* Dubois, 2000). Une fois en contact avec le son, les mécanismes de l'audition se mettent en place chez l'auditeur afin de procéder aux différentes étapes que nous avons exposées plus haut (transduction sensorielle, groupement auditif, et analyse des propriétés). Le résultat de ces différentes étapes de traitement serait alors inscrit dans la mémoire à court terme afin d'être confronté aux hypothèses de départ. Si la confrontation aux hypothèses les infirme, le système cognitivo-auditif de l'auditeur rétrocéderait de quelques étapes et formulerait de nouvelles hypothèses. Si les traitements les confirment, alors, la phase d'association avec le lexique intervient. Selon Carlson (2009:1011) « *La procédure Top-down est un processus qui opère sur l'information qui est générée intérieurement à partir de notre propre système cognitif, constitué par les apprentissages et les expériences* ». <sup>12</sup>

Si les deux modèles que nous venons d'exposer semblent s'opposer, de nombreux auteurs ont intégrés les démarches ascendante et descendante de manière à constituer un modèle plus global de la perception, dans lequel ces deux procédures

<sup>12</sup> (Carlson, 2009:1011) « *Top-down processing is processing that operates on information that is internally generated from within your own cognitive system, formed by learning and experience.* »

interviennent en complémentarité. Ainsi, Carlson (2009:1012) déclare que « *parfois, les informations top-down et bottom-up se complètent. D'un côté, si nous avons de nombreuses attentes top-down, nous recourons certainement moins à l'information bottom-up, et encoderons moins l'input qui nous vient du monde – juste assez pour vérifier la consistance de nos attentes. (...) D'un autre côté, quand il n'y a pas de forte attente top-down, ou peu d'expérience de l'objet perçu, nous aurons besoin de plus s'appuyer sur les informations bottom-up.* »<sup>13</sup>. Selon l'auteure, les procédures ascendante et/ou descendante se mettraient en place en fonction des objectifs de l'écoute, et de son contexte. Ce point est crucial dans notre travail puisqu'il met clairement en lumière que dans le cas de l'apprentissage, qui est un contexte particulier piloté par l'enseignant, une mise en contexte adaptée aux objectifs de l'enseignant entraînera une stratégie perceptive adaptée de la part de l'apprenant.

Si nous nous penchons sur les différences entre les modèles ascendants et descendants, alors une divergence essentielle apparaît néanmoins au niveau du point de départ de la perception. Selon le modèle ascendant, la perception démarrerait lorsque le système auditif de l'individu entre en contact avec le son. En revanche, pour le modèle descendant, la perception débute avant le contact avec le son : elle débiterait par la formulation d'attentes ou d'hypothèses quant à ce qui va être perçu. Dit autrement, dans ces deux modèles, les phénomènes attentionnels n'interviennent pas au même moment : alors que dans le modèle *Top-Down* l'attention intervient dès l'amont du processus perceptif (formulation d'hypothèses, émission d'intentions d'écoute), dans le modèle *Bottom-Up*, les phénomènes attentionnels interviennent dans un deuxième temps, une fois le signal sonore traité, au niveau physique. Malgré ce point de divergence fondamental entre les deux modèles, il apparaît que les connaissances de l'individu jouent un rôle primordial dans la perception, que ce soit au niveau de la formulation des hypothèses (selon le modèle descendant) ou de la confrontation au lexique (dans les deux modèles). Ainsi, Barsalou, (2008:91) affirme que « [les connaissances dans le système conceptuel] *supportent la perception en apportant les connaissances qui complètent les perceptions et qui génèrent l'anticipation* »<sup>14</sup>.

---

<sup>13</sup> (Carlson, 2009:1012) « *Sometimes the top-down and bottom-up information complete. (...) On one hand, if you have a lot of top-down expectation, you may rely less on the bottom-up information, and simply encode less of the input from the world – only enough to verify that it is consistent with your expectations. (...) On the other hand, when there is not a strong top-down expectation or not a lot of experience with given object, you will need to rely more on bottom-up information.* »

<sup>14</sup> « *It supports perception, providing knowledge that completes perceptions and that generates anticipatory* ».

Un second point de divergence entre les deux modèles est l'association au lexique auditif prévue par le modèle descendant. L'individu disposerait, dans ce modèle, d'un « *stock auditif* ». Cette considération impliquerait l'existence de représentations cognitives *modales*, c'est-à-dire dépendantes de la sensorialité mise en jeu lors de l'expérience. Dans cette optique, chaque objet du monde serait représenté dans différents lexiques mettant en jeu différents accès sensoriels. Un autre positionnement est toutefois envisageable : celui de l'existence de représentations cognitives *amodales*, c'est-à-dire composées des différentes facettes sensorielles d'un même objet du monde (pour un rapide état de l'art sur cette discussion voir Barsalou (2008)). Dans ce cadre, nous nous positionnons en faveur de la théorie des représentations amodales, composées de différentes facettes sensorielles.

Pour les expérimentations que nous avons mises en place (voir *Expérimentation Principale* p.165), nous considérons que 1) les consignes des tâches à effectuer mettent les participants en position d'activer en amont de l'écoute, l'ensemble des connaissances qu'ils possèdent à propos de la tâche (compréhension), du support utilisé (dialogue type dialogue pédagogique) et de la langue mise en œuvre (le français) ; 2) la perception d'éléments sonores environnementaux activera, si l'attention partagée le permet, des représentations cognitives associées à ces éléments, dont l'activation sera par la suite renforcée ou non par la redondance complémentaire qu'ils entretiennent avec le dialogue-cible.

## 2. L'analyse des scènes auditives (ASA)

La perception auditive est donc, quel que soit le modèle évoqué, un processus complexe qui aboutit à la compréhension de l'environnement qui l'entoure. Or, en dehors des situations de laboratoire, cet environnement est le plus souvent composé de sources multiples. En effet, dans les situations de la vie quotidienne, nous baignons dans un environnement bruité et la compréhension de cet environnement peut passer par une étape de décomposition de la « mixture sonore » en différents flux, potentiellement associés à différentes sources. Ainsi, comme le dit Alain de Cheveigné (2000) : « *Chaque oreille reçoit des ondes provenant d'une multitude de sources, mais on peut souvent porter son attention sur une source particulière et juger de sa sonie, de sa hauteur, de son timbre, voire comprendre ce qui est dit lorsqu'il s'agit de parole* » (p.171). En effet, nous décomposons (le plus souvent facilement et inconsciemment) la cacophonie qui nous entoure en identifiant certaines sources de manière irrépessible. Ainsi, l'une des

fonctions possibles de notre système auditif est d'analyser les patterns de vibration qui lui parviennent afin, si nécessaire, d'accéder aux sources qui les génèrent (Shamma & Micheyl, 2010). Ces patterns correspondent à « *des descriptions isolées fondées uniquement sur les composants issus d'une même source de l'environnement. Pour cela, il (le système cognitivo-auditif) utilise un processus appelé l' « Analyse des Scènes Auditives » »* (Bregman, 1994:12). L'analyse des scènes auditives (désormais ASA) correspond donc au processus par lequel l'individu auditeur extrait de l'environnement sonore les flux émis par les sources de son entourage (Bregman, 1990). Martin Cooke et Daniel Ellis (2001:142) nous rappellent que « *Bregman fait une distinction entre une source acoustique – un seul système physique qui donne lieu à un pattern particulier d'ondes sonores – et un flux auditif qui représente l'effet abstrait et conceptuel présent dans l'esprit de l'auditeur. Les auditeurs doivent résoudre un problème d'analyse des scènes auditives pour extraire un ou plusieurs flux auditifs pertinents à partir de la cacophonie de sources qui participent à leur environnement acoustique »*<sup>15</sup>. Nous distinguons trois niveaux conceptuels aux unités sonores : la source (objet du monde physique dont la mise en mouvement engendre des vibrations observables auditivement), le flux (ensemble acoustique issu d'une ou plusieurs source(s), associé à une unité cognitivo-représentationnelle par le sujet percevant), et le plan sonore (construction sonore composée d'un ou plusieurs flux). Le plan sonore renvoie à la fois au paysage sonore (*soundscape* (Schafer, 1977)) dans le but de rendre compte d'un environnement mais aussi à l'élaboration d'un paysage représentationnel re-construit par l'auditeur (voir le modèle *landscape* p.103)<sup>16</sup>.

Selon Bregman (1994:14-15), trois processus sont mis en œuvre dans l'ASA. Le premier processus relevé par l'auteur correspond à « *l'activation purement automatique des schémas appris* ». Dit autrement, l'auditeur disposerait de représentations cognitives acoustiques des différents patterns préalablement perçus, construites empiriquement, à partir de ses expériences antérieures. De nombreux auteurs, tels que Husserl, (1950, cité par Barabas, 2009), Merleau-Ponty, (2002), Barsalou (2008) ou Barabas (2009) considèrent la perception comme une expérience à partir de laquelle l'individu extrairait des connaissances sur le monde, connaissances qui, par la suite, participeraient à la construction des représentations cognitives de l'individu. Ainsi, la perception d'une source, dont le flux a d'ores et déjà permis l'élaboration d'une représentation, *i.e.* dont la

---

<sup>15</sup> (Cooke & Ellis, 2001:142): « *Bregman 1984, 1990) draws a distinction between an acoustic source – a single physical system giving rise to a particular pattern of sound waves- and an auditory stream which denotes the abstract, conceptual effect it has in the mind of the listener. Listeners have to solve an auditory scene analysis (ASA) problem in order to extract one or more relevant auditory streams from the mixture of sources which contribute to their acoustic environment* ».

<sup>16</sup> Pour plus de précision, au chapitre sur *l'Élaboration du matériel expérimental* (p.115)

représentation a déjà été élaborée à partir d'expériences perceptives auditives antérieures, impliquerait l'activation de cette représentation de manière à la fois automatique et irrépessible. En outre Alain & Woods (1997:534) affirment que « *dans la perception auditive, les patterns temporeux sont cruciaux pour comprendre et interpréter les signaux complexes, tels que la parole ou la musique* »<sup>17</sup>, laissant penser que l'élaboration de schémas ne s'impose que dans le cadre de la perception de signaux complexes.

Le deuxième processus mis en avant par Bregman (1994) correspond à « *l'utilisation volontaire de schémas* ». Ce deuxième processus serait mis en œuvre lorsque l'auditeur cherche à percevoir quelque chose, faisant ainsi référence aux processus attentionnels et plus spécifiquement à l'attention préparatoire. « *Du point de vue cognitif, on peut considérer que l'orientation de l'attention avant l'apparition d'une cible à laquelle le sujet doit impérativement répondre contribue à réduire partiellement l'incertitude du sujet quant à l'endroit, au moment au type d'information présenté et à la nature de la réponse. L'orientation préalable de l'attention apporte donc une information qui, en réduisant l'éventail des choix, réduit corrélativement la latence des réponses. Moyennant quoi, selon le nombre d'informations préalables délivrées et le caractère pertinent de celles-ci, le temps de réaction mesuré est de moins en moins un temps de réaction de choix et de plus en plus un temps de réaction simple* » (Camus, 1996:26). Dans *l'utilisation volontaire de schémas*, l'individu percevant serait à la recherche d'un pattern correspondant à son attente, *i.e.* à la représentation (cognitive) du phénomène attendu.

Enfin, le troisième processus évoqué par l'auteur correspond à une « *analyse primitive* », c'est-à-dire à une analyse des propriétés physiques acoustiques générales du son lorsque celui-ci ne renvoie à aucun schéma appris. Le troisième processus correspondrait alors à l'apprentissage des schémas, c'est-à-dire l'expérience à la base de l'élaboration de la représentation (cognitive) du phénomène acoustique.

L'identification des sources dans un environnement bruité fait donc intervenir différents processus de haut niveau, tels que la mémoire ou l'attention (Chanquoy *et al.*, 2007) et de bas niveau, tels que l'analyse primitive.

Il apparaît donc, selon les travaux de Bregman que les individus percevants, en situation d'identification des sources qui composent leur environnement sonore, s'appuient en priorité sur leurs connaissances antérieures (activation de schémas appris

---

<sup>17</sup> « *In auditory perception, temporal patterns are critical for understanding and interpreting complex signals such as speech and music* »

ou utilisation volontaire de schéma), et que l'analyse primitive des sons ne se mette en place que dans les situations où les connaissances antérieures sont suffisantes pour la réalisation de la tâche.

Deux manières d'appréhender les sons environnementaux et leur étude coexistent, et ces deux axes se retrouvent dans la littérature. Le premier axe est centré sur les mesures physiques des dimensions du son, et le deuxième, est centré sur leur interprétation. Guyot, Castellengo, Fabre (1997:45), suggèrent que « *la première attitude indique une écoute analytique des sons, alors que la deuxième semble être le fruit d'une écoute globale des événements sonores* » (p.45). Nous pouvons supposer que ces différentes attitudes face au son relèvent de l'expertise. Il semblerait en effet que les *experts* soient plus prompts à un traitement catégoriel en intension, c'est-à-dire un mode de traitement « *centré sur l'analyse des propriétés des objets* » (Castellengo & Dubois, 2007:28) alors que les *novices* préféreraient un traitement en extension (centré sur la source) qui semble correspondre à « *un niveau plus « ordinaire » de traitement* » (*ibidem*). La tâche 1 (catégorisation libre), dont Guyot et ses collaborateurs (1997, pp.43-49) rendent compte, montre des traitements à la fois extensionnels (*a priori* traitement expert) et intensionnels (*a priori* traitement novice) de la part des participants. Toutefois il est important de mettre ici la lumière sur les neuf personnes interrogées, qui relèvent majoritairement de la catégorie « *experts* » (« *membres du Laboratoire d'Acoustique Musicale ayant tous une pratique de l'écoute* » p.45). Nous pouvons donc supposer que si le panel de participants avait correspondu uniquement à des non-musiciens, et/ou plus généralement des individus n'ayant pas un rapport professionnel à l'écoute, les résultats seraient possiblement différents. Aurait-on retrouvé la même proportion de traitements intensionnels et extensionnels, ou au contraire aurions-nous observé des traitements uniquement intensionnels ? En effet, il apparaît (Bregman, 1994; Dubois, 2000) que c'est lorsqu'il ne parvient pas à identifier la source que l'individu novice met en place des stratégies centrées sur les caractéristiques physiques du son. En d'autres termes, les auteurs suggèrent que les novices auraient tendance à caractériser les sources des phénomènes acoustiques lorsqu'ils l'identifient. Ils privilégieraient donc un traitement en extension. Toutefois, lorsque celui-ci est impossible en raison de la non-identification de la source, les novices procéderaient à un traitement en intension en caractérisant le phénomène acoustique principalement selon ses caractéristiques physiques. Toutefois, ces différentes attitudes relèvent également de la tâche, du contexte et des objectifs de l'écoute.

Dans ce cadre, nous pouvons considérer que dans les expérimentations que nous avons mises en place (et plus particulièrement l'identification des scènes sonores (voir p.138) et la tâche de compréhension de dialogues accompagnés de contextes sonores

environnementaux dans le *Classe-Images* (voir p.165)), les participants interrogés (qui *a priori* ne présentent pas d'expertise particulière en traitement des sons environnementaux) procéderont à un traitement en extension des scènes sonores environnementales. Dit autrement, nous pensons que ces participants en situation d'interprétation des sons environnementaux privilégieront une approche basée sur les sources à une approche basée sur les caractéristiques physiques des sons.

Certains auteurs (Barsalou, 1983; Gaillard, Magnen, & Billières, 2006) suggèrent qu'en plus de la notion d'expertise, nous pourrions imaginer la perception comme étant directement liée à l'action, et à son contexte. « *La notion même d'expertise est alors vue non pas comme une « capacité à faire » mais comme une « nécessité de faire ».* » (Gaillard *et al.*, 2006:14) En d'autres termes, dans une certaine mesure et dans certains cas (il s'agissait ici de la perception des sons de la parole), le fait que l'individu soit capable d'une écoute analytique fine n'implique pas que ce type d'écoute soit systématique. Le « choix » entre ces deux types d'écoute se ferait essentiellement en fonction des besoins de l'individu et de ses objectifs. La question des objectifs d'écoute est donc de nouveau posée. Aussi, dans le cadre de l'interprétation des scènes sonores, la question de l'identification des sources est entière : les auditeurs-compreneurs ont-ils besoin d'identifier les sources pour interpréter ces scènes ? S'agit-il pour eux de décomposer le flux en sources ou au contraire sont-ils en situation de compréhension globale de la situation ?

Par ailleurs, au-delà de l'expertise des sujets, nous pouvons considérer que la culture d'un individu influencera sa démarche en tâche de perception, de compréhension et de catégorisation. En effet, selon ses expériences antérieures (constitutives de sa culture), l'individu acquerra des connaissances différentes sur son environnement sonore, développera des capacités particulières dans l'interprétation des *soundscape*s, et dans l'identification des sources, et adoptera donc un positionnement différent face à l'interprétation de son environnement sonore: nous pouvons considérer que la vie quotidienne façonne les connaissances au même titre qu'un apprentissage déclaratif ou qu'une pratique professionnelle (Dubois, 2000).

## **2.1. La perception des sources dans un environnement bruité**

L'analyse des scènes auditives est donc ce qui permet de discriminer les différents flux, les différentes sources qui composent notre environnement sonore. Deux processus sont en œuvre dans cette analyse : **l'intégration**, qui permet de réunir dans un même schéma les différentes parties d'un flux, et la **ségrégation**, qui permet de

différencier les flux entre eux. Dans le cadre des expérimentations que nous avons mises en place et plus particulièrement dans les conditions d'écoute des dialogues accompagnés d'un contexte sonore environnemental, les participants, en fonction des consignes que nous leur donnerons, devront procéder à l'ASA, afin de reconstruire les sons perçus en unités signifiantes.

### **2.1.1. L'intégration (ou fusion)**

D'après Cooke, Brown, Crawford, & Green (1993:186) l'ASA impliquerait plusieurs étapes : « *dans un premier temps, le signal acoustique est décomposé en plusieurs « constituants sensoriels ». Par la suite, les composants qui proviennent du même événement environnemental sont recombinaés dans des structures perceptuelles qui peuvent être interprétées via des processus de haut-niveaux* »<sup>18</sup>. Qu'elle soit ascendante ou descendante (voir schémas pages 56 et 57), l'Analyse de Scènes Auditives passe (à un moment ou à un autre) par une étape d'analyse du signal sonore de manière à extraire les indices qui lui permettent de « reconstruire » les flux de la 'cacophonie' dans une étape d'analyse des propriétés physiques du son. Ainsi, Albert Bregman (1994) a proposé quatre régularités présidant à l'intégration de différents items sonores en un même flux. La régularité n°1 repose sur le fait qu'« *il est extrêmement rare que des sons qui n'ont rien à voir démarrent et s'arrêtent en même temps* ». (p.19). Ainsi, l'attaque et la fin d'un son seront considérés comme des indices importants dans l'ASA. La deuxième régularité présentée par l'auteur concerne la « *progression de transformation* ». En effet, les propriétés d'un son ou d'une séquence de sons issue de la même source tendent à se modifier lentement et de façon continue. La troisième régularité correspond aux harmoniques qui composent le son. En effet, « *lorsqu'un corps vibre à une période répétée, ses vibrations donnent naissance à un pattern acoustique dont les fréquences des composants sont des multiples d'une même fréquence fondamentale* » (p.31). Enfin, la quatrième régularité relevée par l'auteur repose sur l'évolution des sons : « *la plupart des modifications qui surviennent dans un signal affecteront tous les composants du son résultant, de manière identique et simultanée* » (p.33). Nous notons toutefois qu'alors que la première régularité mise en avant par l'auteur se rapporte aux flux<sup>19</sup> (potentiellement composés de plusieurs sources), les trois autres se rapportent spécifiquement aux sources. Ces quatre régularités sont considérées comme à la base de l'ASA et sont plus

---

<sup>18</sup> « *In the first stage, the acoustic signal is decomposed into a number of 'sensory components'. Subsequently, components which are likely to have arisen from the same environmental event are recombined into perceptual structures that can be interpreted by higher-level process* ».

<sup>19</sup> En musique par exemple, il est fréquent que plusieurs instruments (dont chacun se distingue par les trois autres régularités) entament ou terminent une mélodie en même temps

ou moins reprises par Alain de Cheveigné (2000:2) : « *Le premier facteur de cohésion est la simultanéité d'attaque et plus généralement la communauté de variation des composantes fréquentielles du son. Lorsque toutes démarrent en même temps, on tend à percevoir une source unique. Une asynchronie d'attaque évoque la perception d'objets multiples. C'est un exemple plus général de destin commun. Le deuxième facteur important est l'harmonicité (...). Le troisième facteur est la cohésion binaurale* ». Enfin, il est important de ne pas oublier les propriétés du timbre qui semblent influencer fortement l'intégration des sons à un même flux bien que moins fortement que la  $f_0$  (Bregman, 1990; Cooke & Ellis, 2001)<sup>20</sup>.

Ainsi, les différents constituants du son seront regroupés dans le flux obéissant au plus grand nombre de facteurs et principalement aux facteurs qui sont privilégiés par le système auditif, par rapport notamment aux schémas appris. Denham & Winkler (2006:155) nous rappellent que « *de manière générale, trouver ce qui va ensemble au sein d'un input auditif en cours requiert presque toujours des informations sur le comportement préalable de la source active du son* »<sup>21</sup>. Les connaissances antérieures jouent donc un rôle primordial dans l'intégration de différentes unités sonores à un même flux. Lorsque les auditeurs disposent d'éléments de connaissance sur la source, le flux ou plus généralement la situation à laquelle ils sont confrontés, l'analyse des scènes auditives est facilitée. Dans cette perspective et dans le cadre des expérimentations que nous avons mises en place, nous pouvons considérer que les scènes sonores que nous avons construites auront été différemment comprises selon la culture d'origine des participants aux tests (voir p.138). En effet, les sons environnementaux varient d'un environnement à l'autre, c'est-à-dire d'une culture à l'autre.

### ***2.1.2. La ségrégation (ou scission)***

Si l'ASA permet de regrouper les éléments sonores qui vont ensemble (issus de la même source et/ou participant au même flux), elle permet aussi d'isoler chaque groupe de sons.

Ainsi, il apparaît que les sources, lorsqu'elles ne correspondent pas à un pattern appris, sont le plus souvent ségréguées en fonction de paramètres physiques acoustiques tels que leur spatialisation, leurs fréquences fondamentales et leur

---

<sup>20</sup> Rappelons néanmoins que ces régularités ont été démontrées à partir de sons de synthèse, non complexes et répétés, alors que, dans l'élaboration de nos scènes sonores (voir p.117), nous travaillons à partir de scènes sonores plus complexes, élaborées à partir de sons semi-authentiques.

<sup>21</sup> « *In general, finding out what belongs together within the current auditory input almost always requires information about the previous behavior of the active sound sources* ».

enveloppe fréquentielle. L'auditeur procède alors à l'ASA via une procédure *primitive* ou *bottom-up* (Cooke & Ellis, 2001). « *Alors que les [procédures primitives] sont basées sur des capacités innées et reposent sur des principes valides pour la plupart des sons (...) les [procédures basées sur les patterns] requièrent l'apprentissage de règles, sont influencées par les expériences antérieures et pourraient s'appliquer à des sous-ensembles de sons* » (Denham & Winkler, 2006:155)<sup>22</sup>. Selon les auteurs, ces deux procédures s'opposent. Toutefois, nous ne rejoignons pas ce positionnement, et considérons que ces procédures interviennent en complémentarité. En effet, nous pouvons envisager que les procédures primitives sont impliquées dans l'ASA lorsque l'individu est confronté à des sons dont il ne reconnaît pas les patterns ou que la tâche l'amène à effectuer une analyse « *centrée sur les propriétés des objets* » (Castellengo & Dubois, 2007:28). En allant plus loin, nous pouvons même émettre l'hypothèse de démarches primitives qui pourraient correspondre à une étape de l'apprentissage des patterns. Cette hypothèse nous amène à considérer que l'ASA (en tant que processus cognitif) est susceptible de faire intervenir l'expertise de l'individu, en fonction de ses connaissances et expériences antérieures, des patterns qu'il a déjà construits.

Par ailleurs, Cooke & Ellis (2001:166) suggèrent que lorsque les indices physiques ne sont pas suffisants, alors le système auditif a recours à des éléments d'informations extérieurs, tels que le contexte par exemple. Ce point est particulièrement important dans le cadre de notre travail car notre hypothèse principale est que présenter un contexte sonore environnemental peut favoriser la compréhension (de certains aspects au moins) d'un dialogue en langue étrangère, par des apprenants débutants de cette langue. Ainsi, si le système auditif a recours au contexte pour désambiguïser des éléments sonores dont les indices physiques ne permettent pas l'identification, on peut penser que le système cognitif suit les mêmes procédures pour appréhender le sens d'un message verbal difficile à comprendre, à cause du caractère débutant en langue du sujet compreneur.

En outre, il apparaît que lorsque les éléments sonores sont ségrégués en différents flux, différentes populations de neurones sont activées. En revanche, lorsqu'on ne perçoit qu'un seul flux, une seule population est activée (Shamma, Elhilali, & Micheyl, 2011). Ainsi, la détection des flux a-t-elle une importance majeure dans la compréhension de l'environnement sonore et dans l'interprétation des scènes sonores

---

<sup>22</sup> « *Whereas the [primitive processes] are based on innate capabilities and rely on principles that are valid for most sounds (...), the [schema-based processes] require the learning of rules, are influenced by previous experience and may apply to subset of sounds* ».

que nous avons construites (voir p.121) et sur lesquelles nous avons interrogé les participants (voir p.138).

Enfin, d'après Thompson, Carlyon, et Cusack (2011), la ségrégation implique à la fois la séparation des flux et l'« aveuglement » perceptuel des flux non-cibles de manière à focaliser l'interprétation sur le flux visé (généralement de la musique ou de la parole dans la littérature scientifique). Ainsi, la première ségrégation semble être la distinction sons de la parole/autres sons (Belin, Zatorre, & Ahad, 2002; Cooke & Ellis, 2001). Les auteurs nous font remarquer que « (...) *l'étendue des situations dans lesquelles la communication orale est possible alors que d'autres sources sont en concurrence souligne la flexibilité et la robustesse de la perception de la parole humaine* »<sup>23</sup> (p.141). Selon les auteurs, la scission parole vs « autres sources » implique le recours à des patterns appris. Toutefois, la focalisation sur le signal de parole peut aller plus loin que la discrimination parole vs « autres sons ». Nous pouvons également nous concentrer sur un discours, une parole plutôt qu'un autre ou nous centrer sur un flux non linguistique pourtant en concurrence avec des sons de la parole. Cherry (1953) a affirmé que, en situation d'écoute dichotique, les individus peuvent facilement se centrer sur un message cible et rejeter les autres (l'auteur a montré que les individus présentent une capacité réduite à relater le message non-cible). C'est ce qu'il a appelé « l'effet cocktail party ». Ainsi, nous pouvons nous interroger sur la pertinence de la distinction parole/ensemble des autres sons. En effet, cette position laisse entendre que la parole serait intrinsèquement différente des autres flux rencontrés qui, en ce sens, composeraient un ensemble dont le dénominateur commun serait le caractère non-linguistique. Or, selon nous, et comme nous avons pu le constater, notamment dans notre première étude sur les attentes auditives (voir p.121), ce sont essentiellement les processus attentionnels et les procédures de focalisation qui amènent l'individu à séparer la parole, généralement cible, des sons de l'environnement (qui peuvent contenir des sons de parole, par exemple dans l'effet *cocktail party* (Cherry, 1953)).

## 2.2. L'analyse des scènes auditives : l'influence de l'attention

À l'heure actuelle de nombreuses recherches portant sur l'analyse des scènes auditives s'interrogent sur l'influence de l'attention dans le processus de formation des flux (par exemple Shamma *et al.*, 2011; Sussman & Steinschneider, 2009; Thompson *et al.*, 2011). En effet, il s'agit de savoir si l'attention participe à l'élaboration des flux par

---

<sup>23</sup> « (...) *the range of situations in which spoken communication is possible in the presence of competing sources highlights the flexibility and robustness of human speech perception* ».

l'individu ou si elle n'intervient qu'au niveau de leur(s) éventuel(s) maintien(s). Dans le cadre de la tâche à laquelle nous avons soumis les participants des tests sur les scènes sonores (voir p.138) et de notre expérimentation principale (voir p.165), considérer que l'attention concourt à la formation des flux revient à considérer que les connaissances antérieures des individus et leurs objectifs d'écoute jouent un rôle dès la perception (envisagée dans le sens de première interaction entre l'individu et les éléments sonores), alors que si l'attention ne participe qu'au maintien des flux (et non à leur élaboration), alors, cela implique que ces connaissances interviennent principalement dans les processus de compréhension que nous décrirons dans le prochain chapitre (p.81). Dit autrement, la question n'est pas tant de savoir si les processus attentionnels interviennent, mais plutôt de savoir *quand* (en amont ou en aval de la formation des flux): « *Une question d'intérêt grandissant est de savoir si la construction de flux auditifs provient de mécanismes automatiques de bas-niveau ou si elle est connectée à des processus de haut-niveau tels que l'attention. Une telle influence de l'attention pourrait intervenir soit en raison des mécanismes responsables de la ségrégation et de l'intégration dans le système auditif central, et/ou en raison de la modulation « top-down » des processus périphériques (...) Bien que certains processus qui interviennent en dehors de l'attention aient un effet sur la construction de flux, il y a aussi selon nous, des signes d'une potentielle influence de l'attention* »<sup>24</sup> (Thompson *et al.*, 2011:2). En d'autres termes, selon les auteurs, l'attention n'est pas un composant nécessaire de l'ASA mais peut jouer un rôle important dans l'élaboration et le maintien des flux, *i.e.* dans l'ASA en général.

À l'instar de Shamma et ses collaborateurs (Shamma *et al.*, 2011), on peut considérer que la sélection, par le système attentionnel, de certains éléments est une « mesure d'économie ». En effet, comme le soulignent les auteurs, le traitement de l'ensemble des flux par le cerveau impliquerait une forte charge cognitive. Il semble donc logique, dans ce contexte, de considérer que le traitement serait d'abord global, avant de ségréguer les flux, en fonction des objectifs de l'écoute. Ainsi, comme l'affirme Camus (1996:17) « *la sélection d'une petite quantité d'informations parmi toutes celles qui nous environnent au même instant apparaît comme une opération des plus curieuses de notre système cognitif de traitement de l'information. Isoler dans une foule de gens le visage attendu ou ne suivre dans le brouhaha d'une salle de réunion que la conversation des voisins constituent des manifestations familières de cette opération* », cette opération

---

<sup>24</sup> « *A question of increasing interest concerns whether auditory streaming arise solely from automatic, low-level mechanisms, or whether it is intimately connected with higher-level processes such as attention. Such an influence of attention could occur either due to mechanism responsible for segregation and integration operating in the central auditory system and/or due to « top-down » modulation of peripheral processes (...) Although processes that occur in the absence of attention have an effect on streaming, there is also, we believe, good evidence that streaming can be influenced by attention.* ».

relevant avant tout de phénomènes attentionnels. Par ailleurs, Bregman (1994), considère que ce sont les processus attentionnels qui effectuent la sélection des flux poursuivis, c'est-à-dire que ce processus de haut niveau participerait au maintien des flux mais pas forcément à leur formation. Selon Alain & Woods (1997:534) « *En plus des facteurs pilotés par le stimulus, la perception des patterns auditifs est influencée par l'attention et l'expérience antérieure de l'auditeur* »<sup>25</sup>. Sussman & Steinschneider (2009) vont également dans ce sens en suggérant que l'attention interagirait avec les processus de bas niveau en régulant et maintenant les représentations mentales des différentes sources en accord avec les objectifs spécifiques de l'écoute. Pour ces auteurs, l'attention jouerait un rôle important dans l'association des flux avec des représentations mentales (voir le sous chapitre sur le sémantisme des sons de l'environnement p.72) et donc, dans la compréhension des sons environnementaux. Dans ce cadre, l'attention est considérée comme jouant un rôle dans le maintien des flux et la focalisation qu'ils peuvent susciter.

Un autre argument en faveur de l'influence de l'attention sur la formation des flux auditifs est l'effet « cocktail party », évoqué plus haut. Ce phénomène intervient lorsque nous nous trouvons dans un environnement bruyant et que nous focalisons notre attention sur un flux particulier (effet cocktail party, Cherry, 1953). Malgré les différents flux (conversations dans le cas de cette étude) qui nous parviennent en même temps, nous sommes tout à fait capables de nous focaliser sur les propos d'un seul locuteur. L'auteur a en effet mis en évidence l'incapacité des participants à son étude (dans une situation d'écoute binaurale) de rendre compte des éléments contenus dans les flux non-cibles. Il apparaît donc que la focalisation attentionnelle sur un flux donne lieu à une certaine « atténuation » du/des flux non cible(s) même s'ils sont composés de sons de la parole. Il semble donc que c'est bien l'attention qui guide l'ASA et les traitements appliqués aux unités perçues. En outre, Botte, Drake, Brochard, McAdams, & Argenti (1994) ont montré qu'une différence perceptuelle de 7 dB entre la cible et les distracteurs apparaît en situation naturelle d'écoute. Dit autrement, la focalisation attentionnelle sur un flux (cible) engendrerait un rehaussement perceptuel de ce flux, ou, au contraire, une atténuation (perceptuelle toujours) du flux non cible. Ainsi, il apparaît que l'attention portée à un flux a un impact sur la perception globale de l'ensemble de l'environnement sonore. En outre, des études ne faisant intervenir qu'une seule cible ont montré que la détection de cette dernière sera affaiblie quand l'attention en est détournée (Tremblay, Vachon, & Jones, 2005).

---

<sup>25</sup> « *In addition to stimulus-driven factors, the perception of audition patterns is influenced by the listener's attention and prior experience* ».

Enfin, le phénomène de l'aveuglement attentionnel (*attentional blinding*) vient renforcer les hypothèses en faveur d'un rôle de l'attention dans la formation des flux. Il s'agit du phénomène qui intervient lorsque deux cibles se suivent de près (entre 200 et 500ms d'écart). En général, dans cette configuration, la deuxième cible n'est pas identifiée par les participants (Dux & Marois, 2009). Certains auteurs (notamment Tremblay *et al.*, 2005) avancent l'hypothèse selon laquelle la modalité auditive serait dotée d'un mécanisme équivalent à la modalité purement visuelle du calepin visuo-spatial (également appelé VSTM *Visual short-term memory*) ; ce mécanisme est appelé le ESTM (*echoic short-term memory*). Dans ce modèle, la *confusion* qui amène le sujet à ne pas percevoir la deuxième cible serait due à la compétition parmi les stimuli qui doivent être rappelés. Ces observations nous amènent à envisager que l'attention jouerait un rôle à la fois dans la formation des flux et dans la focalisation sur la ou les cible(s). Ainsi, dans le cadre de nos expérimentations, certains sons de la parole, contenus dans ce que nous avons appelé les scènes sonores (voir p.121), seraient considérés comme participants à l'environnement sonore, au 2<sup>ième</sup> ou 3<sup>ième</sup> plan, alors que d'autres sons linguistiques, en l'occurrence le dialogue<sup>26</sup> seraient considérés comme cibles, impliquant une plus forte focalisation et un traitement plus profond. L'ensemble de ces données plaide en faveur de l'hypothèse d'une Analyse des Scènes Auditives pilotée par les phénomènes attentionnels.

Toutefois, certains auteurs (par exemple Shamma *et al.*, 2011) semblent considérer que la formation des flux auditifs serait indépendante des phénomènes attentionnels. Sussman, Horvath, Winkler, & Orr (2007) rappellent que dans ses travaux, Bregman (1990) suggérait que la ségrégation des flux intervenait en dehors de l'attention, dans le cadre d'une perception *bottom-up*. Denham & Winkler (2006:155-156) proposent en outre la même interprétation du modèle de Bregman. Dans ce cadre, encore une fois, les auteurs semblent considérer que les phénomènes attentionnels interviennent en aval de la formation des flux, c'est-à-dire, dans leur maintien.

Ainsi, il apparaît que, quelle que soient les perspectives retenues, la formation des flux et leur maintien est fortement lié aux objectifs de l'écoute et aux phénomènes attentionnels qui seront mis en place par l'individu perceuteur. En outre, il apparaît que dans le cadre des expérimentations que nous avons mises en place (notamment *la construction des scènes sonores* (p.132), *l'identification des scènes sonores* (p.138) et *l'expérimentation principale* (p.165)), les participants ont dû, pour accomplir la tâche,

---

<sup>26</sup> dont l'intensité en dB est plus forte que celle de la scène sonore au deuxième plan

interpréter des scènes sonores environnementales et mettre en place les stratégies que nous avons présentées ici. Enfin, considérant que les phénomènes attentionnels ont un impact fort sur la formation des flux, et donc la compréhension de l'environnement sonore, l'attention sera également prise en compte dans la compréhension de dialogues pédagogiques accompagnés d'un contexte sonore, en situation expérimentale ou de classe.

Comme nous l'avons vu plus haut, nous considérons que les objectifs d'écoute président à l'ensemble des traitements auditifs effectués par l'individu. Ainsi, selon nous, celui-ci effectue une première ségrégation au sein de la « mixture sonore » qui lui parvient en discriminant les sons « cibles » (1<sup>er</sup> plan) des sons « de l'environnement » (2<sup>ième</sup> et/ou 3<sup>ième</sup> plan(s)). Cette discrimination s'effectuera en fonction de ses connaissances et expériences antérieures. Dans le cadre de cette étude et plus particulièrement de notre expérimentation principale (voir p.165) les sons cibles correspondent aux dialogues (voir p.151) et les sons de l'environnement au « contexte sonore environnemental » parfois composés en partie de sons de parole (voir p.121). Ce contexte sonore environnemental correspond donc à l'ensemble des sons perçus par l'individu (quelle que soient leur nature, en dehors des sons cibles). Toutefois, malgré les effets de la focalisation attentionnelle sur un flux particulier sur le(s) flux non-cibles, nous considérons qu'un partage de l'attention permet un traitement simultané de divers flux. Ainsi, *« on peut aussi considérer que l'attention intervient pour déterminer des priorités de traitement. L'attention effectue un choix dans l'ordre des traitements qu'il convient de mettre en œuvre à un moment donné. (...) Les représentations sensorielles qui présentent alors une signification particulière pour l'organisme sont traitées prioritairement et peuvent ainsi accéder à des analyses perceptives et cognitives plus élaborées (notamment conduire à une identification consciente) (...). L'attention, en déterminant l'ordre de traitement, détermine aussi les informations qui seront analysées complètement par le système cognitif, ce qui permet d'expliquer le caractère superficiel des traitements réservés à l'information inattendue »* (Camus, 1996:25). Pour la tâche de compréhension de dialogues que nous avons proposée dans notre expérimentation principale (voir p.165), les participants se sont *a priori* focalisés sur les dialogues linguistiques (au premier plan) tout en traitant les contextes sonores environnementaux (au deuxième plan). Notre hypothèse à ce niveau est que ces scènes sonores, en position contextuelle, sont chargées sémantiquement et que leur perception donne lieu à l'activation de représentations cognitives.

### 3. Le sémantisme des sons de l'environnement

À l'heure actuelle, notre environnement se caractérise de plus en plus par son caractère bruyé, et ce d'autant plus en contexte urbain « *Ainsi, au moins depuis la première révolution industrielle, dans les villes et plus particulièrement dans les métropoles, on constate une multiplication des sources sonores* » (Pecqueux, 2012:5). Dans ce cadre, nous pouvons considérer les sons comme autant d'indices de l'environnement. En outre, l'audition est irrépessible : alors qu'il est possible de fermer les yeux par exemple, et ainsi interrompre toute perception visuelle, il est impossible d'interrompre naturellement la perception auditive. Cet aspect nous semble primordial car ce paramètre, couplé à la multiplication des sons environnementaux met sur le devant de la scène les études portant sur la valeur subjective des sons, parfois considérés comme autant de nuisances (Dubois *et al.*, 2006; Jaworski, 2012), et qui semblent aujourd'hui correspondre à une forme de pollution redoutée (Jaworski, 2012). Selon Schafer, cette *pollution sonore* apparaîtrait « *quand l'homme n'écoute plus, car il a appris à ignorer le bruit* » (Schafer, 2010:24). En d'autres termes, selon l'auteur, nous considérons comme nuisants les sons auxquels nous n'attribuons pas de qualité informationnelle. Cette position nous amène donc à penser que la « valeur » d'un élément sonore est dépendante de la perception de l'individu, de ses objectifs d'écoute et donc son éventuel caractère de nuisance n'est pas une dimension intrinsèque de ses propriétés physiques mais le résultat d'un traitement attentionnel.

#### 3.1. Des nuisances sonores à la sémantique du son

Prenons l'exemple de récentes lois allemandes. Jusque là, l'Allemagne avait statué sur les nuisances sonores à partir d'un seul paramètre physique acoustique : l'intensité (en dB). Mais en 2010, après la multiplication des plaintes, accompagnées d'illustrations sonométriques, relatives aux « *nuisances sonores dommageables* » enregistrées à proximité des crèches, la ville de Berlin (rapporté par Pecqueux, 2012:7-8) a édicté une loi tolérant « *les sons des enfants en tant que signe de leur épanouissement* » (*ibidem* :7-8). Cette loi a depuis été reprise par le Gouvernement fédéral allemand « *déclarant les sons/bruits des enfants acceptables socialement quelle que soit leur mesure en décibels, jusqu'à les promouvoir au rang de « musique vivante de l'avenir »* » (*ibidem* :8). Il apparaît donc que l'approche physique (et considérée comme objective) des sons n'est pas suffisante à l'appréhension de la complexité de l'environnement sonore. Nous nous positionnons donc en faveur d'une approche cognitive de la perception auditive (Dubois *et al.*, 2006; Dubois, 2000), centrée sur l'interprétation des *inputs* sonores en tant qu'éléments sémantiques. Ce positionnement

en faveur d'une approche subjective de la perception est également présent chez Barabas (2009) qui affirme que « *la perception n'est pas un événement objectif, elle est un acte subjectif; la sensation n'est pas présente à la conscience comme une chose est présente dans le monde, c'est plutôt la perception qui se rend l'objet présent, qui se le présente (...)* En tant qu'expérience, en tant que « subjective », la perception suppose un acte par lequel la conscience appréhende tel contenu, c'est-à-dire lui confère un sens. » (p.26).

Le travail sur les nuisances sonores a donc mis le doigt sur les limites de la description purement physique et a donné lieu à une approche cognitive des sons de l'environnement (Dubois *et al.*, 2006). En effet, comme nous le font remarquer les auteurs, les démarches « objectivantes », issues de la physique acoustique par exemple, laissent bien souvent de côté la dimension sémantique du son, prise en charge par les études en psychologie. Par ailleurs, certains de leurs résultats semblent indiquer que les sons environnementaux complexes sont traités comme des événements chargés de signification apportant des informations pertinentes sur les possibles interactions avec l'environnement (Dubois *et al.*, 2006 :868). Il apparaît donc que les sons environnementaux sont susceptibles de donner lieu à une interprétation et qu'ils peuvent effectivement être considérés comme autant d'indices de l'environnement. Jaworski (2012 :85) va dans le même sens en rappelant que « *l'intensité sonore ne peut fournir l'unique critère de définition de la nuisance, car une émission acoustique même faible, si elle est répétée, peut occasionner une gêne. Il existe un sens dans tout bruit influencé par des facteurs de situation, et le bruit est d'autant plus agressif qu'il n'a pas de sens pour celui qui le subit* ». Nous mettons cette affirmation en parallèle à la position adoptée par Schafer (2010, cité p.3) : c'est l'absence de pertinence de l'environnement sonore qui amène l'individu à considérer les sons comme nuisants. Toutefois Jaworski introduit ici une nouvelle dimension subjective : celle d'une nuisance **subie**. Cet aspect nous ramène au caractère irrépressible de l'ouïe. « Subir » son environnement sonore se distingue donc, dans ce paradigme, de l'écoute de ce même environnement. En effet, l'écoute vise l'appréhension du milieu, d'un point de vue sémiologique. « *Ainsi, explorer le paysage sonore revient à rattacher les marques sonores aux activités qui en sont à l'origine, à dérouler ce paysage par ses éléments factuels.(...) Observer le paysage implique donc d'assimiler les formes sonores présentes dans l'environnement à une réalité phénoménale incluant la dimension culturelle et identitaire de ces manifestations à savoir des idéoscènes*» (Woloszyn, 2012 p. 53).

Nous considérons donc qu'un travail sur la dimension perceptive des sons de l'environnement doit avant tout relever d'une démarche qualitative et non pas uniquement basée sur les paramètres physiques des sons. Les sons ne sont pas

considérés « *en eux-mêmes et par eux-mêmes* » mais à travers leur perception par les individus. Selon Dubois et ses collaborateurs (2006), étudier la sémantique des phénomènes acoustiques renvoie à la mise au jour des principes psychologiques de leur catégorisation et conduit à des recherches sur l'élaboration des concepts en tant que représentations collectives, et connaissances. Il convient donc de nous interroger sur les représentations des phénomènes acoustiques. En effet, c'est sur ces représentations que se base une partie de notre méthodologie *d'élaboration du matériel expérimental* (voir p.119).

### **3.2. Qu'est-ce que la représentation d'un phénomène acoustique ?**

Selon Michel Denis et Danièle Dubois (1976), trois définitions de la *représentation* peuvent être proposées, elles sont à considérer en complémentarité. Les auteurs nous proposent en première acception un ensemble d'unités matérielles assurant une certaine « description » d'un ensemble de données de la réalité extérieure (à la cognition) (p.543). À ce niveau, on remarque principalement le caractère modélisateur du concept de représentation. La représentation y est envisagée comme une construction matérielle, schématique et imagée d'une réalité complexe, potentiellement extérieure au système cognitif de l'individu.

En deuxième acception, les auteurs considèrent la *représentation* comme une « *réalité psychologique (...) définie en première approximation comme l'ensemble des acquisitions d'un individu traduites au plan de ses structures mentales* » (*ibidem*). Les *représentations* sont dépendantes des expériences antérieures de l'individu, puisque leur élaboration s'opère à partir des expériences, notamment perceptives et de compréhension. Les représentations peuvent donc être fortement variables d'une personne à l'autre et d'une culture à l'autre<sup>27</sup>. Par ailleurs, dans cette partie de définition, la *représentation* ne serait pas observable mais correspondrait à un « *système latent dont la mise en jeu fonctionnelle serait soumise à des processus (concevables en terme de récupération, d'activation, etc.)* » (*ibidem*). Dans cette deuxième dimension, l'accent est donc mis sur le caractère fluctuant de la représentation. En effet, celle-ci dépend à la fois

---

<sup>27</sup> Cet aspect est particulièrement important dans le cadre de la didactique des langues puisque, la compréhension est dépendante des connaissances antérieures de l'individu et donc des représentations cognitives dont il dispose. Ainsi, en classe de Français Langue Étrangère en milieu homoglotte (en France) où les groupes d'apprenants sont le plus souvent composés d'individus originaires de différentes cultures, nous pouvons considérer que les connaissances et les représentations cognitives dont dispose chaque apprenant varient d'un individu à l'autre. Dès lors, les significations que chacun élaborera seront différentes, sans pour autant qu'on puisse conclure à des non-compréhensions.

de son contexte d'élaboration et de récupération ou d'activation. Dit autrement, nous pouvons considérer que 1) les représentations sont variables d'un individu à l'autre, en fonction de ses expériences antérieures et du contexte de leur élaboration et 2) leur activation dépend à la fois de l'individu et du contexte. Les représentations peuvent donc correspondre à une « reconstitution » des éléments perçus et compris, et sont de ce fait dépendantes des expériences antérieures de l'individu, ainsi que du contexte dans lequel elles sont mises en œuvre. Comme le rappelle Camus (1996:34) les représentations sont élaborées « *au cours des diverses expériences perceptives, intellectuelles et sensori-motrices du sujet. Ces représentations sont organisées en unités cognitives articulées les unes avec les autres* ». Cette remarque nous amène à souligner le caractère organisé des connaissances de l'individu. Dès lors nous pouvons parler d'un système représentationnel fonctionnel et durable.

Enfin, la troisième acception présentée par Denis et Dubois (1976) renverrait à « *une forme subjective et consciente d'activité mentale, souvent mais non nécessairement associée à des phénomènes quasi sensoriels tels que l'imagerie* » (p.544). Dans cette acception, la représentation renvoie à une **activité** mentale et **consciente** du sujet qui cherche à re-crée une « image » d'un objet du monde ou d'une situation. Nous émettons l'hypothèse que cette forme d'imagination, à travers son caractère conscient, peut permettre l'accès aux représentations (plus inconscientes) telles que définies en deuxième acception. C'est cette hypothèse qui a « guidé » notre *étude sur les attentes auditives* (voir p.121).

**Dans le cadre de notre travail, nous considérons donc les représentations comme des unités de connaissances, élaborées grâce à la perception et parce qu'il y a eu compréhension** (donc potentiellement activées lors de ces tâches à propos d'autres objets du monde).

Dubois, Guastavino, & Raimbault, (2006) ont observé huit caractéristiques des représentations cognitives des phénomènes acoustiques. De ces huit caractéristiques, nous retiendrons ici (1) qu'elles sont à la fois **individuelles** puisque dépendantes des expériences antérieures de l'individu, des activités de perception préalables, et des réponses associées, stockées en mémoire ; et **collectives** parce qu'elles sont (plus ou moins) partagées au sein d'une communauté culturelle. Par ailleurs, (2) nous nous attacherons également à leur caractère **global et multimodal**, c'est-à-dire que les représentations cognitives des phénomènes acoustiques peuvent également englober les aspects développés ou perçus *via* une autre forme de sensorialité, visuelle par exemple. Enfin, (3) les représentations sont, d'après les auteures, accessibles et partagées par des **productions linguistiques**. En d'autres termes, les représentations cognitives sont

appariées à des « formes d'expression verbale ». Aramaki *et al.* (2010) proposent des relations directes son/item linguistique qui permettraient à la fois leur observation et leur discussion au sein d'une communauté culturelle et linguistique (partageant les mêmes codes de connotation et dénotation).

En outre, il apparaît qu'à la suite de Pavio (1971, 1986), de nombreux chercheurs considèrent qu'« *il existerait au moins deux types de représentations mentales — ou deux types de codes symboliques pour les informations à stocker en MLT\_—, possédant chacun des règles de fonctionnement précises et des traitements particuliers. Le premier système de codage serait un système de représentations verbales ou propositionnelles — lié au langage — de nature plus abstraite, et le second système de codage figuratif ou imagé — cette fois plutôt lié à la perception de l'environnement — de nature plus concrète. Ces deux types de codage sont postulés être en relation, une information entrante pouvant, si cela est possible, être codée à la fois de manière verbale et de manière imagée (il paraît en effet difficile de se former spontanément une représentation imagée d'un concept abstrait comme « vérité » par rapport à un concept concret comme « baleine »). Ce double codage aurait enfin un effet bénéfique sur les apprentissages* » (Chanquoy *et al.*, 2007:69). Cette observation est particulièrement importante dans le cadre de notre travail expérimental. En effet, notre *étude sur les attentes auditives* (voir p.121) repose sur l'hypothèse que « l'identité sonore » des lieux est connue des individus qui en ont une expérience familière. Nous considérons également que ces connaissances sont organisées à un niveau représentationnel et donc accessibles *via* la parole. Par ailleurs, dans la mesure où elles sont également à la fois individuelles et collectives, nous supposons que les discours des participants présenteront des éléments communs sur lesquels nous pourrions baser la réalisation de scènes sonores.

### **3.3. Sons environnementaux, langage et représentation... un lien sémantique ?**

Les catégories sémantiques peuvent être considérées comme des médiations des expériences sensorielles individuelles, vers les représentations collectives (Dubois *et al.*, 2006 :867) En d'autres termes, pour les auteurs, l'étude des catégories sémantiques élaborées avec des phénomènes acoustiques permet l'étude des représentations collectives et de leur partage *via* le langage. Or cet aspect est primordial dans le cadre de notre étude. En effet, dans *l'élaboration de notre matériel expérimental* tout d'abord, nous sommes appuyée, pour la *construction des scènes sonores* (voir p.132), sur les déclarations d'individus francophones quant à leurs attentes auditives dans des situations de la vie quotidienne. Nous avons donc considéré leur expression linguistique comme l'outil qui nous a permis de prendre en compte leurs représentations cognitives

pour ces situations. Par ailleurs, dans ce recueil de données verbales, nous avons également considéré que ces représentations étaient culturelles et partagées socialement (d'où l'hypothèse d'observer des invariants dans ces discours)<sup>28</sup>.

Par ailleurs certains travaux ont montré une similarité de traitement sémantique entre le langage et les sons environnementaux via des études incluant des techniques d'électro-encéphalographie (EEG) (Aramaki *et al.*, 2010). Les auteurs font le parallèle entre la perception des items linguistiques (mots) et celle des sons environnementaux (de synthèse). Leurs travaux indiquent que des effets similaires apparaissent au niveau de la N400 avec des stimuli sonores environnementaux et avec des mots, au niveau de la perception de l'incongruence et des effets de primings (p.2559). La N400 correspond à l'apparition, sur l'EEG d'un pic négatif environ 400 ms après la perception d'une incongruité. Il semble donc qu'une incongruité sonore donne lieu à une réaction comparable à celle induite par une incongruité lexicale (rappelons que la N400 se rapporte généralement au lexique et à la sémantique, alors que la P600 relève de la syntaxe (Bornkessel-Schlesewsky & Schlewsky, 2008; Frenzel, Schlewsky, & Bornkessel-Schlesewsky, 2011)). Aramaki et ses collaborateurs (2010) émettent donc l'hypothèse d'une similarité dans les processus neuronaux impliqués dans le jugement sémantique, laissant entendre, une fois de plus, que différencier sons de la parole et « autres sons » n'est pas efficient dans la cadre de notre étude et qu'une différence entre sons signifiants et non-signifiants relèverait plus des objectifs de l'écoute et de ses conditions, que de la nature des sons. Dans cette optique, il n'apparaît pas illogique de parler de sémantique du son. Toutefois, une différence importante distingue les deux types de stimuli linguistiques et non-linguistiques : si les unités linguistiques (lexèmes) se caractérisent pas l'arbitraire du signe saussurien, les sons environnementaux (autres que linguistiques et signaux d'alerte) entretiennent généralement une relation causale avec leur source. De plus, alors que la langue est organisée en un système discret, les sons environnementaux (les signaux d'alerte mis à part) n'ont pas une telle organisation, ce qui implique des différences fondamentales dans leur compréhension. Cependant, nous considérons que pour qu'un son soit rapidement identifié et interprété, son pattern de vibration doit avoir été appris par l'individu, cet apprentissage étant fortement dépendant du milieu et pouvant être considéré comme participant à la culture. Dit autrement, pour qu'un phénomène acoustique soit compris par un individu, celui-ci doit au préalable avoir élaboré une représentation cognitive multidimensionnelle (*a minima*

---

<sup>28</sup> Pour plus de précisions, se reporter à *l'étude sur les attentes auditives* p.117.

sonore, et éventuellement linguistique) du concept mis en jeu, cette construction étant le fruit de ses expériences antérieures.

Il nous apparaît donc que les sons environnementaux sont associés à des représentations cognitives. Nos travaux portent sur l'aide qu'un contexte sonore environnemental congruent peut apporter à la compréhension de dialogues pédagogiques par des apprenants débutants en langue. En ce sens, notre hypothèse s'appuie sur l'idée que les sons environnementaux, dotés d'une valeur sémantique, c'est-à-dire qu'ils donnent lieu à une interprétation, peuvent jouer un rôle dans la compréhension d'une langue.

## **Ce qu'il faut retenir...**

Nous avons donc vu dans ce premier chapitre que la perception correspond à l'appréhension d'un environnement ou d'un événement par un individu. Dans ce cadre, plusieurs conceptions de la perception et plusieurs modèles cohabitent. Nous avons fait le choix de nous positionner dans la perspective d'une perception descendante, c'est-à-dire trouvant son point de départ chez l'individu, *via* la formulation d'attentes et d'hypothèses à propos de ce qui l'entoure. Dans le cadre de la perception auditive, de nombreuses recherches ont été menées sur l'ASA et ont proposé différentes règles quant à son fonctionnement. Il en ressort que la perception aboutit à l'identification des sources qui composent notre environnement et est fortement dirigée par les objectifs d'écoute, la mémoire et l'attention. En ce sens, écouter s'apprend et est fortement lié à la compréhension. Dans le cadre des liens qui unissent perception et compréhension, nous avons vu que ce que nous avons appelé les sons environnementaux (c'est-à-dire des éléments sonores sur lesquels ne porte pas spécifiquement l'attention de l'individu percevant) présentent néanmoins une dimension sémantique. En ce sens, ils peuvent donner lieu à une identification et participer à l'élaboration d'un sens, d'une signification *i.e.* être compris.

Dans ce cadre, nous considérons que, confrontés à des dialogues pédagogiques accompagnés d'un Contexte Sonore Environnemental, les individus seront susceptibles d'élaborer du sens non seulement à partir des éléments linguistiques contenus dans les supports mais également à partir des sons environnementaux qui les accompagnent.



## Deuxième Chapitre : ... aux études sur la compréhension

Dans cette première partie théorique, nous avons donc séparé les études sur la perception auditive et les études sur la compréhension linguistique. Or, comme nous l'avons vu, selon nous, ces différents éléments participent à un même phénomène cognitif. En outre, il est apparu que les phénomènes attentionnels sont à l'origine de la profondeur des traitements appliqués aux éléments perçus (voir le sous-chapitre sur le rôle de l'attention dans le chapitre *Des études sur la perception...* p.49) et compris. Dès lors la nature (linguistique ou non) de l'objet n'est plus un facteur déterminant. Il apparaît donc que la distinction théorique que nous avons introduite n'est que partiellement représentative de notre positionnement. Nous introduisons, en effet, une nuance, celle de la prise en compte de l'émission des sons. Deux dimensions fondamentales différencient les sons de la parole de la majorité des autres sons de l'environnement : l'*arbitraire* du signe saussurien et le caractère essentiellement communicatif des énoncés linguistiques.

Alors que ce que nous appelons « *sons environnementaux* » ne relèvent pas toujours d'une intention communicative (par exemple les bruits de circulation automobile), la parole témoigne **nécessairement** de cette intention, particulièrement importante dans le cadre des études sur la compréhension du langage (qu'il soit oral ou écrit). Néanmoins, il est vrai que cette intention peut ne pas s'adresser à l'individu percevant que nous désignons par le terme « *compreneur* » à la suite de Le Ny (2005). C'est notamment le cas dans la compréhension de dialogues pédagogiques où plusieurs types d'intention cohabitent : l'intention des locuteurs entre eux, dans la situation d'énonciation (parfois artificielle comme nous l'avons évoqué dans le chapitre *Contexte didactique*, p.17), et l'intention des auteurs (concepteurs des dialogues), à destination des apprenants. Au-delà de ces « *vouloir dire* » prévus, se construit aussi l'intention de l'enseignant qui utilise le dialogue dans la classe. Ce dernier ajoute son intention en

choisissant tel dialogue plutôt qu'un autre. Dans sa préparation de la leçon, il sait que ce dialogue va lui permettre d'aborder tels vocabulaires, telle question grammaticale, etc. Dans le cadre de notre étude, c'est cette intention qui prime : l'enseignant décide a priori ce que les apprenants doivent comprendre et évaluera en retour leur compréhension en relation avec ses intentions.

Ces deux paramètres (arbitraire du signe et intentionnalité du message) sont, selon nous à l'origine des traitements différenciés de la parole *versus* « l'ensemble des autres sons ». En outre, dans le cadre des expérimentations que nous avons mises en place, et plus particulièrement du Classe-Images (voir p.165), la tâche correspond à la compréhension d'un dialogue linguistique accompagné (ou non) d'un contexte sonore environnemental. Dans ce cadre, le dialogue a été positionné au premier plan du montage (avec une intensité rehaussée par rapport au contexte sonore environnemental), donnant ainsi lieu *a priori* à une focalisation attentionnelle plus importante de la part des participants-auditeurs. En effet, les participants ayant comme objectif déclaré (et imposé par la consigne) de *comprendre* les dialogues, c'est sur cet aspect qu'ils auront *a priori* concentré leur attention. Toutefois conscients que les stimuli complexes (parole + contexte sonore environnemental) que nous leur avons proposés relèvent d'un travail d'élaboration volontaire, ils ont également pu porter leur attention sur le contexte sonore environnemental. Il s'agit là d'une des hypothèses fondatrices de notre travail.

## **1. Éléments définitoires : Qu'est-ce que comprendre ?**

Chanquoy, Tricot, & Sweller (2007:102) définissent la compréhension du langage comme « *le processus d'élaboration d'une représentation mentale du contenu d'un discours* ». La compréhension apparaît alors comme un processus dynamique, dans lequel le compreneur joue un rôle actif. En ce sens la compréhension se distingue de la « simple » récupération d'un élément en mémoire ou de sa mémorisation (comme dans les tâches de rappel par exemple). En effet, alors que la récupération en mémoire implique « uniquement » l'activation d'une représentation et/ou d'un concept d'ores et déjà élaboré(s) et stocké(s), la compréhension concerne l'élaboration, en tant que processus de construction, d'une nouvelle représentation, ou d'un nouveau concept, à partir à la fois du stimulus perçu et interprété, et des connaissances antérieures de l'individu compreneur. Pour Blanc & Brouillet (2003:51) « *le sens (ce qui est compris*

*dans la situation) n'est pas le produit de l'activation de schémas stockés en mémoire (hypothèse cognitiviste) mais l'émergence d'une organisation temporaire contextualisée. Cette façon de considérer le sens nous renvoie à la notion de schéma de Bartlett (1932) : une organisation construite au fur et à mesure. (...) ».* Nous pouvons donc considérer qu'en activité de compréhension le compreneur élabore une (nouvelle) représentation de « l'objet à comprendre » à partir 1) des représentations cognitives activées (potentiellement en phase de perception) et 2) de (potentiellement nouveaux) liens ou un (nouvel) agencement de ces représentations au sein d'une nouvelle entité.

Parallèlement, les mêmes auteurs (2003:47) définissent la compréhension comme « *un processus émergent (...) reflétant moins les caractéristiques particulières de l'objet à mémoriser que les aspects saillants de son contenu en interaction avec les connaissances et/ou les croyances de l'individu, ainsi qu'avec le contexte (...) »*. Deux aspects fondamentaux apparaissent ici : d'une part le caractère partiel de la compréhension est mis en avant (ce n'est pas la totalité de l'objet ou de la situation qui est traitée mais uniquement les saillances, *i.e.*, les éléments sur lesquels s'est focalisée l'attention), et d'autre part les rapports qu'entretiennent perception et compréhension. En effet, à partir de la situation de perception, l'individu encode différents éléments, (re)construit une représentation de la situation et des objets qui la composent, active des représentations (préalablement élaborées) associées à celle mise en jeu à ce moment précis et devient le dépositaire d'une nouvelle connaissance. Nous pourrions dire que la perception est, sous un certain nombre de conditions, ce qui permet la compréhension, ce terme désignant un aboutissement, même s'il n'est pas stable (c'est-à-dire pas nécessairement disponible dans la mémoire à long terme). Toutefois, si nous considérons la compréhension comme l'élaboration d'une (nouvelle) représentation complexe ou d'un schéma, construit(e) à partir des représentations activées préalablement et agencées dans une configuration nouvelle et particulière, alors nous ne pouvons pas considérer les processus inhérents à cette activité en dehors de la mémoire (c'est-à-dire les représentations déjà construites). Ainsi, la compréhension implique non seulement le recours aux connaissances antérieures mais aussi une inscription, au moins temporaire, des éléments compris en mémoire. Tulving (1972) a montré que deux types de mémoire composent la Mémoire à Long Terme. Le premier correspond à la mémoire épisodique, « *subdivisée en mémoire biographique faisant référence aux souvenirs anciens de nature autobiographique, et en mémoire prospective, de nature intentionnelle, permettant la mémorisation d'actions à entreprendre dans le futur »* (Deschamps & Moulignier, 2005:507), qui stocke les événements spécifiques, fortement associés au contexte. Parallèlement, Tulving propose que la mémoire sémantique est « *peu sensible à l'affectivité »* (Deschamps & Moulignier, 2005:507). Elle est considérée comme « *un thésaurus mental et concerne les connaissances organisées que possède une personne au*

*sujet des mots ou d'autres symboles verbaux, leur sens et leur référent, les relations entre eux, ainsi que les connaissances propres aux règles, formules, et algorithmes relatifs à la manipulation de ces symboles, concepts et relations (Tulving 1972, p.386) » (Blanc & Brouillet, 2003:18). Ainsi, l'inscription du compris se fera en mémoire sémantique, lorsque l'objet à comprendre correspond à une nouvelle connaissance sur le monde, ou épisodique, lorsque c'est l'événement en lui-même qui est compris et retenu. Dans cette perspective, la compréhension reste donc intrinsèquement liée à la mémoire où sont stockées les éléments de connaissance déjà élaborés et stockés par l'individu.*

D'après Kekenbosch & Denhière (1988), dans les modèles connexionnistes en réseau, la récupération d'une connaissance stockée en mémoire est à l'origine de l'émergence de la signification. Cette récupération passerait par l'activation d'un nœud et pourrait résulter de la perception d'un objet du monde, ou de la diffusion de l'activation préalable d'un autre nœud, relié sémantiquement (ou d'une autre manière) au premier (Blanc & Brouillet, 2003). Il apparaît donc que deux voies coexistent pour cette activation. La première voie correspond à la perception « directe » d'un objet du monde dont l'identification aboutit à l'activation de la représentation ou d'un concept. La deuxième voie correspond à la propagation de l'activation d'un nœud. L'activation passerait, dans ce cadre, d'un nœud à l'autre, jusqu'à l'activation d'un réseau. John R. Anderson (1983) a proposé la notion de « *fan effect* » pour rendre compte de la dispersion de l'activation entre différents nœuds, lorsqu'elle se diffuse en réseau. Selon l'auteur « *Il y a trois voies par lesquelles un élément peut devenir source d'activation en mémoire de travail. Tout d'abord, un élément qui encode un stimulus de l'environnement deviendra une source. Par exemple, si un mot est présenté, sa représentation en mémoire sera activée. Deuxièmement, quand une production se réalise, les structures de l'élaboration des actions deviennent sources d'activation. Troisièmement, une production peut focaliser sur l'objectif dans la structure en mémoire de travail, et les éléments d'une telle structure focalisée peuvent devenir source d'activation* »<sup>29</sup> (Anderson, 1983:89). Selon l'auteur, la récupération d'une représentation en mémoire dépend donc largement de la propagation de l'activation, de sa dispersion entre plusieurs nœuds. Selon Anderson & Pirolli (1984) trois lois régissent la propagation de l'activation : 1) la quantité d'activation qui se propage décroît à mesure que la distance entre les nœuds

---

<sup>29</sup> « *There are three ways in which an element can become a source of activation in working memory. First, an element that encodes an environmental stimulus will become a source. For example, if a word is presented, its memory representation will be activated. Second, when a production executes, its actions build structures that become sources of activation. Third, a production can focus the goal element on a structure in working memory, and the elements of such a focused structure can become sources of activation* » (Anderson, 1983:89).

augmente, 2) la quantité d'activation qui « part » d'un nœud source est répartie sur l'ensemble des nœuds qui lui sont connectés, et 3) quand un seul nœud reçoit l'activation de différents nœuds sources, les quantités d'activation s'additionnent. Ainsi, il apparaît qu'à partir de la perception d'un item (qu'il soit linguistique –oral ou écrit-, ou non linguistique), différentes représentations cognitives s'activeraient dans l'esprit de l'individu percevant. Cette première activation en entraînerait d'autres et ce processus se poursuivrait jusqu'à épuisement de la charge d'activation. Parallèlement à ce processus d'activation, un second processus, d'inhibition cette fois, interviendrait. L'inhibition permet de ne pas maintenir (ou d'empêcher) l'activation de nœuds inappropriés. Ces principes d'activation et d'inhibition renvoient en outre au modèle de construction-intégration que nous développerons plus loin dans ce chapitre.

Dans ce cadre, la compréhension correspondrait au traitement, en mémoire de travail, de nombreux éléments d'informations relevant à la fois du perçu (présent dans l'environnement et/ou dans le message), mais aussi aux connaissances antérieures de l'individu. Dans cette perspective, la compréhension suppose une forte interactivité entre ces différents éléments. Selon Chanquoy *et al.* (2007) cette interactivité implique le traitement simultané de ces éléments (perçus ou en mémoire). Dit autrement, la compréhension se distingue de la récupération en mémoire par la dimension dynamique de la construction qu'elle implique. Or, cette construction nécessite la manipulation et le traitement profond de différents items (représentations, concepts, ou nœuds), et ces procédures induisent à leur tour un coût cognitif notable. Il apparaît donc que la compréhension est une procédure complexe pour le système cognitif, faisant intervenir de nombreux modules et processus imbriqués. En ce sens, nous pouvons considérer que la compréhension est une activité cognitive de haut-niveau.

Aussi, il apparaît que la compréhension, dans son acception large, représente un domaine de recherches particulièrement vaste, généralement frayé par la psychologie. L'étendue de ce domaine implique donc différents degrés de spécialisation. Dans ce cadre, une première distinction peut apparaître entre la compréhension de la langue et celles des autres formes d'informations en provenance de l'environnement. Toutefois, cette distinction n'est, encore une fois, pas opérationnelle dans l'étude que nous menons où la focalisation attentionnelle joue un rôle plus important que la nature linguistique ou non des sons.

## 2. Des études sur la compréhension orale : de la perception du signal à la (re-) construction de sens

Dans l'étude de la compréhension orale, il apparaît que deux écoles théoriques, qui se distinguent quant à leur conception de la parole, cohabitent. La première, inspirée des travaux de Miller (Miller & Chomsky, 1963; Miller, 1962, dans le paradigme du traitement de l'information) et de Chomsky (notamment pour ses travaux sur la grammaire générative et transformationnelle), considère le **langage en tant que produit**. Cette école met l'accent sur le caractère « individuel » de la compréhension (en dehors de l'interaction), l'objet de la compréhension est alors considéré comme un « objet fini ». Dans cette perspective, les recherches en psycholinguistique ont principalement porté sur les processus *on-line* « *en ayant recours à des mesures précises des temps de réaction et à des stimuli contrôlés précautionneusement* »<sup>30</sup> (Tanenhaus & Brown-Schmidt, 2008:1105). Le contexte est alors considéré soit comme le cadre dans lequel se « réalise » une représentation par ailleurs conçue indépendamment de ce même contexte, soit comme le cadre limitant par lequel s'opère le traitement des unités ambiguës. Il semble donc que dans cette perspective, la « production langagière » soit envisagée comme indépendante de la situation d'énonciation dans laquelle elle émerge. Dans ce cadre, le contexte ne participe pas à l'interprétation du message autrement que de manière secondaire, et l'intention des locuteurs et auteurs n'est pas directement convoquée dans la compréhension. La deuxième école théorique considère le **langage en tant qu'action** et trouve son origine dans les travaux de Austin (1970) et Grice (1957) ; ils ont donné lieu à la pragmatique et à l'analyse conversationnelle. Cette tradition se concentre sur « *la manière dont les personnes ont recours au langage pour réaliser des actions dans la conversation (...). Les recherches psycholinguistiques au sein de la tradition 'action' se concentrent le plus souvent sur les interactions conversationnelles qui engagent au moins deux participants dans une tâche collaborative, généralement avec des références au monde réel et des objectifs comportementaux* »<sup>31</sup> (Tanenhaus & Brown-Schmidt, 2008:1106). L'intentionnalité du discours est donc au centre de cette perspective. Deux types d'intentions semblent diriger la communication : une intention locutoire et une intention perlocutoire (Sperber, Clément, Heintz, Mascaro, Mercier,

---

<sup>30</sup> « *using fine-grained reaction time measures and carefully controlled stimuli* » (Tanenhaus & Brown-Schmidt, 2008:1105).

<sup>31</sup> « *how people use language to perform acts in conversation—the most basic form of language use. Psycholinguistic research within the action tradition typical focuses on interactive conversation involving two or more participants engaged in cooperative task, typically with real-world referents and well-defined behavioral goals* » (Tanenhaus & Brown-Schmidt, 2008:1106).

Origiani, & Wilson, 2010). Dans cette perspective, la majorité des énoncés ne peut être comprise en dehors du contexte d'énonciation (défini notamment par les intentions communicatives et le cadre spatio-temporel de l'interaction) (Thomson & Tulving, 1970; E. Tulving, 1976). En outre, « *Bien que ces notions puissent être conceptualisées comme une forme de contrainte, elles sont intrinsèques au processus de compréhension et non une source d'information qui aiderait à résoudre l'ambiguïté de l'input* »<sup>32</sup> (Tanenhaus & Brown-Schmidt, 2008:1107)). Le contexte devient alors un élément primordial de la compréhension : « comprendre un objet en contexte » ne correspond dès lors plus à comprendre premièrement l'objet et deuxièmement son contexte, pour en élaborer une « synthèse » dans un troisième temps, mais bien une tâche de compréhension unique où objet et contexte constituent un même ensemble. C'est cette perspective que nous adoptons en posant notre hypothèse générale : un contexte sonore environnemental peut favoriser la compréhension d'un dialogue pédagogique. En effet, présenter des éléments contextuels (sonores dans le cas de notre travail), c'est présenter des éléments indispensables à la compréhension, qui ne correspondent pas à la somme de la compréhension des parties qui la composent.

Il apparaît donc que ces deux écoles étudient la compréhension du langage dans deux perspectives complémentaires. Considérer le langage en tant que produit implique que le « compris » ne dépend pas directement de la situation de communication. Or, dans le cadre de la compréhension orale en classe de langue (voir le *Contexte Didactique* p.17), l'apprenant y est même extérieur. En ce sens, dans cette situation de compréhension bien particulière, la langue est bien un produit, si les apprenants peuvent entrer dans la dimension illocutoire des discours qui leur sont présentés, la dimension perlocutoire des discours est rarement abordée avec les niveaux débutants. Toutefois, dans la majorité des situations de compréhension de la parole en contexte naturel (extérieur à la salle de classe), les locuteurs sont généralement des interlocuteurs, tout à la fois producteurs et compreneurs du langage, impliqués dans l'interaction, et il s'agit évidemment pour l'enseignant de former les apprenants à ce type de situations. Nous pouvons donc constater que dans le cadre des recherches en didactique de la compréhension orale, la dichotomie *langage-action versus langage-produit* n'est pas efficiente. Nous considérons le langage à la fois comme *action* et comme

---

<sup>32</sup> « *Although these notions can be conceptualized as a form of constraint, they are intrinsic to the comprehension process rather than a source of information that helps resolve ambiguity in the input* » (Tanenhaus & Brown-Schmidt, 2008:1107).

*produit*. La parole est en effet une action, non seulement parce qu'elle implique l'action de dire (de communiquer) mais aussi parce que, comme le rappelle le CECRL (qui met l'accent sur la perspective *actionnelle* de l'enseignement des langues) « *l'usager et l'apprenant d'une langue (sont des) acteurs sociaux, ayant à accomplir des tâches (qui ne sont pas seulement langagières) dans des circonstances et un environnement donnés, à l'intérieur d'un domaine d'action particulier. Si les actes de parole se réalisent dans des activités langagières, celles-ci s'inscrivent elles-mêmes à l'intérieur d'actions en contexte social qui seules leur donnent leur pleine signification* » (Conseil de l'Europe, 2001:15). Dans ce cadre, la langue est action, le locuteur est acteur et l'apprenant observe et apprend de ce qu'il entend (et voit) afin de pouvoir, à son tour, produire des actions (non seulement langagières). En ce sens, les dialogues pédagogiques sont à envisager comme des « exemplaires » de la « langue en action ». Mais dans ces exemplaires, la parole est aussi *produit* puisque la situation est reproduite, de la même manière que le discours. Dans ce cadre, les apprenants compreneurs ne sont pas les acteurs de la situation, mais en sont les témoins : ils se positionnent face à une parole *produite* en dehors de leur présence. En tant qu'enseignants, notre rôle est donc d'amener les apprenants à développer des compétences dans la compréhension de la langue à la fois en tant *qu'action* (notamment par le biais d'exercices de jeux de rôles ou de la *pédagogie action*), dans ce cas l'apprenant est impliqué dans une interaction et en tant que *produit* (par le biais de la compréhension des dialogues pédagogiques ou de documents authentiques par exemple), dans ce cas l'apprenant est en situation de compréhension. Ainsi, dans le cadre de nos expérimentations, les participants au *Classe-Images* (voir p.182) assistent aux interactions entre les locuteurs. Cette spécificité semble en outre inhérente aux supports pédagogiques oraux en classe de langue : élaborés et construits à des fins d'apprentissage, ils mettent en scène la langue telle qu'elle est parlée par des locuteurs natifs et ne demandent pas aux apprenants d'entrer dans une réelle interaction avec le discours produit.

Que l'individu prenne part à l'échange ou soit en position seule de comprendre, son activité de compréhension témoigne de l'élaboration d'un schéma, d'une représentation complexe et précise du discours en train de se dérouler. Cette élaboration s'appuie sur les éléments identifiés (unités signifiantes) et dépend donc, dans une certaine mesure, de l'attention que l'individu porte au signal. Dans la perception/interprétation des scènes sonores environnementales (comme celles que nous avons réalisées (voir p.121) l'identification des unités signifiantes se fait à partir d'une focalisation attentionnelle et des objectifs d'écoute de l'individu (voir le chapitre sur *Des études de la perception...* p.49). Dans la perception/interprétation de la parole,

l'identification des unités signifiantes trouve également son point de départ dans un processus de focalisation attentionnelle qui permet à l'individu d'organiser le « brouhaha » qui lui parvient en différents plans, en fonction de ses objectifs d'écoute. Si le signal/flux de parole est « au premier plan » (c'est-à-dire si l'attention se focalise sur lui), alors l'individu procédera à une analyse du signal afin d'en identifier les unités signifiantes. Nous allons donc maintenant nous pencher sur certains aspects de cette analyse.

## 2.1. Reconnaissances des unités de la chaîne parlée

Ainsi, la compréhension de la parole passe par le décodage d'un signal sonore, composé d'unités discrètes (définies arbitrairement et par convention) qui rendent compte d'unités de sens mises en œuvre par l'émetteur du message, et qui doivent être associées à des unités signifiantes pour et par le compreneur (Picoche, 1999). « *Le processus de reconnaissances des mots parlés nécessite que l'auditeur établisse un rapport entre une analyse acoustico-phonétique du signal de parole avec une forme des représentations lexicales stockées dans le lexique mental* »<sup>33</sup> (Marslen-Wilson & Warren, 1994:653). En outre, ce décodage passe par l'interprétation des unités discrètes au niveau paradigmatique et qui entretiennent entre elles des rapports syntagmatiques. La lexie, le mot, est l'unité de sens sur laquelle repose l'élaboration d'une signification « *comprendre la parole nécessite la conversion du signal sonore en signification(s). Les mots sont les unités de sens exprimées dans le signal sonore. Dès lors, le processus de reconnaissance des mots est au cœur de la compréhension de la parole* »<sup>34</sup> (Norris, Cutler, McQueen, & Butterfield (2006:147). Nous préférons le terme d'item lexical ou de lexie à celui de mot : une lexie est définie comme « *toute unité lexicale mémorisée en langue* » (Pottier, 2003:171 ; Pottier, 1999) qui comprend au moins un lexème et qui est une unité tout à la fois syntagmatique et paradigmatique. Par exemple *pomme de terre* est une lexie bien que composée de trois mots.

La question de la perception de ces unités lexicales est donc posée et il apparaît que de nombreux modèles s'appuient sur des représentations infra-lexicales. Ces représentations correspondent à des unités plus élémentaires (le plus souvent définies

---

<sup>33</sup> « *The process of spoken word recognition requires the listener to relate an acoustic-phonetic analysis of the speech signal to stored representations of lexical form in the mental recognition lexicon* » (Marslen-Wilson & Warren, 1994:653)

<sup>34</sup> « *Understanding spoken language requires the conversion of sound to meaning. Words are the units of meaning expressed in sound. The process of word recognition is thus the heart of language understanding.* » (Norris et al., 2006, p. 147).

par la phonologie structurale) telles que la syllabe<sup>35</sup>, le phonème<sup>36</sup>, ou encore le trait phonétique<sup>37</sup> (Christophe, Pallier, Bertoncini, & Mehler, 1991; Frauenfelder & Nguyen, 2000). Parmi ces modèles, certains argumentent en faveur de la syllabe comme unité de base (Savin & Bever, 1970 cité par Christophe, Pallier, Bertoncini, & Mehler, 1991; Frauenfelder & Nguyen, 2000). Selon ces modèles, à partir de la syllabe d'attaque de la lexie, l'individu compreneur entamerait une comparaison avec son lexique mental. Cette comparaison se déroulerait tout au long de la perception de l'item lexical, jusqu'au maintien de l'activation sur une unité du lexique mental spécifique et l'inhibition de ses « concurrents ». Dans ce cadre, la syllabe d'attaque jouerait un rôle d'amorce et le contexte phrastique permettrait une reconnaissance rapide (Sivonen *et al.*, 2006). Toutefois, certains résultats expérimentaux (rapportés par Christophe, Pallier, Bertoncini, & Mehler (1991)) montrent que l'unité pré-lexicale utilisée par les locuteurs varie en fonction de la langue ; pour le français, l'unité de base semble être la syllabe. Ces variations nous apparaissent problématiques dans la perspective de l'enseignement des langues étrangères. Si l'unité de base de la segmentation de la langue parlée varie d'une langue à l'autre, alors l'apprenant débutant, locuteur natif d'une langue dont l'unité de base ne serait pas la syllabe, se trouverait en difficulté dans sa perception d'une langue qui, comme le français, aurait une unité de segmentation fondée sur la syllabe. Parallèlement, certains auteurs (notamment Norris, McQueen, & Cutler, 1995) mettent en avant la connaissance des patterns rythmiques de la langue comme support à la reconnaissance de la parole. Encore une fois, cet aspect met le doigt sur les difficultés particulières que rencontrent les apprenants de langue, et plus particulièrement les apprenants débutants puisque les patterns rythmiques varient d'une langue à l'autre. « *Entrer dans une langue, c'est entrer dans un rythme nouveau* ».

Toutefois, il apparaît que, pour nous, ces modèles relèvent plus de la reconnaissance de la parole que de la compréhension telle que nous l'avons décrite plus haut puisqu'ils ne font pas état de la construction du sens par l'individu destinataire du message. En effet, les différentes pistes explicatives que nous venons de présenter se proposent de faire la lumière sur la reconnaissance des lexies, c'est-à-dire la manière

---

<sup>35</sup> « Elle est constituée d'au moins une voyelle qui peut être précédée ou suivie d'une ou plusieurs consonnes ; les structures syllabiques possibles diffèrent selon les langues » (Christophe *et al.*, 1991:62)

<sup>36</sup> « C'est l'unité de construction de la syllabe, et la plus petite unité qui permette une décomposition séquentielle de la chaîne de parole » (Christophe *et al.*, 1991:62-63)

<sup>37</sup> « Chaque phonème est défini par les valeurs qu'il prend pour une dizaine de traits binaires (...) Les traits distinctifs ne représentent pas seulement la manière dont un phonème est articulé, mais ont une valeur perceptive (...) » (Christophe *et al.*, 1991:63)

dont l'individu parvient à encoder le message reçu en unités de sens. « *Les modèles de la reconnaissance des mots parlés supposent généralement que le traitement d'un mot a lieu pendant que le signal acoustique de parole se déroule. Ce traitement commence avec une analyse bottom-up du signal de parole que constitue le début du mot. Cette analyse de départ active un groupe de mots candidats. L'analyse se poursuit jusqu'à ce qu'un seul candidat corresponde au signal (point de reconnaissance) et soit sélectionné (...)* Les procédures top-down et bottom-up avanceraient de manière hautement interactive (...) ou, au moins partiellement, de manière parallèle »<sup>38</sup> (Sivonen, Maess, & Friederici, 2006:220). Dit autrement, cette étape d'identification des unités linguistiques au sein de la chaîne parlée correspond à ce que nous avons décrit comme des étapes perceptives : l'identification des items lexicaux, voire des rapports syntagmatiques qu'ils entretiennent les uns avec les autres, relève selon nous de l'aboutissement de la perception telle que nous l'avons décrite précédemment (association des éléments perçus avec les représentations cognitives d'ores et déjà mémorisés).

Dans ce cadre, nous pouvons proposer un parallèle entre cette identification des unités linguistiques et celle des objets sonores que nous avons décrite dans le chapitre *des études de la perception...* (p. 49). Ce paramètre nous semble en outre particulièrement important car il vient renforcer notre hypothèse de la perception de la parole dépendante des objectifs de l'écoute. Lorsque le flux de parole est considéré comme ne participant pas aux éléments signifiants de l'environnement, aucun décodage/encodage ne se met en place chez l'auditeur. Dit autrement, la perception correspondrait, selon nous à l'identification et à la reconnaissance des stimuli ou des éléments qui le composent. Interviendrait ensuite une étape de compréhension, c'est-à-dire d'élaboration d'une nouvelle représentation, sans toutefois que ces deux étapes puissent intervenir indépendamment l'une de l'autre. En ce sens, nous considérons qu'il est possible de parler de compréhension aussi bien pour les sons de la parole que pour les sons de l'environnement.

## **2.2. Des unités perçues à l'émergence d'une signification**

Dans ce cadre et comme nous avons pu le voir sur les modèles de perception des sons environnementaux (voir p.49), différents modèles ascendants et/ou descendants

---

<sup>38</sup> « *Models of spoken word recognition usually assume that the processing of a word proceeds as the acoustic speech signal develops in time. Processing begins with a bottom-up analysis of the speech signal constituting the word's beginning. This initial analysis activates a group of candidate words. The analysis continues until only one candidate matches the signal (recognition point) and becomes selected (...)* Top-down and bottom-up processes may proceed in a highly interactive manner (...) or, at least partly, in a parallel manner (...) » (Sivonen, Maess, & Friederici, 2006:220)

sont également proposés pour la compréhension de la parole. « *Les modèles psycholinguistiques de traitement de la parole postulent généralement l'existence de traitements de « bas niveau » orientés vers la perception – c'est-à-dire vers le décodage des informations acoustico-phonétiques, et des traitements de « plus haut niveau » impliquant à la fois les informations décodées et les connaissances de l'auditeur, et conduisant à l'interprétation du message* » (Fontan, 2012:65). Marslen-Wilson (1989) propose deux mécanismes de compréhension de la parole. Le premier s'appuierait sur la forme, passerait par une étape de décodage de la forme acoustico-phonémique des sons de la parole, et permettrait l'association des unités de la parole au lexique mental de l'individu percevant (voir plus haut dans ce chapitre). Cette étape correspond donc à une étape d'identification. Parallèlement, le deuxième mécanisme proposé par l'auteur, en impliquant le contexte (et le co-texte), permettrait la mise en relation de ces unités identifiées aux connaissances antérieures, faisant ainsi émerger une signification pour l'individu comprendreur.

Il apparaît que les procédures *top-down* et *bottom-up* interagissent lorsque l'individu cherche à élaborer du sens à partir des éléments qu'il perçoit de son environnement (Grataloup *et al.*, 2009). Sivonen, Maess, & Friederici (2006) ont en outre montré, à partir d'études incluant des techniques d'électro-encéphalographie (plus particulièrement l'observation de la N400) et portant sur la reconnaissance de mots parlés dont le signal a été détérioré, que « *les informations bottom-up et top-down sont l'une et l'autre utilisées efficacement pour la perception on-line rapide de la parole* » (p.224). Toutefois, et comme nous l'avons présenté dans le chapitre *Des études sur la perception...* (p.49), nous nous positionnons en faveur de l'hypothèse de procédures essentiellement *top-down*, c'est-à-dire qui trouvent leur point d'entrée dans la cognition de l'individu, mais incluant une étape de « confrontation » à l'*input*.

Nous avons donc vu dans quelle mesure, malgré les spécificités relatives au caractère arbitraire et à la dimension intrinsèquement communicative du linguistique, le « traitement » du signal de parole reste parallèle (voire en partie similaire) à celui des « éléments sonores environnementaux ». En effet, la perception du signal de parole aboutit à l'identification d'items lexicaux et à l'activation de représentations cognitives ou de concepts, à partir desquels sera élaborée une signification. Dans cette perspective, les unités de sens correspondent aux lexies, dans leur dimension paradigmatique. Or, la langue, et les énoncés produits, se caractérisent également par les liens syntagmatiques que ces unités entretiennent entre elles. Dans le cadre de travaux sur la construction du sens à partir d'un signal de parole, Culicover & Jackendoff (2006) font état de trois théories pour rendre compte de l'appréhension de cette dimension. La première

correspond à la « composition frégienne » : le sens global serait le résultat de la significativité des parties d'une part et des agencements syntaxiques d'autre part. Toutefois, il apparaît que cette hypothèse cesse d'être fonctionnelle face à certains énoncés et qu'une partie du décodage (et de l'encodage) passe par *l'intuition linguistique* qui laisse entendre que la signification d'un énoncé n'est pas réductible à la somme de ses parties (quel que soit le niveau de ces dernières). « *Cette intuition, de même que celles qui nous permettent de juger de la grammaticalité d'une phrase, ne sont pas explicitement apprises* » (Christophe *et al.*, 1991:60). Cette intuition relève des aspects particulièrement difficiles à acquérir pour les apprenants débutants en langue étrangère. Leur caractère novice est généralement corrélé à leur faible expérience de la langue cible. Ils ne disposent donc pas *a priori* d'intuition linguistique dans cette langue. En outre, le caractère implicite de cet apprentissage le rend particulièrement difficile à acquérir par des apprenants « imprégnés » d'habitudes liées à la scolarisation (cadre dans lequel, du moins en Europe, l'apprentissage reste bien souvent explicite).

Parallèlement à l'hypothèse de la composition frégienne, les auteurs (Culicover & Jackendoff, 2006) avancent deux autres hypothèses : « la sémantique autonome » et « l'hypothèse de la syntaxe simple ». D'après l'hypothèse de la sémantique autonome, « *les significations de la phrase et du syntagme sont composées de la signification des mots ainsi que de principes indépendants de la construction de sens, dont seulement certains sont corrélés à la structure sémantique* »<sup>39</sup> (p.413). Dans le cadre de cette hypothèse, lexicale et syntaxe ne seraient plus aussi nettement séparés que dans la composition frégienne. Parallèlement, pour les auteurs et selon l'hypothèse de la syntaxe simple « *la structure syntaxique n'est complexe que dans la mesure où elle a besoin de l'être pour qu'une interprétation soit établie* »<sup>40</sup> (Culicover & Jackendoff, 2006:413). Dans ce cadre, la compartimentation lexicale/sémantique/syntaxe n'apparaît plus aussi clairement. En outre, de nombreux travaux plaident en la défaveur de la séparation des traitements syntaxiques et sémantiques. Par exemple, Bornkessel-Schlesewsky & Schlewsky (2008) rapportent qu' « (...) *un nombre croissant d'expérimentations a commencé à révéler d'une part des effets N400 non-sémantiques et d'autre part des effets P600 non syntaxiques (...)* »<sup>41</sup> (p.56). Ainsi, il semble que les distinctions théoriques apportées

---

<sup>39</sup> « *Phrase and sentence meanings are composed from the meanings of the words plus independent principles for constructing meanings, only some of which correlates with syntactic structure* » (Culicover & Jackendoff, 2006:413).

<sup>40</sup> « *syntactic structure is only as complex as it needs to be to establish interpretation* » (Culicover & Jackendoff, 2006:413).

<sup>41</sup> « (...) *an increasing number of experiments had begun to reveal non-semantic N400 effects on one hand and non-syntactic P600 effects on the other hand (...)* » (Bornkessel-Schlesewsky & Schlewsky, 2008:56).

entre les différents modules de la langue ne soient pas très fonctionnelles dans l'étude de la compréhension orale.

Un autre modèle de la compréhension orale que nous présentent Hagoort & van Berkum (2007) est le modèle *two-steps*. Ce modèle propose comme première étape de la compréhension, l'interprétation du sens littéral du message. Dans un deuxième temps, cette première signification construite serait compilée avec les connaissances antérieures de l'individu (représentations cognitives stockées en mémoire). La compréhension correspondrait donc à un traitement séquentiel. Ce modèle renvoie à la conception modulariste du langage (Fodor, 1983) qui « *se traduit par un fonctionnement essentiellement séquentiel des modules, ce qui ne veut pas dire que, pour qu'un module se mette en activité, il faille que l'information ait été totalement traitée par le module précédent. Quoi qu'il en soit, cette conception exclut toute intervention d'informations contextuelles (i.e., syntaxiques ou sémantiques) avant un certain niveau de traitement du lexique, par exemple* » (Blanc & Brouillet, 2003:53). Ce modèle reprend donc les principes d'un accès au sens **ascendant**, c'est-à-dire trouvant son point d'entrée dans le signal. En opposition à ce modèle, van Berkum, van den Brink, Tesink, Kos, & Hagoort (2008) ont montré, de leur côté, que la prise en compte de la situation de communication (dans le cadre de leurs travaux, les caractéristiques sociolinguistiques des locuteurs) donne lieu à des attentes chez le compreneur... « *Pour le cerveau de l'utilisateur, il n'a pas de sens en dehors du contexte* » (*ibidem* :589)<sup>42</sup>. Ces travaux ont contribué à conforter une théorie opposée au modèle *two-steps* : le modèle *one-step*, dans lequel tous les indices, quels que soit leur niveau, sont pris en compte immédiatement et simultanément (Hagoort & van Berkum, 2007; Kuperberg, 2007; Lotze, Tune, Schlesewsky, & Bornkessel-Schlesewsky, 2011). Ce modèle renvoie quant à lui à l'approche interactionniste du langage. Dans cette approche « *les traitements sont effectués en parallèle à partir d'une information ascendante (bottom-up) qui provient des entrées et d'une information descendante (top-down) qui provient des connaissances en mémoire. (...) Le résultat terminal des traitements effectués sur les mots est un patron d'activation qui représente ce qu'est la phrase. Cette représentation exercera une influence descendante sur les traitements ultérieurs* » (Blanc & Brouillet, 2003:53). Dans ce modèle, le contexte joue un rôle primordial dans la compréhension : il participe directement à l'élaboration du sens, comme dans la théorie du langage en tant qu'action. La

---

<sup>42</sup> « To the brain of the language user, there is no context-free meaning » (Van Berkum *et al.*, 2008, p. 589).

compréhension d'un discours, en tant que représentation émergente de ce dernier, s'élabore à partir des différents éléments perçus et identifiés quel que soit leur niveau (sémantique, lexical ou syntaxique) et des connaissances antérieures de l'individu. Dans cette perspective, les indices signifiants présents dans le contexte sonore environnemental du discours, s'ils font l'objet d'une focalisation attentionnelle suffisante pour impliquer leur identification, peuvent participer à l'élaboration du sens (Bornkessel-Schlesewsky & Schlewsky, 2008; Kuperberg, 2007; pour une revue de littérature, voir (Bornkessel-Schlesewsky & Friederici, 2007).

Cette revue de littérature nous a permis de mettre en évidence deux types de recherches en compréhension des lexies orales. Le premier ensemble met en évidence le traitement du signal (mouvement ascendant de l'objet vers l'individu) alors que l'autre insiste sur les intentions du compreneur (mouvement descendant). Les unes renvoient donc à un modèle séquentiel presque mécanique alors que les autres font une part belle d'une part au contexte et d'autre part aux intentions du compreneur. Notre positionnement scientifique avancé par l'objet de notre recherche nous situe du côté de la seconde approche.

### **3. Des études sur la compréhension de textes narratifs**

Nous avons considéré dans quelle mesure la compréhension orale renvoie à l'émergence d'une représentation de « l'objet à comprendre », réalisée à partir de l'activation de représentations cognitives. Dans ce cadre, il semble intéressant de se pencher sur la compréhension de textes : rien dans cette définition ne semble dépendant de la modalité dans laquelle la langue est mise en œuvre. Parallèlement, il semble également que rien, dans les définitions de la compréhension écrite auxquelles nous nous référerons, n'entre en contradiction avec une application de ces principes à la compréhension de l'oral. Dit autrement, nous considérons que le processus de compréhension, tel que nous l'avons décrit jusqu'ici, implique à la fois des processus généraux à l'activité de compréhension (activation de représentations cognitives et élaboration d'une nouvelle représentation émergente, propre à la situation donnée et à l'objet à comprendre –en contexte), et des processus spécifiques, liés aux conditions de cette activité (présence/absence du stimulus, nature du stimulus ...). A la suite de Blanc & Brouillet (2003:67), nous pouvons affirmer qu'« *étudier la compréhension de textes consiste à s'intéresser aux représentations cognitives, aux procédures et mécanismes que l'individu met en œuvre lorsqu'il est engagé dans ce type d'activité* ». Parallèlement, Rouet,

Coutelet, & Dinet, (2004) disent que « *comprendre, (...) c'est se construire un modèle mental de la situation décrite par le texte* ». Il apparaît donc que comprendre un texte revient à faire émerger une représentation de l'objet à comprendre (ici un texte). Les auteurs (Rouet *et al.*, 2004) précisent que pour qu'il y ait compréhension, trois conditions doivent être remplies « (a) *le texte contient une représentation nécessaire et suffisante d'une situation de référence; (b) la personne qui se saisit d'un texte se donne pour but de comprendre cette situation; (c) la personne lit intégralement le texte et construit une représentation mentale qui dérive en grande partie de la représentation verbale donnée par le texte* ». Ainsi, nous pouvons encore une fois considérer que les éléments relevés par les auteurs restent fonctionnels, même dans le cadre de l'étude de la compréhension de l'oral.

Dans ce cadre, il semble particulièrement difficile de considérer les activités de compréhension orale et écrite comme deux processus distincts.

### **3.1. Les dialogues pédagogiques : une forme de narration ?**

Dans le cadre des travaux sur la compréhension de l'écrit, de nombreuses études, que nous allons présenter dans les pages qui suivent, portent sur la compréhension des textes narratifs. Il est possible de faire un parallèle entre les textes narratifs et les dialogues pédagogiques sur lesquels nous travaillons. La narration peut en effet être envisagée comme un récit, dans le sens où une narration comprend un début (situation initiale), un déroulement (état d'actualisation) et un état final (Charaudeau, 1992). Ces aspects peuvent être partagés par les dialogues pédagogiques (voir p.151), surtout dans le cadre des dialogues accompagnés de sons environnementaux où le contexte sonore environnemental, en position d'amorce, peut avoir pour fonction d'introduire la situation initiale. Par ailleurs, l'auteur s'interroge sur la polysémie du terme de narration et donc du concept de « *narré* » : « (...) *les termes récit, narration, histoire sont-ils équivalents ? Est-ce qu'un récit est une narration de faits ? Est-ce que narrer c'est raconter une histoire ?* » (Charaudeau, 1992: 710). Le concept de *narration* reste relativement vague et flou. Dès lors, considérer les dialogues pédagogiques comme se rapprochant de la narration ne semble pas problématique.

Nous considérerons ici que le caractère narratif d'un discours dépend essentiellement de l'individu émetteur de ce discours, de son intentionnalité. Dans ce cadre, la narration est à opposer à la description : « **Le Descriptif** nous fait découvrir un monde qui est censé exister comme un être-là qui se donne pour tel, de manière immuable. Ce monde qui a seulement besoin d'être reconnu n'a plus qu'à être montré. **Le Narratif**, au contraire, nous fait découvrir un monde qui est construit dans le déroulement même d'une succession

*d'actions qui s'influencent les unes les autres et se transforment dans un enchaînement progressif*» (Charaudeau, 1992:715). En ce sens, les dialogues pédagogiques peuvent être considérés comme des formes de narrations puisqu'ils rendent compte d'une suite logique et cohérente d'événements et d'actions (les tours de parole et certains sons environnementaux « de *transcodage* » (voir *Contexte Didactique* p.17) ; les sons environnementaux *situationnels* assumerait, dans ce cadre une fonction descriptive). Parallèlement, (en tant qu'auteur des dialogues, et avec une double intention, à la fois narrative et pédagogique) nous, avons veillé à présenter, dans ces dialogues les éléments nécessaires à leur cohésion, respectant ainsi l'un des impératifs de la narration. Pour autant, peut-on considérer, dans le cadre d'un discours direct, tel que les dialogues sur lesquels nous nous appuyons, qu'une instance intermédiaire est présente et « *parle* » (en opposition à l'auteur qui « *écrit* »). Ainsi, Bal (1977) affirme que « *à un moment décisif de l'histoire de la théorie du récit, on a découvert l'importance essentielle de ce délégué, l'autonomie de celui que l'auteur a délibérément investi de la fonction narrative dans le récit : le narrateur* ». Cet aspect représente donc une des distinctions majeures entre la narration et nos dialogues.

En outre, et d'un point de vue plus pragmatique, il apparaît que les modèles issus des travaux sur la compréhension de texte narratifs font sens également dans le cadre de l'étude qui nous occupe. En effet, ces travaux participent à notre réflexion et au cadre théorique dans lequel nous nous inscrivons.

## **3.2. Les modèles de compréhension de textes narratifs**

La compréhension de texte, tout comme la compréhension de la parole passe par une étape d'association des unités (lexies) avec une représentation cognitive déjà présente. À partir de l'identification de ces premières unités de sens, l'élaboration d'une nouvelle représentation, correspondant au texte, se met en œuvre. Différents modèles tentent de rendre compte de cette élaboration. Nous retiendrons ici principalement le modèle de *construction-intégration* et le modèle *landscape*.

### **3.2.1. Le modèle de construction-intégration : l'élaboration d'un modèle de situation**

En 1983, puis en 1988 et 1998, Kintsch (Kintsch, 1998; Kintsch, 1988; Van Dijk & Kintsch, 1983) a proposé différents modèles de la compréhension de textes narratifs (introduction du *modèle de situation*, puis élaboration du modèle de *construction-intégration*), s'intéressant essentiellement aux processus cognitifs qui sous-tendent la compréhension de textes (et notamment à l'élaboration d'une représentation du texte

lu). Le modèle de *construction-intégration* rend compte de la manière dont les textes sont représentés pendant le processus de compréhension, puis intégrés aux connaissances de l'individu. Ce modèle propose que la compréhension d'un texte repose sur l'élaboration de deux niveaux de représentation par le lecteur. Le premier correspond à la *base de texte*. Il est composé, d'une part de la *surface de texte* c'est-à-dire des caractéristiques lexico-syntaxiques du texte (dans Kintsch 1983) et d'un *niveau sémantique*, c'est-à-dire de propositions globales ou locales, organisées en micros et macros structures. La proposition apparaît alors comme une unité particulièrement importante dans le cadre de la compréhension du linguistique : Kintsch & Keenan (1973) ont en outre observé qu'une variation du nombre de propositions au sein de phrases de même longueur, impliquait une variation des temps de lecture. En outre, Goetz, Anderson, & Schallert (1981), ont également observé qu'une variation du nombre de propositions impliquait une variation des performances en tâche de rappel. Les propositions et leur nombre jouent un rôle primordial dans l'élaboration du sens des phrases. En effet, plus le nombre de propositions à traiter sera important, plus cette opération impliquera une charge cognitive importante.

La *base de texte* renvoie donc à un réseau propositionnel directement issu du texte. À ce niveau déjà, « *un processus d'intégration va renforcer les éléments précédemment sélectionnés comme étant pertinents et inhiber ceux qui sont inappropriés* » (Chanquoy, Tricot, & Sweller, 2007:201). Toutefois, ce réseau reste très souvent pauvre, et, comme pour l'ensemble des tâches de compréhension que nous avons présentées, le lecteur doit faire intervenir ses connaissances antérieures (représentations cognitives, schémas et nœuds) à ce réseau pour accéder à la construction d'un sens nouveau. À ce stade, sont maintenues les connaissances antérieures directement activées par la base de texte, ainsi qu'un petit nombre de leurs plus proches voisins dans le système de connaissance. C'est dans ce cadre que se construit le *modèle de situation*. Ce modèle correspond à une représentation temporaire dans le sens où elle ne présente pas la cohésion nécessaire à son stockage à plus long terme. Ainsi, le modèle de situation se construit à partir de la base de texte : après son élaboration, les éléments qui la composent sont associés au réseau de connaissances de l'individu. A partir de cette construction, l'individu émet des inférences (activation d'informations qui ne sont pas explicitement données (Sperber *et al.*, 2010)). van Den Broek (1994) va jusqu'à imaginer un réseau de représentations interreliés, composé de nœuds lexicaux activés, de propositions construites et d'inférence. Dans ce cadre, l'élaboration de la signification trouverait donc son point de départ dans l'activation de plusieurs sens potentiels. Intervient ensuite l'inhibition des nœuds les moins forts (les moins interreliés à d'autres nœuds du réseau), parallèlement au renforcement et au maintien d'une signification spécifique au sein d'une structure cohérente : les événements du récit

deviennent les éléments centraux, les nœuds autour desquels s'organise l'ensemble des autres informations. Dans ce cadre, la signification d'un nœud dépend de la place qu'il occupe au sein du réseau, c'est-à-dire de la force des liens qu'il entretient avec les autres nœuds. Ce modèle suggère qu'au fil de la compréhension, les nœuds construits précédemment seront renforcés ou au contraire inhibés en fonction des nouvelles informations contenues dans le texte. L'activation et l'inhibition apparaissent dans ce modèle comme dépendantes de l'efficacité des nœuds. Dit autrement, le réseau de propositions s'adapte à travers un processus de satisfaction de contraintes : la pertinence et la cohérence deviennent alors des facteurs décisifs (Kintsch, 1998; Van Den Broek, Young, Tzeng, Linderholm, 1999). Dans ce cadre, il est parfois impossible d'offrir un traitement profond à chacune des propositions. Toutefois, les auteurs considèrent généralement que plusieurs sous-systèmes participent à l'organisation du *modèle de situation*. Ainsi, selon Van Dijk & Kintsch (1983), la compréhension de texte prend place dans *l'unité centrale* autour de laquelle s'organisent trois sous-systèmes : le registre sensoriel (qui maintient les unités perçues disponibles pour le traitement), la mémoire à long terme (où sont stockées les connaissances antérieures de l'individu) et la mémoire épisodique du texte (c'est-à-dire une représentation en cours d'élaboration). *L'unité centrale de traitement* correspond donc au « noyau actif de la compréhension ». À partir de ce réseau de propositions, la représentation du texte lu (ou d'un de ses fragments) s'organiserait à travers l'inhibition de certains éléments qui composent le réseau et le renforcement de certains nœuds (*a priori* ceux connectés à d'autres nœuds du réseau).

Il apparaît en outre que le maintien de chacun de ces niveaux est différent: alors que *les informations de surface* sont rapidement oubliées, la *base de texte* se maintient plus longtemps, mais décline rapidement si elle n'est pas stimulée. Enfin, le *modèle de situation* se caractérise par son passage en mémoire à long terme.

La distinction introduite entre *base de texte* et *modèle de situation* permet de distinguer clairement mémorisation et compréhension. Dans ce cadre, il apparaît qu'une tâche de résumé pour rendre compte d'un texte renvoie plus à la récupération en mémoire de la *base de texte* alors que vérifier la validité d'énoncés renvoie au *modèle de situation* (Blanc & Brouillet, 2003:74). Parallèlement, Schmalhofer & Glavanov (1986), ont pu observer que « *l'annonce d'une tâche de résumé favorisait la représentation des niveaux de surface et de la base de texte, tandis que la consigne d'acquérir des connaissances accentuait l'attention portée au niveau du modèle de situation* » (cité par Blanc & Brouillet, 2003:75). Par ailleurs, Bransford, Barclay, & Franks (1972) ont montré que le lecteur peut avoir élaboré un *modèle de situation*, sans être en mesure d'effectuer des

tâches de rappel sur la *base de texte*. Il apparaît donc que compréhension et mémorisation, malgré les liens forts qui les unissent restent deux opérations distinctes. Les capacités du système mnésique, et notamment de la mémoire de travail, étant limitées (conception structurelle de la mémoire), le *modèle de situation* est considéré comme fragmentaire et largement dépendant du nombre de propositions à traiter.

En 1995, Zwaan, et ses collaborateurs (Zwaan *et al.*, 1995) introduisent dans leurs travaux la différence entre « *structure du récit* » (que l'on peut rapprocher du *modèle de situation*) et « *structure du discours* » (qui correspond à la présentation, l'organisation du texte et qui, en ce sens, est à rapprocher de la *base de texte*). Le lecteur doit donc reconstruire la « *structure du récit* » à partir de la « *structure du discours* » : il s'agit d'identifier les actants, les actions et la situation spatio-temporelle. Cinq dimensions sont alors définies pour la construction de la représentation (structure) du récit : le temps, l'espace, la causalité, l'intentionnalité et, les protagonistes/objets. Ces indices construits permettraient au lecteur de se représenter à la fois le cadre spatio-temporel de la situation présentée par le texte (à travers les dimensions du temps et de l'espace), mais également les personnages impliqués dans la narration, leurs objectifs, les raisons de ces objectifs ainsi que les événements dont l'*histoire* rend compte. Dans le modèle d'indexage d'événements, la cohérence du texte repose sur la continuité des cinq dimensions évoquées plus haut. Ainsi, lorsque la continuité situationnelle est assurée par le texte, les inférences de connexions suffisent, mais lorsque ce n'est pas le cas, le lecteur émet des inférences nouvelles, basées sur ces connaissances antérieures (Magliano *et al.*, 1999; Zwaan *et al.*, 1995), augmentant ainsi le temps de traitement. Or, dans les recherches en psychologie expérimentale, le temps de traitement est généralement considéré comme un bon indicateur de la complexité de la tâche et de la charge cognitive qui en découle. Il apparaît donc que le cadre situationnel a un impact sur les processus de compréhension.

Toutefois, le modèle de *construction-intégration* n'est pas le seul modèle qui fait état du *modèle de situation*. Le modèle dit « *d'indexage d'événements* » y a également recours pour rendre compte des élaborations de sens successives dans la compréhension de textes. En 1998, Zwaan, en collaboration avec Radvansky et Hillyard ont proposé, à partir du modèle d'*indexage d'événements*, que la tâche de compréhension de textes impliquerait l'élaboration de trois types de modèles : un modèle courant (où sont traitées les représentations en cours d'activation), un modèle intégré (où cohabitent les représentations en cours de traitement et celles maintenues des cycles de lecture précédents), et un modèle complet de la situation, (où participent à la fois les représentations activées par la lecture, les représentations maintenues, les

connaissances antérieures et une forme de méta-représentation du texte. Selon ces auteurs et conformément aux travaux d'Ericsson & Kintsch (1995), le *modèle courant* (c'est-à-dire une forme de *modèle de situation*, élaboré à partir de la proposition lue « à ce moment là ») serait élaboré au sein de la Mémoire de Travail à Court Terme (MDT CT), alors que le *modèle intégré* (informations préalablement traitées, *modèles de situations précédents*) serait élaboré en Mémoire de Travail à Long Terme (MDT LT). À ce niveau, un processus de *mise à jour* (c'est-à-dire la construction de liens entre le *modèle courant* et le *modèle intégré*) permettrait la communication entre les modèles stockés en MDT CT et MDT LT jusqu'à la construction d'un modèle complet, stocké en Mémoire à Long Terme. La proposition la plus activée passerait d'un cycle à l'autre, permettant la construction de liens entre les différentes propositions, assurant ainsi la cohérence de la représentation de l'ensemble. Dit autrement, les nouvelles informations seraient ajoutées à la « structure en développement ». En outre, Zwaan *et al.*, (1995) ont montré, à partir d'éléments empiriques, que les discontinuités temporelles et causales impliquent une augmentation des temps de lecture (augmentation de la charge cognitive), laissant entendre que la continuité situationnelle est une dimension importante pour les lecteurs. Cette continuité entraînerait le lecteur à construire des liens causaux entre les différentes propositions. Lorsqu'aucune relation causale ne peut être réalisée entre deux phrases ou deux propositions, la cohérence du texte est rompue. Il s'agit alors pour le lecteur d'aller chercher en mémoire à long terme des éléments du texte, ou parmi ses connaissances antérieures, des éléments lui permettant de rétablir cette connexion. En outre, Magliano, Zwaan, & Graesser (1999) ont mis en avant une corrélation entre la cohérence perçue par le lecteur et le temps de lecture du texte. En effet, une discontinuité situationnelle amènerait le lecteur à construire des sous-structures dont le passage de l'une à l'autre entraînerait une augmentation de la charge cognitive induite par la tâche de compréhension, notamment par l'intermédiaire de l'augmentation des ressources mobilisées en mémoire de travail (Chanquoy *et al.*, 2007).

Ainsi, la signification d'un texte apparaît comme « *flexible, changeante et temporaire* » (Blanc & Brouillet, 2003:87), dépendante des objectifs de l'individu compreneur (J.D. Bransford & McCarrel, 1974), de ces connaissances antérieures et du contexte (phrastique notamment). Or, dans le cadre du paradigme d'une perception et d'une compréhension descendantes, nous considérons que les connaissances antérieures de l'individu permettent à ce dernier de formuler des hypothèses quant aux informations qu'il recherche dans les procédures de perception et de compréhension. En outre, Zwann, Radvansky, Hillyard, & Curiel (1998) ont obtenu des résultats expérimentaux qui montrent que les objectifs du lecteur interviennent dans le suivi des dimensions situationnelles.

La compréhension apparaît dès lors comme le résultat d'une étape de sélection des informations contenues dans le texte. Le *modèle de situation* rend compte d'une construction dynamique, individuelle et dépendante des connaissances antérieures de l'individu et donc susceptible de varier d'un individu à l'autre. Comme le disent Chanquoy *et al.*, (2007), « *Le modèle de situation correspond à une représentation propositionnelle des contenus du texte reliés et intégrés aux connaissances antérieures de l'apprenant (...). Ce dernier complète, grâce à ses propres connaissances, l'information qui n'est pas explicitement donnée dans le discours ; pour cela, il produit des inférences* » (p.103).

Comme nous l'avons vu, à la fois dans le modèle de *construction-intégration* et dans le modèle d'indexage d'événements, le *modèle de situation* est issu de l'interaction entre le texte et les connaissances antérieures que l'individu active pendant la lecture. Ce « *modèle de situation* », en tant que niveau de représentation est constitué à la fois de certains éléments contenus dans le texte et d'inférences du lecteur.

À ce niveau de représentation, le lecteur irait puiser dans ses connaissances antérieures, lorsque le texte ne lui permet pas d'établir une cohérence suffisante. Il ferait alors appel à la fois aux éléments du texte préalablement traités (*modèle intégré*) et à ses connaissances sur le monde. Il ne s'agit donc plus seulement de reconstruire la cohérence de la base de texte, mais bien une cohérence de ces éléments avec les connaissances antérieures. Chaque fragment de texte traité entraînerait l'élaboration d'un *modèle de situation* intégré et maintenu en mémoire de travail. En ce sens, les *modèles de situation* sont des « *reconstructions cognitives de fragments du monde* » (Blanc & Brouillet, 2003:77). Les auteurs évoquent de constants allers-retours entre le texte (et la base de texte), où apparaissent de nouveaux éléments, et la modélisation de ce texte par le lecteur (*modèle de situation*).

Dans le cadre de ce modèle de la compréhension de textes, les connaissances antérieures du lecteur, quelle que soit leur nature, jouent un rôle primordial. Les auteurs (Van Dijk & Kintsch, 1983) soulignent, en outre, la part active du lecteur dans la compréhension, c'est-à-dire dans l'élaboration d'un *modèle de situation*. En effet, Kintsch part du postulat que ce sont les connaissances (à la fois linguistiques et extralinguistiques) du lecteur qui guident la compréhension de textes. Parallèlement, Zwaan considère que les connaissances antérieures de l'individu jouent également un rôle important dans la recherche de signification par le lecteur, dans ses objectifs de lecture. Selon ses objectifs et ses connaissances antérieures, le lecteur sélectionnera différentes informations du texte pour élaborer son *modèle de situation* qui de ce fait,

variera d'un individu à l'autre. Dès lors, la compréhension devient un acte à la fois irrépressible et volontaire.

### **3.2.2. Le modèle *landscape* : la fluctuation de l'activation**

Dans le prolongement de ces travaux, le modèle *landscape* (Van Den Broek, Young, Tzeng, Linderholm, 1999) propose une approche plus dynamique encore. Ce modèle reprend celui de construction-intégration mais y ajoute la fluctuation des concepts activés au cours de la lecture, en raison des limites des ressources attentionnelles et de la capacité de la mémoire de travail (Gernsbacher, Varner, & Faust, 1990). Dans ce modèle, le produit de la compréhension « *est une structure mentale qui ressemble à un réseau de concepts interreliés (i.e., propositions, phrases...)* » (Blanc & Brouillet, 2003:113). Un paysage d'activation serait alors construit, au sein duquel chaque variation impliquerait une restructuration de la représentation dans son ensemble. Cette conception de la compréhension renvoie à la diffusion de l'activation (Jo (Anderson, 1983). En effet, dans le modèle *landscape*, lorsqu'un concept est activé, c'est sa cohorte qui s'active (voir le sous-chapitre *Éléments définitoires: qu'es-ce que comprendre?* p.82). Dans ce cadre, l'activation et la désactivation d'un concept ne renvoient pas seulement à la compréhension du texte, mais bien à l'interaction entre les informations contenues dans le texte et les connaissances antérieures de l'individu, ses attentes, ses objectifs de lecture. A chaque étape de lecture, de nouveaux concepts seraient activés, maintenus ou écartés, sans toutefois être inhibées. Ainsi, dans ce modèle, les concepts peuvent être activés à différents degrés, laissant une large part aux processus attentionnels et à la focalisation qu'ils peuvent impliquer.

Le caractère linéaire de la lecture impose une compréhension progressive du texte, mais la fluctuation des concepts aboutit à la modification perpétuelle du paysage d'activations (certains concepts sont maintenus plus longtemps que d'autres). Dans ce modèle, l'activation des concepts serait issue d'une part de la portion de texte en cours de lecture, stockée en MDT à CT, d'autre part du cycle précédent (les informations contenues dans chaque cycle sont provisoirement maintenues pour le cycle suivant), stockées MDT à LT, d'autre part encore des informations contenues dans des cycles antérieurs, stockées en MLT que le lecteur peut réactiver, et enfin des connaissances antérieures, stockées en MLT que le lecteur peut réactiver également.

## Ce qu'il faut retenir

Dans ce chapitre, nous avons donc vu que perception et compréhension sont deux processus intrinsèquement liés puisque ce que nous appelons « compréhension » correspond à l'élaboration, par l'individu compreneur, d'une représentation nouvelle à partir à la fois d'éléments perçus dans la situation et de ses connaissances antérieures.

Cette nouvelle représentation passe tout d'abord par la décomposition de l'input en unités de sens. Dans le cadre de la compréhension orale, cette décomposition s'appuie sur la connaissance de la langue parlée et aboutit à l'activation de représentations d'ores et déjà stockées en mémoire. À partir de cette première étape, l'activation se propage vers les autres connaissances de l'individu, jusqu'à aboutir à l'activation d'un réseau, réorganisé en fonction des objectifs de l'écoute et des éléments contenus dans le discours.

C'est dans cette perspective que nous considérons la compréhension orale en classe de langue, où le support linguistique correspond à une production langagière (langage en tant que produit) dans laquelle les locuteurs sont inscrits en tant qu'acteurs (langage en tant qu'action). Dès lors, il est envisageable de considérer que les travaux menés sur la compréhension de textes narratifs nous renseignent sur les mécanismes de la compréhension de dialogues pédagogiques tels que ceux que nous incluons dans notre expérimentation principale. Dans ce cadre, deux modèles ont principalement retenu notre attention : le *modèle de situation* et le *modèle landscape*. Le premier propose que la compréhension d'un texte repose sur l'élaboration d'une représentation de ce texte dont les nœuds qui la composent et les différentes propositions construites fluctueraient à mesure que l'individu poursuit sa lecture. Parallèlement, le *modèle Landscape* reprend l'idée de la fluctuation des concepts et la renforce, considérant ainsi la compréhension comme un processus particulièrement dynamique né de l'interaction des informations contenues dans le texte et des connaissances antérieures de l'individu compreneur.

## Troisième Chapitre : Hypothèses

Comme nous l'avons vu dans les chapitres *Des études sur la Perception... aux études sur la Compréhension* (voir p.49 et p.81), nous considérons que perception et compréhension sont des processus imbriqués, intrinsèquement liés, voire co-dépendants. Selon nous, la **compréhension** correspond à l'élaboration d'une signification à partir des connaissances antérieures de l'individu (stockées en MLT) d'éléments perçus, que cette perception soit proche, ou éloignée dans le temps. Dit autrement, la compréhension peut intervenir plus ou moins longtemps après la perception. Parallèlement, nous considérons que la **perception** correspond à un processus dont l'aboutissement est l'activation d'une représentation associée aux objets de l'environnement identifiés. En ce sens, et bien que nous les ayons distingués, dans l'architecture de ce travail, nous considérons que compréhension et perception participent du même phénomène et que nous pouvons dès lors parler de compréhension des scènes sonores.

Notre travail de recherche, et notamment dans sa dimension expérimentale (voir p.119 et p.165), vise **la mise en lumière des spécificités de la compréhension de courts dialogues pédagogiques par des apprenants débutants en langues et des locuteurs natifs (experts linguistique)**. Comme nous l'avons vu dans le chapitre *Contexte Didactique* de ce travail (voir p.17), ces dialogues, fréquemment utilisés en salle de classe pour le développement des compétences en compréhension orale, peuvent présenter de la parole seule ou accompagnée d'un contexte sonore environnemental. Dans ce cadre, et à la suite d'autres travaux (par exemple Kamber & Skupien, 2009; Spanghero-Gaillard & Gaillard, 2007, entre autres) nous nous interrogeons sur **le rôle que peut assumer ce contexte sonore environnemental dans la compréhension de ces supports**.

Ainsi, notre questionnement s'organise autour de deux grandes questions : **Qu'est-ce qu'être débutant en langue ? et dans quelle mesure la présence d'un contexte sonore environnemental adjacent à un dialogue pédagogique va modifier**

**la compréhension de ce dernier** ? Dit autrement, nous nous interrogerons, dans un premier temps sur les spécificités de la compréhension chez les locuteurs novices en langue cible. Le caractère débutant en langue implique-t-il la mise en place de stratégies particulières ? Dans un deuxième temps, nous nous concentrerons sur le rôle d'un contexte sonore environnemental (CSE) sur la compréhension d'un court dialogue par des apprenants débutants de la langue cible et par des locuteurs natifs de cette même langue. À ce niveau, nous tenterons de voir dans quelle mesure le CSE participe à la compréhension des dialogues et si le niveau d'expertise linguistique des participants implique des modifications significatives dans l'appréhension de ce support complexe.

Dans ce cadre, et comme nous le verrons dans la partie expérimentale de ce travail, nous avons choisi d'évaluer la compréhension de trois dialogues pédagogiques en Français de niveau A (*La préparation d'un repas, la vaisselle, et À la terrasse d'un café* ; voir p.151), présentés seuls ou accompagnés d'un contexte sonore environnemental (CSE ; voir p.121). Deux panels de participants ont été interrogés. Le premier panel est composé de participants francophones natifs, et est scindé en deux groupes : le premier soumis aux dialogues seuls ; le deuxième soumis aux dialogues accompagnés d'un CSE. Le deuxième panel que nous avons interrogé est composé de participants apprenants de FLE de niveau A. À son tour, ce panel a été scindé en deux groupes : le premier soumis aux dialogues seuls ; le deuxième soumis aux accompagnés d'un CSE. Une plus ample présentation des participants sera proposée dans la partie *Expérimentation principale* (p.165). Deux variables indépendantes ont présidé à l'élaboration des groupes de participants : le niveau d'expertise linguistique et la confrontation à des dialogues présentés seuls ou accompagnés d'un contexte sonore environnemental. Afin d'évaluer la compréhension des dialogues, dans leurs différentes configurations (avec ou sans CSE), nous avons proposé aux participants une tâche de sélection de dessins, à partir d'un logiciel *ad hoc* (le Classe-Images). À la suite de leur écoute du dialogue, les participants ont été amenés à sélectionner trois dessins, parmi douze, pour rendre compte du dialogue qu'ils avaient entendu. Dans l'élaboration de ce corpus d'images (voir p.169) différents facteurs de variation ont été introduits de manière à appréhender la compréhension de deux paramètres majeurs dans la compréhension : la *situation d'énonciation* et le *contenu propositionnel*. Une troisième dimension a également retenu notre attention : la capacité des participants à reproduire la dimension chronologique des dialogues. Enfin, certains dessins (trois par dialogue) ne présentent aucun facteur de variation, c'est-à-dire qu'ils correspondent au sens que nous avons cherché à construire dans l'élaboration des dialogues. Dans ce cadre, les participants ont élaboré des suites de trois dessins (voir partie *expérimentation principale* p.165). Dès lors, dans l'analyse des résultats nous pourrions considérer soit les suites élaborées (trois dessins), soit les dessins un à un.

Il apparaît que la tâche du Classe-Images correspond à une tâche complexe. En effet, les participants doivent comprendre les dialogues, analyser les dessins qui leur sont proposés, les comparer avec la représentation du dialogue qu'ils auront élaborée, pour enfin rendre compte de leur représentation à partir de dessins qu'ils sélectionnent dans le corpus.

## 1. Qu'est-ce qu'être débutant en langue ?

Ce premier axe d'hypothèses pourrait sembler superflu au regard des objectifs centraux de notre travail (évaluer l'impact de la présence d'un CSE sur la compréhension d'un dialogue pédagogique). En effet, intuitivement, il apparaît que la comparaison des francophones et des apprenants au niveau de leur compréhension orale en Français ne peut que confirmer le choix de placer les participants dans des panels différents. Toutefois, cette question reste pertinente : les différents niveaux d'expertise linguistique impliquent-ils des différences significatives sur la compréhension des éléments para ou extralinguistiques ? Sur quelles dimensions de la compréhension linguistique le niveau d'expertise intervient-il le plus fortement ?

**Dans ce cadre notre hypothèse théorique est que le niveau d'expertise linguistique implique des performances différentes entre les locuteurs francophones natifs et des apprenants débutants de FLE, dans la compréhension de dialogues pédagogiques, évaluée à partir du Classe-Images** et ce malgré la « simplicité » des dialogues proposés : le vocabulaire, la grammaire et la forme des échanges qui composent ces dialogues ont été contrôlés de manière à correspondre aux connaissances en langue des participants apprenants débutants. En effet, comme nous l'avons vu dans le chapitre sur les *... aux études sur la compréhension* (voir p.81), la compréhension du langage passe par une étape d'association des unités sonores (lexies) avec des concepts stockés en mémoire. Cette association trouve son origine dans la segmentation du flux de parole et repose sur leur reconnaissance. Or ces différentes démarches impliquent la connaissance de la langue cible : c'est parce que l'individu connaît la langue qu'il sera capable d'en identifier les unités. Parallèlement une faible connaissance de cette langue entraînera des difficultés au niveau de la segmentation du signal de parole en unités de sens et de leur reconnaissance. Cette augmentation de la difficulté induite par la tâche de compréhension (en raison du déficit de connaissances en langue cible) peut, dans ce cadre, impliquer une augmentation de la charge cognitive chez les participants apprenants débutants. La charge cognitive correspond « à une

*quantité : c'est une mesure de l'intensité du traitement cognitif engagé par un individu particulier – possédant certaines connaissances et certaines ressources -, pour réaliser une certaine tâche, d'une certaine manière, dans un certain environnement* » (Chanquoy, Tricot, & Sweller, 2007:31). Dans ce cadre, nous ne pouvons dès lors pas exclure les situations de surcharge cognitive (baisse des performances d'un point de vue qualitatif, notamment) chez les participants apprenants débutants, sans que de tels effets apparaissent chez les participants francophones (plus un sujet est expert dans un domaine, moins le traitement d'une situation dans ce domaine va être coûteuse cognitivement (Tricot, 1998). Ces différences s'observeront à différents niveaux, différentes hypothèses opérationnelles seront donc formulées. À ce niveau, nous nous concentrerons sur les résultats obtenus par les groupes ayant entendu les dialogues seuls (sans CSE ; condition contrôle) et comparerons les performances des participants francophones et apprenants débutants.

Ainsi, notre **première hypothèse opérationnelle** est que **le niveau d'expertise linguistique des participants a un impact sur la compréhension des dialogues**. Nous pensons que des locuteurs francophones natifs produiront de « meilleurs résultats » en tâche de compréhension de dialogues pédagogiques de niveau A) que le panel d'apprenants débutants même si les deux ont théoriquement accès au lexique et à la syntaxe employés (voir *Élaboration du matériel linguistique* p.151). Dit autrement, nous supposons que les participants francophones sélectionneront plus de dessins ne présentant aucun facteur de variation, que les participants de niveau A. Si nous considérons la compréhension en termes binaires (compris /non compris) alors nous nous attendons à ce que les francophones natifs présentent un meilleur taux de réussite pour cette tâche particulière que les apprenants débutants.

Pour confirmer cette hypothèse, nous comparerons les dessins sélectionnés par les francophones et apprenants débutants dans la configuration « écoute du dialogue seul » (condition contrôle). Cette comparaison portera sur l'élaboration de suites de trois dessins-cibles, c'est-à-dire sur les constructions qui laissent penser à une adéquation entre le sens envisagé dans l'élaboration des supports et celui (re)construit par les participants. Nous supposons que de telles suites seront plus fréquentes chez les participants francophones que chez les apprenants débutants. Nous observerons également les dessins cibles (sans facteur de variation) sélectionnés par les participants et considérés isolément. Pour confirmer ces hypothèses nous procéderons à un test de comparaison de fréquences (tables de contingence ; pour plus de détails se reporter aux annexes, sur le CD) à partir du nombre de suites de dessins cibles construites. Ce test portera à la fois sur les résultats globaux (pour les trois dialogues considérés

conjointement) et pour chaque dialogue pris isolément. Par ailleurs nous procéderons à un t-test à partir du nombre total de dessins cibles sélectionnés et considérés isolément (données quantitatives). De cette manière nous pourrions évaluer les différentes « parcelles » qui composent la compréhension des participants. Dit autrement, alors que dans la comparaison des suites de trois dessins cibles nous considérerons la compréhension en termes binaires (est-ce que la signification élaborée par les participants correspond à celle que nous avons envisagée dans la construction des supports ?), dans cette deuxième analyse des dessins cibles sélectionnés nous considérerons que chaque dessin rend compte d'une *parcelle* de signification.

Notre **deuxième hypothèse opérationnelle** est que **le niveau d'expertise linguistique des participants a un impact sur leur appréhension de la situation d'énonciation des dialogues**. Or, cette dimension est particulièrement importante dans la compréhension (cf les travaux de Zwaan, présentés dans le chapitre sur ... *aux études sur la compréhension* ; voir p.81). Nous supposons que les participants francophones natifs présenteront moins de difficulté à appréhender cette dimension que les participants apprenants débutants puisqu'en dehors des indices extra-verbaux (comme un contexte sonore environnemental par exemple), la situation d'énonciation est avant tout présentée *via* les éléments paralinguistiques *a priori* difficilement appréhendables par des apprenants débutants. Dit autrement, nous supposons que les participants apprenants débutants sélectionneront plus de dessins présentant une variation de la situation d'énonciation que les participants francophones. Pour vérifier cette hypothèse nous procéderons à un t-test à partir du nombre total de dessins avec une variation de la situation d'énonciation sélectionnés et correctement agencés sur l'axe chronologique/syntagmatique. Toutefois, à ce niveau, nous avançons une nouvelle hypothèse, sur la répartition des dessins non-cibles chez les deux panels de participants. En effet, nous supposons que les participants francophones et apprenants débutants ne présenteront pas les mêmes profils d'erreurs. Dit autrement, selon nous, le niveau d'expertise linguistique des participants impliquera différentes stratégies de compréhension et, par là même, des « sensibilités » différentes aux divers facteurs de variation. Nous nous concentrerons donc également sur les proportions que représente chaque facteur de variation sur l'ensemble des dessins non cibles sélectionnés par chaque groupe de participants. Dans ce cadre, nous discuterons (voir p.224) ces profils d'erreurs et tenterons de voir dans quelle mesure les niveaux d'expertise entraînent des erreurs spécifiques à chaque niveau.

Notre **troisième hypothèse opérationnelle** est que **le niveau d'expertise linguistique des participants a un impact sur leur appréhension du contenu propositionnel des dialogues**. Nous supposons que la capacité à appréhender le contenu propositionnel augmente avec le niveau d'expertise des participants. Dit autrement, nous supposons que le panel de francophones aura sélectionné moins de dessins présentant une variation du contenu propositionnel que le panel d'apprenants débutants. Pour confirmer cette hypothèse nous procéderons à un t-test à partir du nombre de dessins présentant ce facteur de variation, sélectionnés par les participants francophones et les apprenants débutants ayant entendu le dialogue seul (en condition de contrôle). Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément.

Notre **quatrième hypothèse opérationnelle** est que **le niveau d'expertise linguistique des participants aura un impact sur leur appréhension de la dimension chronologique des dialogues**. Cet aspect renvoie selon nous à l'élaboration des cinq dimensions de la structure du récit décrites par Zwaan, Magliano, & Graesser (1995), et impose l'émission de nombreuses inférences et la reconstruction de la cohésion du discours. Dit autrement, nous supposons que le panel de francophones aura rencontré moins de difficultés à positionner les dessins sur l'axe chronologique que le panel d'apprenants débutants. Cet impact ne sera observable qu'à partir des dessins cibles (sans variation) et avec une variation de la situation d'énonciation sélectionnés puisqu'il s'agit des seuls dessins pour lesquels nous avons pu contrôler ce paramètre (voir p.169). Pour confirmer cette hypothèse nous procéderons à un t-test à partir du nombre de dessins cibles et avec variation de la situation d'énonciation, sélectionnés mais mal positionnés sur l'axe chronologique par les participants francophones et les apprenants débutants ayant entendu le dialogue seul (en condition de contrôle). Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément.

Enfin, notre **cinquième hypothèse opérationnelle** est que **le niveau d'expertise linguistique des participants aura un impact sur les temps de réponse** impliqués par la réalisation de la tâche. Les Temps de Réponse (TR) sont considérés comme un indice de la charge cognitive que l'individu investit dans la réalisation de la tâche : ici, rendre compte, à partir de dessins, du dialogue entendu. La charge cognitive peut être définie comme la mesure des ressources mentales qu'un individu investit dans une tâche. Dans ce cadre, la charge cognitive est à la fois dépendante des capacités propres de l'individu (variation interindividuelle), des spécificités de la tâche en

question (variation inter-tâche), et enfin des choix (potentiellement inconscients) de l'individu quant à ses investissements mentaux dans les différentes tâches concomitantes auxquelles il peut être confronté. Notons que la tâche impliquée par ce test est une tâche complexe : il s'agit à la fois de comprendre le dialogue entendu (en construire une représentation mentale, un modèle de situation), d'analyser les dessins qui sont présentés (en traiter les différences et les points communs), de comparer les dessins à la représentation mentale du dialogue et enfin de procéder à la restitution d'une construction rendant compte du dialogue, à partir de dessins. Or, « *la compréhension peut représenter un coût important, notamment par la production d'inférences qu'elle implique* » (Chanquoy, Tricot, & Sweller, 2007:102). Cependant, dans ce cadre, les experts linguistiques, plus familiarisés avec la tâche de compréhension en français (leur langue maternelle) devraient présenter moins de difficultés dans l'exercice demandé. En effet, dans l'activité de compréhension, l'individu doit confronter les informations qu'il extrait du discours et ses connaissances antérieures, et ce dans le cadre de la mémoire de travail (Gaonac'h & Larigauderie, 2000) dont la capacité est limitée. Ces limitations seront d'autant plus importantes que les différents types d'informations (verbales ou environnementales, issues des connaissances antérieures ou extraites du flux) qui rentrent en jeu dans la confrontation sont éloignées, comme c'est le cas dans le cadre de la compréhension en langue étrangère. Nous pensons donc que la tâche nécessitera un plus faible investissement en ressources cognitives chez les francophones natifs que chez les apprenants. Dit autrement, nous supposons que les participants francophones auront réalisé cette tâche en moins de temps que les apprenants débutants. Pour confirmer cette hypothèse nous procéderons à un test U de Mann-Whitney à partir des temps de réponse des participants francophones et apprenants débutants ayant entendu le dialogue seul (condition de contrôle). Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément.

## **2. Le rôle du contexte sonore environnemental dans la compréhension d'un dialogue**

Pour ce deuxième axe, **notre hypothèse théorique est que la présence d'un contexte sonore environnemental adjacent à un dialogue et redondant avec ce dernier implique une modification des performances en compréhension de ce dialogue.** Cette modification s'observera à différents niveaux, sur différents aspects de la compréhension. Tout d'abord, à la lecture de la littérature scientifique sur les études sur la compréhension (voir p.81), nous avons pu voir que certains modèles de

compréhension de textes narratifs posent comme première étape de l'élaboration du sens l'identification du cadre spatio-temporel. Or, comme nous le verrons dans *l'élaboration du matériel sonore environnemental* (voir p.121), les contextes sonores environnementaux (CSE) que nous avons construits présentent une forte dimension situationnelle. Dit autrement, les CSE impliquent une *continuité associative* (Blanc & Brouillet, 2003) avec les dialogues. Dans ce cadre, nous pensons que la présence d'un CSE congruent avec le dialogue, en apportant des indices supplémentaires, va favoriser l'élaboration d'une représentation globale du dialogue. De plus, dans la mesure où le contexte sonore environnemental commence systématiquement avant le début du dialogue, nous pouvons penser qu'il assume une fonction d'amorce à l'écoute « préparant » la compréhension du dialogue, « *et ceci est d'autant plus vrai que l'attention est endogène, c'est-à-dire délibérément déployée à la suite de l'amorce* » (Camus, 1996:32). Dit autrement, l'écoute du CSE serait susceptible de déclencher des attentes chez les participants auditeurs-compreneurs. Ces attentes correspondraient alors à une forme de préparation attentionnelle donnant lieu à l'activation préalable des connaissances associées aux éléments perçus et identifiés dans le CSE<sup>43</sup>. Ainsi, les représentations et concepts associés aux éléments du dialogue seraient, en quelque sorte « pré-activés ». Dans ce cadre, l'écoute d'un CSE en amorce du dialogue permettrait de « *guider de manière proactive l'analyse de l'information en cours de traitement* » (Camus, 1996:34).

Différentes hypothèses opérationnelles seront donc formulées. Notons toutefois que la présence d'un contexte sonore environnemental aura *a priori* un impact différent selon le niveau d'expertise des participants. En effet, la compréhension correspond à la confrontation entre les éléments contenus dans l'objet à comprendre (ici le dialogue accompagné ou non d'un CSE) et les connaissances antérieures de l'individu. Par ailleurs « *plus les connaissances contenues dans le discours et celles que possède le lecteur sont proches, plus la compréhension sera aisée* » (Chanquoy, Tricot, & Sweller, 2007:102-103). Or, si les connaissances linguistiques des participants apprenants débutants sont déficitaires en langues cible (le français), alors que les participants francophones peuvent être considérés comme des experts linguistiques, nous pouvons penser que les CSE auront des implications différentes selon le panel de population interrogé. A partir de ce positionnement, nous pouvons considérer que le CSE offrira aux participants apprenants débutants une base pour l'élaboration du sens, susceptible de pallier certaines difficultés liées au caractère novice de ces participants, alors que ce même CSE jouera un rôle moins important chez les participants francophones. Dit autrement, nous

---

<sup>43</sup> L'identification des scènes sonores élaborées a été vérifiée (voir p.134)

pensons que la présentation duelle (linguistique et non-linguistique), mais monomodale (auditive) des informations sera plus « favorable » aux novices qu'aux experts.

Notre **première hypothèse opérationnelle est que la présence d'un contexte sonore environnemental aura un effet positif sur la compréhension**. Dit autrement, nous pensons que les participants qui auront écouté les dialogues dans la configuration « écoute avec contexte sonore » auront construit plus de suites de dessins sans variation (*i.e.*, dessins conformes aux attentes) et auront sélectionné plus de dessins cibles (sans facteur de variation) considérés isolément, que les participants qui auront entendu les dialogues dans la configuration « écoute sans contexte sonore ». Néanmoins, nous avançons également l'hypothèse d'un effet moindre de la présence d'un CSE sur les performances des francophones que sur celles des membres du panel d'apprenants débutants par rapport à leurs performances respectives sans CSE. En effet, nous supposons que 1) les participants francophones seront plus sensibles à la redondance entre le CSE et le dialogues (moins fortement perçue par les apprenants débutants en raison de leur moindre compréhension du linguistique) et 2) le déficit en compréhension des éléments linguistiques des participants apprenants débutants amènera ces derniers à se « raccrocher » plus fortement aux éléments extralinguistiques.

Nous envisageons également que ces effets varient en fonction du dialogue observé et du degré de redondance qu'il entretient avec le CSE. En effet, il apparaît que si les deux sources d'information (linguistiques et non-linguistiques) sont considérées par les participants comme pouvant être comprises isolément (ce qui *a priori* sera le cas des participants francophones ; voir *l'identification des scènes sonores* p.138) alors dialogue et CSE seront considérés comme redondants. Parallèlement, si les deux sources d'informations sont considérées comme ne pouvant être comprises par les participants (le caractère novice des apprenants débutants implique des difficultés dans l'appréhension de la dimension linguistique des constructions), alors les deux sources d'information ne seront plus considérées comme redondantes mais complémentaires. Dans ce cadre, l'éventuelle augmentation de la charge cognitive que nous pourrions observer ne sera plus liée à un effet de redondance mais à une dissociation de l'attention, auquel les experts sont moins sensibles (Tricot, 1998). Toutefois, dans les situations de dissociation de l'attention, les individus doivent procéder à l'intégration mentale des différentes informations issues de diverses sources, impliquant une augmentation de la charge cognitive.

Pour confirmer ces hypothèses nous procéderons à un test de comparaison de fréquence à partir du nombre de suites de dessins cibles construites par les participants francophones et apprenants débutants ayant entendu le dialogue seul ou accompagné

d'un CSE. Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément. Nous procéderons également à un t-test à partir du nombre total de dessins cibles sélectionnés et correctement agencés sur l'axe chronologique/syntagmatique et sur ces mêmes dessins lorsqu'ils ne participent pas à une suite de trois dessins, c'est-à-dire lorsqu'ils sont accompagnés de dessins appartenant à une catégorie différente. Ces tests porteront à la fois sur les résultats globaux (trois dialogues observés conjointement) et sur chaque dialogue considéré isolément.

Notre **deuxième hypothèse opérationnelle** est que **la présence d'un CSE a un impact positif sur l'appréhension de la situation d'énonciation** des dialogues par les participants francophones et les participants apprenants débutants. Dans la mesure où les CSE construits présentent tous une dimension situationnelle assez marquée (voir p.121) ils devraient tous avoir une influence sur l'appréhension de la situation d'énonciation. Toutefois, nous supposons que cet impact sera différent en fonction du dialogue observé et du degré de redondance introduit entre lui et le CSE. Ce facteur sera plus largement analysé en discussion, dans le sous chapitre dédié aux profils d'erreurs et aux effets d'interaction (p.224). Dit autrement, nous supposons que les participants francophones et apprenants débutants qui auront entendu les dialogues accompagnés d'un CSE auront présenté moins de difficulté à rendre compte de la dimension situationnelle. Cette moindre difficulté devrait s'observer à travers le nombre de dessins avec une variation de la situation d'énonciation sélectionnés par les participants : nous pensons que les participants qui auront écouté les dialogues dans la configuration « écoute des dialogues avec contexte sonore », sélectionneront moins de dessins présentant ce type de variation que les participants qui auront écouté les dialogues dans la configuration « dialogue seul ». Pour confirmer cette hypothèse, nous procéderons à un t-test à partir du nombre total de dessins avec une variation de la situation d'énonciation sélectionnés et correctement agencés sur l'axe chronologique/syntagmatique et sur ces mêmes dessins lorsqu'ils ne participent pas à une suite de trois dessins, c'est-à-dire lorsqu'ils sont accompagnés de dessins appartenant à une catégorie différente. Ces tests porteront à la fois sur les résultats globaux (trois dialogues observés conjointement) et sur chaque dialogue considéré isolément.

Notre **troisième hypothèse opérationnelle** est que **la présence d'un CSE a un impact positif sur l'appréhension du contenu propositionnel des dialogues** par les participants francophones et apprenants débutants. Toutefois, dans la mesure où les CSE construits ne présentent pas tous le même degré de transcodage (voir p.121) ils ne devraient pas tous avoir la même influence sur l'appréhension du contenu propositionnel des dialogues. Cet aspect sera plus largement abordé dans la discussion de ce travail. Dit autrement, nous supposons que les participants francophones et apprenants débutants qui auront entendu les dialogues accompagnés d'un CSE rencontreront moins de difficulté à rendre compte du contenu propositionnel. Cette moindre difficulté devrait s'observer à travers le nombre global de dessins avec une variation du contenu propositionnel sélectionnés et correctement agencés sur l'axe chronologique/syntagmatique. Pour confirmer cette hypothèse, nous procéderons à un t-test à partir du nombre total de dessins avec une variation du contenu propositionnel sélectionnés par les participants des groupes qui ont entendu les dialogues avec et sans contexte sonore environnemental, francophones et apprenants débutants. Ces tests porteront à la fois sur les résultats globaux (trois dialogues observés conjointement) et sur chaque dialogue considéré isolément.

Notre **quatrième hypothèse opérationnelle** est que **la présence d'un CSE n'a pas d'impact sur l'appréhension de la dimension chronologique des dialogues**. Dit autrement, nous supposons que les participants des panels de francophones et d'apprenants débutants auront rencontré autant de difficulté à positionner les dessins sur l'axe chronologique/syntagmatique. Pour confirmer cette hypothèse nous procéderons à un t-test à partir du nombre de dessins cibles et avec une variation de la situation d'énonciation sélectionnés mais mal positionnés par les participants. Nous supposons que les différents groupes de participants présenteront des seuils d'erreurs comparables. Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément.

Enfin, notre **cinquième hypothèse opérationnelle** est que **la présence d'un CSE aura un impact sur les temps de réponse** impliqués par la réalisation de la tâche. Dit autrement, nous supposons que les participants francophones et apprenants débutants qui auront écouté le dialogue accompagné d'un contexte sonore environnemental (condition DCS) réaliseront cette tâche en plus de temps que les participants ayant entendu le dialogue seul. En effet, l'ajout d'un contexte sonore au dialogue implique une plus forte interactivité des éléments qui composent l'objet à comprendre. Or, cette interactivité implique un traitement simultané de chacun de ses

composants pour l'élaboration d'une signification, et ce traitement simultané implique à son tour une augmentation de la charge en Mémoire de Travail, et donc une augmentation de la charge cognitive (Chanquoy *et al.*, 2007). Par ailleurs, le traitement de sources d'informations multiples (linguistiques et non linguistiques) mais passant par le même canal (auditif) impliquera une étape d'intégration mentale qui *a priori* donnera lieu à une augmentation de la charge cognitive. De plus, cette charge peut également être augmentée par la production d'inférences de la part du sujet compreneur et par le fait que l'ensemble des supports qui composent l'objet à comprendre reste au niveau auditif. Dit autrement, une seule modalité est en jeu dans la perception et la compréhension du support, entraînant une proximité des traitements engagés. Cette proximité peut également être à l'origine d'une augmentation de la charge cognitive impliquée dans la tâche. Cet aspect est à rapprocher de l'effet de redondance. Cet effet intervient lorsque l'ajout d'une information (compréhensible isolément) interfère avec le premier traitement, impliquant ainsi une augmentation de la charge cognitive, liée à l'augmentation des ressources nécessaires en mémoire de travail (Chanquoy *et al.*, 2007; Tricot, 1998).

Pour confirmer cette hypothèse nous procéderons à un test U de Mann-Whitney à partir des temps de réponse des participants francophones et apprenants débutants ayant entendu le dialogue seul (condition de contrôle) et accompagné d'un contexte sonore environnemental (condition DCS). Ce test portera à la fois sur les résultats globaux (pour l'ensemble des trois dialogues) et pour chaque dialogue pris isolément.





## Partie 2 : Élaboration du matériel expérimental

Pour l'*expérimentation principale* de ce travail (voir p.165) et dont les *hypothèses* ont été présentées p.105) nous nous sommes fixé comme objectif de contrôler en profondeur notre matériel expérimental. Dans ce cadre, nous avons construit l'ensemble des *stimuli* présent dans le Classe-Images *i.e.*, les scènes sonores environnementales ainsi que les dialogues langagiers qu'elles accompagnent. Ces élaborations reposent sur une démarche expérimentale dont le but était à la fois le contrôle (connaissance et maîtrise) des supports de test et le questionnement empirique de l'élaboration de matériel pédagogique. L'ensemble de ces matériaux est disponible sur le CD joint à ce manuscrit.

Dans cette partie, les protocoles de constitution du matériel auditif environnemental et linguistique seront présentés de manière détaillée. Comme nous le verrons, l'ensemble du matériel a été réalisé et évalué dans le cadre de procédures expérimentales.



# **Premier Chapitre : Élaboration du matériel sonore environnemental**

Nous appelons matériel auditif expérimental les scènes sonores qui occuperont une position contextuelle (par rapport aux dialogues) dans l'expérimentation centrale. La construction du matériel auditif environnemental a fait l'objet d'un travail minutieux allant du recueil de données verbales sur les attentes auditives (voir p.121) à l'évaluation (identification) (voir p.138) des scènes sonores réalisées (voir p.132).

En outre, notre démarche de réalisation s'est appuyée sur les entretiens que nous avons menés auprès des professionnels de Radio France Internationale (en mars 2011) et les conseils, issus de leur pratique professionnelle, qu'ils ont pu nous donner.

## **1. Une étude sur les attentes auditives**

Dans la perspective d'une perception onomasiologique (Carlson, 2009; Gaillard, 2000; Magnen, 2009), l'activité de perception est dirigée par les attentes des individus dans une situation donnée. Ces attentes sont à leur tour le fruit de l'expérience antérieure de l'individu (tant au niveau des apprentissages déclaratifs que des acquisitions liées aux expériences sensibles). Dans ce cadre, les scènes sonores environnementales que nous avons construites à partir de l'étude des attentes auditives seront d'autant plus facilement compréhensibles et comprises qu'elles se rapprochent de la représentation cognitive qu'en ont les individus. En effet, l'élaboration des scènes sonores repose sur une première enquête sur les attentes auditives dont les résultats (listes d'items sources cités) sont à la base des enregistrements ultérieurs.

## 1.1. Méthodologie

### 1.1.1. Objectif, cadre théorique et hypothèses

La constitution de notre matériel sonore environnemental repose sur une étude qui tend à lister les sources à enregistrer prioritairement dans la « *mise en ondes* » de différents lieux.

Alors que la majorité des études sur les *soundscape*s (Schafer, 1977) s'interroge sur la perception des objets du monde extérieur et leur traduction conceptuelle par les individus **en situation**, *i.e.* en présence du stimulus (Bregman, 1990, 1994, 2008; Dubois, 2003; Guyot, Castellengo, Fabre, 1997; Houix, 2003; Niessen, 2009; Niessen, Maanen, & Andringa, 2008; Shamma, Elhilali, & Micheyl, 2011; Shamma & Micheyl, 2010 par exemple), cette étude vise la mise en lumière des éléments constitutifs des représentations auditives que des individus ont de différents lieux **en dehors de toute écoute**. En d'autres termes, il ne s'agit pas d'interroger les participants sur ce qu'ils entendent effectivement mais sur ce qu'ils s'attendent à entendre. Ainsi, 23 participants ont été interrogés sur leurs attentes auditives dans deux lieux cibles : une *cuisine* (lieu clos et privé) et une *terrasse de café* (lieu ouvert et public) ; et deux lieux distracteurs : une *forêt* et un *hall de gare*).

A la suite des différents travaux présentés dans le cadre théorique (notamment Rosch, 1975; Denis & Dubois, 1976, Dubois, 1993, 2000, 2003), trois hypothèses président à cette première étude : 1) les individus disposent d'une représentation mentale *auditive* des lieux de la vie quotidienne, 2) les représentations cognitives des individus sont accessibles à travers le langage, et 3) ces représentations sont fortement marquées culturellement (ce qui implique la présence d'éléments communs à une majorité de la population si celle-ci partage la même culture).

La première hypothèse nous a amenée à sélectionner des lieux de la vie quotidienne des Français que nous avons interrogés. Par conséquent le choix des lieux (*cuisine* et *terrasse de café*) s'est fait en fonction d'un double objectif : répondre aux exigences de l'élaboration de supports d'enseignement du FLE à des apprenants débutants (voir *Contexte didactique* p.17), et impliquer des situations assez fréquentes pour donner lieu à des représentations ancrées, acquises et familières pour les participants à cette première étude.

La deuxième hypothèse nous amène à considérer le discours des participants comme un moyen d'appréhender leurs représentations cognitives. On trouve dans la

littérature des références nous permettant d'étayer cette position (Dubois, 1993, 2003; Niessen, Cance, & Dubois, 2010). Toutefois, ce paradigme nous amène à nous interroger sur le concept de langue (Resche-Rigon 2009). Les discours recueillis sont-ils une représentation fidèle d'une réalité « objective » et appréhendable, ou une construction par l'individu d'une représentation (nécessairement limitante) du monde perçu (Schaff, 1969:197) ?

Enfin, la troisième hypothèse implique que certains éléments seront communs aux discours des différents participants interrogés (puisqu'ils ont tous une culture française métropolitaine). C'est en priorité sur ces éléments que nous appuierons notre élaboration des scènes sonores environnementales.

Selon nous, le discours des participants nous renseigne sur les représentations qu'ils ont des espaces sonores, et nous considérerons les éléments récurrents chez une majorité de participants comme des indices du stéréotype (Rosch & Mervis, 1975) de ces espaces. Ainsi, cette enquête vise la constitution de listes de sources sonores fréquemment citées par les participants. Dans ce cadre, les sources seront à enregistrer prioritairement dans l'élaboration des scènes sonores.

### ***1.1.2. Participants et déroulement***

Vingt-trois participants francophones, âgés de 20 à 65 ans, résidant en France métropolitaine depuis au moins 15 ans ont été interrogés. Ce panel est composé de 9 personnes ayant une pratique professionnelle de la manipulation du son (réalisateurs radio, journalistes radio, Techniciens Chargés de Réalisation (TCR) radio, Ingénieur du son, documentaliste de presse radio et psycho acousticien) et de 14 personnes n'ayant pas de rapports professionnels avec les enjeux de l'enquête (pour plus de précisions, se reporter aux annexes). Dans la retranscription des entretiens (voir CD en annexes) l'interviewer (le même pour l'ensemble des entretiens) est noté INT et les interviewés P + chiffre (de 1 à 14 pour les participants n'ayant pas de pratique professionnelle du son et de 101 à 109 pour les professionnels). Les données personnelles des participants ne sont pas présentées, ni dans la retranscription ni dans les enregistrements fournis (voir CD). Ces éléments sont néanmoins présentés dans un tableau synoptique (voir CD en annexes). La moyenne d'âge des 23 participants (11 hommes et 12 femmes) est de 37 ans (33 en médiane). Les participants résident principalement en milieu urbain ou péri-urbain (Paris, Lyon, Toulouse et leurs alentours).

Les participants ont été interrogés dans le cadre d'entretiens oraux semi-dirigés. Ceux-ci se sont déroulés dans différents lieux : le lieu de travail pour certains participants, c'est-à-dire dans les locaux de Radio France Internationale (à l'époque à la Maison de la Radio, à Paris) ou à l'Université de Toulouse le Mirail (désormais UTM) ; ou à leur domicile, entre janvier et mars 2011.

Quatre situations (une *cuisine* et une *terrasse de café* pour les lieux cibles et une *forêt* et un *hall de gare* pour les distracteurs) ont été présentées aux participants, en reprenant, à chaque fois, la même procédure de questionnement. Le protocole de l'entretien a été reproduit pour chaque situation :

- Un questionnaire préalable de renseignement des participants (âge, activité professionnelle, lieu de résidence...)
- Une mise en situation (*Vous êtes à/dans... Qu'est-ce que vous entendez ?*)
- Un discours du participant

Aucune variation de la consigne n'est intervenue, mais l'ordre de présentation des lieux a varié de manière à pondérer les effets « d'ordre » et d'amorçage.

Les entretiens se sont déroulés sans limite de temps ni de quantité de sources évoquées. Les participants ont ainsi pu construire leur discours, soit en faisant des pauses, soit en procédant à des retours en arrière par exemple. Certaines relances ont été introduites dans l'entretien sous la forme « *autre chose ?* » ou encore « *c'est tout ?* ». Enfin, l'interviewer a parfois demandé aux participants d'explicitier ou de spécifier certains termes afin de désambiguïser certains éléments du discours sans toutefois que ces parties n'aient été prises en compte dans l'analyse du discours (potentielle influence de l'interviewer sur le discours du participant (désirabilité sociale) Mayer & Tormala, 2010).

### ***1.1.3. L'encodage des données verbales***

#### **La norme ICOR**

L'oral se définit par sa linéarité et la dimension temporelle y prend une importance forte. En effet, "on ne peut pas étudier l'oral « par l'oral », en se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle, parcourir l'oral en tous sens et en comparer les morceaux." (Blanche-Benveniste, 2010:33)

Les entretiens ont donc fait l'objet d'une retranscription écrite à partir de laquelle nous avons mené nos analyses. Nous avons choisi le recours à la norme ICOR, proposée par le laboratoire ICAR<sup>44</sup>. Cette norme est largement utilisée et répandue dans l'étude de l'oral et, dans la mesure où nous envisageons de nouvelles études de ce corpus, il nous a semblé intéressant de nous rapprocher, dès le début de cette étude, d'une norme établie et partagée. Toutefois, comme le suggèrent les auteurs de la norme<sup>45</sup>, nous l'avons modifiée afin de l'adapter au mieux à nos besoins et objectifs premiers : les encodages ont été sciemment restreints en ne retranscrivant pas les aspects prosodiques par exemple. En outre, nous invitons le lecteur à écouter les enregistrements (proposés sur le CD annexe ainsi que les modifications apportées à la norme).

### **L'extraction des données**

À partir des retranscriptions, le contenu des entretiens a été organisé en différentes catégories. *Questions ouvertes* (*qu'est-ce que vous entendez ?*) et *questions fermées* (questionnaire final) ont été séparés pour centrer l'analyse sur les premières. Celles-ci ont, par la suite, fait l'objet d'un nouveau travail de catégorisation qui a mené à la définition de six classes : la *Mise en situation* (qui correspond à la consigne), le *Discours* (qui correspond à la partie de l'entretien à analyser dans la recherche d'identification des objets et actions sonores), *l'Influence du milieu* (qui correspond au commentaire lié aux conditions spatio-temporelles de l'entretien), *l'Ajustement discursif* (qui correspond à l'ajustement des interlocuteurs sur les paramètres de l'entretien), la *Relance*, et la *Médiation Énonciative* (qui correspond à la co-construction de l'objet en interaction, *i e* une proposition conjointe de l'item lexical, une invitation à la reformulation).

Les éléments constitutifs de la classe *Discours* ont à leur tour été traités et classés en huit catégories grammaticales : les *Énoncés introductifs* (*modus pré-dictum*), les *Substantifs complément d'objet directe* (objets sonores), les *Compléments du nom* (qui se rapportent au substantif), les *Adjectifs épithètes* (qui se rapportent au substantif), les *Propositions participiales ou infinitives* (qui se rapportent le plus souvent au substantif et qui portent l'action sonore), les *Propositions relatives* (qui se rapportent le plus souvent au substantif et qui portent l'action sonore), les *Compléments circonstanciels* (qui caractérisent généralement les conditions d'apparition de l'objet sonore) et enfin les

---

<sup>44</sup> UMR 5191 - CNRS - Lyon 2 - ENS-LSH

<sup>45</sup> [http://icar.univ-lyon2.fr/projets/corinte/bandeau\\_droit/convention\\_icor.htm](http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/convention_icor.htm)

*Modus post-dictum* (qui portent la modalisation). Cette démarche a en outre permis de quantifier les différents types de développement, du *dictum* et du *modus*.

Il est apparu que les éléments de la catégorie *Substantif complément d'objet direct* répondait à notre question de départ<sup>46</sup>.

Ainsi, dans l'exemple ci-dessous, nous pouvons voir que le huitième objet cité par le participant 7 lors de l'évocation de *Cuisine* présente l'objet *bruit*, accompagné de deux développements du *dictum* (l'un placé en *Complément du nom* et l'autre en *Complément Circonstanciel*) et d'une proposition de *modus* portée par l'*Énoncé introductif*.

De cette manière, nous avons pu observer et comptabiliser les développements de chaque item (*substantif*, objet).

## 1.2. Résultats

L'enquête sur les attentes auditives a pour objectif la constitution de listes de sources sonores à enregistrer en priorité lors de la réalisation des scènes sonores. Dans ce cadre, cette enquête a donné lieu à la constitution d'un corpus. Ce dernier a fait l'objet, dans un premier temps d'une analyse formelle. Cette analyse nous a permis d'observer que les éléments de réponse que nous cherchions (quels éléments enregistrer pour réaliser des scènes sonores étaient, de manière quasi-systématique, en position de complément d'objet direct, COD).

Les listes de sources à enregistrer ont été établies à partir des substantifs COD présents dans le discours des participants (voir p.126). Pour l'évocation *cuisine* les participants ont donné 158 items COD. Pour l'évocation *terrasse de café*, les participants ont donné 151 items COD.

Des listes thématiques et hiérarchiques de ces items ont été construites et organisées de manière à obtenir des ensembles d'items représentatifs (par leur fréquence d'apparition) et fonctionnels (par le petit nombre de listes/catégories). Comme nous le verrons, ces catégories sont à leur tour subdivisées en différents ensembles.

---

<sup>46</sup> Quelles sources enregistrer pour construire des scènes sonores proches des représentations des individus.

### 1.2.1. La cuisine

De manière générale, la situation *cuisine* a donné lieu à six catégories d'items : les *aliments* (leur cuisson, et la préparation des repas), la *présence humaine*, les *médias* (télévision ou radio), l'*eau*, les *ustensiles* (la vaisselle et les couverts), les *électroménagers*.

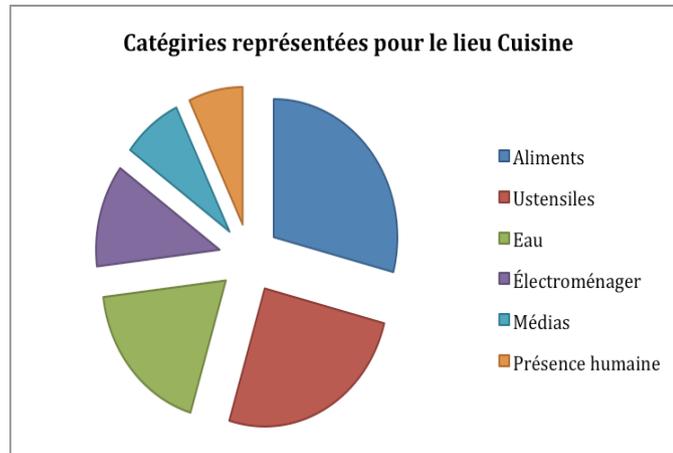


Figure 5 : Catégories d'items construites pour le lieu cuisine

	Nombre d'items	Pourcentage
Aliments	41	25,94
Ustensiles	36	22,78
Eau	26	16,45
Électroménagers	17	10,75
Médias	11	6,96
Présence humaine	10	6,32

Tableau 1 : Catégories d'items construites pour le lieu *cuisine*

Ces catégories ne rassemblent pas tous les objets cités par les participants : onze items non pas été classés (les *violences domestiques*, le *chat*, le *grouillement*, les *bruits de mâchoire*, une *onomatopée*, les *placards*, un *ronflement*, le *tic tac*, la *cafetière*, les *métaux*, les *oiseaux*, la *pendule*, la *porte de placard*, le *bruit*). Ces items, par leur faible représentativité (isolement catégoriel au sein du corpus), ne seront pas pris en compte dans l'élaboration des constructions sonores.

Enfin, chaque catégorie établie peut être décomposée en sous catégories.

Aliments (cuisson, préparation)

Cette catégorie apparaît comme la plus fréquente dans notre corpus. Elle peut à son tour se subdiviser en trois sous catégories :

- La cuisson (17 occurrences)
- Les aliments (13 occurrences)
- La préparation des aliments (11 occurrences).

Ustensiles (dont vaisselle et couverts)

Sur l'ensemble du panel interrogé, la catégorie *ustensiles* (dont vaisselle et couverts) apparaît très fréquemment. Quatre sous catégories apparaissent :

- La vaisselle (15 occurrences)
- Les casseroles et plats (10 occurrences)
- Les couverts (7 occurrences)
- Les ustensiles (4 occurrences)

Eau

La troisième catégorie la plus présente est celle que nous avons nommée *eau*. Cette catégorie regroupe à son tour quatre sous catégories :

- L'eau (lorsque le substantif utilisé par les locuteurs est *eau*, quel que soit le développement) (19 occurrences)
- La vaisselle (lavage) (3 occurrences)
- Le robinet (métonymie) (2 occurrences)
- La bouilloire (métonymie) (2 occurrences)

Électroménagers

Les *électroménagers* représentent la quatrième catégorie de l'évocation *cuisine*. Elle présente à son tour dix sous-catégories dont la majorité ne présente qu'une seule occurrence :

- Le frigo (6 occurrences)
- Le chauffe-eau (2 occurrences)
- Le four (2 occurrences)
- La hotte (2 occurrences)

- Le lave-vaisselle (1 occurrence)
- La machine à laver (1 occurrence)
- La machine à café (1 occurrence)
- Le mixeur (1 occurrence)
- Les électroménagers (1 occurrence)
- La hotte (1 occurrence)

#### Médias

Les *médias* représentent la cinquième catégorie de l'évocation *cuisine*. Elle regroupe trois sous catégories :

- La radio (7 occurrences)
- La télévision (2 occurrences)
- La musique (2 occurrences)

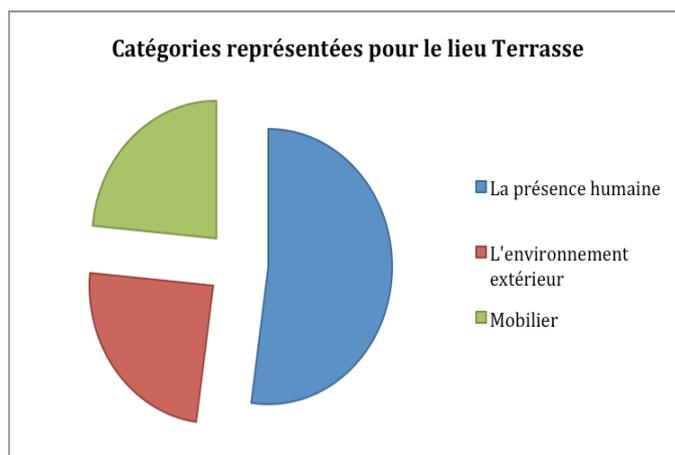
#### Présence humaine

La *présence humaine* représente la dernière catégorie de l'évocation *cuisine*, c'est-à-dire qu'elle est la moins fréquente dans le discours des participants. Comme les autres, elle se subdivise en sous-catégories :

- Les productions vocales (5 occurrences)
- Les personnes (4 occurrences)
- Les parties du corps (1 occurrence)

### **1.2.2. La terrasse de café**

De manière générale, la situation *terrasse de café* a donné lieu à trois grandes catégories d'items : la présence humaine, le mobilier (dont la vaisselle et les couverts), et enfin l'environnement extérieur.



**Figure 6 : Catégories d'items construites pour le lieu *Terrasse de café***

	Nombre d'items	Pourcentage
La présence humaine	71	47,01
L'environnement extérieur	33	21,85
Mobilier	32	21,19

**Tableau 2 : Catégories d'items construites pour la terrasse de café**

Encore une fois, ces catégories ne rassemblent pas tous les objets cités par les participants. quinze items n'ont pas été classés (*bar, brouhaha, chose, espace, monnaie, musique, ambiance, moment, problème, truc, téléphone, liquide*). Ces items, par leur faible représentativité (isolement catégoriel au sein du corpus), ne seront pas pris en compte dans l'élaboration des constructions sonores.

Tout comme pour la situation *cuisine*, chaque catégorie peut être décomposée en sous-catégories.

*Présence humaine*

La Présence Humaine représente, en termes de fréquence, la première catégorie de l'évocation *terrasse de café*. Cette catégorie se subdivise en six sous-catégories :

- Les personnes (30 occurrences)
- Les productions vocales (27 occurrences)
- Le serveur (8 occurrences)
- Les pas (3 occurrences)

- Les cigarettes (2 occurrences)
- Une onomatopée (1 occurrence)

Environnement extérieur

L'environnement extérieur représente, en termes de fréquence, la deuxième catégorie de l'évocation « terrasse de café». Cette catégorie se subdivise en neuf sous-catégories :

- La circulation automobile (17 occurrences)
- La rue (6 occurrences)
- La ville (4 occurrences)
- Les arbres (1 occurrence)
- Un avion (1 occurrence)
- L'environnement (1 occurrence)
- L'extérieur (1 occurrence)
- La nature (1 occurrence)
- Une sonnette (1 occurrence)

Mobilier (dont vaisselle et couverts)

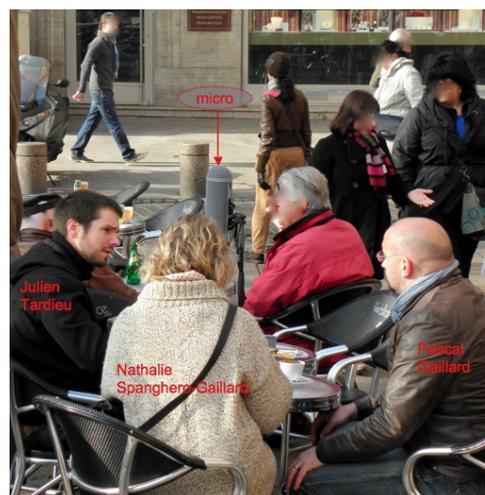
Le Mobilier (dont la vaisselle et les couverts) représente, en termes de fréquence, la dernière catégorie de l'évocation « terrasse de café». Cette catégorie se subdivise en neuf sous-catégories :

- Les verres (14 occurrences)
- Les tasses (5 occurrences)
- Les couverts (5 occurrences)
- Les chaises (2 occurrences)
- Le percolateur (2 occurrences)
- Les glaçons (1 occurrence)
- La paille (1 occurrence)
- Le plancher (1 occurrence)
- La vaisselle (1 occurrence)



La terrasse du « Wallace » se trouve au cœur d'une zone commerciale de la ville. Le passage est principalement pédestre, même si quelques voitures passent sur une rue pavée adjacente, à sens unique. Quelques arbres se trouvent sur la place. Les enregistrements ont été effectués en fin de matinée et malgré la saison, nous avons constaté l'activité de ce lieu.

**Image 3 : Prise de sons au Wallace, pour la terrasse de café**



Le « Florida » se situe place du Capitole, où se trouve la Mairie de la ville de Toulouse. La zone est passante, les espaces sont grands. Des parasols recouvrent les terrasses qui se situent le long de la place, juste derrière une rue pavée, à sens unique qui, encercle la place. Les enregistrements se sont déroulés en début d'après-midi, à l'heure du café.

**Image 4 : Prise de sons au Florida pour la terrasse de café**

La séance d'enregistrement s'est conclue par une prise de sons, rue St Rome afin d'enregistrer des sons de passage à pied, cités dans les questionnaires. La rue St Rome est une rue commerçante et piétonne, au cœur du centre ville, très empruntée. Ces enregistrements se sont déroulés en début d'après-midi.

**Image 5 : Prise de sons rue St Rome, pour la terrasse de café**



La deuxième séance<sup>50</sup> a été consacrée aux enregistrements des éléments attendus dans un contexte de *cuisine*. Ces prises de sons ont eu lieu dans une cuisine privée, au domicile de Nathalie Spanghero-Gaillard et Pascal Gaillard.



**Image 6 : Prise de sons dans un domicile privé, pour la *cuisine***

Au total quarante-trois prises de sons ont été effectuées (seize pour la *terrasse de café* et vingt-sept pour la *cuisine*), visant à rendre compte des sources citées précédemment. L'ensemble de ces pistes (disponibles sur le CD en annexe) a été décrit et étiqueté (description et étiquetage disponibles dans le CD en annexe).

### **2.3. Montage et manipulations du matériau sonore**

Plusieurs scènes sonores ont été construites pour chaque lieu. Pour la situation *cuisine*, deux *histoires sonores* ont été construites : *la préparation d'un repas* (dont nous avons réalisé trois versions) et *la vaisselle* (dont quatre versions ont été réalisées). Pour la situation *terrasse de café* un seul *paysage sonore*<sup>10</sup> a été construit, dont six versions ont été réalisées.

#### **Les constantes**

##### *La durée*

La durée des scènes sonores a été calibrée à 15 secondes. Cette durée répond à un double objectif : d'une part elle est assez longue pour permettre l'insertion d'un court

---

<sup>50</sup> le 23/01/2012

dialogue enregistré (voir chapitre *Élaboration du matériel linguistique* p.151) ; d'autre part elle est assez courte pour permettre des tests « acceptables » pour les sujets.

### La représentativité des sources

Bien que le nombre de sources, ou de séquences, corresponde à une des variables du test, leur sélection n'est pas fortuite mais résulte de l'étude des attentes auditives. Il s'est donc agi d'enregistrer en priorité les sources les plus représentatives des attentes exprimées par les participants à l'étude.

### **Les facteurs de variation**

#### La cuisine : la préparation d'un repas

La situation de la *cuisine* a donné lieu à la réalisation de deux *histoires sonores*. La première donne à entendre **la préparation d'un repas**, la seconde **la vaisselle**.

Pour **la préparation d'un repas**, trois versions (appelées préparation-repas1/2/3) ont été construites en faisant varier le nombre de séquences /étapes constituant de l'histoire sonore.

**Préparation-repas-1** : « *Découpe + cuisson de l'oignon* » : la scène sonore présente deux plages juxtaposées : la découpe de l'oignon et sa cuisson. Rien ne vient aider leur articulation ou leur mise en contexte.

**Préparation-repas-2** : « *Découpe + gaz + cuisson* » : nous avons ajouté « l'allumage » du gaz entre les deux premières plages (*découpe* et *cuisson*). L'hypothèse que cette nouvelle séquence jouera un rôle d'articulateur.

**Préparation-repas-3** : « *Épluchage + découpage + gaz + cuisson* » : Une séquence d'*épluchage* a été ajoutée en préambule de la scène sonore *préparation-repas-2*.

#### La cuisine : la vaisselle

Pour **la vaisselle**, quatre versions de la scène sonore ont été construites (*lavage-vaisselle-1/2/3/4*), en faisant varier à la fois le nombre des séquences et leur ordre d'apparition. Deux séquences ont été manipulées pour ces montages : « *vaisselle* » qui correspond à des sons de vaisselle que l'on nettoie, pendant que l'eau coule du robinet et « *radio* ». Cette deuxième séquence met en scène la recherche d'une station sur un poste de radio analogique, puis le programme de la *bourse*.

**Vaisselle-1** : présente des sons d'eau et de vaisselle que l'on lave (séquence *vaisselle*).

**Vaisselle-2** : « *vaisselle + radio pendant* » les deux séquences (*vaisselle* et *radio*) apparaissent simultanément.

**Vaisselle-3** : « *vaisselle + radio après* » l'écoute de la *radio* apparaît après le début de la séquence *vaisselle*.

**Vaisselle-4** : « *vaisselle + radio avant* » l'écoute de la *radio* apparaît en préambule de la séquence *vaisselle*.

#### À la terrasse d'un café

La situation de la *terrasse de café* a donné lieu à la réalisation d'un seul *paysage sonore*, présentant six versions. Pour la construction de ces différentes versions, une méthodologie différente de celle utilisée pour la situation *cuisine* a été suivie. En effet, chaque version de la scène sonore se singularise par l'ordre d'apparition des séquences qui la composent. Trois séquences sont présentes dans l'ensemble des six versions de la scène sonore : les bruits de *verres et de glaçons*, les bruits de *pas* et les bruits de *voiture*.

L'ensemble de ces sons semble immergé dans un *brouhaha* correspondant aux sons de conversation indistincts (majoritaires) mêlés aux autres sons « de la rue » et « de la ville ». Ces éléments amorphes représentent un aspect particulièrement important du réalisme des scènes sonores.

**Terrasse-1** : « *Verres + glaçons* » + « *Pas* » + « *Voiture* »

**Terrasse-2** : « *Verres + glaçons* » + « *Voiture* » + « *Pas* »

**Terrasse-3** : « *Pas* » + « *Verres + glaçons* » + « *Voiture* »

**Terrasse-4** : « *Pas* » + « *Voiture* » + « *Verres + glaçons* »

**Terrasse-5** : « *Voiture* » + « *Pas* » + « *Verres + glaçons* »

**Terrasse-6** : « *Voiture* » + « *Verres + glaçons* » + « *Pas* »

#### **Remarques empiriques**

La *cuisine* et la *terrasse de café* n'ont pas permis les mêmes prises de sons, ni les mêmes méthodologies de montage.

Les enregistrements de la *cuisine* ont permis d'isoler chaque phénomène acoustique associé à une source avant de procéder au montage. En effet, les enregistrements ont été faits dans un milieu silencieux maîtrisé, sans chevauchement des séquences. La réalisation des scènes sonores de la *cuisine* est donc passée par une

étape de rédaction de *scénarii*. Ceux-ci ont été rédigés de manière à faire apparaître les éléments cités par les participants à l'étude des attentes auditives (p.121). Deux *histoires* ont été construites, lors de la *préparation d'un repas* et lors du lavage de la *vaisselle*, mettant en scène les sources dans une narration cohérente. Dans ce cadre, ces scènes sonores présenteront un plus fort degré de *transcodage* (voir *Contexte Didactique* p.17) par rapports aux dialogues d'elles accompagneront. Dit autrement, les scènes sonores élaborées pour la *cuisine* rendent compte non seulement du lieu (dimension situationnelle) dans lequel se déroulent les scènes, mais aussi des actions qui y sont menées (dimension de *transcodage*).

Une démarche similaire n'a pas été possible avec la *terrasse de café*. En effet, les enregistrements se sont déroulés dans un contexte « écologique » *i.e.* à de réelles terrasses de café. Dans un milieu bruyé, il s'est avéré impossible de procéder à l'enregistrement des sources isolément (cf propos de P101). Ainsi, les enregistrement ont consisté en de longues prises de sons contenant plusieurs sources juxtaposées ou concomitantes, se chevauchant le plus souvent (la durée moyenne d'une piste de la *cuisine* est de 53sec. (médiane 46 sec.) contre 1min36sec. (médiane 1min52sec.) pour la *terrasse de café*). À la suite de ces enregistrements, un étiquetage a été mené mais, même *a posteriori* l'isolement des sources n'était pas compatible avec le niveau de qualité acoustique visé. La méthodologie de rédaction de *scénarii* (telle que suivie pour la cuisine) s'est donc avérée inopérante puisque l'association *a posteriori* d'éléments issus de pistes d'enregistrement distinctes aurait engendré des ruptures saillantes dans l'arrière plan *amorphe*. Dans ce cadre, le choix d'une piste contenant un nombre important de sources citées par les participants à l'enquête sur les attentes s'est imposé. C'est en retravaillant cette piste (suppression de parties, déplacement de séquences...) que nous avons respecté les constantes fixées pour le montage. Dans ce cadre, les scènes sonores que nous avons construites pour la *terrasse de café* sont uniquement *situationnelles*. En effet, elles rendent compte du lieu dans lequel se déroule la scène mais n'entretiennent pas de rapport de redondance avec le dialogue qui s'y inscrira (voir *Élaboration du matériel linguistique* p.151).

Enfin, lors du montage, des différences ont été introduites dans les niveaux d'intensité de sortie des scènes sonores afin que chacune d'elles soit audible dans des conditions d'écoute en classe. Or, ces variations ne rendent pas bien compte des différences d'intensités physiques observées dans les situations effectives d'écoute. En effet, il s'est agi de choisir entre respect de l'authenticité (et des mesures physiques écologiques, « être vrai ») et réalisation de scènes sonores correspondant aux représentations des individus (manipulation des paramètres physiques pour « faire

vrai»). En outre, l'intensité est plus « *une propriété qu'une dimension* » (Dubois, 2000:49). Nous pouvons donc penser que le système cognitif de l'auditeur ré-agencera les séquences perçues de manière à reconstruire un tout cohérent avec ses représentations mentales.

### **3. Identification des scènes sonores**

Une fois réalisées, toutes les versions de chaque scène sonore ont été soumises deux test : un test d'*identification* (les scènes sonores permettent-elles effectivement la reconnaissance des lieux mis en ondes ?) et un protocole de *sélection* (parmi toutes les versions d'une même scène sonore, laquelle est les plus représentative du lieu mis en ondes ?).

#### **3.1. Méthodologie**

##### **3.1.1. Objectifs et hypothèses**

L'identification est ici définie comme le processus qui amène l'individu à considérer un stimulus perçu comme correspondant à une catégorie sémantique générale, préalablement identifiée (Le Ny, 2005).

Comme le note Vanderveer, (1979) en tâche de caractérisation libre, les auditeurs parlent plutôt de l'« *événement à l'origine du son* », ce qui nous laisse penser que les participants, en tâche d'identification, aborderont de préférence les événements et les sources.

Le test proposé correspond à une tâche d'identification du lieu. Les participants ont été amenés à identifier le lieu dans lequel se déroule la scène sonore, toute autre réponse étant considérée comme *non-réponse*. De cette manière, nous nous sommes assurée que la méthodologie de construction des scènes sonores suivie permet effectivement de mettre en ondes les lieux visés.

##### **3.1.2. Participants et déroulement**

###### **Consignes et déroulement**

Le test se déroule en deux étapes. Dans un premier temps, une phase d'entraînement a permis aux participants de se familiariser avec la tâche. Cette phase se compose de trois *sons distracteurs* (présentés plus bas). Dans un deuxième temps, les

participants ont procédé à la phase de test, composée des *sons cibles* et de *sons distracteurs*.

Le support audio du test a été paramétré en amont de manière à garantir l'homogénéité pour l'ensemble des participants. Les étapes du test sont :

- un signal sonore (*bip*, 1,25 secondes) qui prévient l'auditeur du début de la scène sonore,
- la diffusion de la scène sonore (15 secondes)
- un temps de silence de 8 secondes qui permet au participant de répondre par écrit.

L'ensemble des enregistrements, signaux sonores et temps de réponse ont été agencés au sein d'une seule piste afin d'obtenir un support de test (sonore) fixe et stable.

Phase	séquence	Nombre répétitions séquence	Durée séquence (en secondes)	Durée totale	Durée totale
entraînement	bip	4	1,25	74	705,75 secondes (≈ 11min45sec)
	scène sonore	3	15		
	temps de réponse	3	8		
test	bip	27	1,25	631,75	
	scène sonore	26	15		
	temps de réponse	26	8		

**Tableau 3 : Récalitulatif des signaux sonores pour le test d'identification des lieux**

De cette manière, aucune manipulation de l'expérimentateur n'intervient pendant le déroulement du test.

### Les distracteurs

Différents « *sons distracteurs* » ont été créés : les scènes sonores construites contenant intrinsèquement des éléments de comparaison (au niveau des séquences qui les composent notamment), elles ont été mêlées à d'autres scènes afin que la comparaison des différents montages cibles soit moins immédiate.

Treize « *sons distracteurs* » ont été construits, à partir d'enregistrements réalisés par l'intermédiaire de la plateforme PETRA<sup>51</sup>. Ces sons ont été enregistrés dans différentes *gares SNCF*<sup>52</sup>, ainsi que dans un *salon d'exposition*<sup>53</sup>.

Le test d'identification est donc composé de treize sons cibles (sept pour la *cuisine* et six pour la *terrasse de café*) et de treize sons distracteurs (sept pour la *gare SNCF* et six pour le *salon d'exposition*).

Enfin, la phase d'entraînement a été élaborée à partir de sons téléchargés d'Internet<sup>54</sup> qui ne présentent malheureusement pas la même qualité que les sons de la phase de test. Trois scènes sonores ont été sélectionnées pour cette phase.

***Son-distracteur-1*** : « la campagne »

***Son-distracteur-2*** : « un marché »

***Son-distracteur-3*** : « la mer »

## Participants

29 participants français (francophones natifs) ont effectué le test. Différents critères d'inclusion ont participé à leur sélection : l'ensemble des participants a entre 18 et 25 ans (l'âge moyen est de 20 ans, la médiane des âges est de 19 ans), ils sont tous universitaires (même si leur niveau varie de la première année de licence au doctorat). Enfin, aucun problème d'audition n'a été renseigné.

## 3.2. Résultats

Afin d'encoder et d'étiqueter les réponses recueillies dans le test d'identification<sup>55</sup>, l'ensemble des réponses a été organisé en différentes catégories. Toutefois, les différences de nature des sons de la *cuisine* et de la *terrasse de café* nous ont amenée à élaborer différentes classifications en fonction du lieu donné à entendre. Rappelons que ces différences de nature ont impliqué des traitements différents à toutes les étapes de ce travail, des prises de sons jusqu'à l'analyse des données recueillies.

---

<sup>51</sup> <http://petra.univ-tlse2.fr/>

<sup>52</sup> réalisées dans le cadre de travaux de recherche sur l'identification des espaces et leurs « *signatures sonores* ».

<sup>53</sup> étude exploratoire e partenariat avec GL-Events, organisateur de salons <http://www.gl-events.com/>

<sup>54</sup> <http://www.sound-fishing.net/> et <http://www.universal-soundbank.com/ambiances.htm>

<sup>55</sup> Les scènes sonores construites amènent-elle à la reconnaissance du lieu mis en ondes ?

### 3.2.1. La situation Cuisine

#### Étiquetage des réponses

Pour la situation *cuisine*, les réponses obtenues ont donné lieu à cinq catégories.

- *Réponse consensuelle simple* : cette catégorie rassemble les réponses présentant le terme « *cuisine* » avec, au maximum, une structure grammaticale introduisant le terme (ex : *dans une cuisine*).

- *Réponse consensuelle détaillée* : cette catégorie rassemble les réponses contenant la *réponse consensuelle simple*, avec une précision supplémentaire (ex : *dans une cuisine, devant le plan de travail*).<sup>56</sup>

- *Réponse englobante* : qui rassemble les réponses qui relèvent d'une catégorie plus générale, mais à laquelle appartient la réponse attendue (ex : *maison*). Dit autrement, interviennent ici les hyperonymes et les termes génériques.

- *Réponse autre* : qui rassemble les réponses qui font état d'un lieu différent de celui dans lequel nous avons effectué les prises de sons (ex : *salle de bain*)

- *Non-Réponse* : qui rassemble les réponses qui n'apportent pas de renseignement sur le lieu (*i.e.* qui ne répondent pas à la question posée, ex : *en train de faire la vaisselle*)) ou les absences de réponses.

---

<sup>56</sup> Nous les considérons « consensuelles » dans la mesure où ces réponses correspondent à ce que nous attendions (attentes terminologiques), le consensus se faisant alors entre les participants et nous-même.

### Résultats pour *la préparation d'un repas*

Les scènes sonores de *la préparation d'un repas* ont globalement amené les participants à identifier le lieu cuisine. En effet, les réponses consensuelles et englobantes représentent 92% des réponses recueillies pour l'ensemble des scènes sonores de *la préparation d'un repas*.

	nombre d'occurrences		pourcentages	
réponses consensuelles simples	76	80	87%	92%
réponses consensuelles détaillées	4		5%	
réponses englobantes	0		0%	
autre réponse	6	7	7%	8%
non-réponse	1		1%	

**Tableau 4 : Catégorie de réponses au test d'identification pour la scène sonore *la préparation d'un repas* (global)**

Il semble donc que, dans l'ensemble, les scènes sonores relatives à *la préparation d'un repas* aient entraîné une bonne identification des lieux. Toutefois, les scènes sonores *préparation-repas-1* et *2* sont celles qui totalisent le plus grand nombre de réponses consensuelles.

	préparation-repas-1	préparation-repas-2	préparation-repas-3
Autre réponse	1	1	4
Réponse Consensuelle Simple	27	26	23
Réponse consensuelle détaillée	1	2	1
Non réponse	0	0	1
réponse englobante	0	0	0

**Tableau 5 : Répartition des réponses au test d'identification pour la *préparation d'un repas***

Toutefois, ces écarts ne sont pas statistiquement significatifs.

### Résultats pour *la vaisselle*

Les scènes sonores mettant en scène *la vaisselle* ont globalement été bien identifiées aussi.

	nombre d'occurrences		pourcentages	
	réponses consensuelles simples	75	80	65%
réponses consensuelles détaillées	27	23%		
réponses englobantes	2	2%		
autre réponse	6	7	5%	10%
non-réponse	5		4%	

**Tableau 6: Catégorie de réponses au test d'identification pour la scène sonore *la vaisselle* (global)**

Il apparaît en effet que les scènes sonores *lavage-vaisselle- 3et 4* sont celles qui totalisent le plus de *réponses consensuelles*.

	lavage-vaisselle-1	lavage-vaisselle-2	lavage-vaisselle-3	lavage-vaisselle-4
Autre réponse	1	2	1	2
Réponse Consensuelle Simple	20	15	20	20
Réponse consensuelle détaillée	6	6	8	7
Non réponse	2	3	0	0
réponse englobante	0	2	0	0

**Tableau 7 : Répartition des réponses au test d'identification pour la vaisselle**

### 3.2.2. La situation *Terrasse de café*

#### Étiquetage des réponses

Pour la situation *terrasse de café*, les réponses obtenues ont donné lieu à six catégories.

- *Réponse consensuelle*: cette catégorie rassemble les réponses présentant les termes « *terrasse* » et « *café* » (ex : *à la terrasse d'un café*). Les synonymes de type « *bistrot* » ont été comptabilisés dans cette catégorie.<sup>57</sup>

- *Réponse partielle « café »*: cette catégorie rassemble les réponses faisant état des sèmes du café (« *Lieu public où l'on consomme des boissons* », *Le Grand Robert de la Langue Française*). Nous considérons ces réponses comme partielles parce qu'elles ne font pas état de l'ensemble des 2 sèmes (*terrasse* et *café*) qui composent la *terrasse de café* (ex : *dans un café*).

- *Réponse partielle « extérieur »*: cette catégorie rassemble les réponses faisant état de la dimension extérieure de la *terrasse de café*. Elles sont considérées comme partielles parce qu'elles ne font pas état des deux sèmes (ex : *dans une rue*).

- *Réponse relative*: cette catégorie rassemble les réponses rendant compte des deux dimensions « *café* » et « *extérieur* », sans pour autant présenter les termes attendus (ex : *terrasse de restaurant*)<sup>58</sup>.

- *Réponse autre*: qui rassemble les réponses qui font état d'un lieu différent de celui dans lequel nous avons effectué les prises de sons (ex : *à une caisse*)

- *Non-Réponse*: qui rassemble les réponses qui n'apportent pas directement de renseignement sur le lieu ou les absences de réponses (ex : *un serveur prenant les assiettes sales*).

#### Résultats

Tout comme les sons de *la cuisine*, les scènes sonores relatives au lieu *Terrasse de Café* ont été globalement identifiés. Le total des réponses consensuelles, partielles et relatives représente 97% des réponses.

---

<sup>57</sup> Ici encore, nous appelons « *consensuelles* » les réponses qui correspondent aux réponses que nous attendions (attentes terminologiques), le consensus se faisant alors entre les participants et nous.

<sup>58</sup> Le terme *relatif* renvoie à la fois au caractère partiel et au caractère de mise en relation des termes des réponses.

	nombre d'occurrences		pourcentages	
	réponses consensuelles	36	170	21%
réponses partielles café	95	55%		
réponses partielles extérieur	20	11%		
Réponses relatives	19	10%		
autres réponses	3	4	2%	3%
non-réponses	1		1%	

**Tableau 8 : Catégories de réponses au test d'identification pour la scène sonore la terrasse de café (global)**

Il apparaît donc que l'ensemble des scènes sonores construites pour la *terrasse de café* ait été bien identifié. Il apparaît également que peu de différences interviennent entre chaque version, d'après notre mode de calcul.

	terrasse-1	terrasse-2	terrasse-3	terrasse-4	terrasse-5	terrasse-6
Réponse consensuelle	5	9	7	5	4	6
Réponse partielle café	14	0	16	17	18	17
Réponse partielle extérieur	7	6	6	2	2	1
relative Café + extérieur	2	14	0	4	4	5
réponse autre	1	1	0	0	1	0
non réponse	0	0	0	1	0	0

**Tableau 9 : Répartition des réponses au test d'identification pour la terrasse de café**

Nous pouvons donc considérer que les scènes sonores que nous avons construites, pour la *cuisine* ou pour la *terrasse de café* ont globalement donné lieu à une bonne identification de l'espace mis en onde.

## 4. Sélection des scènes sonores

### 4.1. Méthodologie

#### 4.1.1. Objectifs et hypothèses

Dans la mesure où différentes versions de chaque scène sonore ont été construites, un protocole de sélection a été mis en place. Il vise le choix, par des participants, de la scène sonore la plus typique, c'est-à-dire la plus représentative de la catégorie (ici l'ensemble des versions d'une même scène sonore).

En effet, il semble que plus la scène sonore sera prototypique, plus les auditeurs seront en mesure de l'identifier aisément. Or, pour les tests finaux (test de compréhension de dialogues avec ou sans contexte sonore environnemental, voir p.165), il apparaît primordial que les scènes sonores présentent le moins d'ambiguïté possible.

#### 4.1.2. Participants et déroulement

Quatre hommes et dix-huit femmes, tous âgés de 22 à 35 ans (moyenne d'âge 28 ans, médiane 27,5 ans) ont participé à ce test. L'ensemble de ces participants est francophone natif et tous travaillent ou étudient à l'Université de Toulouse le Mirail (3 ATER, 11 doctorants en SHS, 4 étudiants entre la L3 et le M2, 2 ingénieurs d'étude et 2 maîtres de conférences). Aucun déficit auditif ne nous a été signalé.

Les participants ont successivement évalué les versions 1) de la *préparation d'un repas*, 2) de la *vaisselle* et 3) de la *terrasse de café*. « À l'intérieur » de chaque ensemble (scène sonore) les participants ont pu procéder à autant d'écoutes que désiré et dans un ordre non imposé.

Encore une fois, le test a débuté par une phase d'entraînement avec les scènes sonores relatives aux *gares SNCF*.

La consigne était systématiquement la même : « *Laquelle de ces scènes sonores vous semble la plus naturelle ?* », et les réponses des participants ont été recueillies par écrit (voir questionnaire en annexe sur le CD)

Le matériel utilisé pour cette sélection correspond à un ordinateur et un casque. Les « fichiers son » était accessibles sur le bureau de l'ordinateur (organisés par nom de catégorie *préparation-repas*, *vaisselle* et *terrasse de café*). Pour procéder aux écoutes, les

participants n'avaient qu'à sélectionner la piste et lancer la lecture en appuyant sur la barre d'espace du clavier. Ce procédé a été expliqué à chaque participant.

## 4.2. Résultats

Le test d'identification des lieux a montré que les scènes sonores construites permettent la reconnaissance du lieu mis en ondes. Toutefois, aucun écart statistiquement significatif n'a été observé et ce test ne permet donc pas de déterminer laquelle des versions de chaque scène sonore est la plus représentative.

### 4.2.1. La préparation d'un repas

Les scènes sonores mettant en scène la *préparation d'un repas* ont obtenu de bons scores. En effet, nous constatons une répartition quasiment égale du choix des participants :

<b>Préparation-repas-1</b> : « Découpe + cuisson de l'oignon »	6
<b>Préparation-repas-2</b> : « Découpe + gaz + cuisson »	8
<b>Préparation-repas-3</b> : « Épluchage + découpage + gaz + cuisson »	8

Les participants semblent avoir été *légèrement sensibles* à la présence de « l'allumage du gaz ». Si certains nous ont déclaré y voir une marque de réalisme, d'autres en revanche nous ont affirmé avoir été perturbés par le manque de réalisme : « *il est impossible que la poêle chauffe aussi vite* ». Deux *manières d'entendre* s'opposent donc mais la première semble plus répandue. En revanche, l'épluchage semble également avoir perturbé certains participants qui se sont arrêtés à tenter de découvrir la source (un oignon).

Dans ce cadre, la scène sonore retenue pour la suite des tests a été *préparation-repas-2* (Découpe + gaz + cuisson).

#### 4.2.2. La vaisselle

La situation *vaisselle* présente de plus forts écarts de score entre les différentes scènes :

<b>Lavage-vaisselle-1</b> : « lavage vaisselle »	6
<b>Lavage-vaisselle -2</b> : « vaisselle + radio pendant »	5
<b>Lavage-vaisselle -3</b> : « vaisselle + radio après »	2
<b>Lavage-vaisselle -4</b> : « vaisselle + radio avant »	9

Tout comme pour la *préparation d'un repas*, les échanges que nous avons eus avec les participants nous aident à identifier les éléments qui ont joué un rôle dans la sélection. Le *lavage-vaisselle-3*, qui a obtenu seulement 2 points de score de sélection, est souvent apparu aux participants comme non-naturel puisque impliquant l'intervention d'une deuxième personne (« *on n'allume pas la radio quand on a les mains mouillées* »). *Lavage-vaisselle-2* qui présente l'apparition simultanée des différents flux n'est pas apparu aussi « *non-naturel* » mais présente « *quelque chose d'étrange* » dans l'enchaînement des actions. Parallèlement, le *lavage-vaisselle-4* est apparu comme le plus naturel puisque présentant un « *ordre cohérent* » dans l'enchaînement des actions. Enfin, le *lavage-vaisselle-1*, par sa plus grande *simplicité* (présence des sons de vaisselle uniquement) a entraîné le choix de 6 participants.

Nous avons donc choisi, pour la suite des tests la scène sonore *lavage-vaisselle-4* (*vaisselle + radio avant*).

#### 4.2.3. La terrasse de café

Pour la terrasse de café nous observons que trois des six scènes sonores ont obtenu de bons scores de sélection alors que les trois autres ont obtenu des scores moins élevés.

<b>Sons-terrasse-1</b> : « Verres + glaçons » + « Pas » + « Voiture »	5
<b>Sons-terrasse-2</b> : « Verres + glaçons » + « Voiture » + « Pas »	6
<b>Sons-terrasse-3</b> : « Pas » + « Verres + glaçons » + « Voiture »	2
<b>Sons-terrasse-4</b> : « Pas » + « Voiture » + « Verres + glaçons »	2
<b>Sons-terrasse-5</b> : « Voiture » + « Pas » + « Verres + glaçons »	2
<b>Sons-terrasse-6</b> : « Voiture » + « Verres + glaçons » + « Pas »	5

Les échanges oraux que nous avons pu avoir avec les participants à la fin du test nous amènent quelques pistes d'interprétation de ces données. En effet, ils nous ont à plusieurs reprises indiqué que l'ordre chronologique d'apparition des éléments sonores saillants (le critère de variation introduit lors de la réalisation) a constitué un critère de sélection, et ce plus particulièrement pour les « *bruits de verres et de glaçons* ».

Ainsi, en observant la position de ces « *bruits de verres et de glaçons* » par rapport aux scores recueillis pour chaque scène sonore, nous pouvons considérer que ces bruits, positionnés en début d'écoute impliquent de bons scores : 5 (*sons-terrasse-1*) et 6 (*sons-terrasse-2*) participants sur 22. Nous pouvons supposer qu'ils déclenchent un effet d'amorçage (*priming* Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010) sur le reste de la scène sonore. Parallèlement, nous voyons que pour les scènes sonores où les « *bruits de verres et de glaçons* » sont placés à la fin de l'écoute (*sons-terrasse-4 et 5*), les scores de sélection sont relativement mauvais : 2 participants sur 22 dans les deux cas. Enfin, lorsque les « *bruits de verres et de glaçons* » sont placés au milieu de l'écoute (*sons-terrasse-3 et 6*) les scores sont variables : 2 participants sur 22 pour *sons-terrasse-3* et 5 pour *sons-terrasse-6*.

La scène sonore choisie pour la suite des tests est donc *son-terrasse-2*.



## Deuxième Chapitre : Élaboration du matériel linguistique

L'expérimentation centrale de ce travail porte sur l'évaluation de la compréhension de courts dialogues pédagogiques en français, présentés seuls ou avec un contexte sonore environnemental, par des locuteurs natifs et des apprenants débutants en FLE. Dans ce cadre, un matériel expérimental *ad hoc* a été construit à partir de l'élaboration 1) de scènes sonores environnementales (voir p.121) qui assumeront une fonction contextuelle situationnelle, et 2) des dialogues pédagogiques sur lesquels portera l'expérimentation centrale (voir p.151). C'est sur la construction du matériel linguistique que porte ce chapitre.

### 1. Méthode

La méthodologie d'élaboration de ces supports langagiers vise une adéquation avec les dialogues pédagogiques disponibles pour les niveaux A, c'est-à-dire le niveau débutant dans l'apprentissage des langues étrangères. Cet objectif implique un contrôle dans la rédaction des scripts, aux niveaux syntaxique et lexical. Il semble également important de veiller à la thématique des dialogues : ils doivent correspondre à des sujets de la vie quotidienne afin de coïncider avec les recommandations du CECRL (Conseil de l'Europe, 2001).

#### 1.1. Rédaction et enregistrement des dialogues

Le Cadre décrit l'apprenant débutant comme :

« A1 : Peut comprendre et utiliser des expressions familières et quotidiennes ainsi que des énoncés très simples qui visent à satisfaire des besoins concrets. (...) Peut communiquer de façon simple si l'interlocuteur parle lentement et distinctement et se montre coopératif.

A2 : *Peut comprendre des phrases isolées et des expressions fréquemment utilisées en relation avec des domaines immédiats de priorité (par exemple, informations personnelles et familiales simples, achats, environnement proche, travail). Peut communiquer lors de tâches simples et habituelles ne demandant qu'un échange d'informations simple et direct sur des sujets familiers et habituels. Peut décrire avec des moyens simples sa formation, son environnement immédiat et évoquer des sujets qui correspondent à des besoins immédiats.* » (p.25) » (pour plus de précision voir le *Contexte didactique*, p.17)

Dans ce cadre, différents critères ont présidé à l'élaboration des dialogues pédagogiques : 1) le caractère stéréotypé des situations, 2) le niveau de langue (lexique et syntaxe des dialogues), 3) la durée, et 4) les conditions d'enregistrement.

### **1.1.1. Des situations stéréotypées**

Le CECRL (Conseil de l'Europe, 2001), recommande, avec les niveaux élémentaires, de centrer l'enseignement/apprentissage sur des situations familières de la vie quotidienne. Les scènes sonores construites (voir p.121) mettent en ondes des lieux fréquents dans la vie quotidienne des français (une *cuisine* et une *terrasse de café*). Les dialogues ont donc pour cadre ces deux lieux.

Parallèlement à la construction des deux histoires sonores relatives à la *cuisine*, deux dialogues qui leur correspondent ont été élaborés. Le premier met en scène deux interlocuteurs qui échangent autour de *la préparation d'un repas* (voir p. 155) et correspond à la scène sonore du même nom. Le deuxième met en scène les deux mêmes interlocuteurs qui échangent cette fois autour de *la vaisselle* (voir p.155). Il est dès lors envisageable de considérer ces deux dialogues comme participant à une même séquence pédagogique. En effet, s'ils sont envisagés comme participant au même ensemble, ils respectent entre eux une unité de lieu (la même *cuisine*), de temps (les deux dialogues se *déroulent* le même jour), et une unité d'intrigue (les deux scènes sont liées sémantiquement).

Nous pouvons constater que ces dialogues sont très stéréotypés dans les rapports homme/femme qu'ils présentent. En effet, dans le premier, la femme prépare le repas et, dans le deuxième, alors que l'homme fait la vaisselle, la femme le remercie. Comme nous l'ont fait remarquer différents participants à l'évaluation des dialogues

(voir p.158) ou à l'expérimentation principale (voir p.182), une lecture de ces dialogues pourrait donc amener à penser à une répartition des tâches ménagères figée dans des rôles déséquilibrés. Toutefois, il apparaît tout d'abord que cette distribution n'est pas rare dans notre société (l'Observatoire des inégalités<sup>59</sup>). Par ailleurs, il n'est pas exclu de travailler plus particulièrement sur cet aspect du dialogue dans les séances pédagogiques afin, éventuellement, d'aborder l'évolution sociétale qui s'opère à ce niveau.

Un seul dialogue a été construit pour la *terrasse de café* (voir p. 156). Il met en scène deux amies qui commandent une consommation. Ce dialogue met en scène des rapports sociaux « classiques » dans nos sociétés, aborde le tutoiement des amies, (vs le vouvoiement du serveur) et le caractère social fortement ritualisé du lieu.

### **1.1.2. Le niveau de langue**

Nous appelons « niveau de langue des dialogues » la complexité linguistique qu'ils présentent. Deux dimensions contrôlées interviennent : le niveau lexical et le niveau syntaxique. En effet, le CECRL recommande, pour les niveaux débutants (Conseil de l'Europe, 2001 p.26) :

Compréhension Orale -A1 : *« Je peux comprendre des mots familiers et des expressions courantes au sujet de moi-même, de ma famille et de l'environnement concret et immédiat, si les gens parlent lentement et distinctement. »*

Compréhension Orale -A2 : *« Je peux comprendre des expressions et un vocabulaire très fréquent relatifs à ce qui me concerne de très près (par exemple moi-même, ma famille, mes proches, le travail). Je peux saisir l'essentiel d'annonces et de messages simples et clairs. »*

Dès lors, il apparaît primordial de veiller au niveau syntaxique et lexical. Au plan syntaxique, les temps et modes verbaux ont été contrôlés. Ainsi, la majorité des verbes

---

<sup>59</sup> <http://www.inegalites.fr/spip.php?article245>

conjugués apparaît au présent de l'indicatif (un verbe est au futur proche et un autre à l'infinitif), élément de conjugaison généralement abordé en tout début d'apprentissage du FLE. Une seule proposition subordonnée apparaît, introduite par *que* (qui fait partie des premiers pronoms relatifs enseignés et appris). Les autres phrases sont des phrases simples ou des énumérations. Parallèlement, le lexique a fait l'objet d'une attention spécifique. Afin de répondre aux impératifs d'enseignement proposés par le Français Fondamental (Gougenheim *et al.*, 1956; Ministère de l'éducation nationale, 1970, 1972), nous avons contrôlé la fréquence de tous les items lexicaux en procédant par étape.

Dans un premier temps, les items lexicaux ont été confrontés à la *liste des mots les plus fréquents de la langue écrite française (XIXe et XXe siècle) d'Éduscol*<sup>60</sup>. Cette liste est destinée en premier lieu aux enseignants de l'école élémentaire française donc à un public différent de celui avec lequel nous travaillons. Cependant, les liens entre FLM et FLE ne sont plus à démontrer et il est apparu intéressant de vérifier l'intérêt pédagogique des items lexicaux introduits dans les dialogues construits.

Dans un second temps, les items lexicaux des dialogues ont été confrontés au sous-corpus *freqlemfilms2* de la base de données *Lexique*<sup>61</sup>. Cette partie du corpus est composée de sous-titrages de films. Elle est composée de 4 sous-corpus issus de sous-titrages de films anglais (cinéma) (26,5 millions de mots), de films anglais (télévision) (19,5 millions de mots), de films français (1,9 millions de mots) et de films du reste de l'Europe (2,5 millions de mots). La recherche que nous avons effectuée à partir de *freqlemfilms2* nous renseigne sur la moyenne des fréquences par millions d'occurrences pour l'ensemble de ces sous-corpus. Ont été retenus les lemmes, de manière à se concentrer sur la forme la plus courante des mots (entrées dans le dictionnaire).

---

<sup>60</sup> portail national des professionnels de l'éducation, créé et géré par le ministère de l'éducation nationale  
<http://eduscol.education.fr/cid47915/liste-des-mots-classee-par-ordre-alphabetique.html>

<sup>61</sup> Base de données lexicographique constituée à partir d'un corpus de livres (*Frantext*, composé de 218 romans publiés entre 1950 et 2000), et à un corpus de sous-titrages de films (9474 sous-titrages).  
<http://www.lexique.org>

### **La préparation d'un repas**

Le premier dialogue comporte 5 répliques, 3 dites par l'homme (A), et 2 dites par la femme (B).

- A- *Qu'est-ce qu'on mange ?*
- B- *Je prépare des légumes qu'on va manger avec des œufs. Ça te va ?*
- A- *Oui, très bien, quels légumes tu mets ?*
- B- *Des champignons, des pommes de terre et des oignons... Tu veux autre chose ?*
- A- *Non, non, c'est très bien !*

Ce dialogue se compose de 46 mots. La consultation de la *liste Éduscol* nous montre que 34 items de ce dialogue y sont présents, soit 74 %. Cette première vérification garantit l'intérêt pédagogique du dialogue.

Parallèlement, d'après la base de donnée *Lexique*, nous constatons que la fréquence moyenne des items est de 5317,5, la médiane est de 1589,9 (par million d'occurrences). Le mot le moins fréquent, « *champignon* », a une fréquence de 11,03 ; le mot le plus fréquent « *être* » a une fréquence de 32236,5. Cette deuxième vérification confirme l'intérêt du lexique présenté : les mots les plus fréquents sont à enseigner en priorité.

### **La vaisselle**

Le deuxième dialogue compte 4 répliques également distribuées entre les locuteurs.

- B- *Tu as besoin d'aide?*
- A- *Non, c'est bon, il n'y a pas grand-chose : deux verres, deux assiettes, quelques couverts et une poêle...*
- B- *C'est sympa de m'aider... merci !*
- A- *C'est normal, tu as préparé le repas, alors je fais la vaisselle !*

Ce dialogue compte 45 mots. La consultation de la *liste Éduscol* nous montre que 36 items du dialogue y sont présents, soit 80%. Là encore, l'intérêt pédagogique du dialogue est garanti.

Parallèlement, d'après la base de donnée *Lexique*, la fréquence moyenne des items est de 6921,28, la médiane est 1280,72 (par million d'occurrences). Le mot le moins fréquent, « *de* (art.) », a une fréquence de 0,1 ; le mot le plus fréquent, « *ne* », a une fréquence de 22287,83. Cette deuxième vérification confirme elle aussi l'intérêt pédagogique du dialogue.

### ***À la terrasse d'un café***

Le troisième dialogue compte 4 répliques également réparties entre les 2 locutrices.

- C- *Qu'est-ce que tu prends ?*
- D- *Mmmm, je vais prendre un café avec un verre d'eau. Et toi ?*
- C- *Je ne sais pas trop... comme toi je crois...*
- D- *Bon, alors j'appelle le serveur ! (.) S'il vous plait ?*

Ce dialogue rassemble 34 mots. « *S'il vous plait* » est considéré comme un seul item, une seule lexie du fait de sa *lexicalisation*. De ce fait, il n'apparaît pas comme tel dans la liste *Éduscol*, ni dans la base de données *Lexique*. Malgré cela, 29 items de notre dialogue sont présents dans la liste *Éduscol*, soit 85% ce qui, encore une fois, nous assure de son intérêt pédagogique.

Parallèlement, d'après *Lexique*, la fréquence moyenne des items est de 7456,32, la médiane est 2119,95 (par million d'occurrences). Le mot le moins fréquent, « *serveur* », a une fréquence de 32,21 ; le mot le plus fréquent, « *être* », a une fréquence de 32236,5.

Nous pouvons donc considérer que les dialogues construits répondent aux objectifs pédagogiques définis à la fois par *Éduscol* et le Français Fondamental (Gougenheim *et al.*, 1956).

### **1.1.3. Durée**

De la même manière que le lexique (principalement) et la syntaxe ont été contrôlés, la durée des dialogues (sur l'enregistrement) a été limitée. Deux raisons principales expliquent ce choix.

Tout d'abord, il s'est agi de construire des dialogues d'une durée inférieure à la durée des scènes sonores (15 secondes), afin que le Contexte Sonore Environnemental (CSE) commence avant et se termine après le dialogue.

Par ailleurs, un dialogue court permet une charge cognitive moins importante en tâche de compréhension et d'intégration, plus encore avec un public débutant en langue (Tricot, 1998). De la même manière, la mémorisation se voit facilitée par une faible durée du stimulus.

Ainsi, la durée du premier dialogue *la préparation d'un repas* est de 12 secondes. Le deuxième dialogue *la vaisselle* dure 11 secondes. Enfin, le troisième *À la terrasse d'un café* dure également 11 secondes.

#### **1.1.4. Enregistrements**

Pour procéder aux enregistrements des dialogues, nous avons fait participer 3 locuteurs : *Charlotte, Pascal et Cynthia*. *Charlotte* et *Pascal* ont enregistré les deux dialogues relatifs à la *cuisine*. *Cynthia* et *Charlotte* ont enregistré le dialogue relatif à la *terrasse de café*. Lors des enregistrements, les 3 locuteurs, avertis des enjeux prosodiques liés au public débutant en langue ont veillé à maintenir un débit lent et une prosodie marquée afin de faciliter l'intelligibilité des énoncés. Les enregistrements se sont faits grâce à la cabine PÉTRA<sup>62</sup> le 13 septembre 2012.

## **1.2. Évaluation des dialogues**

Au-delà des précautions suivies dans leur élaboration, il est apparu indispensable de confronter les dialogues à des juges. En effet, ces derniers ont pu évaluer l'adéquation des dialogues au niveau des apprenants de FLE que nous interrogerons dans *l'expérimentation principale* (voir p.165).

Douze enseignants de FLE travaillant dans différentes institutions d'enseignement du Français Langue Étrangère ont été interrogés. Ces enseignants exercent en France, 6 d'entre eux travaillent au sein de l'Alliance Française de Toulouse<sup>63</sup> (où le choix du manuel est décidé par l'institution) et 6 autres au Département d'Étude du FLE de l'Université de Toulouse 2 Le Mirail<sup>64</sup> (DEFLE, où les enseignants sont généralement libres de choisir leurs supports). L'ensemble de ces enseignants a en moyenne 10,5 ans d'expérience (8 de médiane) (voir description des

---

<sup>62</sup> [petra.univ-tlse2.fr/](http://petra.univ-tlse2.fr/)

<sup>63</sup> <http://www.alliance-toulouse.org/>

<sup>64</sup> <http://defle.univ-tlse2.fr/>

participants sur le CD en annexe). Un questionnaire auto-administré leur a été soumis. Le premier volet de ce questionnaire vise à savoir à quel(s) niveau(x) (selon la définition du CECRL (Conseil de l'Europe, 2001)) les dialogues proposés pourraient être utilisés. Le deuxième volet cherche à obtenir des éléments de réponses sur une éventuelle utilisation en salle de classe : les enseignants interrogés utiliseraient-ils ces supports ?

### **1.2.1. L'évaluation du niveau de langue**

Pour chaque dialogue, les participants ont pu évaluer le niveau à partir d'une échelle présentant les 6 niveaux du CECRL (*ibidem*). La consigne était systématiquement « *Pour chacun d'eux merci de cocher le ou les niveaux (échelle de niveaux du CECRL) au(x)quel(s) vous proposeriez le dialogue pour une activité de compréhension orale* » (voir questionnaire, dans les annexe).

Le niveau sera alors défini en fonction des scores obtenus par chaque dialogue.

### **1.2.2. Une éventuelle utilisation en classe**

Pour cette question, il s'est agi pour les enseignants de répondre dans un champ libre qui leur était proposé. La question, pour l'ensemble des dialogues était : « *Utiliseriez-vous ce support de cours ?* ».

Les résultats obtenus à partir de ce questionnaire nous renseignent sur un aspect fondamental des dialogues : correspondent-ils aux attentes des enseignants ?

## **2. Résultats**

### **2.1. L'évaluation du niveau de langue**

Pour l'évaluation du niveau de langue des dialogues, les participants pouvaient sélectionner plusieurs réponses. Cet aspect explique que les totaux soient supérieurs à 12.

#### ***Dialogue A : la préparation d'un repas***

Le premier dialogue *la préparation d'un repas*, a globalement été évalué pour un niveau A1-A2. Aucun des enseignants n'a évalué ce dialogue à un niveau C, ni à un niveau B2. Un enseignant a jugé que ce dialogue pouvait correspondre à un niveau B1, sept participants l'ont évalué à un niveau A2 et huit à un niveau A1. En outre, un participant a

évalué le dialogue aux niveaux A1-A2-B1 (en précisant que l'exploitation du dialogue varierait en fonction du niveau des apprenants), et deux participants ont jugé que le dialogue pouvait être utilisé aux niveaux A1 ou A2.

	La préparation d'un repas					
	Niveau					
	A1	A2	B1	B2	C1	C2
part.1	0	1	0			
part.2	0	1	0			
part.3	1	1	0			
part.4	1	1	1			
part.5	1	1	0			
part.6	1	0	0			
part.7	1	0	0			
part.8	1	0	0			
part.9	1	0	0			
part.10	1	0	0			
part.11	0	1	0			
part.12	0	1	0			
TOTAL	8	7	1			

**Tableau 10 : Évaluation du niveau du dialogue la préparation d'un repas**

L'objectif de correspondance avec les niveaux A a donc été atteint pour ce dialogue, avec une légère prévalence pour le niveau A1.

**Dialogue B : la vaisselle**

Le deuxième dialogue, *la vaisselle* a globalement été évalué au niveau A2. Encore une fois, aucun participant n'a évalué le dialogue à un niveau C, ni au niveau B2. Un seul participant a évalué le dialogue au niveau B1, onze participants ont jugé le dialogue comme correspondant à un niveau A2 et sept à un niveau A1. Pour ce deuxième dialogue, un seul enseignant a estimé que le dialogue pouvait correspondre aux trois premiers niveaux et 4 participants ont jugé que le dialogue pouvait intervenir dans les deux niveaux A.

	La vaisselle					
	Niveau					
	A1	A2	B1	B2	C1	C2
part.1	0	1	0			
part.2	1	1	0			
part.3	1	1	0			
part.4	1	1	1			
part.5	1	1	0			
part.6	0	1	0			
part.7	1	1	0			
part.8	1	0	0			
part.9	0	1	0			
part.10	0	1	0			
part.11	0	1	0			
part.12	1	1	0			
TOTAL	7	11	1			

**Tableau 11 : Évaluation du niveau du dialogue *la vaisselle***

Là aussi, le niveau A est respecté, avec cette fois une prévalence pour le niveau A2.

**Dialogue C : À la terrasse d'un café**

Enfin, le dialogue *À la terrasse d'un café*, a globalement été évalué aux niveaux A1-A2. Aucun des évaluateurs n'a jugé que ce dialogue pouvait correspondre au niveau C, ni au niveau B2. Deux participants ont évalué ce dialogue au niveau B1, neuf participants l'ont évalué au niveau A2 et sept au niveau A1. Deux participants ont évalué que ce dialogue pouvait correspondre aux niveaux A2 et B1 et deux participants ont évalué le dialogue aux niveaux A1 et A2.

	À la terrasse d'un café					
	Niveau					
	A1	A2	B1	B2	C1	C2
part.1	0	1	0			
part.2	0	1	1			
part.3	1	0	0			
part.4	0	1	1			
part.5	1	1	0			
part.6	1	0	0			
part.7	1	1	0			
part.8	1	0	0			
part.9	1	1	0			
part.10	0	1	0			
part.11	0	1	0			
part.12	1	1	0			
TOTAL	7	9	2			

**Tableau 12 : Évaluation du niveau du dialogue à la terrasse d'un café**

Une adéquation du dialogue avec le niveau A (avec, encore une fois une prévalence pour le niveau A2) est donc observée.

En conclusion, les dialogues que nous avons construits en visant une correspondance avec les niveaux A, concordent avec les niveaux visés. La variation que nous pouvons observer vers les niveaux A1 et B1 s'explique par la porosité des niveaux et les méthodologies d'exploitation potentiellement envisagées par les participants au test.

Ainsi, dans *l'expérimentation principale* (voir p.165), les participants du panel d'apprenants seront sélectionnés en fonction de leur niveau de français : niveau A1 ou A2.

## **2.2. Une éventuelle exploitation en salle de classe**

La question portant sur une utilisation de ces supports en classe présentait un champ libre pour le recueil des réponses. Néanmoins, la majorité des enseignants n'a répondu que par un mot (généralement *oui*, *non*, ou probablement).

Pour le dialogue *la préparation d'un repas*, onze participants ont déclaré qu'ils utiliseraient le dialogue en classe (réponse *oui*), et un participant a déclaré qu'il l'utiliserait *probablement*. Pour le dialogue *la vaisselle*, dix participants ont déclaré qu'ils utiliseraient ce support (réponse *oui*), et deux enseignants ont déclaré qu'ils n'exploiteraient pas ce dialogue (réponse *non*). Enfin, pour le dialogue *À la terrasse d'un café*, dix participants ont déclaré qu'ils utiliseraient ce support (réponse *oui*), et deux enseignants ont déclaré qu'ils n'exploiteraient pas ce dialogue (réponse *non*). Il apparaît donc que, globalement, les dialogues construits seraient susceptibles d'être exploités en salle de classe.





# Partie 3 : Expérimentation principale

Dans le cadre du chapitre précédent, la constitution de notre matériel expérimental a été présentée. Il s'est agi de construire l'ensemble de nos stimuli de manière à (1) développer des connaissances empiriques sur la méthodologie de construction de scènes sonores environnementales et de dialogues pédagogiques et, (2) contrôler les paramètres définitoires de ces stimuli.

L'ensemble du matériel ainsi constitué a participé à l'expérimentation principale de cette étude : le Classe-Images. Ce test, que nous avons entièrement élaboré, vise, en premier lieu, la vérification de l'hypothèse centrale de notre travail : la présence d'un contexte sonore environnemental modifie la compréhension d'un dialogue utilisé en classe de langue par des apprenants débutants.

## 1. Méthode

Généralement, en classe de langue étrangère, la compréhension d'un document sonore par des apprenants est évaluée *via* un questionnaire. Celui-ci peut être de natures diverses (questions ouvertes ou fermées) et mettre en jeu différentes modalités (le plus souvent orale ou écrite). Or, dans ce cadre, comment interpréter les éventuelles difficultés des apprenants ? S'agit-il de difficultés de compréhension ou de production (dans le cadre des réponses aux questions ouvertes) ? Et même lorsque l'on parvient à éliminer les erreurs de production (en passant par un QCM ou par une réponse en langue maternelle par exemple), les difficultés de compréhension relèvent-elles du support cible ou de la question/consigne ? Que dire aussi d'une évaluation de la compréhension orale à partir d'un questionnaire écrit ? Ces interrogations sont,

aujourd'hui encore, au centre des préoccupations des enseignants de FLE (cf. *Contexte Didactique* p.17).

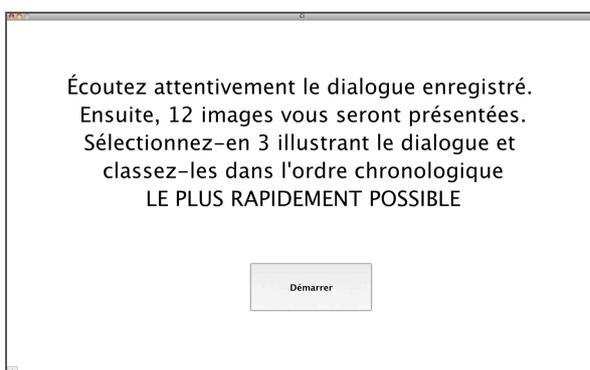
Afin d'éviter ces écueils, dans le cadre de cette étude nous avons choisi de recourir à une tâche de sélection de dessins pour rendre compte des dialogues entendus. Ce choix nous a amenée à élaborer un logiciel *ad hoc* pour cette expérimentation, ainsi qu'à constituer un corpus d'images contrôlé dont les constituants répondent à un maximum d'impératifs de recherche.

## 1.1. Le Classe-Images

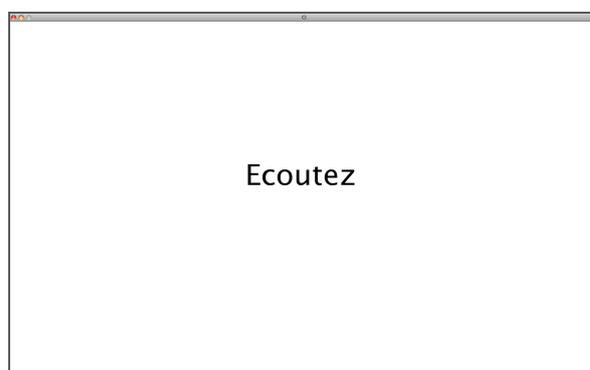
### 1.1.1. Déroulement des passations

Afin de systématiser le test et de recueillir automatiquement les données relatives à la sélection des dessins, nous avons conçu le Classe-Images. Ce logiciel nous permet d'automatiser les passations et de garantir un environnement similaire à tous les participants. Afin de garantir une bonne visibilité des dessins, le test a été réalisé sur un ordinateur 27 pouces. Les passations se sont déroulées individuellement et les écoutes se sont faites au casque à la fois pour garantir une bonne qualité sonore, une « isolation » vis à vis des bruits extérieurs, et une homogénéité des conditions d'écoute.

Dans la mesure où nous avons interrogé des apprenants débutants en Français Langue Étrangère, les consignes ont tout d'abord été explicitées à l'oral. Il s'agissait de vérifier de la bonne compréhension des consignes avant de laisser les participants face à l'écran où la consigne était écrite. Après cette première étape, l'expérimentateur n'est plus intervenu dans le déroulement du test.



**Image 7 : Capture d'écran de la consigne sur le Classe-Images**

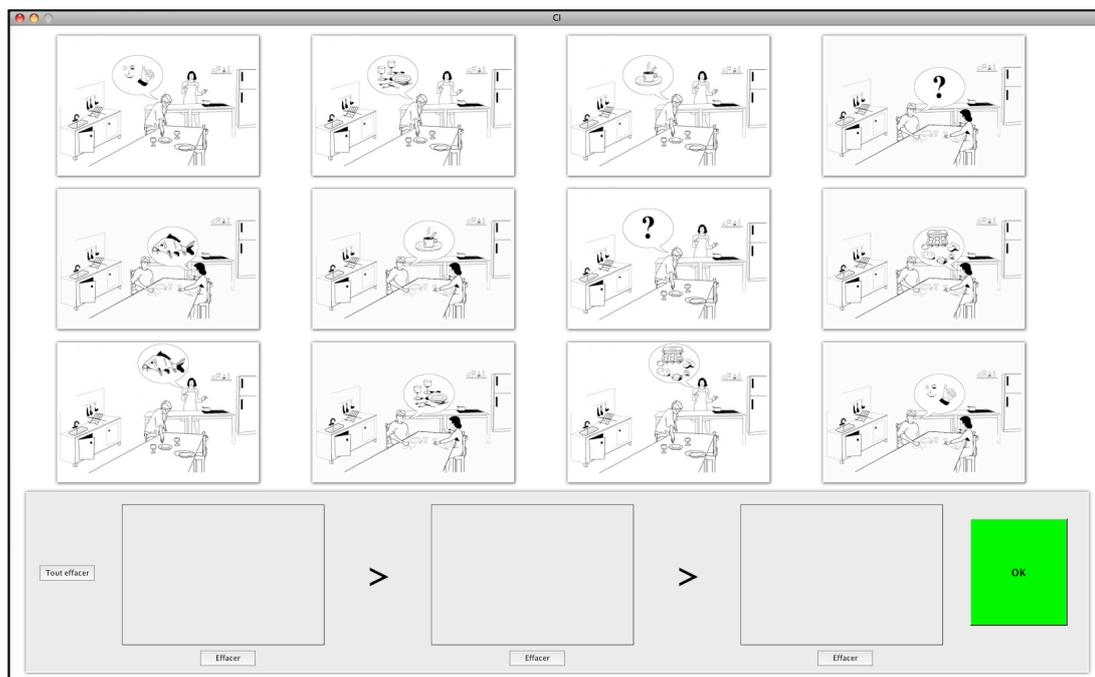


**Image 8 : Capture d'écran de l'écoute sur le Classe-Images**

Chaque participant pouvait prendre son temps pour lire et assimiler la consigne, ainsi il commençait le test quand il le décidait en cliquant sur « *démarrer* ». Durant cette deuxième phase, les participants sont face à un écran « épuré ».

À la suite de l'écoute du dialogue, douze dessins apparaissent sur l'écran. À cet instant un chronomètre se met en marche de manière à recueillir les temps de réponse.

Chaque dessin a une dimension de 8\*5,5 cm, ce qui permet au participant de « bien voir » les détails. Les dessins apparaissent à l'écran dans un ordre aléatoire de manière à « écraser » les éventuels effets de positionnement.



**Image 9 : Capture d'écran de la présentation des dessins pour le dialogue *La préparation d'un repas* sur le Classe-Images**

Ces douze dessins renvoient à quatre catégories dont les facteurs de variation ont été contrôlés (voir p.169)

Les 3 cases situées en bas de l'écran servent à recueillir les dessins sélectionnés et déplacés par les participants. Les flèches qui les séparent marquent l'axe chronologique. Pour placer les dessins dans les cases de réception, les participants devaient effectuer un « cliquer/ glisser ». Ces manipulations ont pu impliquer un léger temps d'apprentissage, susceptible de modifier les temps de réponses enregistrés sur le

premier dialogue entendu. Pour cette raison, et afin de pallier les éventuels effets d'ordre, les dialogues ont été présentés dans un ordre varié, de manière à ce que chacun d'eux apparaisse en position initiale, intermédiaire, et finale.

Les participants ont procédé à leur sélection en temps illimité mais chronométré et ils pouvaient apporter différentes modifications (effacer, recouvrir ou déplacer les dessins). En outre, il apparaît que l'observation des dessins a pu influencer la compréhension des dialogues. En effet, dans la mesure où les éléments de réponse étaient contrôlés et limités (en quelque sorte, comme dans un QCM), ils ont pu aiguiller les participants dans leur élaboration d'une signification. Une fois satisfaits, ils ont validé leur sélection en cliquant sur « ok », arrêtant ainsi le chronomètre.

### 1.1.2. Les données recueillies

Comme nous pouvons le voir sur la capture d'écran ci-dessous, le logiciel Classe-Images récupère l'ensemble des données de manière automatique. Sont ainsi enregistrés :

- Les choix définitifs (dessins choisis et validés par le participant)
- Le Temps de Réponse (TR) global
- La position de chaque dessin sélectionné (case de réception)

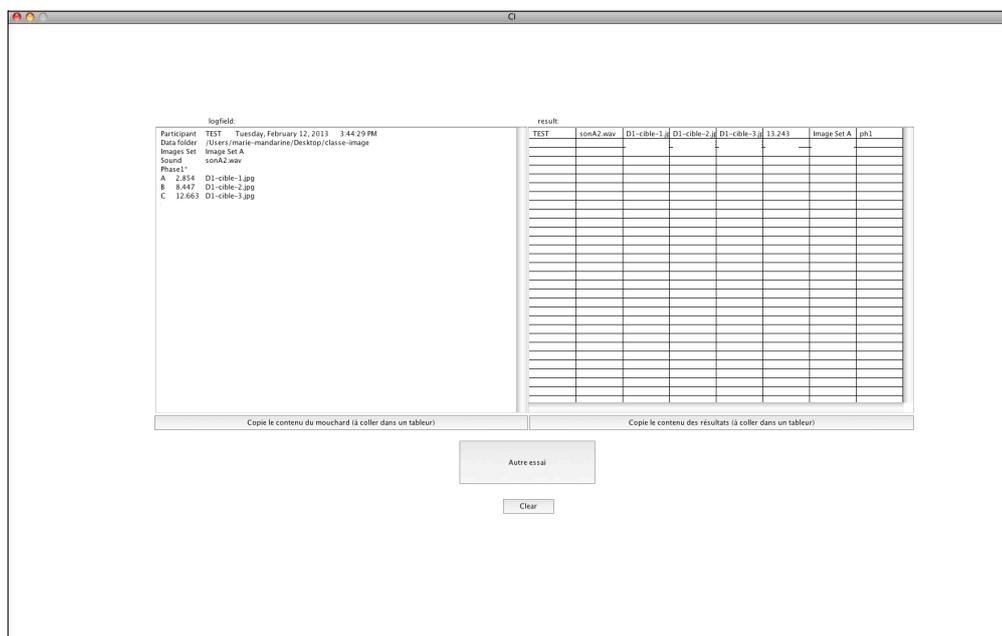


Image 10 : Capture d'écran des résultats sur le Classe-Images

## 1.2. Constitution du corpus d'images

Ce test a pour objectif d'évaluer la compréhension d'un dialogue accompagné ou non d'un contexte sonore environnemental. Cette évaluation s'est effectuée à partir d'images présentant différents facteurs de variation, au niveau de la situation d'énonciation ou du contenu propositionnel. Les participants devaient ainsi sélectionner trois dessins parmi les douze qui leur étaient présentés pour rendre compte du dialogue entendu. Les douze dessins proposés correspondent à quatre groupes de trois dessins présentant différents facteurs de variation :

- 3 dessins correspondent aux dessins cibles, c'est-à-dire qu'ils rendent compte du dialogue. Ces dessins correspondent aux « réponses attendues » et s'ils sont correctement positionnés sur l'axe chronologique laissent entendre une adéquation entre le sens construit dans l'élaboration du dialogue et celui élaboré par les participants en phase de production (construction de la suite de dessins) ;
- 3 dessins présentent une variation de la situation d'énonciation, c'est-à-dire que le contenu des phylactères (contenu propositionnel) reste inchangé, alors que les protagonistes ne sont plus dans les mêmes positions ;
- 3 dessins présentent une variation du contenu propositionnel, c'est-à-dire que la situation d'énonciation (position des protagonistes) reste inchangée par rapport aux dessins cibles, alors que le contenu propositionnel (phylactères) est modifié ;
- 3 dessins présentent à la fois une variation du contenu propositionnel (dans les phylactères) et de la situation d'énonciation (position des protagonistes).

Ces dessins ont été réalisés par une artiste plasticienne confirmée<sup>65</sup>, doctorante en Arts Plastiques, à l'Université de Montpellier 3<sup>66</sup>, en collaboration avec six juges experts, en majorité enseignants de Français Langue Étrangère et au fait des enjeux de cette recherche (voir CD en annexe).

### 1.2.1. La préparation d'un repas

Les dessins pour le dialogue *La préparation d'un repas* sont :

- 3 **dessins cibles** présentent une femme et un homme. Sur ces dessins la protagoniste est derrière le plan de travail dans la cuisine pendant que le protagoniste met la table. Celui-ci demande à la femme ce qu'ils s'apprêtent à

---

<sup>65</sup> <http://www.imiteme.com/>

<sup>66</sup> (Colle, *in prep.*)

manger (« *Qu'est-ce qu'on mange ?* »). En réponse, elle lui présente ce qu'elle cuisine (« *Je prépare des champignons qu'on va manger avec des œufs* »). Enfin, l'homme exprime sa satisfaction (« *C'est très bien !* »)

Ces dessins sont nommés **cible-1**, **cible-2** et **cible-3**.



**Cible 1**

**Cible 2**



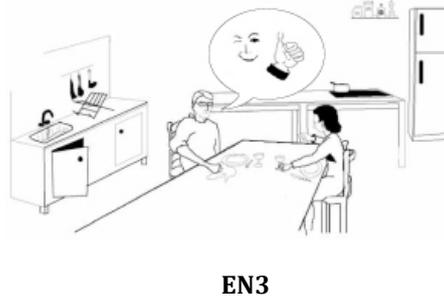
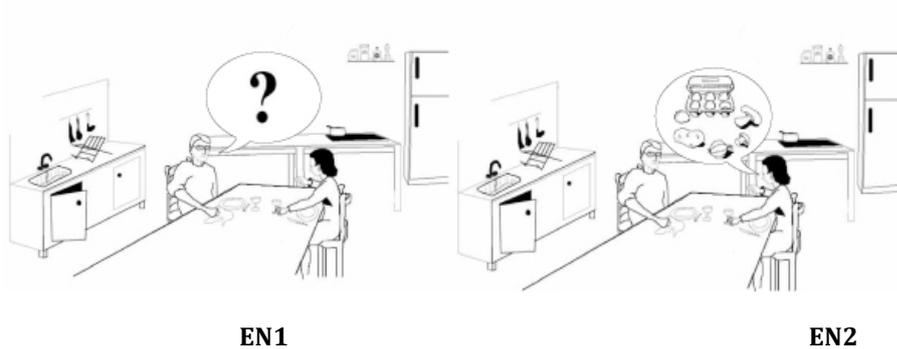
**Cible 3**

**Dessins 1 : Dessins cibles pour le dialogue *La préparation du repas***

Ces dessins se déroulent sur l'axe chronologique construit avec les emplacements prévus pour le recueil des dessins sélectionnés.

- 3 dessins présentent une variation de la **situation d'énonciation** sur lesquels le contenu des phylactères reste inchangé mais les positions des protagonistes varient. Alors que sur les dessins cibles, l'homme mettait la table pendant que la femme préparait le repas, sur cette variante les deux protagonistes sont à table. Ici est ciblée l'évaluation de la compréhension du verbe « *je prépare* » (versus « j'ai préparé », « je préparerai » ou « je vais préparer »). Les autres énoncés du dialogue ne permettent pas de discriminer les deux situations d'énonciation. En effet, nous pourrions tout-à-fait considérer que ce dialogue (« *je prépare* » mis à part) se déroule « à table ».

Ces dessins sont nommés **EN-1**, **EN-2** et **EN-3**.



**Dessins 2 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue *La préparation d'un repas***

Ces dessins présentent le même potentiel chronologique (basé sur le contenu des phylactères) que les dessins cibles.

- 3 dessins dont la situation d'énonciation reste inchangée par rapport aux dessins-cibles alors que le **contenu propositionnel** est modifié. Pour ces dessins, le contenu des phylactères ne correspond, selon nous à aucun énoncé du dialogue. Dès lors, leur interprétation n'a pas été anticipée.

Ces dessins sont nommés **CONT-1**, **CONT-2** et **CONT-3**.



**CONT1**

**CONT2**



**CONT3**

**Dessins 3 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue *La préparation d'un repas***

Pour ces dessins aucun ordre chronologique n'est attendu. Ce critère ne sera donc pas pris en compte dans l'analyse des résultats.

- Enfin, 3 dessins pour lesquels les **2 facteurs** de variation sont en jeu : les protagonistes ne sont pas dans la même position que dans les dessins cibles (position équivalente aux dessins EN) et le contenu propositionnel ne correspond pas non plus à celui des dialogues (contenu des phylactère équivalent aux dessins CONT).

Ces dessins sont nommés **EN-CONT-1**, **EN-CONT-2**, et **EN-CONT-3**.



**EN-CONT1**

**EN-CONT2**



**EN-CONT3**

**Dessins 4 : Dessins avec les 2 facteurs de variation (EN-CONT) pour le dialogue *La préparation d'un repas***

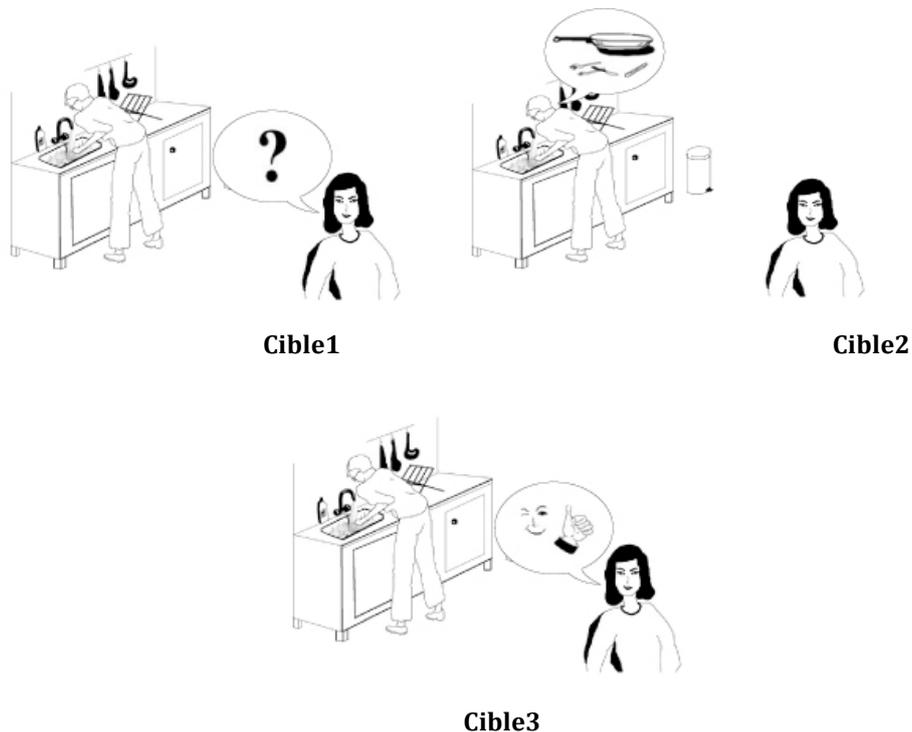
Ces dessins ne présentent pas de potentiel chronologique.

### **1.2.2. La vaisselle**

Pour le dialogue *la vaisselle*, les mêmes facteurs de variation apparaissent. Les dessins présentés aux participants sont alors :

- 3 **dessins cibles**, qui rendent compte à la fois de la situation d'énonciation et du contenu propositionnel du dialogue. Ainsi, le protagoniste lave la vaisselle pendant que la protagoniste l'interroge (« *Tu as besoin d'aide ?* »). Toujours en lavant la vaisselle il lui répond (« *Non, c'est bon, il n'y a pas grand-chose : deux verres, deux assiettes, quelques couverts et une poêle...* »). Enfin, la femme exprime sa gratitude (« *C'est sympa de m'aider... merci !* »).

Ces dessins sont nommés **cible-1**, **cible-2** et **cible-3**.



**Dessins 5 : Dessins cibles pour le dialogue *La vaisselle***

Ces dessins peuvent respecter l'ordre chronologique du dialogue. Un mauvais positionnement sur l'axe chronologique sera interprété comme une difficulté du participant à appréhender la dimension chronologique du dialogue.

- 3 dessins présentent une variation de la **situation d'énonciation** sur lesquels le contenu des phylactères reste inchangé alors que les protagonistes sont à table plutôt que dans leurs positions initiales. Notons toutefois que ces dessins sont valides dans la mesure où on peut parler d'une activité avant de la faire.

Ces dessins sont nommés **EN-1**, **EN-2** et **EN-3**.



**EN1**

**EN2**



**EN3**

**Dessins 6 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue *La vaisselle***

Ces dessins présentent le même potentiel chronologique que les dessins cibles.

- 3 dessins pour lesquels intervient une variation du **contenu propositionnel** sans changement de la situation d'énonciation.

Ces dessins sont nommés **CONT-1**, **CONT-2** et **CONT-3**.



**CONT1**

**CONT2**



**CONT3**

**Dessins 7 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue *La vaisselle***

Aucune interprétation du contenu des phylactères n'a pas été anticipée et l'ordre chronologique de ces dessins ne sera pas pris en compte.

- Enfin, 3 dessins pour lesquels varient à la fois le **contenu propositionnel** et la **situation d'énonciation**.

Ces dessins sont nommés **EN-CONT-1**, **EN-CONT-2** et **EN-CONT-3**.



**EN-CONT1**

**EN-CONT2**



**EN-CONT3**

**Dessins 8 : Dessins avec les 2 facteurs de variation (EN-CONT) pour le dialogue *La vaisselle***

Ces dessins ne présentent pas de potentiel chronologique.

### ***1.2.3. À la terrasse d'un café***

Enfin, pour le dialogue *À la terrasse d'un café* les mêmes facteurs de variation interviennent. Les dessins relatifs à ce dialogue sont donc :

- 3 **dessins cibles** qui présentent la situation d'énonciation : deux jeunes femmes sont attablées à la terrasse d'un café et discutent. La première interroge l'autre sur sa commande (« *Qu'est-ce que tu prends ?* »), l'autre lui répond (« *Mmmm, je vais prendre un café avec un verre d'eau. Et toi ?* »). Enfin, la première décide de commander la même chose (« *comme toi je crois...* »).

Ces dessins sont nommés **cible-1**, **cible-2** et **cible-3**.



**Cible1**



**Cible2**

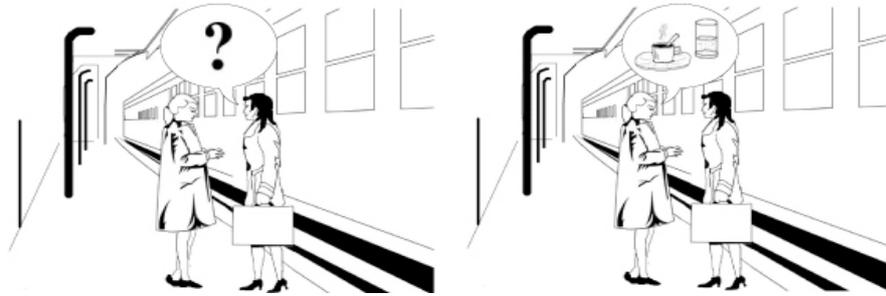


**Cible3**

**Dessins 9 : Dessins cibles pour le dialogue *À la terrasse d'un café***

Ces dessins présentent un potentiel chronologique.

- 3 dessins présentent une variation de la **situation d'énonciation**. Ces dessins sont nommés **EN-1**, **EN-2**, et **EN-3**. Les protagonistes sont, sur ces dessins, sur un quai de gare ou de métro. Cette fois, ce n'est donc plus seulement la position des locuteurs qui varie, mais bien le lieu où se déroule la scène.



EN1

EN2



EN3

**Dessins 10 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue *À la terrasse d'un café***

Ces dessins présentent le même potentiel chronologique que les dessins cibles.

- 3 dessins présentant une variation du **contenu propositionnel** (contenu des phylactères), sans modification de la situation d'énonciation.

Ces dessins sont nommés **CONT-1**, **CONT-2** et **CONT-3**.



**CONT1**



**CONT2**

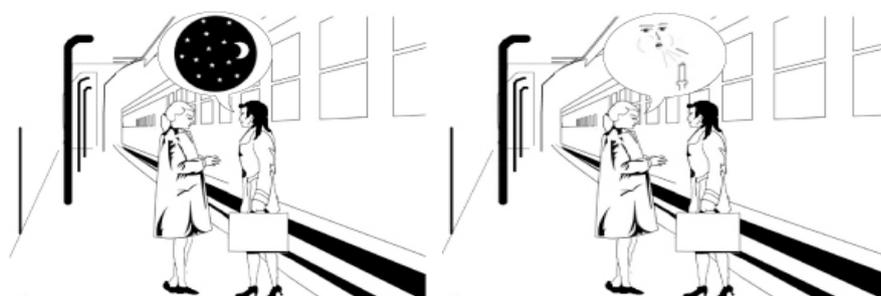


**CONT3**

**Dessins 11 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue *À la terrasse d'un café***

Ces dessins ne présentent pas de potentiel chronologique.

- Enfin, 3 dessins présentent les 2 facteurs de variation. Ces dessins sont nommés **EN-CONT-1**, **EN-CONT-2** et **EN-CONT-3**.



**EN-CONT1**

**EN-CONT2**



**EN-CONT3**

**Dessins 12 : Dessins avec les 2 facteurs de variation (EN-CONT)  
pour le dialogue *À la terrasse d'un café***

Ces dessins ne présentent pas de potentiel chronologique.

#### **1.2.4. Remarques**

Deux remarques sont à apporter à propos de ces dessins. Tout d'abord, au niveau de leur style graphique, *La préparation d'un repas* et *La vaisselle* d'une part et *A la terrasse d'un café d'autre part* ont chacune, mais différemment, une dimension stéréotypée. *La cuisine* est « de style scandinave » et « épurée ». Ce genre de mobilier est à la fois commun et anonyme, de manière à ce que les participants au test du Classe-Images (tout comme ceux du premier recueil de données verbales de l'*Étude sur les attentes auditives* », voir p.121) aient la possibilité de se projeter dans « leur cuisine ». *La terrasse de café* est dans un style esthétique plus « rétro », renvoyant aux terrasses parisiennes telles que Doisneau aurait pu les photographier. Ce style fait écho à une

autre dimension du stéréotype de la France, plus proche de celui que les apprenants ont pu développer avant d'arriver dans l'Hexagone.

La deuxième remarque porte sur les variations de la situation d'énonciation, introduites dans les dessins EN. Alors que dans la *cuisine* cette variation implique un déplacement au sein d'une même pièce, pour la situation *terrasse de café* les protagonistes se déplacent vers un tout autre lieu : un quai de gare ou de métro. En outre, il apparaît que pour les dialogues qui relèvent de la *Cuisine*, les variations de la situation d'énonciation introduites dans les dessins restent plausibles (à l'exception de l'énoncé « *je prépare* » dans le dialogue *La préparation d'un repas*) : il est envisageable d'évoquer des actions en dehors de leur réalisation effective. En revanche, le dialogue *À la terrasse d'un café* est fortement improbable sur un quai de gare. Une plus grande distance sépare les dessins cibles et présentant une variation de la situation d'énonciation pour le dialogue *À la terrasse d'un café* que pour les dialogues *La préparation d'un repas* et *La vaisselle*. La différence qui sépare les dessins cibles des dessins EN est donc plus importante pour la *Terrasse* que pour la *Cuisine*. Cette distinction est en outre parallèle à la différence de « nature » des lieux que nous avons choisis de mettre en scène (voir la *Méthodologie de l'étude sur les attentes auditives* p.122) : la cuisine est un lieu privé et clos (où les interactions relèvent généralement de l'intime), alors que la terrasse de café est un lieu ouvert et public (où les interactions relèvent parfois du domaine social).

### **1.3. Participants**

Pour ce test, deux panels de population ont été interrogés. De manière générale, le public que nous visons correspond à un public d'apprenants débutants dans l'apprentissage du Français Langue Étrangère. Néanmoins, le premier panel est composé d'individus adultes francophones et peut être considéré comme le panel témoin. Le deuxième se compose de participants apprenants de Français Langue Étrangère de niveau débutant (A1-A2 selon le CECRL, Conseil de l'Europe, 2001). Les différents critères d'inclusion et d'exclusion qui ont présidé à la sélection des participants sont explicités ci-après.

#### ***1.3.1. Le panel de participants francophones***

Le groupe francophone est composé de 58 personnes âgées de 18 à 32 ans. La moyenne d'âge de ce panel est de 24,86 ans (avec une médiane de 25 ans). Le statut socio-professionnel de ces participants correspond au public d'apprenants que nous visons : ils sont tous étudiants (de la L3 au doctorat). Ce panel se compose de quarante-

neuf femmes et neuf hommes. Ce déséquilibre peut s'expliquer par le mode de recrutement des participants : *via* l'Université, dans les départements d'enseignement des Sciences du Langage et de la Psychologie où les femmes sont généralement majoritaires. Les critères d'exclusion que nous avons appliqués sont les suivants : les participants présentant un bilinguisme précoce, des troubles du langage (tels que la dysphonie ou l'aphasie), ou des troubles de l'audition ont été exclus.

Ce premier panel a été réparti en deux groupes. Le premier a passé le test en écoutant les Dialogues Seuls (groupe DS). Le deuxième (groupe DCS) a écouté les Dialogues avec leur Contexte Sonore.

Le groupe DS est composé de 29 participants ayant entre 18 et 32 ans. La moyenne d'âge de ce groupe est de 24,8 ans (avec une médiane de 25 ans). Trois hommes font partie de ce groupe contre 26 femmes. Parallèlement, les 29 participants qui composent le groupe DCS ont également entre 18 et 32 ans et la moyenne d'âge de ce groupe est de 24,9 ans (avec une médiane de 25 ans). Ce groupe est composé de 6 hommes et 23 femmes.

### ***1.3.2. Le panel d'apprenants de niveau A***

Le panel d'étudiants étrangers est composé de 42 apprenants de FLE de niveau A, de différentes nationalités et parlant différentes langues maternelles (voir CD en annexe), âgés de 18 à 45 ans. La grande diversité des langues et cultures maternelles des participants pourrait être considérée comme un écueil potentiel de l'étude puisque la distance entre celles-ci et la langue et culture cible est une source de variation importante dans l'apprentissage (Galisson, 1990, 1994; Robert, 2004). Toutefois, cette diversité est représentative des groupes-classes généralement présents en France où une multitude de langues et cultures se côtoient.

Ce panel est composé d'un plus petit nombre de participants que le panel de francophones. Cet écart est lié aux difficultés de recrutement de la population apprenante débutante imposée par les critères d'inclusion que nous avons définis. En effet, nous visons des apprenants résidant en France depuis 3 mois minimum et 1 an maximum. D'autre part, le caractère débutant des apprenants implique (d'après des déclarations de bon nombre de ceux que nous avons rencontrés et qui ont refusé de participer au test) une angoisse face au test. Cet aspect est certainement encore renforcé dans la mesure où l'expérimentatrice n'est pas l'enseignante de référence de ces apprenants et où, dans ce cadre, aucun lien de confiance n'est créé.

La moyenne d'âge de ce panel est de 27,23 ans (avec une médiane de 27). Ils ont tous un statut social « étudiant » puisqu'ils étudient à l'Alliance Française (3 d'entre eux)

ou au Département d'Étude du Français Langue Étrangère à l'Université de Toulouse 2 le Mirail (39 d'entre eux). Comme dans le panel des participants francophones mais dans des proportions moindres, nous avons eu plus de difficultés à recruter des hommes que des femmes : 12 hommes et 30 femmes composent le panel d'apprenants étrangers. Tout comme le premier panel, les quarante-deux apprenants de niveau A de FLE ont été répartis en 2 groupes, selon les stimuli qui leur ont été proposés. Le premier groupe (DS) est composé des 21 participants qui ont écouté le Dialogue Seul *i.e.* sans que celui-ci ne soit accompagné d'une scène sonore en position contextuelle. Le deuxième groupe, de 21 participants aussi, s'est vu proposer l'écoute des Dialogues accompagnés de leur Contexte Sonore (groupe DCS). La moyenne d'âge des participants du groupe DS est de 27,2 ans (avec 27 ans de médiane), et celle du groupe DCS est de 27,3 ans (avec 26 de médiane). Enfin, la répartition des participants selon leur sexe présente un équilibre intergroupe puisque il y a 6 hommes pour 15 femmes dans les 2 groupes.

Ainsi, deux variables indépendantes du test (expertise linguistique et écoute du dialogue avec ou sans contexte sonore) ont présidé à la constitution des groupes.

## **1.4. Variables du test**

Pour le test du Classe-Images, nous avons observé un panel de francophones natifs et un panel d'apprenants débutants dans une tâche de compréhension orale, suivie d'une production/construction à partir de dessins. Dans ce cadre, les dessins sélectionnés par les participants sont considérés comme une production dirigée rendant compte de leur compréhension. Un grand nombre d'études sur la compréhension renvoie à des tâches de reformulation ou de résumé des supports lus ou entendus (voir le chapitre ... *aux études sur la compréhension* p.81, dans le cadre théorique). Pour le Classe-Images, afin de contourner l'écueil de la production linguistique, nous avons choisi une réponse sous forme de sélection et de présentation chronologique de dessins que nous appellerons *la construction dessinée*. Cette construction est « contrainte » dans la mesure où les dessins sont imposés par la procédure. Toutefois, les choix offerts aux participants sont multiples car les combinaisons possibles sont très nombreuses (12 dessins, 3 cases de réception, possibilité de dupliquer et de laisser une ou plusieurs cases vides...).

### **1.4.1. Les variables indépendantes**

Deux grandes questions prévalent dans notre questionnement général : Qu'est-ce qu'être débutant en langue implique en compréhension orale ? Dans quelle mesure la

présence d'un contexte sonore environnemental a-t-elle un impact sur la compréhension d'un dialogue ?

L'objectif du Classe-Images est d'apporter des éléments de réponse à ces questions. Celles-ci ont amené à la formulation de trois hypothèses générales auxquelles nous répondrons en observant deux variables indépendantes :

- (1) **le niveau d'expertise des participants** (francophones natifs /vs/ apprenants débutants) aura un impact sur les performances en compréhension (variable dépendante), observable à partir des résultats du Classe-Images. Pour valider (ou infirmer) cette hypothèse, nous procéderons à des tests statistiques à partir des réponses produites par les participants qui ont entendu le dialogue seul, *i.e.* sans contexte sonore environnemental. Cette condition correspond en effet à la « condition contrôle » du test. Nous comparerons donc les performances du panel des francophones natifs à celles du panel des apprenants débutants, à différents niveaux de la construction dessinée, à partir de la compréhension des dialogues et des corpus de dessins prédéfinis. L'effet attendu est une augmentation des performances en compréhension avec l'augmentation du niveau linguistique des participants ;
- (2) **La présence ou l'absence d'un contexte sonore environnemental** aura un impact sur les performances dans cette tâche (variable dépendante), observable à partir des résultats du Classe-Images. Pour valider (ou infirmer) cette hypothèse, nous procéderons à des tests statistiques à partir des réponses produites par les participants francophones natifs et apprenants débutants dans les conditions **avec** (DCS) et **sans contexte sonore environnemental** (DS). L'effet attendu est une amélioration des performances dans la condition DCS : selon nous, le *contexte sonore environnemental*, en offrant des indices supplémentaires pour la construction de sens, participe favorablement à celle-ci ;

Il s'agira donc, dans la présentation des résultats de cette étude d'observer les effets des différentes variables indépendantes présentées sur différentes variables dépendantes. En effet, ce que nous avons ici appelé « performances dans la tâche » renvoie à différents paramètres qu'il s'agira d'étudier séparément et en interaction.

#### **1.4.2. Les variables dépendantes**

L'hypothèse centrale de notre travail est que les différentes variables identifiées ci-avant ont toutes un impact sur les performances dans la tâche. Cette hypothèse suppose une définition opérante de la compréhension. Dans le cadre de cette

expérimentation, nous considérons trois facteurs qui entrent en jeu dans la compréhension :

- l'appréhension de la situation d'énonciation (aspects essentiellement extralinguistiques) ;
- l'appréhension du contenu propositionnel (aspects essentiellement linguistiques) ;
- l'appréhension de la dimension chronologique (aspects essentiellement paralinguistiques).

Ces différents paramètres sont appréhendables à partir du Classe-Images, *via* l'observation de trois ensembles de résultats :

- (1) les suites de trois dessins cibles sélectionnés et correctement agencés sur l'axe chronologique, qui permettent de poser un jugement binaire de compréhension : *Les participants ont-ils compris ce que nous attendions qu'ils comprennent (oui ou non) ?* Ces suites impliquent la prise en compte des trois paramètres de la compréhension exposés plus haut (*la situation d'énonciation, le contenu propositionnel et la chronologie*) sur la totalité de la construction dessinée. Ainsi, leur production permettra-t-elle de poser un jugement de « bonne compréhension » *i.e.*, en situation de classe, ces productions correspondraient aux réponses attendues par l'enseignant. Le caractère catégoriel de cette partie des données (les participants ont compris ou non) nous a amenée à procéder à un test de comparaison de fréquences (table de contingence, pour plus de précisions, se reporter aux annexes) ;
- (2) les dessins sélectionnés qui, considérés un à un, permettent de rendre compte d'une partie du sens élaboré par les participants. Ces dessins sont envisagés comme des « parcelles » de la signification construite. En effet, nous pouvons considérer que chaque dessin sélectionné renvoie à une micro-proposition, participant au schéma d'action global produit par le participant. Cet ensemble de résultats correspond donc non pas à une considération binaire de la compréhension (*le participant a-t-il compris ou non*) mais à une attention portée aux éléments qui composent la signification construite. Dans ce cadre, les dessins présentant un (ou plusieurs) facteur(s) de variation seront considérés comme des indices d'une inadéquation entre le sens construit dans l'élaboration des supports (dialogues et dessins) et celui interprété et construit par le(s) participant(s) (à partir du dialogue et des dessins). Le caractère quantitatif (et parfois ordinal et quantitatif continu, pour la chronologie) des données traitées

dans cette partie des résultats nous a amenée à procéder à des tests paramétriques (t-tests pour la robustesse à l'inégalité des variances (Schmider, Ziegler, Danay, Beyer, & Bühner, 2010)) ;

- (3) les temps de réponse, que nous considérons comme des indices de la charge cognitive induite par la tâche de compréhension du support sonore, l'analyse et le traitement des supports visuels (dessins) ainsi que la sélection et le positionnement des dessins finalement retenus par les participants. Pour ce paramètre nous avons choisi de procéder à un test non-paramétrique : un U de Mann-Whitney pour la robustesse qu'il présente à la non normalité des données.

## 2. Résultats

Dans le chapitre hypothèses (p.105), comme dans les lignes qui précèdent, nous avons vu que trois hypothèses ont dirigé nos travaux expérimentaux en nous amenant à définir trois variables indépendantes. L'effet de ces variables indépendantes a été observé sur trois variables dépendantes : les suites de trois dessins cibles (voir p.187), les dessins sélectionnés (parcelles de compréhension ; voir p.190), et les temps de réponse (indices de la charge cognitive ; voir p.203).

### 2.1. La construction de suites de trois dessins cibles : la compréhension en termes binaires

Comme dit précédemment, les suites de trois dessins cibles, sélectionnés et correctement agencés sur l'axe chronologique, permettent de poser un diagnostic de « bonne compréhension ». Dit autrement, ces suites correspondent aux réponses attendues, c'est-à-dire celles qui ont été construites en amont par l'expérimentateur reposant sur une adéquation explicite entre le sens construit à partir des dialogues et celui des dessins.

Pour traiter ces résultats, nous avons procédé à des **comparaisons de fréquences** (le nombre de fois où une série de dessins a été sélectionnée) de manière à voir si les différences constatées sont statistiquement significatives.

#### 2.1.1. La variable « expertise linguistique »

Notre hypothèse opérationnelle est que plus le participant sera expert en langue française, plus la sélection des dessins sera conforme à celle attendue. Dit autrement, la variable « *expertise linguistique* » (VI) impliquera une quantité de suites de trois dessins

cibles (sélectionnés et correctement agencés sur l'axe chronologique) construites (VD) plus importante pour le panel de participants francophones que pour le panel d'apprenants débutants. Nous comparerons ces suites construites par les groupes DS (condition d'écoute sans contexte sonore environnemental – condition contrôle) du panel des francophones (experts) et du panel des apprenants débutants (non experts).

Les résultats obtenus par un test de **comparaison de fréquence** nous amènent à considérer que le niveau d'expertise linguistique a systématiquement impliqué une différence significative ( $p < 0,05$ ) entre les panels de francophones et d'apprenants débutants. Si nous considérons l'influence de cette variable sur l'ensemble des trois dialogues, c'est-à-dire en ne prenant pas en compte leurs spécificités, alors il apparaît que la différence entre le nombre de suites construites par les francophones ( $n=87$ , pourcentage de suites de 3 dessins cibles = 66,67) et les apprenants débutants ( $n=63$ , pourcentage de suites de 3 dessins cibles = 26,98) est significative  $Z= 4,797$   $p=8,03E-07$ .

En outre, si nous considérons les dialogues un à un, il apparaît que les écarts restent tout aussi significatifs.

	Panel de francophones (groupe DS)		Panel d'apprenants débutants (groupe DS)		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
dialogue A	29	48,27	21	19,05	2,125	0,017
dialogue B	29	79,31	21	28,57	3,588	0
dialogue C	29	72,41	21	33,33	2,745	0,003

**Tableau 13: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur la construction de suites de 3 dessins cibles (VD)**

Les participants francophones ont construit systématiquement et significativement plus de suites de trois dessins cibles (c'est-à-dire attendues) que les apprenants débutants. L'expertise linguistique est donc un facteur primordial dans la compréhension des dialogues construits, quelles que soient leurs particularités.

### **2.1.2. La variable « contexte sonore environnemental »**

Notre deuxième hypothèse opérationnelle au niveau de la compréhension en termes binaires est que la variable « *contexte sonore environnemental* » (VI) aura une

influence sur le nombre de suites de trois dessins cibles (sélectionnés et correctement agencés sur l'axe chronologique) construites (VD) par les participants. Nous comparerons ces suites construites par les groupes DS (condition d'écoute sans contexte sonore environnemental) et DCS (condition d'écoute avec contexte sonore environnemental), pour les panels des francophones et des apprenants débutants.

Les résultats obtenus par un test de **comparaison de fréquence** nous amènent à considérer que la présence d'un contexte sonore environnemental n'a pas systématiquement impliqué de différence significative ( $p < 0,05$ ) entre les groupes DS et DCS. Si nous observons l'effet de cette variable sur l'ensemble des trois dialogues chez les **francophones**, alors il apparaît que la différence entre le nombre de suites construites par les groupes DS ( $n=87$ , pourcentage de suites de 3 dessins cibles = 66,67) et DCS ( $n=87$ , pourcentage de suites de 3 dessins cibles = 77,011) n'est pas statistiquement significative  $Z=-1,517$   $p= 0,065$ .

Si nous considérons les dialogues un à un, alors il apparaît que l'écart n'est en effet significatif que pour le dialogue A (*la préparation d'un repas*).

	Panel de francophones GROUPE DS		Panel de francophones GROUPE DCS		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
dialogue A	29	48,276	29	72,414	-1,879	0,03
dialogue B	29	79,31	29	86,207	-0,695	0,243
dialogue C	29	72,414	29	72,414	0	0,5

**Tableau 14: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur la construction de suites de 3 dessins cibles (VD) pour le panel de francophones**

Il apparaît que chez les experts (au niveau linguistique), la variable *contexte sonore environnemental* n'implique globalement pas de différence significative entre les groupes au niveau de la compréhension considérée en termes binaires.

Avec le panel d'apprenants débutants, en considérant les trois dialogues comme un ensemble, la différence est significative entre les groupes DS ( $n=63$ , pourcentage de suites de 3 dessins cibles = 26,984) et DCS ( $n=63$ , pourcentage de suites de 3 dessins cibles = 46,032)  $Z=-2,22$   $p=0,013$ .

Pour ce même panel, si nous considérons les dialogues un à un, seul le dialogue A a donné lieu à des résultats avec une différence significative au niveau de la construction de suites de trois dessins cibles.

	Panel d'apprenants débutants GROUPE DS		Panel d'apprenants débutants GROUPE DCS		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
dialogue A	21	19,048	21	47,619	-1,964	0,024
dialogue B	21	28,571	21	42,857	-0,966	0,167
dialogue C	21	33,333	21	47,619	-0,943	0,173

**Tableau 15: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur la construction de suites de 3 dessins cibles (VD) pour le panel d'apprenants débutants**

## 2.2. Les dessins sélectionnés: évaluation des parcelles de compréhension.

Comme dit plus haut, nous considérerons, dans cette partie des résultats, les dessins sélectionnés comme des parcelles de la signification construite par les participants. Selon les dessins sélectionnés, différents points d'accord (le sens construit correspond, en partie, au sens attendu) ou l'inadéquation (le sens construit par les participants diverge de celui attendu) pourront être observés. Il ne s'agit donc plus ici d'évaluer la compréhension en termes binaires (« *compris/pas compris* »), mais bien de tenter de s'approcher du sens construit par les participants, à partir des dessins qu'ils ont sélectionnés. Dit autrement, la question n'est plus de savoir s'ils ont compris mais « *ce qu'ils ont construit ou compris ?* ».

Pour traiter ces résultats, nous avons majoritairement procédé à des **t-tests** (variables « expertise linguistique » et « contexte sonore environnemental ») de manière à voir si les différences constatées sont statistiquement significatives. Notons que, bien que les données recueillies ne respectent pas de loi normale, les tests choisis sont robustes à ce paramètre (Schmider *et al.*, 2010), notamment en suivant l'hypothèse de variances inégales.

Dans ces résultats, nous observerons donc les dessins sélectionnés par les participants : les dessins cibles qui renvoient à une « compréhension en conformité », quelque soit leur placement sur l'axe chronologique, les dessins EN qui présentent une

variation de la situation d'énonciation», quelque soit leur placement sur l'axe chronologique, les dessins CONT qui présentent une variation du contenu propositionnel, et les dessins EN-CONT pour lesquels varient la situation d'énonciation et le contenu propositionnel. La sélection des trois derniers types de dessins montre une inadéquation entre le sens envisagé lors de l'élaboration des dialogues et des dessins et celui construit par les participants. Les dessins cibles et EN mal positionnés laissent apparaître une divergence entre la chronologie envisagée dans l'élaboration des supports et celle produite par les participants. Les absences de réponses seront également prises en compte. Celles-ci peuvent indiquer soit que les participants ont considéré qu'un ou deux dessin(s) étai(en)t suffisant(s) pour rendre compte du dialogue, soit qu'aucun des dessins proposés ne rendait compte de leur compréhension.

### **2.2.1. La variable « expertise linguistique »**

Notre première hypothèse opérationnelle à ce niveau est que la variable « *expertise linguistique* » (VI) impliquera une différence dans les dessins sélectionnés par les participants (VD. Nous comparerons les sélections des groupes DS (condition d'écoute sans contexte sonore environnemental – condition contrôle) du panel de francophones et du panel d'apprenants débutants.

Si l'on considère les trois dialogues ensemble, c'est-à-dire en ne tenant pas compte des spécificités de chacun, alors il apparaît que la variable « *expertise linguistique* » a donné lieu à des écarts statistiquement significatifs ( $p < 0,05$ ) sur la quasi-totalité des paramètres observés.

La différence significative entre les dessins cibles sélectionnés par le panel de francophones et ceux sélectionnés par le panel d'apprenants débutants montre que les participants francophones ont globalement présenté une meilleure compréhension parcellaire. Les différences significatives de sélection de dessins EN, CONT et EN-CONT montrent que les apprenants débutants ont globalement présenté plus de difficulté à appréhender les dimensions situationnelles et propositionnelles que les participants francophones. L'observation des dessins cibles et EN mal positionnés nous renseigne sur l'appréhension de la dimension chronologique : les apprenants débutants ont eu plus de difficulté à appréhender ce paramètre que les natifs. Au niveau des absences de réponse, nous observons que le niveau d'expertise a donné lieu à une différence significative : les participants apprenants débutants ont plus fréquemment présenté ce type de réponse.

	Panel de francophones (groupe DS)			Panel d'apprenants débutants (groupe DS)			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	87	2,52	0,874	63	1,56	1,241	5,274	104,78	0,000
dessins EN	87	0,34	0,804	63	0,73	1,050	-2,439	111,36	0,016
dessins CONT	87	0,07	0,255	63	0,19	0,435	-1,985	92,58	0,050
dessins EN-CONT	87	0,00	0,000	63	0,10	0,296	-2,555	62,00	0,013
dessins cibles mal positionnés	87	0,16	0,568	63	0,37	0,789	-1,751	106,45	0,083
dessins EN mal positionnés	87	0,01	0,107	63	0,19	0,592	-2,372	64,95	0,021
absence de réponse	87	0,07	0,367	63	0,38	0,923	-2,541	76,28	0,013

**Tableau 16: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire-VD) pour l'ensemble des dialogues**

Si l'on considère le dialogue A, alors nous pouvons constater que la variable « niveau d'expertise » n'a donné lieu à des différences significatives qu'aux niveaux de la sélection des dessins cibles (« bonne compréhension ») et des dessins EN mal positionnés (inadéquation au niveau de la chronologie). Toutefois, il apparaît que sur l'ensemble des autres aspects (facteurs de variation dans la constitution du corpus d'images), les participants francophones ont présenté de meilleures performances que les participants apprenants débutants (moindre sélection de dessins présentant des facteurs de variation), sans pour autant que les écarts soient significatifs.

	Panel de francophones (groupe DS)			Panel d'apprenants débutants (groupe DS)			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	29	2,17	1,037	21	1,24	1,221	2,842	38,81	0,007
dessins EN	29	0,72	1,032	21	1,19	1,209	-1,430	38,92	0,161
dessins CONT	29	0,03	0,186	21	0,10	0,301	-0,819	30,88	0,419
dessins EN-CONT	29	0,00	0,000	21	0,14	0,359	-1,826	20,00	0,083
dessins cibles mal positionnés	29	0,07	0,371	21	0,19	0,680	-0,743	28,63	0,464
dessins EN mal positionnés	29	0,00	0,000	21	0,38	0,740	-2,369	20,00	0,029
absence de réponse	29	0,07	0,371	21	0,19	0,680	-0,743	28,63	0,464

**Tableau 17: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue A**

Si l'on considère le dialogue B, alors on constate que l'expertise linguistique a donné lieu à des différences significatives aux niveaux de la sélection des dessins cibles (bonne compréhension), des dessins EN (variation de la situation d'énonciation) et des dessins CONT (variation du contenu propositionnel). Pour les autres facteurs observés (dessins EN-CONT, mauvais positionnement et absences de réponse), aucun des écarts observés n'est significatif.

	Panel de francophones (groupe DS)			Panel d'apprenants débutants (groupe DS)			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	29	2,72	0,649	21	1,38	1,284	4,405	27,42	0,000
dessins EN	29	0,10	0,310	21	0,67	1,017	-2,438	22,71	0,022
dessins CONT	29	0,03	0,186	21	0,33	0,577	-2,288	23,02	0,032
dessins EN-CONT	29	0,00	0,000	21	0,10	0,301	-1,451	20,00	0,162
dessins cibles mal positionnés	29	0,14	0,581	21	0,10	0,301	0,338	44,11	0,737
dessins EN mal positionnés	29	0,03	0,186	21	0,14	0,655	-0,737	22,34	0,469
absence de réponse	29	0,14	0,516	21	0,52	1,123	-1,466	26,14	0,155

**Tableau 18: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue B**

Enfin, si l'on considère le dialogue C, alors on constate que l'expertise linguistique a donné lieu à une différence significative au niveau de la sélection des dessins cibles, de leur mauvais positionnement, ainsi qu'au niveau des absences de réponses. Toutefois, nous pouvons tout de même constater que, sur l'ensemble des facteurs observés, les participants francophones présentent de meilleures performances (moins de sélection de dessins présentant un ou plusieurs facteur(s) de variation) que les participants apprenants débutants.

	Panel de francophones (groupe DS)			Panel d'apprenants débutants (groupe DS)			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	29	2,66	0,814	21	2,05	1,117	2,118	34,67	0,041
dessins EN	29	0,21	0,774	21	0,33	0,730	-0,589	44,65	0,559
dessins CONT	29	0,14	0,351	21	0,14	0,359	-0,048	42,69	0,962
dessins EN-CONT	29	0,00	0,000	21	0,05	0,218	-1,000	20,00	0,329
dessins cibles mal positionnés	29	0,28	0,702	21	0,81	1,030	-2,053	33,03	0,048
dessins EN mal positionnés	29	0,00	0,000	21	0,05	0,218	-1,000	20,00	0,329
absence de réponse	29	0,00	0,000	21	0,43	0,926	-2,121	20,00	0,047

**Tableau 19: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue C**

Notons toutefois que, si l'on considère les dessins cibles correctement positionnés, la différence entre ceux sélectionnés par le panel de francophones (groupe DS n=29 M=2,38 SD=1,115) et le panel d'apprenants débutants (groupe DS n=21 M=1,24 SD=1,375) est également significative  $t(37,515)=3,130$   $p=0,003$ . Cette observation laisse penser que la variable « *expertise linguistique* » a impliqué une différence significative sur la sélection globale de dessins cibles indépendamment de leur positionnement.

Ainsi, il apparaît que, pour l'ensemble des dialogues, **le niveau d'expertise linguistique a donné lieu à des différences significatives aux niveaux de la sélection de dessins cibles, ainsi que d'autres facteurs**. Il apparaît également que la

variabilité inter-dialogue est importante lorsque les francophones sont comparés aux apprenants débutants. Cette variabilité sera traitée en discussion (voir p.211).

### 2.2.2. La variable « contexte sonore environnemental »

Notre hypothèse opérationnelle est que la variable « *contexte sonore environnemental* » (VI) impliquera une différence dans les dessins sélectionnés par les participants (VD). Nous comparerons les sélections des groupes DS (condition d'écoute sans contexte sonore environnemental) à celles des groupes DCS pour les panels de francophones et d'apprenants débutants. Notons que les membres du panel de francophones n'ont sélectionné aucun dessin EN-CONT.

Si nous considérons l'ensemble des trois dialogues, c'est-à-dire en ne tenant pas compte de leurs spécificités, et sur le panel de francophones alors nous pouvons constater que cette variable n'a pas donné lieu à des différences entre les groupes DS et DCS.

	GROUPE DS Panel de francophones			GROUPE DCS Panel de francophones			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	87	2,52	0,874	87	2,72	0,623	-1,798	155,44	0,074
dessins EN	87	0,34	0,804	87	0,17	0,533	1,667	149,27	0,098
dessins CONT	87	0,07	0,255	87	0,09	0,291	-0,555	169,12	0,580
dessins EN-CONT	87	0,00	0,000	87	0,00	0,000	-	-	-
dessins cibles mal positionnés	87	0,16	0,568	87	0,11	0,387	0,624	151,59	0,533
dessins EN mal positionnés	87	0,01	0,107	87	0,03	0,184	-1,009	138,57	0,315
absence de réponse	87	0,07	0,367	87	0,01	0,107	1,402	100,57	0,164

**Tableau 20: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour l'ensemble des dialogues**

Si on considère le dialogue A, et toujours pour le panel de **francophones**, on peut observer que la variable « *contexte sonore environnemental* » n'a pas donné lieu à des différences significatives entre les deux groupes. En outre, nous pouvons aussi constater que les membres de ce panel n'ont présenté aucun dessin EN mal positionné.

	GROUPE DS Panel de francophones			GROUPE DCS Panel de francophones			t	ddl	Sig. (bilaté- -rale)
	n	M	Écart- type	n	M	Écart- type			
dessins cibles	29	2,17	1,037	29	2,59	0,867	-1,648	54,28	0,105
dessins EN	29	0,72	1,032	29	0,38	0,820	1,409	53,29	0,165
dessins CONT	29	0,03	0,186	29	0,03	0,186	0,000	56,00	1,000
dessins cibles mal positionnés	29	0,07	0,371	29	0,03	0,186	0,447	41,18	0,657
dessins EN mal positionnés	29	0,00	0,000	29	0,00	0,000	-	-	-
absence de réponse	29	0,07	0,371	29	0,00	0,000	1,000	28,00	0,326

**Tableau 21: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue A**

Si l'on considère le dialogue B, on peut également observer que la variable « *contexte sonore environnemental* » n'a pas donné lieu à des différences significatives entre les deux groupes.

	GROUPE DS Panel de francophones			GROUPE DCS Panel de francophones			t	ddl	Sig. (bilaté-rale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	29	2,72	0,649	29	2,86	0,441	-0,947	49,32	0,348
dessins EN	29	0,10	0,310	29	0,10	0,310	0,000	56,00	1,000
dessins CONT	29	0,03	0,186	29	0,03	0,186	0,000	56,00	1,000
dessins cibles mal positionnés	29	0,14	0,581	29	0,14	0,441	0,000	52,23	1,000
dessins EN mal positionnés	29	0,03	0,186	29	0,07	0,258	-0,584	50,88	0,562
absence de réponse	29	0,14	0,516	29	0,00	0,000	1,440	28,00	0,161

**Tableau 22 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue B**

Enfin, si l'on considère le dialogue C, encore une fois, aucune différence significative n'est observée entre les groupes DS et DCS, pour le panel de francophones. La variable « *contexte sonore environnemental* » n'a donc pas eu d'impact chez les francophones.

	GROUPE DS Panel de francophones			GROUPE DCS Panel de francophones			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	29	2,66	0,814	29	2,72	0,455	-0,398	43,933	0,692
dessins EN	29	0,21	0,774	29	0,03	0,186	1,167	31,216	0,252
dessins CONT	29	0,14	0,351	29	0,21	0,412	-0,686	54,608	0,496
dessins cibles mal positionnés	29	0,28	0,702	29	0,17	0,468	0,660	48,801	0,512
dessins EN mal positionnés	29	0,00	0,000	29	0,03	0,186	-1,000	28,000	0,326
absence de réponse	29	0,00	0,000	29	0,03	0,186	-1,000	28,000	0,326

**Tableau 23 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue C**

**Il semble donc que la présence (ou l'absence) d'un CSE n'ait pas impliqué de modification significative dans la sélection des dessins par le panel de francophones, quel que soit le dialogue.**

Rappelons toutefois que l'observation des suites de trois dessins cibles avait amené à observer une influence significative du « contexte sonore environnemental » sur ce paramètre, pour le dialogue A.

Si nous nous concentrons maintenant sur le panel d'apprenants débutants et sur les dessins sélectionnés pour l'ensemble des trois dialogues, alors il apparaît que la variable « *contexte sonore environnemental* » a impliqué des différences significatives aux niveaux de la sélection de dessins cibles et de dessins EN (présentant une variation de la situation d'énonciation)<sup>67</sup>.

	GROUPE DS Panel d'apprenants débutants			GROUPE DCS Panel d'apprenants débutants			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	63	1,56	1,241	63	2,41	0,854	-4,514	109,98	0,000
dessins EN	63	0,73	1,050	63	0,29	0,633	2,876	101,80	0,005
dessins CONT	63	0,19	0,435	63	0,11	0,317	1,171	113,37	0,244
dessins EN-CONT	63	0,10	0,296	63	0,06	0,246	0,655	119,96	0,514
dessins cibles mal positionnés	63	0,37	0,789	63	0,56	0,876	-1,282	122,68	0,202
dessins EN mal positionnés	63	0,19	0,592	63	0,08	0,272	1,354	87,16	0,179
absence de réponse	63	0,38	0,923	63	0,11	0,479	2,059	93,11	0,042

**Tableau 24: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - DV) chez le panel d'apprenants débutants et pour l'ensemble des dialogues**

<sup>67</sup> Rappelons en outre que les contextes sonores environnementaux relèvent plus d'une fonction situationnelle que de transcodage.

Si nous nous concentrons maintenant sur les dialogues un à un, nous pouvons constater que pour le dialogue A, la variable a donné lieu à un écart significatif au seul niveau de la sélection de dessins cibles (ne présentant pas de variation).

	GROUPE DS Panel d'apprenants débutants			GROUPE DCS Panel d'apprenants débutants			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	21	1,24	1,221	21	2,19	1,030	-2,732	38,90	0,009
dessins EN	21	1,19	1,209	21	0,62	0,921	1,723	37,36	0,093
dessins CONT	21	0,10	0,301	21	0,10	0,301	0,000	40,00	1,000
dessins EN-CONT	21	0,14	0,359	21	0,10	0,301	0,466	38,83	0,644
dessins cibles mal positionnés	21	0,19	0,680	21	0,33	0,796	-0,626	39,04	0,535
dessins EN mal positionnés	21	0,38	0,740	21	0,10	0,301	1,639	26,43	0,113
absence de réponse	21	0,19	0,680	21	0,00	0,000	1,284	20,000	0,214

**Tableau 25 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue A**

Pour le dialogue B, nous pouvons observer que la variable « *contexte sonore environnemental* » a donné lieu à des différences significatives ( $p < 0,05$ ) aux niveaux de la sélection globale des dessins cibles (sans variation, quel que soit le positionnement sur l'axe chronologique), comme pour le dialogue A et du mauvais positionnement de certains dessins cibles. En effet, les membres du groupe DCS ont sélectionné plus de dessins cibles, mais en ont aussi mal positionnés un plus grand nombre. En outre, il apparaît que si l'on compare les dessins cibles bien positionnés par les groupes DS ( $n=21$   $M= 1,29$   $SD=1,309$ ) et DCS ( $n=21$   $M=1,86$   $SD=1,153$ ), aucune différence significative n'apparaît ( $t(39,367)=1,501$   $p=0,141$ ). Cette observation laisse entendre que la variable « *contexte sonore environnemental* » a eu un impact sur les dessins cibles uniquement dans le cadre de leur mauvais positionnement mais amenant aussi à leur plus grande sélection.

	GROUPE DS Panel d'apprenants débutants			GROUPE DCS Panel d'apprenants débutants			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	21	1,38	1,284	21	2,33	0,856	-2,828	34,86	0,008
dessins EN	21	0,67	1,017	21	0,19	0,402	1,996	26,12	0,560
dessins CONT	21	0,33	0,577	21	0,10	0,301	1,676	30,11	0,104
dessins EN-CONT	21	0,10	0,301	21	0,10	0,301	0,000	40,00	1,000
dessins cibles mal positionnés	21	0,10	0,301	21	0,48	0,680	-2,349	27,56	0,026
dessins EN mal positionnés	21	0,14	0,655	21	0,10	0,301	0,303	28,08	0,764
absence de réponse	21	0,52	1,123	21	0,29	0,784	0,797	35,74	0,431

**Tableau 26 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue B**

Pour le dialogue C (comme pour le dialogue A), une différence significative ( $p < 0,05$ ) est observée au niveau de la sélection de dessins cibles (les membres du groupe DCS en ont sélectionné plus que ceux du groupe DS).

	GROUPE DS Panel d'apprenants débutants			GROUPE DCS Panel d'apprenants débutants			t	ddl	Sig. (bilatérale)
	n	M	Écart-type	n	M	Écart-type			
dessins cibles	21	2,05	1,117	21	2,71	0,561	-2,445	29,47	0,021
dessins EN	21	0,33	0,730	21	0,05	0,218	1,718	23,54	0,099
dessins CONT	21	0,14	0,359	21	0,14	0,359	0,000	40,00	1,000
dessins EN-CONT	21	0,05	0,218	21	0,00	0,000	1,000	20,00	0,329
dessins cibles mal positionnés	21	0,81	1,030	21	0,86	1,062	-0,147	39,96	0,884
dessins EN mal positionnés	21	0,05	0,218	21	0,05	0,218	0,000	40,00	1,000
absence de réponse	21	0,43	0,926	21	0,05	0,218	1,835	22,21	0,080

**Tableau 27 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue C**

**Il semble donc que la présence (ou l'absence) d'un contexte sonore environnemental a donné lieu à des différences significatives dans la sélection des dessins cibles pour le panel d'apprenants débutants, et ce, pour les trois dialogues.**

### 2.3. Compréhension et charge cognitive :

Comme nous l'avons dit plus haut, les temps de réponse sont considérés comme des indices de la charge cognitive impliquée dans la tâche. Pour le Classe-Images le chronomètre s'est enclenché au moment de l'apparition des dessins, pour s'arrêter à la validation des participants. Plusieurs étapes de la tâche sont donc comptabilisées dans ces temps de réponse : l'analyse des dessins, le rappel (mental) du dialogue et la construction de la série de dessins. Plus les temps de réponse seront longs, plus nous considérerons que la charge cognitive aura été importante. À l'inverse un temps de réponse court sera un indice d'une moindre charge cognitive engendrée par l'analyse

des éléments à disposition pour réaliser la tâche demandée. Les différences de temps de réponse seront examinées autour de trois variables : le niveau d'expertise linguistique des sujets, la présence (ou l'absence) d'un contexte sonore environnemental, le dialogue A, B ou C. Afin de vérifier la significativité statistique des différences observées, des tests U de Mann-Whitney ont été réalisés.

### 2.3.1. La variable « expertise linguistique »

Notre première hypothèse opérationnelle est que la variable « *expertise linguistique* » (VI) impliquera une différence dans les temps de réponse des participants (VD). Le temps de réponse est considéré comme un indice de la charge cognitive. Nous comparerons les TR enregistrés par les groupes DS (condition d'écoute sans contexte sonore environnemental – condition contrôle) du panel de francophones et du panel d'apprenants débutants. Nous supposons que les temps de réponse enregistrés par les francophones seront inférieurs à ceux des apprenants débutants du fait de leur expertise linguistique.

Si l'on considère l'ensemble des trois dialogues, alors il apparaît que l'expertise linguistique a donné lieu à une différence significative entre les temps de réponse enregistrés par le panel de francophones (n=87, Rang moyen = 65,14) et d'apprenants débutants (n=63 Rang moyen = 89,81)  $Z = -3,433$   $p = 0,001$  puisque **le panel d'apprenants a enregistré des TR plus longs que le panel de francophones.**

Ainsi, malgré cela, il apparaît que seul le dialogue A présente un écart significatif. Il semble que pour ce dialogue la différence de niveau d'expertise a impliqué une différence dans la charge cognitive impliquée dans la tâche. Toutefois, pour les autres dialogues, aucune différence significative n'a été observée.

	francophones DS		apprenants DS		U de Mann-Whitney	Z	Sig. (bilatérale)
	n	Rang moyen	n	Rang moyen			
dialogue A	29	20,79	21	32,00	168,00	-2,683	0,007
dialogue B	29	22,62	21	29,48	221,00	-1,641	0,101
dialogue C	29	22,55	21	29,57	219,00	-1,681	0,093

**Tableau 28: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les temps de réponses (VD) enregistrés par les panels de francophones et d'apprenants débutants**

Parallèlement, lorsque l'on considère les temps de réponses associés à la réalisation de suites de trois dessins cibles, c'est-à-dire lorsque les participants ont « bien compris », alors il apparaît qu'aucun dialogue ne présente d'écart significatif. En outre, si l'on considère l'ensemble des trois dialogues, il apparaît que l'écart entre les francophones ( $n=58$ , Rang moyen = 34,55) et les apprenants débutants ( $n=17$  Rang moyen = 49,76) n'est pas significatif  $Z = -2,531$   $p = 0,110$  pour la réalisation de suites de trois dessins cibles.

Dans ces situations de « réussite » du test, les différents niveaux d'expertise linguistique ne semblent pas avoir eu d'impact sur la charge cognitive impliquée dans la tâche.

	francophones DS		apprenants DS		U de Mann- Whitney	Z	Sig. (bilaté- rale)
	n	Rang moyen	n	Rang moyen			
dialogue A	14	7,79	4	15,50	4,00	-2,549	0,110
dialogue B	23	14,04	6	18,67	47,00	-1,184	0,236
dialogue C	21	13,48	7	17,57	52,00	-1,141	0,254

**Tableau 29 : Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les temps de réponse (VD) enregistrés par les panels de francophones et d'apprenants débutants et associés aux suites de 3 dessins cibles**

### **2.3.2. La variable « contexte sonore environnemental »**

Notre deuxième hypothèse opérationnelle, au niveau de la charge cognitive, est que la variable « *contexte sonore environnemental* », en apportant un plus grand nombre d'indices à traiter, et en présentant un certain degré de redondance complémentaire entre le dialogue et les sons environnementaux impliquera une différence dans la charge cognitive impliquée dans la tâche par les groupes DS et DCS. Cette différence se traduira, selon nous, par une augmentation des temps de réponse (indices de la charge cognitive impliquée dans la tâche) pour les groupes DCS.

Pour le panel de francophones, si nous considérons les trois dialogues ensemble (sans tenir compte de leurs spécificités), alors aucune différence significative n'apparaît entre les groupes DS (n=87, Rang moyen = 83,44) et DCS (n=87 Rang moyen = 91,56) pour le panel de francophones  $Z = -1,064$   $p = 0,287$

Parallèlement, nous pouvons observer que cette variable a impliqué une différence significative ( $p < 0,05$ ) au niveau du dialogue A. Pour les autres dialogues, cette variable n'a pas eu d'incidence significative sur les temps de réponse globaux des participants francophones.

	francophones DS		francophones DCS		U de Mann-Whitney	Z	Sig. (bilatérale)
	n	Rang moyen	n	Rang moyen			
dialogue A	29	23,34	29	35,66	242,00	-2,776	0,006
dialogue B	29	32,38	29	26,62	337,00	-1,299	0,194
dialogue C	29	28,55	29	30,45	393,00	-0,428	0,669

**Tableau 30: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel de francophones**

Si nous nous concentrons, toujours pour le panel de **francophones**, sur les temps de réponse associés aux seules constructions de suites de trois dessins cibles (« bonnes réponses »), alors, nous observons que la comparaison des groupes DS (n=58, Rang moyen = 57,86) et DCS (n=67 Rang moyen = 67,45) pour les trois dialogues considérés ensemble n'a pas non plus donné lieu à une différence significative  $Z = -1,475$   $p = 0,140$ .

Si nous considérons les dialogues un à un, il apparaît qu'encore une fois, seul le dialogue A a présenté une différence significative entre les temps de réponse enregistrés par les membres du groupes et ceux du groupes DCS.

	francophones DS		francophones DCS		U de Mann-Whitney	Z	Sig. (bilatérale)
	n	Rang moyen	n	Rang moyen			
dialogue A	14	11,50	21	22,33	56,00	-3,064	0,002
dialogue B	23	26,30	25	22,84	246,00	-0,856	0,392
dialogue C	21	21,05	21	21,95	211,00	-0,239	0,811

**Tableau 31: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel de francophones et associés aux suites de 3 dessins cibles**

Pour le panel **d'apprenants débutants**, la variable « *contexte sonore environnemental* » n'a pas eu d'impact sur les temps de réponse globaux impliqués dans la tâche. Les trois dialogues considérés ensemble n'ont pas donné lieu à une différence significative  $Z = -1,193$   $p = 0,233$  entre les groupes DS ( $n = 63$ , Rang moyen = 59,62) et DCS ( $n = 63$  Rang moyen = 67,38).

Les dialogues considérés un à un n'ont pas non plus donné lieu à des différences significatives.

	apprenants DS		apprenants DCS		U de Mann-Whitney	Z	Sig. (bilatérale)
	n	Rang moyen	n	Rang moyen			
dialogue A	21	18,05	21	24,95	148,00	-1,824	0,068
dialogue B	21	21,29	21	21,71	216,00	-0,113	0,910
dialogue C	21	20,95	21	22,05	209,00	-0,289	0,772

**Tableau 32: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel d'apprenants débutants**

Parallèlement, si l'on considère uniquement les temps de réponse associés aux constructions de suites de trois dessins cibles (*bonnes réponses*), alors on constate également que la variable observée n'a pas impliqué de différence significative.

Si l'on considère les trois dialogues ensemble, alors il apparaît que la variable « contexte sonore environnemental » n'a pas eu d'impact significatif sur les temps de réponse enregistrés par le groupe DS (n=17, Rang moyen = 23,71) et DCS (n=29 Rang moyen = 23,38) :  $Z = -0,800$   $p = 0,937$ .

Pris un à un, les dialogues n'ont pas non plus présenté de différence significative à ce niveau.

	apprenants DS		apprenants DCS		U de Mann-Whitney	Z	Sig. (bilatérale)
	n	Rang moyen	n	Rang moyen			
dialogue A	4	7,50	10	7,50	20,00	0,000	1,000
dialogue B	6	8,50	9	7,67	24,00	-0,354	0,724
dialogue C	7	8,57	10	9,30	32,00	-0,293	0,770

**Tableau 33 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel d'apprenants débutants et associés aux suites de 3 dessins cibles**





## Partie 4 : Discussion

Notre recherche s'inscrit dans la démarche de la didactique cognitive des langues étrangères (Billières & Spanghero-Gaillard, 2005); elles portent plus spécifiquement sur la compréhension orale de situations dialogales. Cette compétence, primordiale dans la communication humaine est, comme nous l'avons exposé (voir *Contexte didactique* p.17), bien souvent sous-investie, tant dans l'enseignement/apprentissage des langues que dans les recherches en didactique. Dans ce cadre, et face aux besoins que nous avons identifiés dans notre pratique enseignante, nous avons choisi de nous intéresser aux supports de CO les plus fréquemment utilisés en classe de langue : les dialogues pédagogiques. Cette démarche nous a amenée à élaborer un protocole expérimental nouveau, tant au niveau du matériau de test qu'il fait intervenir (voir *Élaboration du matériel expérimental* p.121), qu'au niveau de sa méthodologie (voir p.165). Nous avons précédemment présenté les résultats que ce test nous a permis de recueillir. Nous allons maintenant les discuter.

Dans la deuxième partie de ce chapitre, nous nous concentrerons sur le matériel (dont l'élaboration a été présentée dans le chapitre *Expérimentation principale* p.165) utilisé dans le Classe-Images. Nous regarderons si ces supports (linguistiques, sonores environnementaux et visuels), et leurs spécificités, ont eu des implications sur les performances enregistrées.

Enfin, nous terminerons ce chapitre par une discussion théorique et didactique portant sur l'ensemble de notre démarche scientifique et ses implications dans notre pratique enseignante.

## 1. Discussion des résultats du Classe-Images

Dans le cadre de cette discussion, nous aborderons à la fois les résultats recueillis et la méthodologie que nous avons suivie. Au niveau des résultats, différents aspects appellent, selon nous, une discussion.

Dans la mesure où les résultats ont été présentés (voir p.187) de manière détaillée dans la partie expérimentale de ce travail, nous en ferons ici une synthèse afin de faire ressortir les aspects les plus importants pour nos objectifs de recherche. À la suite de cette synthèse (basée sur les *Hypothèses* formulées p.105), nous nous interrogerons sur les implications théoriques et didactiques de ces résultats.

### 1.1. Synthèse des principaux résultats

Comme nous l'avons vu dans le chapitre *Hypothèses* (voir p.105), deux principaux axes de questionnement se dégagent dans notre partie expérimentale : nous avons cherché à mieux comprendre ce qu'être débutant en langue implique au niveau de la compréhension orale, et dans quelle mesure les contextes sonores environnementaux, qui peuvent accompagner les dialogues pédagogiques, participent à la compréhension de ces derniers.

Le premier axe de questionnement renvoie aux spécificités du public apprenant sur lequel nous avons centré nos observations. Nous nous attendons à ce que les apprenants de FLE présentent des difficultés de compréhension en langue cible, différentes de celles de francophones natifs (considérés comme des experts linguistiques), en raison, notamment, du faible niveau de compétences en langue des premiers. Toutefois, nous ne saurions considérer les apprenants débutants (et *a fortiori* les adultes, comme dans nos expérimentations) comme des individus ne disposant d'aucune connaissance antérieure, fondamentale dans le processus de compréhension (voir chapitre ... *aux études sur la compréhension* p.81). Ainsi, il s'est agi de voir dans quelle mesure le déficit de connaissances et de compétences en langue cible entraîne des difficultés de compréhension dans cette langue, lorsque l'apprenant dispose néanmoins d'un certain nombre de connaissances sur le monde et de savoir-faire linguistiques et pragmatiques dans la (ou les) langue(s) qu'il maîtrise déjà.

Le deuxième axe renvoie à notre observation et questionnement de départ : de nombreux dialogues pédagogiques proposent, en accompagnement de la dimension purement linguistique, un contexte sonore environnemental (désigné par le sigle CSE). Ce CSE reste néanmoins trop souvent inexploité, faute d'indications

quant à son rôle et sa potentielle exploitation dans le cadre d'une séance d'enseignement/apprentissage. Par conséquent, nous nous sommes interrogée sur le rôle de ces éléments non-linguistiques dans la compréhension des dialogues, tant par des apprenants débutants que par des francophones natifs, considérés, eux, comme des experts linguistiques.

Dans la partie résultats de notre expérimentation principale (voir p.187), nous nous sommes concentrée sur les différents types de résultats que nous avons pu obtenir. Dans cette nouvelle partie, nous allons en faire une synthèse de manière à répondre à nos questionnements de départ.

### ***1.1.1. Qu'est-ce qu'être apprenant débutant en langue***

L'objectif de cet axe de questionnement était de voir dans quelle mesure des locuteurs natifs de la langue cible et des locuteurs apprenants débutants dans cette même langue présentent des performances différentes en tâche de compréhension orale. Pour ce faire, nous avons observé les résultats obtenus *via* le Classe-Images, et comparé les performances des groupes de francophones et d'apprenants dans la tâche de compréhension des dialogues entendus seuls, *i.e.*, sans contexte sonore environnemental. Notre hypothèse théorique était que le niveau d'expertise linguistique (VI) impliquerait de meilleures performances pour le groupe de participants francophones natifs que pour le groupe de participants apprenants débutants. Globalement, il apparaît que cette hypothèse a été vérifiée : on a pu observer que les participants francophones ont globalement mieux « réussi » la tâche de recomposition des dialogues à partir des dessins présentés que les participants apprenants débutants.

Notre première hypothèse opérationnelle portait sur la compréhension des dialogues en termes binaires (compris/pas compris). Comme dit plus haut, nous nous attendions à ce que **les participants francophones comprennent mieux les dialogues que les participants apprenants débutants**, c'est-à-dire que les premiers élaborent plus de suites de trois dessins cibles que les seconds. Rappelons dans ce cadre que nous considérons la compréhension comme « bonne » lorsqu'elle correspond à la signification envisagée lors de l'élaboration du matériel. Dès lors, la « bonne compréhension » renvoie à la coïncidence des sens élaborés par l'enseignant/expérimentateur et par l'apprenant/participant. En outre, les écarts enregistrés sont systématiquement significatifs au plan statistique c'est-à-dire à la fois pour les trois dialogues considérés ensemble et pour chacun, isolément. Ainsi,

**notre première hypothèse opérationnelle a bien été validée.** Par ailleurs, considérant les dessins cibles sélectionnés (et correctement positionnés sur l'axe chronologique), qu'ils participent ou non à des suites de trois, nous avons également pu observer des écarts statistiquement significatifs et systématiques (pour l'ensemble des trois dialogues comme pour chaque dialogue considéré isolément) en faveur des participants francophones : ces derniers ont « mieux compris la suite des actions » que les participants apprenants débutants.

Nos deuxième et troisième hypothèses opérationnelles portaient sur l'appréhension de la situation d'énonciation et du contenu propositionnel par les participants. **Nous avons avancé l'hypothèse de plus grandes difficultés pour le groupe d'apprenants débutants que pour le groupe de participants francophones natifs.** Nous avons pu observer que les participants francophones ont globalement sélectionné moins de dessins présentant des facteurs de variation. Si nous considérons l'ensemble des trois dialogues, c'est-à-dire sans prendre en compte les différences inter-dialogues, alors des écarts statistiquement significatifs apparaissent au niveau de la sélection des dessins EN, CONT, EN-CONT. **Les hypothèses opérationnelles avancées ont donc été validées.** Toutefois, ces écarts ne s'observent pas de manière systématiquement significative, dès lors que l'on considère les dialogues un à un. Nous pouvons donc conclure que, **si le niveau d'expertise linguistique semble avoir un impact sur l'appréhension de la situation d'énonciation et du contenu propositionnel, celui-ci semble variable selon l'objet à comprendre.** En effet, les dialogues A et C n'ont pas présenté d'écarts significatifs au niveau de la sélection des dessins EN, CONT et EN-CONT. En revanche pour le dialogue B, nous avons pu observer des écarts significatifs pour la sélection des dessins EN et CONT. Dans ce cadre, et bien que les trois dialogues aient été considérés d'un niveau de difficulté équivalent, peut-on considérer qu'ils n'impliquent pas la même complexité en tâche de compréhension ? ou au contraire des différences de performances dans le Classe-Images relèvent-elles des supports utilisés (corpus d'images notamment) ? Ces aspects seront développés dans le paragraphe *Matériel utilisé* de ce chapitre (p.230)

Bien que cela n'ait pas fait partie de nos hypothèses, nous avons également pu observer un écart significatif au niveau des absences de réponse, c'est-à-dire quand les participants ont préféré laisser une case vide à la sélection d'un dessin. Ce comportement a été observé plus fréquemment chez les participants apprenants débutants que chez les francophones. Pourtant de fortes variations inter-dialogues restent présentes. Ces absences de réponses peuvent être interprétées de différentes manières. Nous pouvons avancer deux hypothèses explicatives : 1) les participants

ont considéré, au vu des dessins, qu'un seul ou deux suffisaient à reconstituer le dialogue ; ou que 2) aucun des dessins proposés ne correspondait à la signification qu'ils voulaient exprimer. Il semble donc que ces « erreurs » (dans le sens de non-adéquation entre les attentes formulées dans l'élaboration du test et les productions des participants) ne correspondent pas nécessairement à des difficultés de compréhension des dialogues mais sont potentiellement liées au corpus d'images que nous avons constitué. Nous pouvons également envisager qu'à mesure qu'augmente l'expertise linguistique, le besoin de détailler la reformulation (ici élaborée avec des dessins) diminue. Cette interprétation serait à rapprocher des processus de schémas d'action : à mesure que l'individu devient coutumier d'une série d'actions, plus il est en capacité de la reconstituer à partir d'une seule d'entre elles. Dit autrement, le seuil d'activation des actions baisserait en même temps qu'augmente leur fréquence (il s'agirait, en quelque sorte d'une forme de potentialisation à long terme des connaissances procédurales). Cette hypothèse pourrait expliquer le fait que les experts linguistiques ont montré une plus grande propension aux non-réponses (cases laissées vides) que les apprenants débutants qui ont généralement sélectionné trois dessins (conformément à la consigne). En outre, ce phénomène pourrait être lié aux habitudes d'apprentissage de ces derniers puisqu'en classe de langue, l'enseignant encourage le plus souvent les apprenants à préciser leurs propos.

Notre quatrième hypothèse opérationnelle portait sur l'appréhension de la dimension chronologique des dialogues. A nouveau, nous avons supposé que les participants francophones présenteraient moins de difficulté(s) à réaliser cette tâche que les participants apprenants débutants. Cette hypothèse a été vérifiée en comparant les dessins (cibles et avec une variation de la situation d'énonciation) mal positionnés, au niveau des trois dialogues considérés ensemble. Nous avons pu observer que l'écart à ce niveau était significatif pour le mauvais positionnement des dessins EN, mais pas pour les dessins cibles. Cette hypothèse est donc partiellement validée. En outre, si nous considérons l'ensemble des dessins mal positionnés (cibles et EN), l'écart entre les groupes francophones et le groupe d'apprenants est également validé ( $t(99,280)=-2,916$  ;  $p=0,004$  ; avec une hypothèse de variances inégales). **Il apparaît donc que le niveau d'expertise linguistique a également eu une influence sur le positionnement des dessins sélectionnés** (les participants francophones ont présenté moins de difficultés que les participants apprenants débutants). **Toutefois, cet écart semble dépendant du dialogue mis en images.** En effet, si nous nous concentrons sur le dialogue A, nous pouvons observer une différence significative au niveau du positionnement des dessins EN. Pour le dialogue C, nous observons une différence significative au niveau du positionnement des

dessins cibles. En revanche, pour le dialogue B, les écarts observés ne sont pas statistiquement significatifs, mais ils laissent apparaître (comme pour les dialogues A et C) que les francophones ont éprouvé moins de difficulté avec le positionnement sur l'axe chronologique. Les variations inter-dialogues que nous observons seront abordées plus loin dans ce chapitre (voir p.237)

Enfin, notre cinquième hypothèse opérationnelle portait que les temps de réponse. Nous avons avancé l'hypothèse de TR plus courts pour les participants francophones, notamment en raison de la diminution de la charge cognitive induite par l'augmentation du niveau d'expertise linguistique. Nous avons observé un écart significatif au niveau des temps de réponses enregistrés entre les deux panels de population interrogés, pour les trois dialogues considérés ensemble. Considérés isolément, seul le dialogue A a impliqué une augmentation significative de la durée de traitement pour le groupe d'apprenants débutants par rapport à celle des francophones natifs. Ce résultat nous amène à penser que **l'augmentation de la charge cognitive induite par le déficit de connaissances en langue cible des apprenants débutants a entraîné une augmentation de la durée des traitements**. Ainsi, nous pouvons avancer l'hypothèse que cette augmentation a impliqué une surcharge cognitive puisqu'elle s'est traduite à la fois au niveau de la qualité des réponses produites, et des TR. Encore une fois, le fait que cet écart ne soit visible que pour le dialogue A ou pour les trois dialogues considérés ensemble sera discuté plus bas dans ce chapitre. Toutefois, nous pouvons également avancer l'hypothèse interprétative que les tendances que nous observons auraient pu représenter des écarts statistiquement significatifs si nous avions interrogé un plus grand nombre de participants.

**Globalement, nous avons donc vu que le caractère novice en français des apprenants débutants a impliqué de plus nombreuses difficultés dans la tâche du Classe-Images pour ce groupe que pour le groupe de participants francophones natifs, et ce sur la totalité des aspects observés (appréhension de la situation d'énonciation, du contenu propositionnel, de la dimension chronologique et des temps de réponse).** Cette partie des résultats nous a amenée à nous interroger sur les *profils d'erreurs* (voir p.224). En effet, il ne s'agit plus de savoir qui des francophones ou des apprenants débutants a présenté les meilleurs résultats mais plutôt de s'interroger sur les types d'erreurs commises. Dit autrement, les sources de difficultés sont-elles les mêmes quel que soit le niveau d'expertise des individus? En effet, bien que notre première hypothèse théorique (que les francophones présentent de meilleures performances que les apprenants) soit

validée, nous avons pu observer que les membres du groupes de participants francophones présentaient également des difficultés dans cette tâche. Effectivement, alors que la tâche de compréhension aurait *a priori* dû être simple pour les participants francophones, nous observons que leur taux de réussite (si l'on considère les suites de trois dessins cibles) n'est que de 66,67% sur l'ensemble des trois dialogues (48,27% pour le dialogue A, 79,31% pour le dialogue B et 72,41% pour le dialogue C). Nous pouvons donc conclure à un effet de la tâche sur les performances. Il semble en effet que la tâche, ait été complexe à réaliser, même pour les participants francophones. Dans ce cadre, nous pensons qu'une des difficultés majeures se situe dans la tâche, pas nécessairement dans les dialogues, mais peut-être plutôt dans l'interaction des différents supports qui composent le Classe-Images (langagiers, sonores environnementaux et visuels, voir p.230).

**En conclusion, nous pouvons affirmer que le niveau d'expertise linguistique a eu une influence sur l'ensemble des variables que nous avons observées.** Cette observation semble aller dans le sens de nos hypothèses et des travaux réalisés dans ce domaine, toutefois, certains résultats appellent la formulation de nouvelles hypothèses, notamment explicatives. Par exemple, nous pouvons remarquer que l'appréhension de la situation d'énonciation a présenté une forte variation en fonction du dialogue. Or, cette dimension peut être considérée comme la compréhension d'éléments para, voire extralinguistiques. Dès lors, dans le cadre de l'écoute de dialogues présentés sans contexte sonore environnemental, cette partie de l'élaboration de la signification repose essentiellement sur la formulation d'inférences faisant fortement intervenir des connaissances (pragmatiques) antérieures du compreneur, et la micro-proposition élaborée à partir des indices linguistiques repérés et interprétés. Il semble donc que chacun des dialogues que nous avons proposés présente des spécificités qui ont entraîné des modifications dans cette formulation d'inférences par les individus compreneurs. Ces spécificités seront discutées plus bas dans ce chapitre, dans le cadre de la *Discussion méthodologique* (voir p.229). Par ailleurs, nous pouvons également penser que la présence d'un CSE (et à plus forte raison lorsqu'il a un fort caractère situationnel), permettra une plus grande facilité dans l'appréhension de cette dimension, en impliquant de plus nombreux indices extralinguistiques (transmis de manière non-linguistique).

### ***1.1.2. Le rôle du contexte sonore environnemental***

Conformément à nos objectifs de départ, **notre hypothèse théorique était qu'une scène sonore, telle que celles que nous avons réalisées, en position**

**contextuelle par rapport à un court dialogue avec lequel elle entretient un rapport de redondance complémentaire (Spanghero-Gaillard, 2008), pourrait faciliter la compréhension globale du support, et impliquer une amélioration des performances en compréhension** par rapport à la confrontation avec un dialogue seul. Dans ce cadre, nous avons également avancé la sous-hypothèse selon laquelle cette amélioration des performances serait observable chez les groupes de francophones et d'apprenants débutants, sans pour autant impliquer nécessairement la même marge de progression. Dit autrement, notre hypothèse était que la variable indépendante « présence ou absence d'un contexte sonore environnemental » n'entraînerait pas les mêmes effets sur la variable dépendante *performances*. Afin de vérifier cette hypothèse, nous avons comparé les performances des participants qui ont entendu les dialogues seuls à celles des participants qui ont entendu les dialogues accompagnés d'un CSE pour les panels de francophones et d'apprenants débutants.

**Notre première hypothèse opérationnelle à ce niveau était que la présence d'un CSE aurait un effet positif sur la compréhension des dialogues si celle-ci est considérée en termes binaires (compris / pas compris).** Dit autrement, nous avons avancé l'hypothèse d'une augmentation du nombre de Réponses Attendues (RA suites de trois dessins cibles construites) par les participants lorsque ceux-ci ont pu écouter les dialogues accompagnés d'un CSE. **Pour le panel de francophones, nous avons pu observer que les participants qui ont entendu les dialogues avec leurs contextes sonores ont globalement produit plus RA que les participants qui ont entendu les dialogues seuls, sans pour autant que ces écarts soient statistiquement significatifs** pour l'ensemble des dialogues (pour les trois dialogues considérés ensemble, l'écart n'est pas significatif ; les écarts sont significatifs pour les dialogues A et C, mais pas pour le dialogue B). **Parallèlement, pour le panel d'apprenants débutants, les écarts entre le groupe qui a entendu les dialogues seuls et celui qui a entendu les dialogues accompagnés d'un contexte sonore environnemental sont plus importants : les participants qui ont entendu les dialogues avec un CSE ont construit significativement plus de suites de trois dessins cibles que ceux qui ont entendu les dialogues seuls.** Ces écarts sont statistiquement significatifs à la fois pour les trois dialogues considérés ensemble et isolément. **Nous pouvons donc conclure que pour les apprenants débutants la présence d'un CSE représente une aide à la compréhension** (c'est-à-dire que les significations construites concordent plus fréquemment avec les significations envisagées dans l'élaboration des supports lorsque les participants ont pu écouter les dialogues avec un CSE). **Cette aide est en outre plus importante pour les apprenants débutants que pour les natifs de la langue cible, ce qui**

**pourrait laisser penser que le niveau d'expertise linguistique et la présence (ou l'absence) d'un CSE présentent des effets d'interaction** (voir p.224). Nous pouvons, dans ce cadre, émettre l'hypothèse explicative selon laquelle le niveau d'expertise des participants impliquerait une perception-compréhension différente du support. Lorsque l'individu est expert, c'est-à-dire lorsqu'il peut comprendre les différentes sources (ici le dialogue et le CSE) séparément, il perçoit une certaine redondance dans le support complexe. Or, la redondance entraîne souvent une augmentation de la charge cognitive (Chanquoy *et al.*, 2007; Tricot, 1998). Parallèlement, l'individu novice doit passer par une nouvelle étape d'intégration mentale des différentes sources, elle aussi coûteuse cognitivement mais qui peut néanmoins permettre de meilleures performances (notamment au niveau qualitatif). Dès lors, nous pouvons penser que le CSE supplée le déficit linguistique des apprenants, en entretenant un rapport de redondance complémentaire avec les éléments linguistiques du dialogue. Ces différentes stratégies de compréhension impliqueraient différents types d'opérations et donc différentes charges cognitives impliquées dans ces traitements.

Par ailleurs, les dessins cibles sélectionnés par les participants peuvent également être considérés en dehors des constructions des suites de trois, c'est-à-dire comme des parcelles de réponse. En effet, nous observons que **pour le panel de francophones, la présence d'un CSE n'a pas impliqué d'écart significatif, alors que pour le panel d'apprenants débutants, la présence de scènes sonores en position contextuelle a impliqué des écarts statistiquement significatifs à la fois pour chaque dialogue considéré isolément, et pour les trois dialogues considérés ensemble**. Nous pouvons donc conclure que la présence d'un contexte sonore environnemental a facilité la compréhension des dialogues sans que, pour autant, l'effet de cette variable (présence ou absence d'un CSE) ne soit systématique et/ou équivalent pour chacun des panels. **Il semble donc que la présence d'un contexte sonore environnemental ait un effet différent en fonction du niveau d'expertise linguistique des participants et implique un degré de redondance complémentaire, entre les dialogues et les CSE, différent dans son rôle d'aide.**

Dans le prolongement de ces hypothèses, nos deuxième et troisième hypothèses opérationnelles portaient sur l'appréhension de la situation d'énonciation et du contenu propositionnel. Nous avons avancé l'hypothèse que le contexte sonore faciliterait l'appréhension de ces deux paramètres. **Nous avons pu constater que, pour le panel de francophones, la présence d'un CSE n'a pas impliqué de différence statistiquement significative dans l'appréhension de ces deux**

**dimensions, que ce soit en considérant les trois dialogues ensemble ou séparément. Toutefois, considérant les dessins sélectionnés par les participants du panel d'apprenants débutants, nous pouvons observer que la présence d'un CSE a permis aux participants de « mieux appréhender » la situation d'énonciation.** Si nous considérons les trois dialogues ensemble, il apparaît que l'écart du nombre de dessins avec une variation de la situation d'énonciation sélectionnés par le groupe qui a entendu les dialogues seuls et celui qui a entendu les dialogues accompagnés d'un CSE est statistiquement significatif. **Cet impact est a priori fortement lié au caractère situationnel des scènes sonores en position contextuelle** (voir le paragraphe *Matériel utilisé* dans la *Discussion Méthodologique*, p.230). Toutefois, cet écart n'est plus observable dès lors que l'on considère les dialogues un à un.

Parallèlement, si nous nous penchons sur la sélection de dessins avec une variation du contenu propositionnel, par le groupe qui a écouté les dialogues seuls, et celui qui a écouté les dialogues avec un CSE, alors nous notons que les écarts ne sont pas statistiquement significatifs, que ce soit au niveau des trois dialogues considérés ensemble ou de chaque dialogue considéré isolément. **Il semble donc que la présence d'un CSE n'ait pas entraîné de modifications majeures dans l'appréhension du contenu propositionnel.** Encore une fois, cet aspect peut être lié au caractère situationnel des scènes sonores.

**Il apparaît donc que la présence d'un CSE a eu une influence sur l'appréhension de la situation d'énonciation par le panel d'apprenants débutants.** Nous pouvons imaginer que, dans une situation d'insécurité linguistique, les individus se « raccrochent » à ces éléments pour élaborer une signification. Cette stratégie est, selon nous, fortement liée au fait que l'humain est en perpétuelle recherche de sens, *i.e.*, qu'il cherche toujours à percevoir, comprendre et interpréter son environnement. Aussi, nous pouvons émettre l'hypothèse que les sons environnementaux représentent une aide à l'élaboration du sens en offrant un cadre à cette construction, notamment en proposant des indices sur lesquels l'individu peut s'appuyer<sup>68</sup>, en facilitant l'élaboration d'un modèle de situation. En outre, il semble que le CSE participe à l'élaboration des cinq dimensions que Zwaan et ses collaborateurs (Magliano *et al.*, 1999; Zwaan *et al.*, 1995) ont mis en avant pour l'élaboration de la structure du récit, indispensable à la compréhension (le temps –

---

<sup>68</sup> (sans nécessairement faire de distinction entre le linguistique et le non-linguistique)

qui pourrait être rendu par le CSE ; l'espace – qui pourrait être rendu par le CSE ; la causalité – qui pourrait être rendu par le CSE ; l'intentionnalité ; et les protagonistes objets – qui pourrait être rendu par le CSE) (voir chapitre sur *...aux études sur la compréhension* p.81).

Notre quatrième hypothèse opérationnelle portait sur l'appréhension de la dimension chronologique des dialogues. À ce niveau, il semble que la présence d'un contexte sonore environnemental n'ait impliqué d'effet significatif pour aucun des panels (à l'exception du positionnement des dessins cibles par les participants apprenants débutants, pour le dialogue B). **Nous pouvons donc conclure que la présence d'un CSE ne représente pas nécessairement une aide pour l'appréhension de la chronologie dans une tâche de compréhension orale, de dialogue.** Toutefois, il apparaît aussi que cet aspect reste peu documenté dans la littérature scientifique.

Enfin, notre dernière hypothèse quant à l'influence d'un contexte sonore environnemental portait sur les temps de réponse, indices de la charge cognitive impliquée dans la réalisation d'une tâche. **Pour le panel de francophones, il apparaît que la variable CSE a donné lieu à une différence significative sur les TR pour le dialogue A** (la condition d'écoute des dialogues avec un CSE a impliqué des TR plus longs), **mais pas pour les dialogues B et C ou pour les trois dialogues considérés ensemble. Pour le panel d'apprenants débutants, aucun écart significatif n'est apparu à ce niveau, que l'on considère les dialogues ensemble ou séparément.** Toutefois, nous avons pu observer que les deux panels ont globalement enregistré des augmentations des TR avec la présence d'un contexte sonore environnemental, *i.e.*, que la présence d'un CSE a impliqué une augmentation de la charge cognitive associée à la tâche. Dès lors, de nouveaux questionnements émergent : cette augmentation induit-elle un effet néfaste sur la compréhension ? sur l'apprentissage ? sur la mémorisation ? Dans le cadre du Classe-Images, nous avons vu que **la présence d'un CSE est susceptible de représenter une aide à la compréhension des dialogues. Cette aide peut être imputable à deux fonctions qu'assume ce CSE : d'une part, en proposant des informations complémentaires, il offre de nouveaux indices à l'élaboration du sens, d'un modèle de situation ou d'un paysage sonore (Kintsch, 1998; Kintsch, 1988; Magliano, Zwaan, & Graesser, 1999; Van Den Broek, Young, M., Tzeng, Y., Linderholm, T., 1999; Van Dijk & Kintsch, 1983) ; et d'autre part en permettant une préparation attentionnelle.** L'attention préparatoire (Laberge, 1995) *« contribue aussi à rendre compte de la diminution de la latence des réponses correspondant aux traitements des informations attendues. Elle se manifeste dans une*

*variété de situations durant lesquelles le sujet dispose d'une information préalable sur l'endroit, l'événement ou le moment associés à une cible à laquelle il faudra répondre* » (Camus, 1996:26). Toutefois, il semble ici que cette diminution de la latence soit inférieure à l'augmentation des temps de réponse impliquée par l'augmentation des inputs à traiter.

Enfin, il est apparu, tout au long de cette synthèse que les dialogues, considérés isolément, n'ont pas tous donné lieu aux mêmes résultats. Ainsi, nous nous sommes interrogée sur les spécificités de ces dialogues. Cet aspect sera présenté dans le cadre de notre *discussion méthodologique* (voir p.229). Dans cette partie, nous verrons également que les comparaisons inter-dialogues n'ont pas toujours donné lieu à des écarts significatifs.

## **1.2. Temps de réponse, charge cognitive et réussite**

**Si on se réfère aux temps de réponse, il apparaît que la présence d'un contexte sonore environnemental adjacent a impliqué une augmentation de la charge cognitive induite par la tâche de compréhension et de reconstitution des dialogues à partir de dessins imposés.**

Dans cette partie des résultats, nous avons présenté deux types de temps de réponse. Les premiers correspondent aux temps de réponse enregistrés quelles que soient les réponses produites par les participants. En d'autres termes, ces temps de réponses sont associés à la réalisation de la tâche, que cette réalisation corresponde à nos attentes ou non. Les deuxièmes correspondent aux temps de réponse enregistrés lorsque les participants ont reconstruit des suites de trois dessins cibles, c'est-à-dire lorsqu'ils ont produit des réponses correspondant à nos attentes.

Il s'agit pour nous de savoir si l'augmentation de la charge cognitive est similaire quel que soit le type de réponse produit. En effet, d'un côté l'augmentation de la charge cognitive induite par la tâche pourrait être préjudiciable pour l'apprentissage. D'un autre côté, une telle augmentation indiquerait une plus grande implication de l'individu dans la tâche, une plus grande attention portée à sa réalisation. Par conséquent, nous ne pouvons plus parler de charge cognitive aux conséquences néfastes. Différentes catégories de charge cognitive sont en effet considérées dans les recherches sur la question (Paas, Renkl, & Sweller, 2003, 2004) : la charge intrinsèque, la charge extrinsèque et la charge pertinente. La charge intrinsèque est imposée par la structure de la tâche ou des objets qui la composent. Dans ce cadre, elle ne peut être réduite sans modification de la tâche (ici, la compréhension d'un dialogue et la sélection de dessins). En situation expérimentale

comme en situation de classe, cette charge est donc dépendante de l'objet à comprendre (et/ou apprendre/mémoriser). La charge pertinente, quant à elle, correspond à l'effort fourni pour « *acquérir et automatiser les schémas pertinents* » (Chanquoy *et al.*, 2007:162), cette charge est donc dépendante à la fois de l'objet à comprendre et du sujet compreneur, de ses connaissances antérieures et de ses savoir-faire. Enfin, la charge extrinsèque peut provenir de la tâche ou des interférences que les supports impliquent. Cette charge est donc fortement dépendante de la méthodologie choisie par l'expérimentateur en situation de laboratoire ou par l'enseignant, en situation de classe. En situation d'enseignement, cette charge peut parfois s'avérer inutile, si elle n'entraîne pas d'amélioration des performances, ou de l'apprentissage. Dans le cadre du Classe-Images, nous pouvons considérer que la « lourdeur » de la méthodologie (observée notamment à partir des seuils présentés par les participants francophones) a entraîné une forte charge extrinsèque, peut-être inutile pour les participants. Dit autrement, **on ne peut pas affirmer qu'une augmentation de la charge cognitive est nécessairement néfaste à l'apprentissage ou à la réalisation d'une tâche**. Il revient donc à l'enseignant, en situation d'enseignement/apprentissage de s'interroger sur les différents types de charges induits par la tâche. En effet, les ressources cognitives de chaque individu sont limitées, il faut donc veiller à ce qu'une augmentation de la charge ne soit pas synonyme de surcharge cognitive (notamment en mémoire de travail).

Dans le cadre des résultats du Classe-Images, et en ce qui concerne plus précisément la comparaison des temps de réponse associés à la totalité des productions des francophones natifs et des apprenants débutants, nous avons pu observer que l'augmentation de la durée de la tâche liée au caractère novice de certains participants (*VI niveau d'expertise linguistique des participants*) n'a été significative que dans le cadre du dialogue A considéré isolément ou des trois dialogues considérés ensemble. **Cet aspect nous a amenée à conclure à une augmentation de la charge cognitive inversement proportionnelle au niveau d'expertise des participants**. Si nous nous penchons maintenant sur cette même comparaison mais cette fois restreinte aux seules productions de suites de trois dessins cibles, alors nous observons que ces écarts ne sont plus significatifs, ni au niveau des dialogues considérés isolément, ni au niveau des trois dialogues considérés ensemble. **Nous observons que les temps de réponse des constructions de suites de trois dessins cibles sont inférieurs à ceux associés à l'ensemble des productions**. Cette observation nous amène à penser qu'une bonne

maîtrise de la compréhension orale (observée à partir de la qualité des réponses fournies, pour les participants francophones comme pour les participants apprenants débutants) implique une moindre charge cognitive.

Parallèlement, dans le cadre des comparaisons des groupes qui ont écouté les dialogues avec et sans contexte sonore environnemental (VI *présence ou absence d'un CSE*), nous avons pu observer que **la présence d'un CSE n'a pas entraîné d'augmentation significative des temps de réponse chez les participants francophones**, sur les trois dialogues considérés ensemble. Néanmoins, un écart significatif est apparu au niveau du dialogue A. Si nous nous concentrons maintenant sur les temps de réponse enregistrés pour les seules suites de trois dessins cibles construites par ces participants, alors nous n'observons pas non plus de différences significatives pour les trois dialogues considérés ensemble mais un écart significatif apparaît néanmoins pour le dialogue A considéré isolément. **Au niveau des temps de réponse enregistrés par le panel d'apprenants débutants, nous pouvons constater que la présence d'un contexte sonore environnemental n'a pas entraîné de modification significative des temps de réponse, que ce soit pour les trois dialogues considérés ensemble ou pour chaque dialogue considéré isolément.** Parallèlement, si on considère uniquement les temps de réponse associés aux constructions de suites de trois dessins cibles, on peut constater qu'aucun écart significatif n'apparaît non plus. Comme pour le panel de francophones, nous avons également pu observer que les temps de réponse associés aux constructions de suites de trois dessins cibles sont inférieurs à ceux associés à l'ensemble des productions. **Encore une fois, il semble donc que charge cognitive impliquée dans la tâche et niveau d'expertise des participants sont inversement liés.**

### 1.3. Profils d'erreurs et effets d'interaction

Nous avons vu, dans la synthèse, deux axes de résultats correspondant à nos deux axes d'hypothèses. Il apparaît **que la différence des niveaux d'expertise linguistiques entraîne une différence importante dans la compréhension** telle qu'elle a été évaluée avec le Classe-Images. Nous avons également pu constater que **la présence d'un contexte sonore environnemental représentait une aide à la compréhension.** Il semble en outre que cet impact est particulièrement visible au niveau de l'appréhension de la situation d'énonciation. Dès lors, il convient de s'interroger sur la petite taille des effets que nous avons observés, tant chez les participants francophones que chez les participants apprenants débutants. En effet, les hypothèses que nous avons formulées (voir p.111) laissaient entendre que cette variable indépendante impliquerait de plus fortes différences entre les groupes qui

ont écouté les dialogues avec et/ou sans contexte sonore environnemental. Dans ce cadre il semble opportun de s'interroger sur les *profils d'erreurs* des différents groupes de participants. Nous avons avancé une nouvelle question : si nous avons pu constater que les participants francophones présentaient de meilleures performances que les apprenants débutants et que les participants qui ont écouté le dialogue avec un contexte sonore environnemental ont effectivement présenté de meilleurs scores de compréhension, ces meilleurs scores et performances impliquent-ils que les difficultés rencontrées *in fine* soient les mêmes ? Dit autrement, jusque là, nous avons considéré les sélections de dessins présentant un facteur de variation sur la totalité des dessins sélectionnés, de manière à voir, dans une dimension large, « ce qu'avait compris » les participants. Ainsi, la compréhension était considérée dans sa globalité. Mais si nous considérons les sélections de dessins présentant un facteur de variation (par catégorie) non plus sur la totalité des dessins sélectionnés mais *uniquement* sur le nombre de dessins non cibles sélectionnés, alors nous pourrions nous concentrer sur les sources de difficultés (*i.e.* sur les inadéquations entre les résultats attendus et ceux obtenus). C'est ce que nous appelons les *profils d'erreurs*. Au-delà du fait que certaines configurations de présentation du support (avec ou sans CSE) et que certains profils d'apprenants (niveau d'expertise linguistique) impliquent des différences dans les performances en compréhension, les *profils d'erreurs* nous renseignent sur la nature des difficultés affectées par ces deux variables indépendantes. Les profils d'erreurs seront observés à partir de comparaisons de fréquences, obtenues à partir de tables de contingence.

**Les *profils d'erreurs* nous permettent de comparer les erreurs des participants francophones à celles des participants apprenants débutants, en situation d'écoute des dialogues sans contexte sonore environnemental.** Dans ce cadre, nous avons pu observer que, si l'on considère les trois dialogues ensemble, la seule différence significative qui ressort des *profils d'erreurs* se trouve au niveau de l'appréhension de la situation d'énonciation. En effet, sur l'ensemble des dessins présentant un facteur de variation sélectionnés par les participants francophones ( $n_1=56$ ) le pourcentage de dessins EN ( $p_2=51,79$ ) est significativement supérieur ( $p = 0,038$  ;  $Z = 1,771$ ) à celui des participants apprenants débutants ( $n_2=114$  ;  $p_2=32,46$ ). **Ces résultats nous amènent à conclure que, sur l'ensemble des dialogues, les participants francophones ont accordé moins d'attention à la situation d'énonciation et ont donc commis proportionnellement plus d'erreurs à ce niveau.** Nous n'avons pas pu constater d'écart significatif au niveau des autres

dessins présentant un facteur de variation, pour les trois dialogues considérés ensemble.

Si nous considérons les dialogues isolément, alors nous observons que pour **le dialogue A on enregistre également un écart significatif ( $p=5,818 \times 10^{-5}$  ;  $Z=3,853$ ) entre les deux groupes au niveau de l'appréhension de la situation d'énonciation** puisque le groupe francophone ( $n_1=13$ ) a sélectionné proportionnellement moins de dessins EN ( $p_1=48,78$ ) que le groupe d'apprenants débutants ( $n_2=26$  ;  $p_2=80,77$ ). **Nous observons donc que l'écart s'est inversé par rapport à celui que nous avons observé pour l'ensemble des trois dialogues.** Pour le dialogue B nous avons pu observer un écart significatif, au seul niveau du positionnement des dessins cibles et EN puisque le panel de participants francophones ( $n_1=12$  ;  $p_1=41,67$ ) a proportionnellement mal positionné moins de dessins cibles ou EN ( $p=0,013$  ;  $Z=2,207$ ) que les participants apprenants débutants ( $n_2=36$  ;  $p_2=13,89$ ). Au niveau de la comparaison des dessins avec une variation de la situation d'énonciation, nous avons pu toutefois observer que les participants francophones ont sélectionné proportionnellement moins de dessins EN ( $n_1=12$  ;  $p_1=16,67$ ) que les participants apprenants débutants ( $n_2=36$  ;  $p_2=30,56$ ) sans que cet écart ne soit statistiquement significatif. Le dialogue C n'a pas présenté d'écarts significatifs dans les *profils d'erreurs* au niveau de la comparaison des participants francophones et apprenants débutants dans l'écoute des dialogues sans CSE. Pour ce dialogue, nous pouvons observer que les participants francophones en ont sélectionné proportionnellement ( $n_1=18$  ;  $p_1=33,33$ ) moins que les participants apprenants débutants ( $n_2=37$  ;  $p_2=16,22$ ) sans que cet écart ne soit statistiquement significatif. **Cet ensemble de résultats, et à la lumière des résultats que nous avons présentés plus haut, nous amène à penser que l'appréhension de la situation d'énonciation est un des facteurs les plus déterminants en situation de non-expertise linguistique.** En effet ce paramètre fait l'objet de différences entre les francophones et les apprenants débutants quel que soit le mode de calcul statistique. La situation d'énonciation correspond en outre à une dimension particulièrement importante de la compréhension (Magliano, Zwaan, & Graesser, 1999; Zwaan, 1993; Zwaan, Magliano, & Graesser, 1995; Zwaan, Langston, & Graesser, 1995). Dans le cadre des dialogues oraux cette dimension est principalement construite à travers les inférences, que les compreneurs élaborent à partir d'indices para et extralinguistiques (potentiellement accessibles *via* le contexte sonore environnemental lorsqu'il y en a un). **La question du rôle du contexte sonore environnemental redevient dès lors primordiale puisque son caractère non-linguistique est susceptible de le rendre plus accessible pour les individus**

**débutants en langue.** Rappelons en outre que les scènes sonores que nous avons réalisées ont une fonction situationnelle importante : il s'agissait avant tout de rendre compte du lieu dans lequel se déroule le dialogue. Cet aspect est susceptible d'avoir entraîné des différences dans les profils d'erreurs des groupes qui les ont entendu avec ou sans contexte sonore environnemental.

Nous avons ensuite comparé les *profils d'erreurs* liés à la variable contexte sonore environnemental. Nous avons procédé à la comparaison des profils des participants qui ont entendu les dialogues sans contexte sonore environnemental à ceux des participants qui les ont entendu accompagnés d'un CSE, pour les panels de francophones et d'apprenants débutants. **Pour les participants francophones, et si on considère les trois dialogues comme un ensemble, alors il apparaît que la présence (ou l'absence) d'un CSE n'a pas entraîné de différence significative au niveau des *profils d'erreurs*.** Toutefois, considérant les dialogues un à un, on peut observer que la présence d'un contexte sonore environnemental, pour le dialogue B a entraîné un écart significatif ( $p=0,002$  ;  $Z= -3,467$ ) au niveau du positionnement des dessins cibles et EN puisque le groupe qui a écouté le dialogue seul ( $n1=12$  ;  $p1=41,67\%$ ) a proportionnellement mal positionné moins de dessins cibles et/ou EN que le groupe qui a entendu le dialogue avec un CSE ( $n2=8$  ;  $p2=75\%$ ). Notons que pour ce dialogue, la VI *expertise* linguistique avait également entraîné une modification des profils d'erreurs. Pour le dialogue C, nous pouvons constater que la présence d'un CSE a impliqué un effet significatif au niveau de la sélection des dessins avec une variation de contenu propositionnel. En effet, le groupe qui a entendu le dialogue sans contexte sonore a sélectionné proportionnellement moins ( $p=0,023$  ;  $Z= -1,982$ ) de dessins CONT ( $n1=18$  ;  $p1=22,22$ ) que le groupe qui a entendu le dialogue avec un CSE ( $n2=13$  ;  $p2=46,15$ ). **Nous pouvons donc conclure à une forte variabilité inter-dialogue dans la compréhension et les profils d'erreurs des participants francophones.** Cet aspect sera également traité dans le cadre de la *discussion méthodologique*.

**Cependant, globalement, il semble que pour le panel de francophones, la variable contexte sonore environnemental n'a pas donné lieu à un écart significatif et/ou systématique dans les *profils d'erreurs*.** Dès lors, nous pouvons nous interroger sur les phénomènes attentionnels qui ont régi l'activité de perception/compréhension de ces supports complexes (faisant intervenir à la fois du matériel linguistique et non linguistique), et avancer l'hypothèse explicative que ces participants, dans une situation de « sécurité linguistique » ne se sont pas concentrés sur les aspects environnementaux des supports et ont plutôt porté leur attention sur

les aspects linguistiques, directement visés par la consigne qui leur a été présentée (« Écoutez attentivement le dialogue enregistré. Ensuite, 12 images vous seront présentées. Sélectionnez-en 3 illustrant le dialogue et classez-les dans l'ordre chronologique LE PLUS RAPIDEMENT POSSIBLE »).

**Parallèlement, pour le panel d'apprenants débutants, nous avons pu constater que la variable contexte sonore environnemental n'a pas non plus impliqué d'écart significatif systématique.** En effet, au niveau de la comparaison des trois dialogues considérés ensemble, la variable CSE a entraîné une différence significative au seul niveau du positionnement des dessins cibles et EN ( $p=0,027$  ;  $Z=-1,916$ ) puisque le groupe qui a entendu les dialogues sans contexte sonore environnemental a proportionnellement mal positionné moins de dessins cibles et/ou EN ( $n_1=114$  ;  $p_1=30,7$ ) que le groupe qui a écouté les dialogues avec un CSE ( $n_2=72$  ;  $p_2=55,56$ ). Au niveau des dialogues considérés isolément, nous avons pu constater que pour le dialogue A, la présence d'un CSE n'a pas entraîné d'écart significatif au niveau des *profils d'erreurs*. Pour le dialogue B, la variable « présence/absence d'un CSE » a entraîné un écart statistiquement significatif au niveau du mauvais positionnement des dessins cibles et EN ( $p=1,6811 \times 10^{-5}$  ;  $Z=-4,147$ ) puisque le groupe qui a entendu le dialogue sans contexte sonore a proportionnellement mal positionné moins de dessins cibles ou EN ( $n_1=36$  ;  $p_1=13,89$ ) que le groupe qui a entendu le dialogue avec un CSE ( $n_2=24$  ;  $p_2=50$ ). Enfin, le dialogue C a également présenté une différence significative ( $p=0,001$  ;  $Z=-2,972$ ) des *profils d'erreurs* à ce niveau. En effet, le groupe qui a entendu le dialogue sans contexte sonore a proportionnellement mal positionné moins de dessins cibles ou EN ( $n_1=37$  ;  $p_1=48,65$ ) que le groupe qui a entendu le dialogue avec un CSE ( $n_2=24$  ;  $p_2=79,17$ ). **Malgré une forte variabilité inter-dialogues, mais pouvons conclure que la présence d'un CSE a globalement engendré une diminution des erreurs liées à l'appréhension de la situation d'énonciation et du contenu propositionnel, sans pour autant entraîner une modification systématique du profil d'erreurs.** Nous pouvons par ailleurs constater que la présence d'un CSE n'a pas entraîné de modification significative des *profils d'erreurs* au niveau de l'appréhension de la situation d'énonciation. Dit autrement, si la présence d'un CSE a globalement impliqué moins d'erreurs à ce niveau, celles-ci représentent néanmoins la même proportion des erreurs que dans le cadre de l'écoute sans CSE. Cette remarque nous amène à penser que la présence d'un CSE n'entraîne pas de modification majeure dans les stratégies d'élaboration du sens mises en place.

Étudier les *profils d'erreurs* nous a permis de voir que :

- la principale différence entre les participants francophones et apprenants débutants, c'est-à-dire entre des experts linguistiques et des novices, se situe au niveau de l'appréhension de la situation d'énonciation ;
- la variable *présence ou absence d'un contexte sonore environnemental* implique surtout une différence au niveau de la proportion de dessins mal positionnés. Les comparaisons des résultats calculés sur la totalité des dessins sélectionnés (présentées plus haut) n'avaient pas permis de mettre cet aspect au jour.

Nous avons également pu constater, à la lumière des différents ensembles de résultats enregistrés que la question des effets d'interaction entre les variables *niveau d'expertise* et *présence d'un CSE* s'est posée à plusieurs reprises. Nous avons donc procédé à des ANOVA multifactorielles. Pour les trois dialogues considérés ensemble, ces tests statistiques ont mis en lumière des écarts significatifs au niveau des dessins cibles sélectionnés, qu'ils soient bien ou mal positionnés ( $F(1,296) = 9,53 ; p=0,002$ ). Considéré isolément, le dialogue A n'a pas présenté d'effet d'interaction entre les variables *expertise linguistique* et *présence d'un CSE*. Pour le dialogue B, nous avons pu observer un écart significatif au niveau de la sélection de dessins EN bien positionnés sur l'axe chronologique ( $F(1,96) = 4,568, p=0,035$ ) et pour ces mêmes dessins quel que soit leur positionnement sur l'axe chronologique ( $F(1,96) = 4,527, p=0,036$ ). Enfin, pour le dialogue C, nous avons pu observer un écart significatif au niveau des absences de réponses, c'est-à-dire des cases laissées vides par les participants ( $F(1,96) = 5,294, p=0,024$ ). Toutefois, le caractère non systématique de ces effets nous empêche d'interpréter cette partie des résultats. Il semblerait donc que ces deux variables aient présenté peu d'interactions. Dit autrement, si effectivement nous avons pu constater que la variable *présence d'un CSE* (VI 2) n'a pas donné lieu aux mêmes implications au niveau de la compréhension en fonction du *niveau d'expertise linguistique* (VI 1) des participants, ces écarts ne sont pas assez importants pour que l'on puisse conclure à des effets d'interaction statistiquement significatifs.

## 2. Discussion méthodologique

Nous avons présenté jusqu'ici une discussion des résultats du Classe-Images. Nous en avons fait une rapide synthèse (présentés p.212) qui nous a permis de nous questionner sur les implications de l'augmentation de la charge cognitive impliquée par le caractère débutant de certains participants et la présence d'un CSE. Nous avons également présenté les *profils d'erreurs* associés aux différentes variables

indépendantes introduites dans le test et que les résultats de Classe-Images ont mis en lumière. Dans cette nouvelle partie, nous allons nous attacher à discuter les aspects plus méthodologiques de ce test. En effet, nous aborderons ici le matériel que nous avons introduit dans le Classe-Images (les éléments linguistiques, sonores environnementaux et visuels), les différents dialogues que nous avons choisi de présenter aux participants, ainsi que les participants eux-mêmes. Dans un second temps, la question des interactions de ces différentes composantes du Classe-Images sera discutée afin, notamment, de voir dans quelles mesures les spécificités de chaque dialogue ont pu avoir une incidence sur les résultats. Cette question sera abordée dans le cadre de la partie *différences inter-dialogues*. Enfin, nous reconsidérerons les groupes des participants au test et verrons dans quelle mesure leur composition a pu avoir une influence sur les résultats.

## **2.1. Le matériel utilisé**

Les résultats que nous avons obtenus grâce au Classe-Images nous ont renseignée sur différents aspects de la compréhension orale par des natifs de la langue cible et des apprenants débutants de cette langue. Toutefois, il apparaît que ces résultats sont spécifiques au cadre de cette expérimentation particulière. Il semble donc nécessaire ici de présenter une rapide discussion des éléments qui, selon nous, les ont conditionnés. Nous nous concentrerons donc sur les éléments linguistiques, sonores environnementaux et visuels des supports que nous avons utilisés, ainsi que sur leur interaction mutuelle.

### **2.1.1. Le matériel linguistique**

En ce qui concerne les dialogues sans contexte sonore environnemental, nous pouvons rappeler que la première opposition se trouve au niveau des locuteurs et des lieux où ils se déroulent. En effet, alors que les dialogues A (*la préparation d'un repas*) et B (*la vaisselle*) prennent place dans une cuisine (c'est-à-dire un lieu clos et privé), le dialogue C (*à la terrasse d'un café*) se déroule dans un lieu ouvert et public. Par ailleurs, ces dialogues s'opposent, toujours sur un schéma A et B *versus* C sur le sexe des locuteurs. En effet, les dialogues A et B mettent en scène deux locuteurs de sexe opposé, alors que le dialogue C met deux femmes en scène. Nous pouvons dès lors nous interroger sur l'impact de ce changement sur la compréhension des dialogues par les participants au test : lorsque les locuteurs d'un dialogue ne sont pas du même

sexe, la discrimination entre leurs deux voix peut-elle représenter une aide significative à la compréhension du dialogue ?

En outre, considérés isolément, et en dehors des oppositions liées à la discrétion des thématiques et du vocabulaire utilisé, nous pouvons voir que les dialogues s'opposent au niveau de la concomitance des actions évoquées dans le discours et du discours en lui-même. Dit autrement, pour le dialogue C, la présence de l'énoncé, « *S'il vous plait ?* » implique un déroulement des actions nécessairement parallèle à la tenue du reste du discours. En d'autres termes, ce dialogue peut être considéré comme relevant du *présent actuel* (bien que l'ensemble des verbes qui le composent ne relèvent pas nécessairement de cette « valeur »). Cette « valeur » du présent renvoie à un déroulement des actions parallèle au discours. C'est un « *processus (qui se) réalise nécessairement au moment où le locuteur parle ; il se trouve dans l'actualité de celui-ci* » (Charaudeau, 1992:452). En revanche, pour les dialogues A et B, cette concomitance n'est pas aussi nécessaire. En effet, pour ces dialogues, nous pourrions tout-à-fait imaginer un autre contexte que celui que nous avons envisagé lors de l'élaboration (la locutrice prépare **effectivement** le repas et le locuteur fait **effectivement** la vaisselle, pendant les échanges verbaux). Dans ce cadre, nous pouvons envisager qu'à la seule écoute des éléments linguistiques de ces dialogues, c'est-à-dire sans CSE ni images, les compreneurs élaborent une signification dans laquelle le *présent* n'aurait pas forcément « valeur » de *présent actuel*. En effet, le présent peut également être interprété comme un présent *générique* pour lequel « *le processus ne se réalise pas nécessairement au moment même où le locuteur parle, mais, du fait qu'il se répète à l'infini, lorsque certaines conditions sont réunies, il acquiert une valeur de réalisation pantemporelle (à la fois virtuelle et effective)* » (*ibidem* p.453). Dans ce cadre, il apparaît que le présent exprimé dans le dialogue A peut, selon nous, assumer une « valeur » de présent *générique* (bien que l'énoncé « *Qu'est-ce qu'on mange ?* » laisse cette interprétation peu probable et donc peu plausible). Le présent du dialogue B peut assumer la même « valeur » de présent *générique* puisque l'énoncé « *C'est normal, tu as préparé le repas alors je fais la vaisselle* » peut laisser le compreneur interpréter qu'il s'agit d'une habitude.

Enfin, notons que, bien que chaque dialogue présente un vocabulaire fréquent et donc en adéquation avec les impératifs du niveau débutant en langue, les dialogues A et B présentent des lexies plus spécifiques que le dialogue C. Dès lors, nous pouvons supposer que l'appréhension du contenu propositionnel de ces dialogues impliquera une plus grande focalisation attentionnelle que pour le dialogue C.

Toutefois, comme nous le verrons plus loin, les autres éléments supports du Classe-Images (visuels et sonores environnementaux) sont à prendre en considération dans nos interprétations.

### **2.1.2. Le corpus d'images**

L'objectif de ces dessins était de rendre compte des dialogues et de présenter les facteurs de variation que nous avons choisi d'analyser (variation de la situation d'énonciation et variation du contenu propositionnel). Ainsi, douze dessins ont été proposés pour chaque dialogue voir *Expérimentation principale* p.165).

Dans la mesure où les dessins visaient la représentation des dialogues rédigés en amont, certaines de leurs caractéristiques et des contextes sonores environnementaux qui peuvent les accompagner apparaissent dans le corpus d'images. Ainsi, les dialogues A et B se déroulent dans une *cuisine* (lieu clos et privé) et l'ensemble des dessins qui ont été élaborés pour les représenter mettent cet espace en scène, même les dessins avec une variation de la situation d'énonciation. En revanche, le dialogue C se déroule à une *terrasse de café* (lieu ouvert et public). Les dessins qui ne présentent pas de variation de la situation d'énonciation (c'est-à-dire les dessins cibles et CONT) se déroulent donc dans ce lieu, mais les dessins EN et EN-CONT (avec variation de la situation d'énonciation) mettent en scène les protagonistes sur un quai de gare. Dans ce cadre, nous pouvons constater que, alors que pour les deux premiers dialogues la variation de la situation d'énonciation est représentée à travers un changement de positionnement des locuteurs à l'intérieur d'un même espace, pour le troisième dialogue, cette variation est représentée par un changement du lieu même, donc plus important. En outre, cette différence a eu, selon nous, un impact sur la plausibilité des dessins avec une variation de la situation d'énonciation, qui, dès lors est plus importante pour les dialogues qui se déroulent dans la *cuisine* que pour celui qui se déroule à une *terrasse de café*. La plausibilité renvoie ici selon nous à la probabilité pour ces dialogues (dimension linguistique) de se dérouler dans ces lieux (une *cuisine*, une *terrasse de café* ou un *quai de gare*).

Toutefois, il apparaît que l'ensemble de ces dessins a pu avoir une influence sur l'interprétation des dialogues. En effet, il nous apparaît difficile d'envisager que le processus d'élaboration de la signification se soit arrêté et que la signification se soit stabilisée dès la fin de l'écoute. Dans ce cadre, il apparaît que les dessins ont pu modifier, plus ou moins profondément la compréhension des participants, et ce notamment au niveau de la « valeur » du présent (présentée plus haut) et ce d'autant plus quand les dialogues ont été présentés sans contexte sonore environnemental.

Pour le dialogue A, il semble en effet que les deux « valeurs » du présent restent envisageables compte tenu des dessins (voir p.169). En effet, les dessins cibles montrent la femme **effectivement** en train de préparer le repas (présent *actuel* pour l'énoncé « *je prépare* »), et les dessins avec une variation de la situation d'énonciation montrent les deux protagonistes à table. Dans le cadre de ces derniers dessins, les compreneurs ont pu maintenir une interprétation *générique* du présent, et ce d'autant plus qu'une casserole apparaît sur la table de cuisson. Cependant, il apparaît que cette interprétation semble moins probable, moins plausible que celle d'un présent *actuel*, notamment avec la présentation des dessins cibles. Enfin, pour le dialogue C, il semble que les deux « valeurs » du présent restent également viables et plausibles. En effet, les dessins cibles peuvent renforcer l'hypothèse d'une « valeur » actuelle du présent, et les dessins avec une variation de la situation d'énonciation sont, selon nous, tout-à-fait compatibles avec une interprétation *générique* du présent. Toutefois, l'énoncé « *Tu as besoin d'aide ?* » participe à l'« *actualisation* » du présent et nous amène à considérer que sa « valeur » *générique* reste peu probable et peu plausible. Nous considérons donc que la « valeur » *actuelle* du présent est plus marquée pour le dialogue B que pour le dialogue A.

Par ailleurs, nous avons noté plus haut que les dialogues A et B présentaient des spécificités dans le vocabulaire qui les compose. Il apparaît, dans l'interaction des éléments linguistiques avec les dessins, que cette spécificité rend les dessins avec une variation du contenu propositionnel du dialogue B moins plausibles que ceux du dialogue A. Dit autrement, alors que pour le dialogue *La préparation d'un repas*, les dessins avec une variation du contenu propositionnel restent valides pour les compreneurs qui n'auraient pas prêté attention à ces spécificités, ce n'est pas le cas pour le dialogue *La vaisselle* dans lequel les dessins CONT (et EN-CONT) sont incompatibles avec la signification générale de l'échange.

Enfin, nous pouvons également rappeler que les dessins que nous avons fait réaliser n'ont pas été évalués/interprétés seuls par les participants au Classe-Images. Dès lors, il semble particulièrement complexe d'interpréter leurs choix et, dans les cas de non coïncidence entre le choix des participants et nos « anticipations », nous ne pouvons exclure que ces différences soient imputables non pas à une « mauvaise » compréhension des dialogues, mais à une « mauvaise » interprétation des dessins.

### ***2.1.3. Le matériel sonore environnemental***

Comme nous l'avons vu dans le chapitre *Élaboration du matériel expérimental*, nous avons entièrement construit les éléments sonores environnementaux inclus

dans le Classe-Images. Pour cette élaboration nous avons fait le choix d'une méthodologie expérimentale qui trouve son point de départ dans une enquête que nous avons effectuée auprès de participants francophones (non nécessairement natifs). Dans cette première étape, nous avons demandé aux participants de s'imaginer dans des lieux (une forêt<sup>69</sup>, un hall de gare<sup>1</sup>, une terrasse de café et une cuisine (privée)) et d'exprimer les éléments qu'ils entendaient (entretiens oraux semi-dirigés). À partir des éléments de réponse extraits du corpus que nous avons constitué, nous avons procédé aux enregistrements (et au montage) des scènes sonores (validées par la suite par une nouvelle étude). À ce niveau, nous pouvons donc nous interroger sur notre choix de recourir à une méthodologie aussi « lourde » pour l'élaboration de ces *ambiances sonores*. En effet, de nombreuses alternatives s'offraient à nous. Nous aurions pu recourir à des ambiances réalisées par des professionnels telles que celles qui sont disponibles sur internet par exemple. Toutefois, au-delà des difficultés inhérentes à la question de droits d'auteurs qu'une telle procédure aurait impliquées, se seraient alors posées les questions de la maîtrise du matériel expérimental, de son authenticité et de sa typicalité. En réalisant les scènes sonores à partir des enregistrements réalisés à la suite de l'étude du corpus, nous avons fait le choix d'une semi-authenticité et probablement d'un environnement sonore avec des éléments véritablement typiques de celui-ci. Ce choix pourrait sembler questionnable eu égard au grand nombre de documents authentiques qui sont quotidiennement utilisés en salle de classe. Néanmoins, nous pouvons également considérer la salle de classe comme un lieu où la réalité est de toute façon reproduite et simulée. En effet, dans le cadre du travail de la CO, il s'agit le plus souvent de reproduire des situations susceptibles d'être rencontrées dans la vie de tous les jours. En ce sens, la salle de classe peut être considérée comme un espace où l'artificialité des procédures n'est pas un problème tant que les objectifs d'enseignement et d'apprentissage correspondent à des savoirs ou à des savoir-faire transférables en dehors de ce contexte bien spécifique. Le recours à des enregistrements semi-authentiques nous semble être un bon compromis entre les différents impératifs d'authenticité, de réalisme et de stéréotypage. En effet, c'est dans ce cadre que nous avons fait le choix de bâtir notre procédure expérimentale sur les représentations que les individus se font des lieux que nous avons mis en ondes. Ce choix méthodologique repose sur l'idée que ce qui est connu est plus facilement identifié que ce qui est inconnu. Nous avons donc cherché à nous rapprocher au maximum des représentations que les individus se font des lieux visés (voir le chapitre *Élaboration*

---

<sup>69</sup> (distracteurs)

*du matériel sonore environnemental*, p.121), c'est-à-dire du prototype (noyau de la catégorie) que les individus ont construit à partir de leurs connaissances et expériences. Nous nous sommes concentrée sur les connaissances d'individus de culture française métropolitaine, c'est-à-dire sur un public différent de notre public cible (apprenants débutants en FLE). Cette position s'appuie sur un double choix : d'une part sur la nécessité de s'appuyer sur un groupe de participants homogène du point de vue de la culture, et d'autre part sur l'objectif d'enseignement qui vise la transmission de certains éléments de la culture francophone (voir *Contexte didactique*, p.17).

Afin de nous assurer que cette méthodologie de réalisation donnait lieu à des scènes sonores facilement reconnaissables, nous avons procédé à des tests d'identification des lieux que nous avons mis en scènes (voir p.138) par des Français. Toutefois, nous pouvons nous interroger sur la sélection des participants interrogés. La première remarque que nous pourrions avancer repose sur la nationalité (et donc, indirectement la culture) des participants, tous Français. En effet, si les dialogues que nous avons construits sont, à terme, destinés à des apprenants débutants en Français Langue Étrangère, il peut sembler surprenant de faire valider une partie des supports qui les composent (ici les scènes sonores environnementales qui assumeront par la suite une position contextuelle) par des individus de culture française métropolitaine. Ce choix méthodologique a été fait en fonction de différentes contraintes. Tout d'abord, le fait que la totalité de notre démarche expérimentale s'est déroulée en France a impliqué un « recrutement » de Français plus aisé que celui de participants étrangers. Par ailleurs, si nous avions fait le choix d'interroger un public étranger, la question de la culture de ces participants se serait alors posée. En effet, l'enseignement du FLE en France se déroule le plus souvent auprès de groupes-classes composés d'apprenants de différentes langues-cultures d'origine. Dans ce cadre comment déterminer le choix de la culture d'origine des participants à un nouveau test d'identification des scènes sonores ? En effet, dans le Classe-Images, nous avons interrogé des participants de différentes nationalités parlant différentes langues. Cette première remarque en appelle une seconde. En effet, une partie du Classe-Images visait à observer l'impact de la présence d'un contexte sonore environnemental sur la compréhension de courts dialogues par ces participants. Dans ces conditions, nous pouvons nous interroger sur la compréhension des scènes sonores environnementales (présentées seules) par les participants au Classe-Images. En effet, nous ne nous sommes pas assurée que ces scènes sonores étaient correctement interprétées. Dès lors, il nous est difficile de conclure à une éventuelle redondance des supports. Rappelons que face à un support complexe, on considère

qu'il y a redondance lorsque l'individu peut comprendre les différentes sources qui le composent séparément. Lorsque l'individu ne peut pas comprendre les sources séparément, il doit procéder à une étape d'intégration mentale (Tricot, 1998). Il aurait donc été intéressant de nous pencher sur la compréhension des scènes sonores présentées sans dialogues, c'est-à-dire en soi.

En outre, il apparaît que les scènes sonores que nous avons élaborées et retenues assument différentes fonctions vis à vis des dialogues (aspects linguistiques) (voir *Contexte didactique* p.17). Ainsi, les CSE associés aux dialogues A et B assument des fonctions de transcodage (redondance avec les actions menées et décrites dans le discours) et situationnelle (mise en ondes du lieu où se déroule la scène), alors que le CSE associé au dialogue C n'assure qu'une fonction situationnelle. Comme nous l'avons vu dans l'élaboration des scènes sonores environnementales (voir p.121), ces différentes fonctions ont été déterminées par notre méthodologie de réalisation.

Par ailleurs, un autre aspect permet d'opposer les dialogues, et les scènes sonores environnementales : la présence de parole dans les scènes sonores. En effet, dans les CSE des dialogues B et C (*La vaisselle* et *À la terrasse d'un café*), des sons de la parole apparaissent dans les scènes sonores, *i.e.* constituent des sons environnementaux. Nous avons vu dans le chapitre « *des études sur la perception* » (voir p. 49) que les sons de la parole peuvent être considérés comme des sons de l'environnement lorsque les processus attentionnels sont focalisés sur une autre cible. Néanmoins, la nature spécifique de la parole et son intention communicative impliquent une attention particulière sur ce type de sons. Dès lors, nous pouvons supposer que cette spécificité des dialogues B et C a pu impliquer des traitements spécifiques des sons associés. Le dialogue A (*La préparation d'un repas*) ne présente pas de sons de la parole dans le CSE. Dans le CSE du dialogue B (*La vaisselle*), les sons de la parole sont représentés par la radio (flash d'informations sur la bourse, diffusé par France Inter). Dans le dialogue C (*À la terrasse d'un café*), les sons de la parole proviennent des conversations des tables avoisinantes.

Enfin, il apparaît que l'ajout d'un CSE a pu, tout comme nous l'avons imputé aux dessins, modifier la compréhension des dialogues en permettant à une « valeur » particulière du présent d'être maintenue, ou en en inhibant une autre. Selon notre interprétation des scènes sonores, la présence d'un CSE a pu participer à une interprétation *actuelle* des verbes au présent, tant pour le dialogue A que pour le dialogue B. En effet, pour le dialogue A, il apparaît que, s'ils ont été perçus et

interprétés, les sons de cuisson, associés aux sons d'« ustensile en bois contre la poêle » excluent la possibilité d'un présent *générique*, qui, dans cette interaction entre dialogue linguistique, dessins et CSE devient fortement improbable, voire impossible. De la même manière pour le dialogue B, la présence d'un CSE a dû, selon nous, inhiber les hypothèses d'interprétation faisant intervenir un présent *générique*, en raison, notamment des sons « d'eau qui coule » et de « couverts qui s'entrechoquent ».

## 2.2. Des différences inter-dialogues

Dans la synthèse des résultats, (présentée plus haut), nous avons pu voir que certains dialogues, considérés isolément, ne présentaient pas les mêmes résultats que l'ensemble des trois dialogues considérés ensemble. Ces particularités sont observables au niveau de chaque variable indépendante (*niveau d'expertise linguistique et présence ou absence d'un contexte sonore environnemental*) et pour les différents modes de calcul statistiques que nous avons appliqués (calculés sur la totalité des réponses recueillies comme dans les profils d'erreurs). Elles peuvent, selon nous, s'expliquer par les spécificités de chacun des supports (envisagés dans leur dimension linguistique, sonore, et visuelle). Toutefois, nous ne pouvons pas savoir, à ce stade, si ces différences sont imputables aux dialogues (et leurs composants) ou s'il s'agit de différences individuelles marginales non significatives d'un point de vue statistique. En effet, le faible nombre de participants interrogés ne nous permet pas de tirer de telles conclusions. Nous avons néanmoins voulu examiner ces différences inter-dialogues à l'aide d'un traitement statistique approprié.

Les suites de trois dessins cibles sont considérées comme des indices de « bonne compréhension » des dialogues par les participants. Rappelons que nous considérons que ces constructions correspondent à de « bonnes compréhensions » dans la mesure où nous pouvons considérer que le sens construit par les participants correspond à celui élaboré par le concepteur des supports (linguistique, sonores environnementaux ou visuels). Nous comparerons ces suites construites pour le dialogue A à celles construites pour le dialogue B et le dialogue C, ainsi que celles construites pour le dialogue B à celles construites pour le dialogue C. Nous effectuerons ces comparaisons au sein de chaque panel (les francophones et les apprenants débutants) pour chaque modalité du test (avec et sans contexte sonore). Ces comparaisons tiennent compte du nombre de participants dans chaque groupe (29 pour les groupes de francophones et 21 pour les groupes d'apprenants

débutants) et les pourcentages de suites de trois dessins cibles (c'est-à-dire les Réponses Attendues), calculés sur l'ensemble des constructions élaborées.

La comparaison du nombre de suites de trois dessins cibles construites par le **panel des francophones** fait apparaître une différence significative entre les dialogues A (*La préparation d'un repas*) et B (*La vaisselle*) : le dialogue B a donné lieu à significativement plus de suites correctes que le dialogue A (voir tableau 2). En revanche, pour le **panel d'apprenants débutants**, il apparaît que les écarts ne sont pas significatifs et que l'orientation des tendances varie : pour le groupe DS le dialogue B a été mieux compris que le A, alors que pour le groupe DCS, l'écart est moindre mais le dialogue A a été mieux compris que le B (voir tableau 2). Toutefois, ces écarts ne sont pas statistiquement significatifs et cela peut être dû au faible nombre de participants.

	DIALOGUE A		DIALOGUE B		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
francophones DS	29	48,279	29	79,31	-3,537	0,000
francophones DCS	29	72,414	29	86,207	-2,136	0,016
apprenants DS	21	19,048	21	28,571	-1,225	0,110
apprenants DCS	21	47,619	21	42,857	0,6	0,274

**Tableau 34: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues A et B - VI) sur la construction de suites de 3 dessins cibles, pour les panels de francophones et d'apprenants débutants**

Nous avons pu observer, **au niveau de l'élaboration des suites de trois dessins cibles (c'est-à-dire de la compréhension envisagée en termes binaires) et pour le panel de francophones, des différences significatives, tant pour le groupe qui a écouté les dialogues seuls que pour le groupe qui a écouté les dialogues accompagnés d'un contexte sonore environnemental.** Cette remarque nous amène à penser que le dialogue B était plus « facile à comprendre »<sup>70</sup> que le dialogue A. Nous pouvons supposer que la principale différence réside plus dans la dimension linguistique et visuelle de ces dialogues que dans les CSE. Nous pouvons

<sup>70</sup> Encore une fois, ici la compréhension est considérée comme l'adéquation du sens élaboré par les compreneurs et dans la construction du dialogue.

avancer l'hypothèse qu'au-delà de la discrétion des unités linguistiques qui composent les dialogues et des spécificités des thématiques, les « valeurs » du présent de l'indicatif et la spécificité du vocabulaire ont constitué des facteurs déterminants. Nous pouvons en effet penser que le caractère potentiellement plus *générique* du présent du dialogue A a donné lieu à une plus grande complexité dans l'élaboration d'une signification « stabilisée » et que les participants ont éprouvé des difficultés à choisir une des deux interprétations possibles. Toutefois, cette difficulté ne semble pas avoir été palliée par la présence d'un CSE, c'est-à-dire que l'« actualisation » du présent que nous avons imaginée *via* la présence du CSE n'a pas eu lieu. Enfin, nous pouvons nous étonner que les différences observées pour le panel de francophones ne se retrouvent pas chez le panel d'apprenants débutants. Ce résultat pourrait nous amener à considérer que la « valeur » potentiellement *générique* du présent n'a peut-être pas été perçue par ces derniers. Dès lors, nous pouvons penser que la compréhension du *présent générique* présente plus de difficultés que celle du *présent actuel*, et que celles-ci seraient proportionnelles au niveau d'expertise du locuteur. Dit autrement, on peut considérer que les apprenants débutants sont moins sensibles à cette difficulté, peut-être en raison du recours aux autres indices (linguistiques, para ou extralinguistiques) sur lesquels s'appuient les novices en langues (cf. les résultats et interprétations présentées plus haut).

Les résultats obtenus en comparant les dialogues A (*La préparation d'un repas*) et C (*À la terrasse d'un café*) indiquent qu'il y a un écart significatif ( $p < 0,05$ ) lorsque l'on compare les suites de trois dessins cibles (RA) réalisées en condition « dialogue seul » pour les deux panels. Aucun effet n'apparaît lorsque l'on procède à la même comparaison en condition « écoute avec contexte sonore ».

	DIALOGUE A		DIALOGUE C		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
francophones DS	29	48,279	29	72,414	-2,655	0,004
francophones DCS	29	72,414	29	72,414	0	0,5
apprenants DS	21	19,048	21	33,33	-1,749	0,040
apprenants DCS	21	47,619	21	47,619	0	0,5

**Tableau 35: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues A et C - VI) sur la construction de suites de 3 dessins cibles, pour les panels francophones et apprenants débutants**

Encore une fois, le dialogue A semble plus difficile à comprendre que le dialogue C. Ces deux dialogues s'opposent sur la quasi-totalité des paramètres que nous avons présentés plus haut (sexe des locuteurs, lieu ouvert ou clos, public ou privé) mais selon nous, la différence majeure (ou du moins celle qui a le plus d'impact sur la compréhension) réside dans les « valeurs » du présent. En effet, alors que pour le dialogue C, le présent *actuel* est le plus probable (donc *a priori* le plus activé par les compreneurs), pour le dialogue A une ambiguïté subsiste entre le présent *générique* et le présent *actuel*. En outre, il apparaît que la présence d'un contexte sonore environnemental a, en quelque sorte, permis d'inhiber la « valeur » *générique* du présent du dialogue A puisqu'aucun écart significatif n'apparaît chez les groupes qui ont écouté les dialogues avec leurs contextes sonores. Enfin, notons que l'écart observé entre les groupes qui ont entendu les dialogues seuls est plus important pour le panel de francophones que d'apprenants débutants, ce qui nous laisse penser, encore une fois que la potentielle « valeur » *générique* du présent ne semble pas avoir entraîné de difficulté pour le panel d'apprenants débutants. Dès lors, nous pouvons envisager que ces participants n'ont pas été sensibles à la plus grande ambiguïté du présent *générique*, en raison de leur faible niveau d'expertise linguistique en français.

Enfin, les résultats obtenus en comparant les dialogues B et C indiquent qu'aucun effet n'apparaît lorsque l'on procède à la comparaison des suites de trois dessins cibles réalisées par les panels de francophones et d'apprenants débutants, en condition DS (dialogue seul) comme DCS (dialogue avec contexte sonore environnemental).

	DIALOGUE B		DIALOGUE C		Z	p
	n	% de suites de 3 dessins cibles	n	% de suites de 3 dessins cibles		
francophones DS	29	79,31	29	72,414	0,991	0,161
francophones DCS	29	86,207	29	72,414	2,136	0,163
apprenants DS	21	28,571	21	33,33	0,633	0,263
apprenants DCS	21	42,857	21	47,619	-0,600	0,274

**Tableau 36: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues B et C - VI) sur la construction de suites de 3 dessins cibles, pour les panels de francophones et d'apprenants débutants**

Ainsi, nous pouvons finalement considérer que le dialogue A est celui qui se distingue et qui présente les spécificités les plus marquantes.

Au niveau de la dispersion des réponses, c'est-à-dire de la considération des dessins un à un, chacun des dialogues que nous avons soumis au test a donné lieu à des résultats spécifiques, bien que non fondamentalement différents. Il apparaît donc primordial de se concentrer sur les éléments de différenciation, au sein des réponses, de chacun d'eux. Pour ce faire, nous avons procédé à des tests ANOVA<sup>71</sup>. Ces tests ont révélé que, si nous considérons les dialogues dans leur configuration « **sans contexte sonore** » et pour le **panel de francophones**, alors la variable « *dialogue* » a donné lieu à des différences significatives au niveau de la sélection de dessins **cibles**  $f(2)=2,621$  ,  $p=0,030$  et de **dessins EN** (variation de la situation d'énonciation)  $f(2)=3,207$  ,  $p=0,006$ . Une analyse post-hoc de Scheffé a en outre montré que les différences au niveau de la sélection de dessins **cibles** se retrouvent entre les dialogues A (*la préparation d'un repas*) et B (*le lavage de la vaisselle*) :  $p=0,006$  et entre les dialogues A et C (*à la terrasse d'un café*) :  $p=0,039$ . Les participants ont sélectionné plus de dessins cibles pour les dialogues B et C que pour le dialogue A. Au niveau de la sélection de dessins **EN**, les différences se retrouvent entre les dialogues A et B :  $p=0,011$  et entre les dialogues A et C :  $p=0,041$ . **Les participants ont sélectionné plus de dessins EN pour le dialogue A que pour les dialogues B et C.** En revanche aucune différence significative n'apparaît entre les dialogues B et C. **Encore une fois, il semble donc que le dialogue A soit celui qui se distingue plus particulièrement pour le panel de francophones (sans CSE), notamment au niveau de l'appréhension de la situation d'énonciation.** Cet aspect nous amène à considérer les « valeurs » du présent. En effet, nous avons vu plus haut que pour le dialogue A la « valeur » *générique* du présent semble être plus probable que pour les dialogues B et C pour lesquels cette « valeur » semble peu plausible. Dans ce cadre, nous pouvons envisager que les participants francophones ont pu activer deux types d'interprétation qui se distinguent par la « valeur » du présent mise en œuvre et la situation d'énonciation dans laquelle se « réalise » cette valeur.

Pour le panel de francophones toujours, mais si nous considérons cette fois les dialogues dans leur configuration « **avec contexte sonore** », alors les tests

---

<sup>71</sup> en raison de la nature quantitative des données (dessins sélectionnés, considérés en dehors des suites, en tant que parcelles de réponse)

ANOVA ont montré que la variable « *dialogue* » a présenté un impact sur la sélection de dessins **EN** (variation de la situation d'énonciation)  $f(2)= 3,607$  ,  $p=0,031$ . Les analyses post-hoc de Scheffé ont montré que les différences significatives se retrouvent entre les dialogues A et C, sans que les différences entre les dialogues A et B, d'une part et B et C, d'autre part ne soient significatives. **Une fois de plus, le dialogue A a donné lieu à une plus grande sélection de dessins EN.** Cette remarque nous amène à considérer que le cadre spatio-temporel de ce dialogue (envisagé dans l'élaboration des supports et du test) était plus difficile à appréhender que celui des autres. Nous pouvons tout d'abord envisager que les compreneurs n'aient pas été sensibles à l'« *actualisation* » du présent que permettait le CSE et que, dans ce cadre, les deux interprétations possibles aient été maintenues. Par ailleurs, nous pouvons également envisager que les participants francophones (dans une situation de « sécurité linguistique ») se soient concentrés, pour ce dialogue plus que pour les autres, sur les aspects propositionnels en raison, peut-être, de la spécificité du vocabulaire du dialogue A. Cette éventualité nous amène à considérer que, peut-être, en situation de « sécurité linguistique », le compreneur se concentre sur les aspects verbaux et laisse de côté les indices extralinguistiques, d'autant plus que la consigne n'attirait pas spécifiquement l'attention sur eux.

Si nous considérons les dialogues dans leur configuration « **sans contexte sonore** » pour le panel d'**apprenants débutants**, les tests ANOVA ont révélé des différences significatives aux niveaux des dessins **EN** sélectionnés (variation de la situation d'énonciation)  $f(2)=3,884$  ,  $p= 0,026$  et des dessins **cibles** mal positionnés (variation de la chronologie)  $f(2)=5,870$  ,  $p=0,005$ . Les analyses post-hoc de Scheffé ont montré que, au niveau de la sélection de dessins **EN** une différence significative apparaît entre les dialogues A et C  $p=0,027$  (une fois de plus, les participants ont sélectionné plus de dessins EN pour le dialogue A), mais pas entre les dialogues A et B, d'une part ni B et C, d'autre part. Au niveau de la sélection de dessins **cibles** mal positionnés, il apparaît que les différences sont significatives entre les dessins C et A, d'une part et C et B, d'autre part. Le dialogue C a donné lieu à un plus grand nombre de dessins cibles mal positionnés que les dialogues A ou B. En revanche, aucun écart significatif n'apparaît entre les dialogues A et B, ce qui laisse entendre que **le dialogue C est celui qui se distingue plus particulièrement pour le panel d'apprenants débutants, notamment au niveau de l'appréhension de la situation d'énonciation et du positionnement des dessins cibles. Ces résultats insistent sur le fait que** la situation d'énonciation est une dimension particulièrement difficile à appréhender pour le dialogue A, mais pas pour le dialogue

C. Pour ce dernier, nous pouvons remarquer que les dessins avec une variation de la situation d'énonciation étaient peu plausibles car le lien entre le contenu du discours et le cadre spatio-temporel est particulièrement important (notamment en raison de la « valeur » *actuelle* du présent du dialogue C), alors que pour le dialogue A, les deux « valeurs » du présent restent plausibles et laissent donc plus de possibilités quant à la situation dans laquelle se déroule le dialogue. Dès lors, nous pouvons penser que c'est pour cette raison que les participants n'ont presque pas sélectionné de dessins EN pour ce dialogue. Du point de vue de l'appréhension de la chronologie (principalement pour les dessins cibles), il semble qu'elle ait été plus problématique pour le dialogue C que pour les autres. Nous pouvons envisager que cela tient au fait que les protagonistes, dans ce dialogue, sont de même sexe, ou que, pour cette situation, l'ordre chronologique n'apparaît pas, pour les participants, comme une dimension qui fait sens et donc sur laquelle il est important de porter son attention.

Toujours pour le panel d'**apprenants débutants**, mais dans une configuration d'écoute « **avec contexte sonore** », les tests ANOVA ont présenté une différence significative au niveau de la sélection de dessins EN  $f(2)=5,270$ ,  $p=0,008$ . Une analyse post-hoc de Scheffé a révélé que cette différence se retrouve entre les dialogues A et C,  $p=0,011$  (une fois de plus, les participants ont sélectionné plus de dessins EN pour le dialogue A), mais pas entre les dialogues A et B, et B et C. **Là aussi, il apparaît que la situation d'énonciation correspond à une dimension particulière puisqu'il s'agit du paramètre qui varie le plus dans l'analyse des variations inter-dialogues.** Nous pouvons penser que les compreneurs (quel que soit leur niveau d'expertise linguistique) n'accordent pas toujours la même importance à ce paramètre, notamment en fonction de la difficulté qui se présente au moment de l'interprétation du contenu propositionnel (spécificité du vocabulaire par exemple) et des potentielles « valeurs » du présent. En effet, les résultats que nous avons présentés jusque-là suggèrent que lorsque le compreneur se concentre sur le contenu du discours, il accorde moins d'importance au cadre spatio-temporel dans lequel celui-ci se réalise. Cette remarque nous renvoie d'une part à la théorie de la charge cognitive (les ressources étant limitées, leur répartition est dépendante des processus attentionnels), ainsi qu'à la conception du langage en tant que produit. Si le linguistique est facilement appréhendable, le contexte joue un rôle mineur. Dès lors, en situation pédagogique, il convient, pour l'enseignant, de s'interroger sur la place qu'il accorde au langage et au contexte, notamment dans la formulation des consignes et objectifs, et dans le travail de guidage qu'il peut mettre en place (voir *Comprendre et apprendre*, p.250) selon le niveau des apprenants, notamment dans le cas de classe de niveaux hétérogènes.

Enfin, il apparaît, après avoir procédé à des tests ANOVA que la variable « *dialogue* » n'a pas impliqué de différences significatives entre les TR produits par les panels de francophones et d'apprenants débutants, quelle que soit la configuration (DS ou DCS). »

Au niveau des comparaisons des *profils d'erreurs* nous n'avons pas pu observer d'écarts significatifs entre les dialogues, quel que ce soit le panel (francophone ou apprenant débutant) ou la condition d'écoute (avec ou sans contexte sonore environnemental). Enfin, au niveau des effets d'interaction, nous avons pu observer que les variables *dialogues\_et présence ou absence d'un contexte sonore environnemental* présentent une interaction au seul niveau des temps de réponse, pour le panel de francophones.

### **2.3. Les participants au test**

Enfin, un dernier point du Classe-Images reste selon nous à discuter. En effet, nous avons présenté jusque-là les différents paramètres qui ont pu avoir une influence sur les résultats, ainsi que les résultats eux-mêmes. Nous souhaitons à présent commenter les groupes des participants aux tests.

Tout d'abord, nous pouvons constater que de nombreux résultats ont présenté des tendances (au niveau des différentes variables que nous avons observées). Nous avons fait le choix dans ce travail de ne nous appuyer que sur les résultats statistiquement significatifs ( $p < 0,05$ ). Dans ce cadre, nous pouvons supposer que si nous avions interrogé de plus larges panels de populations, certains des effets marginaux que nous avons observés auraient peut-être entraîné des différences statistiquement significatives.

Par exemple, nous avons vu plus haut que choisir d'interroger des panels de population similaires à la fois pour les tests de validation du matériel expérimental (aspects linguistiques des dialogues, aspects sonores environnementaux des CSE) aurait pu avoir des implications sur les résultats que nous avons recueillis. En effet, si les éléments linguistiques (les dialogues) ont été traités isolément par les participants au Classe-Images (notamment dans le cadre de l'étude de la variable *expertise linguistique*), les aspects sonores environnementaux ont été testés par un panel de population francophone, différent de celui du Classe-Images. Parallèlement,

nous avons également relevé que les corpus d'images que nous avons fait faire ont été évalués par des juges experts. En effet, nous avons vu que certains résultats s'expliquent par l'interaction entre les composantes des supports qui participent au Classe-Images. Toutefois, dans la mesure où chacune de ces composantes (linguistique, sonore environnemental et visuel) n'a pas été testée séparément et par les mêmes participants, il est difficile d'interpréter l'ensemble des résultats recueillis. Nous aurions pu alors imaginer de procéder à des épreuves annexes au Classe-Images, visant la validation des différents éléments qui composent les supports définitifs.

Par ailleurs, sans ajouter de nouvelles tâches au Classe-Images, différentes remarques peuvent être formulées, notamment par rapport au panel d'apprenants étrangers que nous avons interrogé. En effet, nous pouvons tout d'abord noter que les participants à ce panel n'ont pas tous la même nationalité, langue maternelle ou culture d'origine. Or, ces différences culturelles peuvent avoir donné lieu à des interprétations différentes à la fois des scènes sonores et des dialogues. En effet, il se peut que les situations que nous avons mises en scène, si elles correspondent à des situations de la vie quotidienne pour des Français métropolitains, soient beaucoup plus éloignées du quotidien des participants que nous avons interrogés. De la même manière, au niveau des scènes sonores que nous avons réalisées, nous ne pouvons pas savoir si les objets que nous avons enregistrés sont bien connus des participants, et si, de ce fait, ils sont bien reconnus et identifiés. Il aurait pu être intéressant d'interroger différents panels de populations, issues de différentes régions du monde afin de mener une étude comparative selon ce facteur de variation. Cette démarche aurait pu nous permettre de créer des supports spécifiques pour différentes cultures. Toutefois, il apparaît que créer des supports dans la langue-culture cible (ici le français), reste une option à la fois pertinente et potentiellement généralisable : les publics cibles cherchent tous à développer des compétences en langue cible.

### **3. Discussion didactique et théorique**

Nous avons discuté les résultats recueillis à partir de notre expérience menée grâce au Classe-images. Ce test nous a permis d'évaluer les performances des participants (francophones natifs ou apprenants débutants) dans une tâche bien spécifique : la compréhension et restitution, à partir de dessins, de courts dialogues entendus seuls ou accompagnés d'un contexte sonore environnemental. L'origine de

notre réflexion se trouve dans nos questionnements d'enseignante de français langue étrangère et nos observations des difficultés rencontrées par des apprenants allophones lorsqu'un enregistrement sonore à comprendre leur est présenté. Aussi, le matériel que nous avons élaboré pour le test tient-il compte des contraintes d'une situation d'apprentissage en classe de français langue étrangère. Il nous paraît donc important de nous interroger sur la pertinence à utiliser un tel matériel comme support de cours.

### **3.1. Apprendre à comprendre ?**

#### ***3.1.1. Une éventuelle utilisation du Classe-Images en classe de langue***

Il apparaît, suite aux travaux que nous avons menés et présentés dans le cadre de cette étude, que l'évaluation est un outil important pour l'enseignement (voir *Contexte didactique* p.17). En effet, sans la connaissance du niveau des apprenants (en termes de compétences et de performances), il est particulièrement difficile pour l'enseignant de programmer des cours qui répondent à leurs besoins. En ce sens, le Classe-Images pourrait convenir à des objectifs pédagogiques puisqu'il permet non seulement de savoir si un dialogue a été compris (dans le sens où la signification élaborée par l'apprenant/compreneur correspond à celle envisagée par l'enseignant/expérimentateur), mais il renseigne également sur les difficultés rencontrées par l'apprenant (points de divergence entre la signification élaborée par l'apprenant/compreneur et celle envisagée par l'enseignant/expérimentateur). Ainsi, bien que la méthodologie du Classe-Images nous semble particulièrement lourde pour une transposition directe dans la salle de classe, il est important de nous interroger sur les outils de l'évaluation de la compréhension orale en FLE. Dans ce cadre, le Classe-Images permet, comme nous l'avons présenté, de recueillir les performances des apprenants-participants. Nous pouvons observer que la signification de la *performance* est différente dans le contexte expérimental et didactique. En effet, en contexte didactique, il ne s'agit pas seulement de savoir si les apprenants comprennent (dans le sens d'élaborer une signification) ou ce qu'ils comprennent, mais bien de « mesurer » la distance qui sépare ce que le compreneur a compris (la signification qu'il a élaborée) de ce qui était attendu de lui (l'élaboration d'une signification particulière) afin de pouvoir le guider et le « corriger ».

Il semble que le Classe-Images serait un outil utile pour la classe de langue. Toutefois, différentes limites à sa transposition apparaissent. En effet, le recours à une tâche de sélection de dessins a été décidé pour pallier le caractère novice des

participants apprenants de FLE : avec ces publics spécifiques, les tâches de résumé ou de reformulation en langue cible sont particulièrement difficiles à mettre en place et à évaluer (les erreurs sont-elles dues à la compréhension ou à la production ?). Cette première remarque nous amène à penser que le Classe-Images est donc adapté à l'enseignement auprès de publics apprenants débutants. Avec les apprenants avancés, les tâches de reformulation ou de résumé permettent en effet à la fois de vérifier que le support de compréhension orale a bien été compris, mais donne également l'occasion à l'apprenant de produire un discours en langue cible, qui pourra également être évalué par l'enseignant. Le recours à la méthodologie du Classe-Images pourrait donc sembler superflu. Par ailleurs, notre choix d'une tâche de sélection de dessins a été inspiré des activités pédagogiques proposées aux enfants. Nous pouvons donc nous interroger sur la perception que des adultes auraient de cette tâche en situation de classe. En situation de laboratoire, nous n'avons pas constaté d'appréhension particulière mais en situation d'enseignement/apprentissage, nous pouvons craindre que le Classe-Images soit perçu comme une activité infantilisante.

Enfin, le Classe-Images tel que nous l'avons mis en place dans cette étude correspond à une activité individuelle : les participants/apprenants sont seuls face à l'ordinateur qui leur permet de passer le test. Dans ce cadre, une éventuelle transposition de cette méthodologie en salle de classe suppose que les apprenants disposent du matériel nécessaire à la pratique de l'activité. Dit autrement, le Classe-Images ne peut être utilisé que dans le cadre de salles de classe équipées de matériel informatique adapté (un ordinateur et un casque par apprenant).

Le travail à partir du Classe-Images nous semble donc directement transférable en situations d'apprentissage guidé, pour des apprenants débutants, et avec quelques aménagements.

### ***3.1.2. Vers un enseignement stratégique de la compréhension***

Travailler la compréhension orale en classe de langue correspond donc à la poursuite d'un double objectif : il s'agit d'amener les apprenants à développer à la fois des connaissances en langues, et des savoir-faire procéduraux quant aux *stratégies d'écoute et de compréhension*. Développer ses connaissances en langue permet, progressivement à l'apprenant de constituer un « stock » de connaissances qui, si elles sont fréquemment utilisées, présenteront un seuil d'activation relativement bas. Développer des compétences en *stratégies d'écoute et de compréhension* revient, parallèlement, à construire et renforcer des modes de *confrontation* aux supports

oraux. Dans ce cadre, habituer l'apprenant à appréhender des dialogues (qu'ils soient pédagogiques ou qu'ils correspondent à des documents authentiques) de manière globale, et donc à prendre en compte les indices contextuels (potentiellement portés par le CSE) l'amène à mettre en place des automatismes de traitement. Ces automatismes correspondent à des « *programmes, potentiellement conscients, qu'un individu adopte pour résoudre ce qu'il se représente comme un problème pour atteindre un objectif communicationnel particulier* » (Faerch & Kasper, 1983:18). En ce sens, nous pouvons bien parler de stratégies d'écoute et de compréhension qui permettent à l'apprenant de développer ses compétences en compréhension (Spanghero-Gaillard, 2008).

Dans ce cadre, nous pensons que l'enseignant doit, comme préconisé dans la méthodologie SGAV (Metz, 1970; Rivenc & Guberina, 1962), guider progressivement l'attention des apprenants sur différents éléments contextuels ou linguistiques de manière à ce que ces derniers intègrent ces éléments à leur élaboration de sens. Dit autrement, si l'enseignant met en évidence l'importance des éléments contextuels, ceux-ci pourraient jouer un rôle important dans la compréhension du dialogue, et les apprenants pourraient également « prendre l'habitude » de se référer à ce type d'indices dans toutes les situations de compréhension. À partir des résultats obtenus avec le Classe-Images, nous avons vu que les apprenants présentent des difficultés aux niveaux 1) de l'appréhension de la situation d'énonciation et du contenu propositionnel, et 2) des temps de réponse, considérés comme indice de la charge cognitive impliquée dans la tâche (voir synthèse des résultats sur *qu'est-ce qu'être apprenant en langue ?* p.213).

Du point de vue de la charge cognitive impliquée dans la tâche, nous pensons qu'une segmentation de celle-ci pourrait permettre une diminution de la charge en la démultipliant. Pour ce faire, l'enseignant pourrait commencer l'activité par une confrontation à la scène sonore seule. De cette manière, les apprenants pourraient inférer le lieu où se déroule(ra) l'action représentée par le dialogue. Cette première *mise en situation* les amènerait à (pré)activer leurs connaissances linguistiques, pragmatiques et générales et permettrait une meilleure appréhension de la situation d'énonciation. Cette démarche pourrait passer par des questions plus ou moins générales telles que, dans l'exemple du dialogue A *La préparation du repas : Où se déroule la scène que nous avons entendue ?* (Réponse Attendue (désormais RA) : dans une cuisine) ; *Que fait-on **généralement** dans une cuisine ?* (RA : on prépare le(s) repas, on mange) ; *D'après ce qu'on a entendu, quelles actions sont menées dans cette cuisine ?* (RA : quelqu'un prépare un repas) ; *Selon vous, **généralement**, qui prépare les repas ?* (RA : les femmes, les mères) ; *Pouvez-vous deviner ce qui est préparé dans ce*

*qu'on a entendu ?* (RA : quelque chose qui cuit à la poêle, des légumes, de la viande, des œufs). Cette procédure assumerait ainsi (et potentiellement après plusieurs écoutes) la fonction de préparation attentionnelle. Cette confrontation pourrait en outre présenter différents degrés de granularité, c'est-à-dire qu'il reviendrait à l'enseignant de déterminer « jusqu'où » aller dans l'analyse de ce support. Dans un second temps, l'enseignant pourrait présenter le dialogue accompagné du CSE. À ce stade, il nous semble que la reprise des questions classiques dans la compréhension de documents oraux permettrait de poursuivre le guidage des apprenants : *Qui parle ?* (RA : un homme et une femme) ; *De quoi parlent-ils ?* (RA de la préparation du repas) ; *Qui prépare le repas ?* (RA : la femme) ; *Que vont-ils manger ?* (RA : des légumes et des œufs) ; *Quels légumes vont-ils manger ?* (RA : des pomme de terre, des champignons et des oignons) ; *Ce menu convient-il à l'homme ?* (RA : oui). Ces questions permettraient de guider l'attention des apprenants sur les aspects du contenu propositionnel sélectionnés comme prioritaires par l'enseignant. Encore une fois, nous n'excluons pas que plusieurs écoutes soient nécessaires pour répondre à l'ensemble des questions. Enfin, une fois que l'ensemble des questions aura reçu une réponse satisfaisante (que le dialogue sera considéré comme « compris »), l'enseignant pourra proposer une dernière écoute du dialogue afin que les apprenants puissent prendre conscience de la globalité du document et de la présence des différents indices qui ont permis la compréhension. Dans ce cadre nous pouvons envisager que la cohésion du dialogue sera reconstruite par les apprenants et que, indirectement, la chronologie du dialogue apparaîtra plus distinctement. Par ailleurs, ces supports pourront également être utilisés en phase de transposition, afin de permettre le travail de différents aspects de l'apprentissage de la langue. Par exemple, pour travailler la production orale, nous pouvons envisager de réutiliser les scènes sonores dans les jeux de rôles. Ces activités, très fréquemment utilisées dans les approches communicatives visent le réinvestissement des structures (lexicales, syntaxiques, pragmatiques) vues pendant le cours dans une production de discours plus ou moins spontanée, plus ou moins guidée. Nous pensons que *mettre en contexte* ces exercices, à partir de sons environnementaux, pourrait permettre aux apprenants de se sentir investis dans une tâche plus réaliste (les sons environnementaux seraient des « accessoires »).

Enfin, il nous semble que les situations de compréhension peuvent être envisagées comme des situations de *résolutions de problème*. En effet, comme nous l'avons vu dans notre cadre théorique, les activités de perception et de compréhension sont guidées par les phénomènes attentionnels, eux-mêmes

dépendants des intentions et des objectifs d'écoute et de compréhension. Dans ce cadre, nous pouvons considérer que la compréhension ne correspond pas à une fin en soit mais au moyen d'accéder à la résolution d'un problème. Il peut s'agir de la réponse à des questions de compréhension qui, si elles sont données en amont, guideront l'écoute. On peut également envisager que la compréhension du support oral n'est qu'une étape d'une programmation plus vaste. Dit autrement, il s'agit de faire de l'activité de compréhension orale une activité intégrée à une séquence plus vaste, dans une perspective actionnelle. Dès lors comprendre devient le moyen de *résoudre un problème*. Dans ces situations l'individu (l'apprenant) « *apprend à reconnaître dans une situation-problème les connaissances qui vont être utiles à sa résolution (puis) l'individu apprend à créer des règles qui consistent en l'appariement entre une condition et une action. Ce appariement pourra devenir un automatisme s'il est utilisé très fréquemment et longtemps* » (Chanquoy et al., 2007:107). Il s'agirait donc en quelque sorte d'une analyse moyens-fin, dans laquelle l'apprenant mesurerait les écarts entre la situation de départ et l'objectif et envisagerait la compréhension comme une partie de la stratégie à mettre en place pour atteindre le (ou les) objectif(s).

### 3.2. Comprendre et apprendre?

Dans les premières parties de ce travail, nous avons défini la compréhension comme l'élaboration d'une signification nouvelle par un individu, à partir d'éléments perçus et de ses connaissances antérieures. Les recherches indiquent en outre que la compréhension est une activité irrépressible et qu'elle correspond à un **besoin** des individus qui cherchent continuellement à interpréter leur environnement. Toutefois, ce besoin d'appréhender notre environnement est détourné dans un cours de langue en ce sens que l'individu à qui est présenté un document sonore doit comprendre ce que l'enseignant a lui même repéré et sélectionné comme étant les éléments les plus importants. Ainsi, l'enseignant est généralement amené à évaluer la compréhension des apprenants, non pas dans une perspective « absolue » comme celle définie plus haut et dans un cadre disons, naturel, mais plutôt en termes binaires (compris/pas compris) en considérant que les apprenants ont compris **uniquement lorsque la signification qu'ils ont élaborée correspond à celle attendue (prévue par l'enseignant)**. Dans ce cadre, ce n'est plus le processus cognitif de la compréhension qui est considérée, mais bien son résultat. En ce sens, ne peut-on considérer que les évaluations de compréhension ou non-compréhension prévues par l'enseignant correspondent à des « jugements de valeurs » ?

### 3.2.1. Qu'est-ce qu'apprendre ?

Nous avons vu que nous nous positionnons dans le domaine de la didactique cognitive des langues. Aussi, nos recherches sur la compréhension orale visent la description des comportements langagiers qui rendent compte de l'activité cognitive engendrée lorsque l'apprenant est soumis à une tâche envisagée comme une activité d'apprentissage sollicitant une expression de sa compréhension. Bien souvent, dans une telle activité, l'apprenant se concentre sur les éléments linguistiques qui lui sont proposés à l'écoute. Il cherche à repérer un mot, qu'il connaît, et à partir duquel il va convoquer une série de connaissances (notamment linguistiques mais pas exclusivement) pour inférer un sens à ce qu'il entend. De plus, comme il est en situation d'apprentissage, il sait que l'enseignant attend de lui qu'il manifeste sa compréhension. Dans ce contexte, très souvent non explicité, nous nous interrogeons sur les moyens que l'enseignant peut mettre en œuvre pour aider l'apprenant à apprendre à comprendre. La compréhension correspond à un processus, à un savoir-faire procédural. Comment l'apprendre et comment l'enseigner ? Selon nous, apprendre à comprendre revient à développer de nouvelles compétences, à automatiser différentes procédures pour les rendre plus efficaces, à partir de la pratique. Dit autrement, on apprend à comprendre en comprenant. « (...) *tout comme un apprenant (quelle que soit la nature de son apprentissage) devient de plus en plus performant au fur et à mesure qu'il s'entraîne, progresse et apprend, le système cognitif deviendrait également plus performant et cette performance se traduirait, entre autres, par un moindre besoin en ressources cognitives dû à l'automatisation de certains traitements* » (Chanquoy, Tricot, & Sweller, 2007:42). Il s'agit donc, dans une situation d'enseignement/apprentissage de confronter les individus-apprenants à différentes situations dans lesquelles ils devront élaborer une signification efficiente, c'est-à-dire en adéquation avec le sens élaboré dans la production du message et avec les attentes de l'enseignant. C'est dans ce cadre, et à force de répétitions, que l'apprenant développera des automatismes, renforcera des procédures. Ces automatismes correspondent à des schémas d'activation, peu coûteux cognitivement, « *mis en jeu dans des conditions définies qui ont déjà permis de traiter une information semblable* » (Sieroff, 1992:16). Selon nous, ces automatismes peuvent intervenir à deux niveaux : à la fois au niveau de la connaissance de la langue (qui permettra l'identification et l'association des unités de la parole avec des unités de sens) et au niveau des *stratégies d'écoute et de compréhension*.

Il apparaît donc que l'apprentissage de la compréhension correspond à l'élaboration, à force de pratique, d'automatismes (traitements inconscients et difficilement contrôlables) dont le développement permettrait une diminution de la

consommation de ressources cognitives, permettant l'exécution de plusieurs traitements parallèles (Schneider & Shiffrin, 1977). Ainsi, ils s'opposent aux traitements attentionnels des tâches qui s'effectuent en pleine conscience et sont plus coûteux cognitivement. Pour Cohen, Servan-Schreiber, & McClelland (1992), il existerait un continuum d'automatisme qui permettrait aux individus de passer d'un traitement automatique à un traitement contrôlé en fonction de la tâche. Ainsi, on peut même imaginer que le recours à des processus automatisés permette de dégager des ressources cognitives et attentionnelles disponibles pour effectuer en parallèle d'autres tâches, cette fois plus contrôlées. « *En effet, l'acquisition d'automatismes correspondrait à la soustraction graduelle de la demande attentionnelle* » (Sieroff, 1992:13). Ainsi, dans la mesure où l'activation de connaissances déjà élaborées peut correspondre à un traitement automatique, cette procédure laisserait des ressources disponibles pour l'élaboration de nouvelles connaissances, élaboration qui « *requiert la participation attentionnelle et consciente du sujet* » (Perruchet, 1988:92). Dit autrement, la compréhension d'items connus *i.e.* qui ont déjà fait l'objet d'un apprentissage (implicite ou explicite) est moins coûteuse cognitivement que la compréhension d'éléments en cours d'apprentissage mais l'automatisme de la première rend le traitement parallèle des deux types de compréhension possible. Ainsi, l'enseignant stratégique s'appuierait sur les automatismes d'ores et déjà en place chez les apprenants. Dans le cadre de l'enseignement auprès d'apprenants débutants, on peut supposer qu'un certain nombre d'éléments para et extralinguistiques sont déjà maîtrisés, et ce d'autant plus avec des adultes. Dans ces situations, l'enseignant peut amener les apprenants à expliciter (par la verbalisation par exemple) ce qu'ils ont compris à partir, notamment, du CSE, des gestes ou de la prosodie par exemple. Cette méthode pourrait ainsi permettre aux apprenants de construire des procédures de repérage et même d'automatiser cette approche. À la suite de ce travail à partir des éléments connus de l'apprenant, l'enseignant, et l'apprenant pourront alors se concentrer sur les items moins connus du support. Comme nous l'avons explicité, nous considérons qu'à partir d'un questionnement ciblé, l'enseignant peut guider les inférences des apprenants, les amenant à élaborer des significations nouvelles, en association avec de nouvelles formes linguistiques ou culturelles. Ces nouvelles significations pourront par la suite être stabilisées sous forme de nouvelle(s) connaissance(s) dans le cadre de leur transposition et réutilisation dans des contextes plus ou moins similaires. Dès lors, une programmation « en spirale » de l'enseignement et des objets à comprendre et à apprendre entraînera de nouvelles itérations de la nouvelle connaissance, et par la même occasion, son automatisation. Par cette approche nous affirmons que l'apprentissage correspond, au moins en partie à une modification de la « *structure*

*attentionnelle liée à l'expertise* » (Camus, 1996:37). **Dans ce cadre, nous considérons l'apprentissage de la compréhension comme le développement par l'apprenant d'automatismes qui lui permettront, dans les situations extérieures à la salle de classe, de pouvoir élaborer une signification à partir des discours perçus.** En outre, si la démarche (repérage des éléments connus, formulation d'inférences, élaboration d'une nouvelle connaissance et réemploi) que nous avons exposée ci-avant est fréquemment répétée dans les situations d'apprentissage, nous pouvons envisager qu'elle soit (elle aussi) automatisée par l'apprenant et de ce fait également mise en place dans les situations écologiques (extérieures à la situation didactique) et naturelles d'utilisation de la langue.

Dans cette perspective, nous considérons que les dialogues pédagogiques avec contexte sonore environnemental et leur utilisation telle que nous l'avons exposée participent à l'apprentissage de la langue à deux niveaux : d'une part en donnant des exemples de la *langue en situation* et, d'autre part en amenant les apprenants à mettre en place des stratégies d'écoute et de compréhension.

À ces deux niveaux, les exercices de compréhension orale à partir de tels supports permettront selon nous, le développement d'automatismes. Présenter la langue en situation amène l'apprenant à (re)construire des significations pour les unités de la langue qu'il perçoit. Dit autrement, l'apprenant « comprend » la langue et peut dès lors mémoriser ce qu'il a compris. Parallèlement, confronter l'apprenant à des productions langagières potentiellement complexes pour son niveau et, dans le cadre de nos travaux, accompagnées d'éléments sonores environnementaux, peut l'amener à développer et renforcer des comportements (nouveaux) dans cette confrontation. Notre positionnement didactique est donc qu'à travers une pratique répétée d'exercices de compréhension orale, en l'accompagnant et en l'amenant à prendre conscience de sa démarche (implicite et utilisée habituellement en langue maternelle) l'apprenant développera et renforcera des schémas d'activation et des procédures automatiques lui permettant de développer des savoir-faire dans ce domaine. En outre, accompagner l'apprenant sur le cheminement de l'élaboration d'une signification et d'une connaissance lui permet de développer ces nouveaux savoirs dans un cadre privilégié, qui l'amènera, peut-être, à asseoir sa confiance en lui et renforcera (parfois) le contrat didactique qui unit l'enseignant et l'apprenant.

**Il ne s'agit pas, pour la compréhension orale d'amener les apprenants à connaître la totalité des énoncés qu'ils seraient susceptibles de rencontrer dans des conditions écologiques de pratiques de la langue, mais de les amener à**

**développer des stratégies de confrontation à la langue parlée.** Dans ce cadre, les exercices de compréhension orale correspondent à des mises en situations qui permettront à l'apprenant de développer ses connaissances *de* et *sur* la langue (« *L'apprenant élabore le système de la langue cible à partir de contacts avec des échantillons plus ou moins représentatifs de la norme* » (Py, 2000:80)). En effet, « *s'agissant d'une mise en situation de compréhension, un certain nombre de connaissances de l'individu sont mobilisées, et pas uniquement linguistiques. Dans ses activités de communication, l'individu a recours cognitivement et simultanément à des représentations verbales et non verbales (en références à des objets, événements ou attitudes)* » (Spanghero-Gaillard & Garnier, 2013:119).

En ce sens, aider les apprenants à développer des compétences en compréhension orale, c'est les amener à développer et à prendre conscience de leur comportement de compreneurs, c'est-à-dire de mise en œuvre de stratégies de repérage d'informations, tant dans leur langue maternelle que dans la langue cible. Dans le cadre de nos travaux, nous avons mis en avant le repérage d'indices sonores environnementaux qui, dans les dialogues pédagogiques, peuvent être portés par les CSE, mais qui accompagnent systématiquement la parole en condition naturelle. Toutefois, ces indices peuvent également être portés (et donc repérés et interprétés) dans les images (telles que celles du Classe-Images par exemple), les gestes ou les mimiques. Dès lors, la langue devient un mode d'expression parmi d'autres et ne détient plus le monopole du cours de langue. Développer des compétences en compréhension, pour les apprenants, revient à adopter un comportement de compreneur, attentif à son environnement. Pour l'enseignant il s'agit alors de guider ce comportement et d'amener l'apprenant à en prendre conscience. Nous pensons en effet que c'est en conscientisant ce comportement que l'apprenant le repèrera et pourra, *a priori*, le transposer de sa langue maternelle vers la langue cible.

### **3.2.2. Comprendre pour apprendre ...**

Dans le cadre de la fonction d'exemple de production langagière, l'objectif principal du travail de la compréhension orale, à partir de dialogues pédagogiques, est d'amener l'apprenant à développer ses connaissances en langues, notamment au niveau du vocabulaire, de la grammaire et de la pragmatique. Deux paramètres primordiaux semblent participer à l'élaboration, au stockage et à l'automatisme des connaissances : la fréquence et la profondeur. La **fréquence** se définit par l'itération : plus la présentation et le traitement d'un item seront répétés, plus la représentation cognitive qui lui est associée sera fréquemment activée, et de ce fait disponible. Il est donc primordial de présenter, le plus fréquemment possible l'objet compris et en

cours de mémorisation dans des situations à la fois différentes (nouveau contexte d'apparition ou d'utilisation) mais partageant assez de points communs avec la première présentation de l'objet pour que l'intégration de la nouvelle occurrence à une connaissance plus générale soit facilitée.

La **profondeur** du traitement désigne l'implication attentionnelle et la nature sémantique des opérations effectuées (Craik & Lockhart, 1972). Ainsi, passer par une « formulation non linguistique » telle que le permettent les sons environnementaux, les images ou les gestes par exemple nous semble particulièrement intéressant puisque cette démarche rend possible le traitement sémantique et est propice à l'émission d'inférences (guidée par l'enseignant). Il apparaît, en outre, que la nature de la tâche, lors de l'élaboration de la représentation, et donc la profondeur des traitements impliqués, sera plus déterminante pour le stockage que les caractéristiques formelles de l'objet. Selon Mandler (2002), un traitement profond correspond à un traitement sémantique et concourt à l'assimilation du stimulus dans une structure signifiante, débouchant ainsi sur de meilleures performances en tâche de rappel. Dès lors, la compréhension d'un item *via* un traitement profond aura des conséquences positives sur sa mémorisation. C'est l'une des raisons qui incitent certains méthodologues à prôner de ne jamais traduire en classe de langue. En outre, il apparaît que l'effort fourni pour « comprendre » un message a un impact sur la mémorisation des objets à comprendre et à mémoriser (Auble & Franks, 1978). Dit autrement, lorsque l'apprenant est confronté à des éléments linguistiques qu'il ne maîtrise pas, il doit mettre en place des traitements profonds qui donneront lieu à de meilleurs résultats en tâche de récupération. Dans le cadre de la compréhension de dialogues accompagnés d'un contexte sonore environnemental, l'apprenant pourra (s'il y est incité) recourir aux indices extralinguistiques pour élaborer une signification à partir de l'*input*, et ces indices participeront au stockage de l'information en MLT. En effet « (1) les informations sont encodées en MLT avec des indices contextuels pour former des représentations complètes et complexes ; (2) les indices contextuels stockés avec l'information font partie intégrante de la représentation mnésique et constituent des indices lors de la récupération ; (3) ces indices vont faciliter largement la recherche et l'accès à l'information (...) » Chanquoy, Tricot, & Sweller, 2007:71). Dans cette perspective, une procédure de compréhension à partir des éléments contenus dans le support, amènerait les apprenants à effectuer un traitement plus profond des objets linguistiques à comprendre, auquel les éléments contextuels participeraient. Cette démarche faciliterait la mémorisation. En outre, Deschamps & Moulignier (2005:509) nous rappellent que « *le contexte, au sens large, influence considérablement l'enregistrement des données. Il peut s'agir d'un*

*indiciage volontaire (sémantique ou selon l'apparence de la donnée) ou automatique (en particulier environnemental : heure de la journée, lieu, environnement subjectif, humeur ...)* ». Dans la conception *structurale* de la mémoire, la différence entre information liée au *stimulus* (stockée en mémoire sémantique) et information liée au *contexte d'apparition du stimulus* (stockée en mémoire épisodique) est importante. Toutefois, comme nous le font remarquer Blanc & Brouillet (2003), cette distinction est difficile à maintenir : « *Les données recueillies depuis ces vingt dernières années militent en faveur de l'idée selon laquelle stimulus et environnement forment un tout indissociable : c'est de leur co-dépendance qu'émerge le sens, ainsi que le souvenir* » (Blanc & Brouillet, 2003:63-64). En effet, il apparaît que les informations contextuelles permettent de mieux caractériser l'information, d'effectuer un codage plus profond dans plusieurs systèmes mnésiques et, de ce fait, de proposer de plus nombreux indices de récupération. En outre, la mémorisation se fait toujours en contexte. Tulving (1962, 1972) a mis en évidence que la qualité du contexte d'encodage a un impact fort sur les tâches de rappel : plus les contextes d'encodage et de récupération sont proches, plus les performances en rappel sont élevées. Dit autrement, en situation de récupération, un indice (« *qui devra être un fragment de la situation d'apprentissage, c'est-à-dire une information contextuelle stockée en même temps que l'information à mémoriser* » (Chanquoy, Tricot, & Sweller, 2007:73)) facilitera l'accès (rappel) de la connaissance stockée en mémoire à long terme. En ce sens, nous pouvons penser que la présence de scènes sonores pour accompagner les dialogues pédagogiques jouerait un rôle important dans l'apprentissage (et pas seulement dans la compréhension). **En effet, les indices contextuels devraient permettre à l'apprenant non seulement de « mieux comprendre » les éléments linguistiques auxquels il est confronté, mais aussi de mieux mémoriser les éléments de signification construits.**

Selon Chanquoy, Tricot, & Sweller (2007:70) « *Il existerait trois grands types d'informations contextuelles qui peuvent être codées avec l'information à mémoriser : (a) les informations contextuelles externes (ou environnementales qui correspondent aux informations sans rapport apparent avec la tâche mais présentes dans l'environnement) ; (b) les informations internes (qui correspondent aux informations là encore non pertinentes pour la tâche mais caractérisant l'état émotionnel et physiologique général de l'individu) ; (c) les informations intrinsèques (qui sont des caractéristiques propres du matériel à mémoriser)* ». Si l'on met en perspective ces trois types d'information contextuelle avec les objectifs de notre démarche expérimentale, alors on peut supposer que le contexte sonore environnemental que l'on a ajouté aux dialogues pédagogiques propose des informations *externes* (conçues pour assumer une fonction *situationnelle* (voir dans le *Contexte didactique* p.17), pour

rendre compte de la situation d'énonciation des discours, à travers la mise en onde des lieux dans lesquels se déroulent les échanges). Nous pouvons également considérer que le contexte sonore environnemental, lorsqu'il assume une fonction de *transcodage* (voir dans le *Contexte didactique* p.17), représente des informations *internes*. Enfin, les informations *intrinsèques* sont contenues dans les aspects paralinguistiques des discours.

En outre, les informations contextuelles prises en compte par l'apprenant lors de l'activité de compréhension orale le renvoie au contexte écologique des situations mises en scènes. Dès lors, la situation et le cadre de l'interaction correspondent à un élément connu de l'apprenant et auquel, s'il y est incité par l'enseignant, il pourra s'attacher pour élaborer la nouvelle signification, voire la nouvelle connaissance. En outre, rapprocher la situation d'apprentissage de la vie (quotidienne) de l'apprenant peut faciliter sa motivation, l'aider à définir des objectifs (d'action et/ou d'apprentissage) et focaliser son attention sur des aspects spécifiques de la situation mise en scène. Ces objectifs et cette focalisation attentionnelle pourront à leur tour guider les processus de perception et de compréhension, permettant ainsi un apprentissage non seulement plus aisé, mais aussi correspondant mieux aux objectifs de l'apprenant.

Enfin, accéder à la signification d'un item (une lexie par exemple) par une procédure d'induction, de déduction ou d'abduction telle que l'élaboration de la signification à partir du contexte (qu'il soit phrastique, discursif ou extralinguistique) correspond à ce que l'on appelle *l'apprentissage implicite* et est moins coûteux cognitivement. Cet apprentissage correspond à un « *mode d'adaptation dans lequel le comportement d'un sujet apparaît sensible à la structure d'une situation, sans que cette adaptation ne soit imputable à l'exploitation intentionnelle de la connaissance explicite de cette structure. Il ne s'agit pas d'affirmer l'absence de toute connaissance explicite, mais seulement de souligner que l'adaptation comportementale ne repose pas sur l'exploitation intentionnelle de cette connaissance* » (Perruchet & Nicolas, 1998:15). Toutefois, ce mode d'apprentissage reste difficile à mettre en place et à contrôler en classe où le rôle de l'enseignant est, comme nous l'avons vu, d'amener l'apprenant à expliciter sa compréhension et parfois ses comportements. Cette explicitation, dans le cadre de la compréhension orale, porte sur les relations qu'entretiennent les éléments présentés (dans le dialogue notamment) avec ses connaissances antérieures, qu'elles relèvent de sa langue maternelle, d'une autre langue ou des connaissances générales sur le monde. Il s'agit donc pour l'enseignant de trouver sa place dans ce mode de relation, à la fois guide (dans les procédures) et modèle (pour

l'usage de la langue cible), il doit savoir repérer les indices de la compréhension et de la non-compréhension, amener les apprenants à développer des savoir-faire procéduraux et à s'approprier la langue.

Ainsi, il apparaît qu'au-delà des connaissances en langue que l'apprenant pourra développer dans les activités de compréhension orale, il sera également amené à développer un nouveau mode de comportement face à la langue cible. En ce sens, comprendre serait un comportement, un savoir-faire, une connaissance procédurale.





## Conclusion générale

Dans notre étude, nous nous sommes interrogée sur la compréhension orale des apprenants débutants en langue étrangère et plus particulièrement en français. Cette réflexion, issue de notre pratique enseignante, s'est organisée autour de deux questions principales : *Qu'est-ce qu'être apprenant débutant en langue ? Dans quelle mesure un contexte sonore environnemental peut faciliter la compréhension d'un dialogue pédagogique ?*

Pour répondre à ces questions, nous avons tout d'abord convoqué un cadre théorique qui nous a amenée à définir les termes de *compréhension* et de *perception*. Dans cette perspective, nous avons, dans un premier temps, défini ces deux processus de manière indépendante. Nous avons défini la compréhension comme l'élaboration d'une nouvelle signification, à partir du perçu et des connaissances antérieures du comprendre. Parallèlement, nous avons caractérisé la perception comme « *l'acte par lequel un sujet se rapporte à la réalité* » (Barabas, 2009: 44). Ainsi, la perception fonderait le processus de compréhension : perception et compréhension seraient deux processus fortement imbriqués, intrinsèquement liés et co-dépendants. Nous sommes parvenue à la conclusion qu'il n'y a pas de compréhension sans perception et que réciproquement la perception conduit, le plus souvent, à une forme de compréhension. Les études sur la perception mettent également en évidence que les processus perceptifs démarrent en amont de la « rencontre » avec le stimulus, dans la formulation d'hypothèses et d'attentes. En outre, la compréhension a été décrite comme un phénomène irrépressible. Dit autrement, l'individu est en perpétuelle recherche de sens et cherche toujours à interpréter (comprendre) son environnement (perçu), pour y prendre place. La perception est donc un processus

de haut niveau, guidé par les connaissances antérieures de l'individu et qui participe, selon nous, au processus plus vaste qu'est la compréhension. Dans ce cadre, nous pouvons aujourd'hui, et à la lumière des nombreux travaux qui portent sur ces deux processus (voir cadre théorique p.49), nous interroger sur la pertinence de maintenir une distinction entre ces deux champs d'investigation, au demeurant si proches. En effet, nous proposons d'envisager la perception comme une partie de l'élaboration générale de sens, dans laquelle s'inscrit l'individu. Nous considérons que perception et compréhension, telles qu'elles sont définies aujourd'hui, participent conjointement à l'interaction entre l'individu (et son univers conceptuel, élaboré au fil des apprentissages) et le monde extérieur. Ainsi, le questionnement originel de notre étude dépasse-t-il la seule situation d'enseignement/apprentissage d'une langue étrangère.

Notre questionnement sur la compréhension orale des apprenants débutants nous a amenée à élaborer un test *ad hoc* : le Classe-Images. Ce protocole particulier permet d'observer certains aspects de la compréhension de courts dialogues oraux, *via* une tâche de sélection de dessins (contrôlés) pour rendre compte du support sonore. Trois dialogues ont été construits, conformément aux supports d'enseignement utilisés dans les salles de classe de français langue étrangère. Trois scènes sonores ont également été construites, afin d'être associées aux dialogues, dans une position contextuelle. Ces élaborations ont fait l'objet d'une présentation spécifique dans ce mémoire (voir p.121). Nous avons vu, que chacun de ces types de supports (sonore environnemental et linguistique) a fait l'objet d'une procédure empirique et d'une validation expérimentale.

L'élaboration de dialogues pédagogiques nous a amenée à réinterroger les considérations de fréquence dans le choix du vocabulaire à enseigner (voir p.153). En effet, depuis 1950 et l'avènement du Français Fondamental, nous considérons que le vocabulaire fréquent et disponible est à enseigner en priorité. Depuis, des bases de données lexicales sont disponibles et peuvent nous renseigner sur la fréquence des items lexicaux dans différents type de productions langagières (nous avons fait le choix de travailler avec la base de données *Lexique*, et plus spécifiquement du sous-corpus composé de sous-titrages de films). Néanmoins, nous pouvons aujourd'hui nous interroger sur ce critère de sélection et de définition de cette partie des objectifs d'enseignement et d'apprentissage. Un enseignant peut-il fonder la progression de son cours sur le seul critère de la fréquence ? Ne devrions-nous pas prendre en

compte également la dimension utilitaire du vocabulaire ? En effet, certains items lexicaux peuvent être à la fois spécifiques et fréquents. La question de leur disponibilité est alors posée. Dans ce cadre, nous considérons qu'à la suite de Gougenheim, Michéa, Rivenc, et Sauvageot, en 1956, et de la publication de la liste *Éduscol*, une nouvelle étude sur le "vocabulaire utile" du français (et spécialement en langue étrangère et seconde) pourrait aujourd'hui s'avérer intéressante, tant dans le cadre de l'enseignement du français général que pour le français de spécialité. Il s'agirait dans ce cadre, non pas d'amener l'apprenant de Français Langue Étrangère à s'approcher du locuteur idéal (tel qu'envisagé et décrit par Chomsky), mais de cibler les compétences à développer en priorité pour augmenter de manière efficiente les performances communicationnelles et actionnelles du locuteur novice dans une démarche encouragée par le conseil de l'Europe à travers le Cadre Européen Commun de Référence pour les Langues (Conseil de l'Europe, 2001)

Parallèlement à l'élaboration du matériel linguistique, nous avons construit le matériel sonore environnemental utilisé dans le Classe-Images. Comme nous l'avons présenté, cette élaboration s'est appuyée sur différentes enquêtes et études empiriques (voir p.121). Notre objectif à ce niveau était la réalisation de scènes sonores rendant compte non seulement des environnements ciblés, mais plus spécifiquement, des représentations que des individus de culture française en ont. Cette démarche a pour objectif de permettre la réalisation de scènes sonores facilement identifiables et compréhensibles pour les participants au Classe-Images. Cette approche repose sur l'idée que si la perception du monde est pilotée par les connaissances antérieures de l'individu (théorie du *knowledge driven*), alors ce qui est connu est plus facilement identifié (re-connu). Dès lors, construire des supports qui se rapprochent de ce que connaissent les individus, ou du moins de la représentation qu'ils en ont élaborée, permet une interprétation/compréhension plus aisée. Cette démarche nous a en outre permis d'intégrer au Classe-Images des supports fortement contrôlés, et ainsi de réfléchir aux impératifs liés à l'élaboration de matériel pédagogique pour les niveaux débutants. En effet, dans ce type de réalisations, il ne s'agit pas nécessairement de rendre compte de la réalité des choses (que ce soit au niveau des éléments linguistiques ou extralinguistiques) mais plutôt de re-créeer un environnement stéréotypé dont les éléments saillants « rappellent » cette réalité et les connaissances (ou représentations) que les apprenants en ont. Dit autrement, il ne s'agit pas d'« être » vrai, mais de « *faire vrai* » et d'amener, à partir de situations reconstruites artificiellement, les apprenants à développer des comportements et des connaissances transférables en dehors de la classe, tout en respectant des impératifs de progression. Il s'agit de s'appuyer sur les savoirs et

savoir-faire des apprenants (qu'ils soient linguistiques ou non), pour introduire de nouveaux éléments de connaissances et de nouveaux comportements.

Dans ce cadre, nous avons pu observer que, lorsque les éléments linguistiques et sonores environnementaux entretiennent un rapport de redondance complémentaire, la présence d'un Contexte Sonore Environnemental représente une aide à la compréhension du support (et notamment des éléments linguistiques). Se pose alors la question de l'impact que pourrait avoir un CSE non-congruent avec le dialogue. En effet, nous pensons que les CSE ont facilité la compréhension des apprenants débutants, notamment parce que les supports « complexes » (c'est-à-dire composés de parole et de sons environnementaux) restaient cohésifs et cohérents. Nous ne considérons pas que tout CSE permet une amélioration des performances. Ainsi, nous pensons que la présence de scènes sonores environnementales en position contextuelle par rapport à un dialogue n'est bénéfique que lorsque les deux composantes entretiennent des rapports complémentaires. Ajouter un CSE qui ne respecterait pas ces caractéristiques entraînerait des difficultés supplémentaires. Cette remarque nous amène à envisager des prolongements à ce travail, qui permettraient, entre autres, de faire la lumière sur les éventuels obstacles à la compréhension, engendrés par des éléments extralinguistiques mal construits ou mal sélectionnés.

Dans le cadre du Classe-Images, nous avons interrogé des participants francophones (considérés comme experts linguistiques) et des apprenants débutants en Français Langue Étrangère. Ces participants ont pu entendre et rendre compte des dialogues, présentés seuls ou accompagnés d'un contexte sonore environnemental. La comparaison des résultats obtenus dans ces différentes configurations (francophones soumis aux dialogues seuls ou accompagnés d'un CSE ; apprenants débutants soumis aux dialogues seuls ou accompagnés d'un CSE) nous a permis d'apporter des éléments de réponses à nos questions de départ.

En comparant les performances des francophones natifs et des apprenants débutants, nous avons pu observer certaines spécificités liées à l'expertise linguistique et à son déficit. Il apparaît donc que, *a minima* dans une tâche aussi spéciale que celle du Classe-Images, les novices en langue rencontrent des difficultés principalement au niveau de l'appréhension de la situation d'énonciation et du contenu propositionnel. En outre, le déficit de connaissances linguistiques a également entraîné une augmentation de la charge cognitive impliquée dans la tâche (voir p.204). Ces résultats nous amènent à penser qu'en situation de classe

l'enseignant doit privilégier une approche guidée et stratégique des supports de compréhension orale, notamment en insistant sur les aspects situationnels (cadre spatio-temporel, intentionnalité des messages). Selon nous, et conformément à notre hypothèse de départ, le contexte sonore environnemental peut participer à ce repérage d'indices situationnels et, en ce sens, faciliter la compréhension.

Afin de vérifier cette deuxième hypothèse, nous avons comparé les performances des participants (francophones et apprenants débutants) qui ont écouté les dialogues seuls et accompagnés d'un contexte sonore environnemental (CSE). Cette deuxième comparaison nous a permis d'observer que, pour les apprenants débutants, la présence d'indices sonores environnementaux contextuels peut effectivement faciliter l'appréhension de la situation d'énonciation et, de manière plus générale, faciliter la compréhension du dialogue. Nous avons également pu constater que la présence d'un contexte sonore environnemental est plus favorable (représente une aide plus importante) aux participants novices en langue qu'aux participants experts. Néanmoins, parallèlement à l'amélioration des performances, nous avons pu observer que la présence d'un CSE a entraîné une augmentation de la charge cognitive. Toutefois, il apparaît également que cette augmentation n'a pas impliqué d'effet négatif sur la qualité des réponses proposées par les participants. À partir de ces différents résultats nous pouvons donc nous interroger sur les implications de la complexification des supports de test (à l'origine de l'augmentation de la charge cognitive) dans la compréhension des apprenants de la langue cible (les résultats semblent montrer que la compréhension n'a pas été affectée par l'intensification de la charge cognitive). Pour expliquer cela nous avons avancé que les novices ont recours aux indices extralinguistiques pour « pallier » leurs difficultés au niveau linguistique. Le CSE représenterait, pour les apprenants débutants, une aide à la compréhension en suppléant les éléments inconnus dans la composante linguistique. Dans ce cadre, nous pouvons penser que, pour les novices en langue, les composantes linguistiques et sonores environnementales entretiennent un rapport de redondance complémentaire, alors que pour les experts linguistiques, la présence d'un CSE peut être perçue comme « simplement » redondante avec le dialogue, entraînant ainsi une surcharge cognitive (par rapport à la compréhension du dialogue seul) potentiellement inutile.

Ces différents résultats nous ont amenée à conclure que la présence d'un CSE, en fournissant divers indices extralinguistiques situationnels, peut faciliter la compréhension des apprenants débutants. Ce public spécifique a retenu notre attention pour différentes raisons : d'une part parce que, dans notre pratique

enseignante, ce public est celui avec lequel nous travaillons le plus fréquemment (tant par choix que par concours de circonstances), et d'autre part parce que, situé en début d'apprentissage, les habitudes didactiques et pédagogiques construites à ce stade auront des implications sur l'ensemble de ses apprentissages futurs. Ainsi, notre positionnement didactique est en faveur d'un accompagnement des apprenants dans la tâche de compréhension orale, surtout avec les apprenants débutants. Nous considérons en effet qu'à travers des exercices de repérage d'indices extralinguistiques l'apprenant peut développer des comportements stratégiques de compréhension, et ce au-delà des cours de Français Langue Étrangère. En outre, ce type de repérage est susceptible de correspondre à des démarches d'ores et déjà en place en langue maternelle et/ou dans d'autres langues parlées (et comprises). Dit autrement, nous avons pu constater, à travers les résultats du Classe-Images, que les apprenants débutants ont recours aux éléments non-linguistiques pour comprendre le dialogue et que cette démarche leur a permis d'enregistrer de meilleures performances. Dès lors nous pouvons considérer que cette procédure s'est mise en place implicitement (aucune indication dans ce sens n'était donnée aux participants) et renvoie, de ce fait, à un savoir-faire d'ores et déjà acquis (pour une majorité), peut-être même automatisé. Il s'agit donc pour l'enseignant de s'appuyer sur les connaissances antérieures de l'apprenant, d'amener ce dernier à conscientiser ces comportements réflexifs afin de pouvoir les modifier ou les renforcer. Si la prise en compte des éléments extralinguistiques correspond à une activité guidée par l'enseignant, alors l'apprenant automatisera progressivement cette démarche et pourra par la suite la mettre en place plus facilement dans différents contextes, d'apprentissage scolaire ou autre. Cette procédure renvoie à la mise en place des schémas d'action, c'est-à-dire au fait que plus l'individu est coutumier d'une série d'actions plus celle-ci se met en place rapidement, à partir de peu d'indices (inputs), et avec un seuil d'activation bas. On peut donc considérer que, dans ce cadre, l'enseignant qui prend appui sur les savoir-faire de l'apprenant, l'amène à renforcer un comportement *a priori* déjà en place (mais pas nécessairement conscientisé) afin de le rendre plus systématique et efficace. Il s'agirait alors d'un enseignement stratégique visant un apprentissage stratégique.

Enfin, dans l'étude que nous avons réalisée à partir du Classe-Images centrée sur l'analyse de l'impact de la présence (ou de l'absence) d'un contexte sonore environnemental, nous avons travaillé à partir de supports différents (présentation duelle de l'information) mais néanmoins uniquement sonore (donc monomodale). Dans ce cadre, le rôle des dessins dans l'élaboration du sens n'a pas pu être analysé,

puisque ce facteur n'a pas été isolé. Or la théorie de la charge cognitive nous laisse penser que la mobilisation de diverses modalités sensorielles pourrait donner lieu à une diminution de celle-ci (Chanquoy *et al.*, 2007). Dès lors, il pourrait être intéressant de poursuivre l'exploitation du Classe-Images dans le cadre de recherche sur le rôle de la représentation visuelle dans la compréhension.



# Bibliographie

- Abe, D., Carton, F., Cembalo, M., & Regent, O. (1979). Didactique et authentique: du document à la pédagogie. *Mélanges CRAPEL*. Retrieved from <http://www.atilf.fr/spip.php?article3580>
- Abry, D., Daâs, Y., Fert, C., Deschamps, H., Richaud, F., & Sperandio, C. (2007). *Ici 2*. Paris: Clé International.
- Adami, H. (2009). Les documents authentiques dans la formation des adultes migrants: pratiques pédagogiques et contraintes institutionnelles. *Mélanges CRAPEL*, 31, 159–172. Retrieved from <http://www.atilf.fr/spip.php?article3732>
- Al-Saleh, C. (2011). *Note sur le sensoriel, le sensationnel et le perceptuel*. HAL SHS. Retrieved from [halshs.archives-ouvertes.fr/docs/00/55/32/16/PDF/SSP.pdf](http://halshs.archives-ouvertes.fr/docs/00/55/32/16/PDF/SSP.pdf)
- Alain, C., & Woods, D. L. (1997). Attention modulates auditory pattern memory as indexed by event-related brain potentials. *Psychophysiology*, 34(5), 534–46. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.1997.tb01740.x/full>
- Anderson, J. R. (1983). *The architecture of cognition* (1996th ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Anderson, J. R., & Pirolli, P. L. (1984). Spread of activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4), 791–798. doi:10.1037/0278-7393.10.4.791
- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22(11), 2555–69. doi:10.1162/jocn.2009.21398

- Arditty, J., & Vasseur, M.-T. (1999). Interaction et langue étrangère : Présentation. *Langages*, 33(134), 3–19. doi:10.3406/lgge.1999.2189
- Auble, P. M., & Franks, J. J. (1978). The effects of effort toward comprehension on recall. *Memory & Cognition*, 6(1), 20–25. doi:10.3758/BF03197424
- Auroux, S. (2000). Le langage n'est pas dans le cerveaux. Entretien avec Sylvain Auroux recueilli par J.F. Dortier et M. Fagot. *Le Langage : Origine, Nature et Diversité. Sciences Humaines*, 20–22.
- Austin, J. L. (1970). *Quand Dire c'est faire*. Paris: Seuil.
- Avram, C. (2006). Voix et images de France. *Dialogos*, 14, 57–59.
- Baglieto, D., Girardeau, B., & Mistichelli, M. (2011). *Agenda 1*. Paris: Hachette FLE.
- Bal, M. (1977). Narration et focalisation. Pour une théorie des instances du récit. *Poétique : Revue de Théâtre et D'analyse Littéraire*, 29, 107–127. Retrieved from <http://dare.uva.nl/document/15290>
- Barabas, R. (1994). *La perception. Essai sur le sensible* (2009th ed.). Paris: Librairie philosophique J. Vrin.
- Barsalou, L. W. (2008). Cognitive and Neural Contributions to Understanding the Conceptual System. *Current Directions in Psychological Science*, 17(2), 91–95. doi:10.1111/j.1467-8721.2008.00555.x
- Barsalou, L. W. (2013). Ad hoc categories. *Memory & Cognition*, 11(3), 211–227. doi:10.3758/BF03196968
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13(1), 17–26. doi:10.1016/S0926-6410(01)00084-2
- Berdal-Masuy, F., & Briet, G. (2010). Stratégie pour une écoute efficace. In *Le français de demain: enjeux éducatifs et professionnels* (pp. 16–38). Sofia: CREFECO. Retrieved from [http://crefec.org/fr\\_version/pages/8@Berdal-Masuy et Briet.pdf](http://crefec.org/fr_version/pages/8@Berdal-Masuy%20et%20Briet.pdf)
- Bergeron, R. (2000). Pour un apprentissage réussi de l'oral en classe. *Québec Français*, 118, 30–33. Retrieved from <http://quebec.synergiescanada.org/culture/qf1076656/qf1197351/56055ac.pdf>
- Berkeley, G. (1970). *Trois dialogues entre Hylas et Philonous*. Paris: Aubier.
- Berthet, A., Hugot, C., Kizirian, V. M., Sampsonis, B., & Waendendries, M. (2006). *Alter Ego 1*. Paris: Hachette FLE.

- Besse, H. (1974). Signes iconiques, signes linguistiques. *Langue Française*, 24(1), 27–54. doi:10.3406/lfr.1974.5692
- Billières, M. (2009). *Apprentissage et cognition III; cours de master Apprentissage/didactique du Français Langue Étrangère et Seconde*. Université de Toulouse 2 le Mirail.
- Billières, M., & Spanghero-Gaillard, N. (2005). La didactique cognitive des langues : regards croisés de disciplines pour comprendre le comment. *Revue PARole*, 34/35/36, 101–135.
- Blanc, N., & Brouillet, D. (2003). *Mémoire et compréhension: Lire pour comprendre*. Paris: Armand Colin.
- Blanche-Benveniste, C. (1985). Quelques caractères de l'oralité. *Boletim de Filologia*, 30, 87–95. Retrieved from [http://cvc.instituto-camoes.pt/bdc/lingua/boletimfilologia/30/boletim30\\_pag87\\_95.pdf](http://cvc.instituto-camoes.pt/bdc/lingua/boletimfilologia/30/boletim30_pag87_95.pdf)
- Blanche-Benveniste, C. (1990). Un modèle d'analyse syntaxique "en grilles" pour les productions orales. *Anuario de Psicologia*, 47, 11–28. Retrieved from [http://lattice.prod.lamp.cnrs.fr/IMG/pdf/Blanche-Benveniste\\_1990\\_Modele\\_syntaxique\\_en\\_grilles.pdf](http://lattice.prod.lamp.cnrs.fr/IMG/pdf/Blanche-Benveniste_1990_Modele_syntaxique_en_grilles.pdf)
- Blanche-Benveniste, C. (2010). *Approche de la langue parlée en français*. Paris: Editions Ophrys.
- Blanche-Benveniste, C., Bergounioux, G., Chevalier, J.-C., Dumont, C., Encrevé, P., Perdue, C. M., & Simoni-Arembou, M.-R. (1992). Actualité de l'enquête et des études sur l'oral. *Langue Française*, 93, 94–119. doi:10.3406/lfr.1992.5813
- Bornkessel-Schlesewski, I. D., & Friederici, A. D. (2007). Neuroimaging studies of sentence and discourse comprehension. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 407–424). Oxford: Oxford University Press.
- Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews*, 59(1), 55–73. doi:10.1016/j.brainresrev.2008.05.003
- Botte, M.-C., Drake, C., Brochard, R., McAdams, S., & Argenti, A.-M. (1994). Focalisation attentionnelle appliquée aux flux auditifs. Premières mesures. *Le Journal de Physique IV*, 4(C5), C5–407–C5–410. doi:10.1051/jp4:1994585
- Botteldooren, D., Lavandier, C., Preis, A., Dubois, D., Aspuru, I., Guastavino, C., ... Andringa, T. C. (2011). Understanding urban and natural soundscapes. In *Forum Acusticum 2011*. Aalborg (danmark). Retrieved from <https://biblio.ugent.be/input/download?func=downloadFile&recordId=2041806&fileId=2041808>

- Boulton, A. (2009). Documents authentiques, oral, corpus. *Mélanges CRAPEL*, 31, 5–13. Retrieved from [http://194.214.124.224/IMG/pdf/melanges/02\\_Debaisieux.pdf](http://194.214.124.224/IMG/pdf/melanges/02_Debaisieux.pdf)
- Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 3(2), 193–209. doi:10.1016/0010-0285(72)90003-5
- Bransford, J. D., & McCarrel, N. S. (1974). A sketch of cognitive approach to comprehension. In W. Weimer & D. S. Palermo (Eds.), *Perceiving, acting and knowing: Toward an ecological psychology*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bregman, A. S. (1990). *Auditory scene analysis: the perceptual organization of sound*. Cambridge MA: The MIT press.
- Bregman, A. S. (1994). L'analyse des scènes auditives: l'audition dans les environnements complexes. In S. Mc Adams & E. Bigand (Eds.), *Penser les sons, Psychologie cognitive de l'audition* (Vol. 1–9, pp. 11–39). Paris: Presses Universitaires de France.
- Camus, J.-F. (1996). *La psychologie cognitive de l'attention*. Paris: Masson & Armand Colin.
- Cance, C. (2009). Stimuli, dispositifs expérimentaux et mondes construits. In D. Dubois (Ed.), *Le sentir et le dire*. Paris: L'Harmattan.
- Capelle, G., & Menand. (2009). *Le nouveau Taxi 1*. Paris: Hachette FLE.
- Carlson, L. A. (2010). *Encyclopedia of Perception*. (B. E. Goldstein, Ed.) *Encyclopedia of perception* (pp. 1012–1015). 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc. doi:10.4135/9781412972000
- Carton, F. (1995). L'apprentissage différencié des quatre aptitudes. In M.-J. Gremmo (Ed.), *Didactique du Français Langue Étrangère*. Nancy: Verbum, Presses Universitaires de Nancy.
- Castellengo, M., & Dubois, D. (2005). Timbre ou Timbres? Propriété du signal, de l'instrument, ou construction cognitive. *Les Cahiers de La Société Québécoise de Recherche En Musique*, 9(1-2), 25–38. Retrieved from [oicm.umontreal.ca/cim05](http://oicm.umontreal.ca/cim05)
- Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La Charge Cognitive: Théorie et applications*. Paris: Arman Colin.
- Charaudeau, P. (1992). *Grammaire du sens et de l'expression*. Paris: Hachette éducation.
- Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, 25(5), 975. doi:10.1121/1.1907229

- Cheveigné (de), A. (2000). Analyse de scène auditive computationnelle. In J. Mariani (Ed.), *La parole, des modèles cognitifs aux machines communicantes - Analyse, synthèse et codage de la parole* (pp. 175–196). Paris: Hermès.
- Christophe, A., Pallier, C., Bertoncini, J., & Mehler, J. (1991). A la recherche d'une unité : segmentation et traitement de la parole. *L'année Psychologique*, 91(1), 59–86. doi:10.3406/psy.1991.29445
- Clerc, M. (1999). La compréhension orale en langue voisine. Espagnol pour francophones: analyse d'erreurs et conséquences méthodologiques. *Les Langues Modernes*, 2, 48–58. Retrieved from <http://www.galanet.eu/publication/fichiers/Clerc1999.pdf>
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. L. (1992). A parallel distributed processing approach to automaticity. *The American Journal of Psychology*, 105(2), 239–69. Retrieved from <http://psych.stanford.edu/~jlm/papers/CohenServanSchreiberMcC92.pdf>
- Colle, C. (n.d.). *La courbe du diable, vers une mésologie de l'art, pour l'art et par l'art*. Université Paul Valéry, Montpellier 3.
- Colle-Quesada, M.-M., & Spanghero-Gaillard, N. (2014). Étude de l'impact du contexte sonore environnemental sur la compréhension d'un dialogue verbal : une démarche expérimentale. *Cellule de Recherche En Linguistique*.
- Conseil de l'Europe. (2001). *Cadre Européen Commun de Référence pour les Langues, Apprendre, enseigner, Evaluer*. Paris: Didier.
- Cooke, M., Brown, G. J., Crawford, M., & Green, P. (1993). Computational auditory scene analysis: listening to several things at once. *Endeavour*, 17(4), 186–90. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7510225>
- Cooke, M., & Ellis, D. P. . (2001). The auditory organization of speech and other sources in listeners and computational models. *Speech Communication*, 35(3-4), 141–177. doi:10.1016/S0167-6393(00)00078-9
- Coste, D. (1970). Le renouvellement méthodologique dans l'enseignement du français langue étrangère : remarques sur les années 1955-1970. *Langue Française*, 8(1), 7–23. doi:10.3406/lfr.1970.5525
- Coste, D., Courtillon, J., Ferenczi, V., Martins-Baltar, M., Papo, E., & Roulet, E. (1976). *Pour un niveau Seuil*. Paris: Hatier.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671–684. doi:10.1016/S0022-5371(72)80001-X
- Culicover, P. W., & Jackendoff, R. (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences*, 10(9), 413–8. doi:10.1016/j.tics.2006.07.007

- Cuq, J.-P., & Gruca, I. (2002). *Cours de didactique du Français Langue étrangère et seconde* (2005th ed.). Grenoble: Presses Universitaires de Grenoble.
- Cuq, J.-P., & Gruca, I. (2005). *Cours de didactique du français langue étrangère et seconde*. Grenoble: Presses Universitaires de Grenoble.
- De Sanctis, P., Ritter, W., Molholm, S., Kelly, S. P., & Foxe, J. J. (2008). Auditory scene analysis: the interaction of stimulation rate and frequency separation on pre-attentive grouping. *The European Journal of Neuroscience*, 27(5), 1271–6. doi:10.1111/j.1460-9568.2008.06080.x
- Denham, S. L., & Winkler, I. (2006). The role of predictive models in the formation of auditory streams. *Journal of Physiology, Paris*, 100(1-3), 154–70. doi:10.1016/j.jphysparis.2006.09.012
- Denis, M., & Dubois, D. (1976). La représentation cognitive: quelques modèles récents. *L'année Psychologique*, 76(2), 541–562. doi:10.3406/psy.1976.28161
- Denyer, M., Garmendia, A., & Lions-Olivieri, M.-L. (2009). *Version originale 1*. (C. Puren, Ed.). Paris-Barcelone: Maison des Langues- diffusion.
- Deschamps, R., & Moulignier, A. (2005). La mémoire et ses troubles. *EMC - Neurologie*, 2(4), 505–525. doi:10.1016/j.emcn.2005.07.003
- Detey, S. (2007). Transcription, translittération et didactique de l'oral en FLE au Japon: Katakana, rōmaji et orthographe française. *Revue Japonaise de Didactique Du Français*, 2(1), 19–36. Retrieved from <http://sjdf.org/publication/documents/Revue/vol2/1-19sylvain.pdf>
- Di Giura, M., & Beacco, J.-C. (2007). *Alors? niveau A2*. St Etienne: Didier.
- Dolz, J., & Schneuwly, B. (1998). *Pour un enseignement de l'oral. Initiation aux genres formels à l'école*. Paris: ESF.
- Dubois, D. (1993). Catégorisation et cognition: "10 ans après", une évaluation des concepts de Rosch. In D. Dubois (Ed.), *Sémantique et cognition - Catégories, prototypes, typicalité* (pp. 31–54). Paris: CNRS. doi:10.1234/12345678
- Dubois, D. (2000). Categories as Acts of Meaning: The Case of Categories in Olfaction and audition. *Cognitive Science Quarterly*, 1, pp. 35–68. Retrieved from <http://cognition.iig.uni-freiburg.de/csq/pdf-files/dubois.pdf>
- Dubois, D. (2003). Perception, representation and Knowledge: acoustic phenomena between noise and sounds. In *TecniAcustica*. Bilbao.
- Dubois, D., Guastavino, C., & Raimbault, M. (2006). A cognitive Approach to Urban Soundscapes: Using Verbal Data to Access Everyday Life Auditory Categories. *Acta Acustica United with Acustica*, 92, 865–874. Retrieved from [http://www.torontosoundmap.com/papers/Dubois\\_Guastavino\\_Raimbault\\_2006.pdf](http://www.torontosoundmap.com/papers/Dubois_Guastavino_Raimbault_2006.pdf)

- Durand, J. (2000). Oral, écrit et faculté de langage. In M.-N. Guilot & M.-M. Kenning (Eds.), *Changing Landscapes in Language and Language Pedagogy* (pp. 40–72). Londres: AFLS/CILT.
- Dux, P. E., & Marois, R. (2009). The attentional blink: a review of data and theory. *Attention, Perception & Psychophysics*, 71(8), 1683–700. doi:10.3758/APP.71.8.1683
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102(2), 211–245. doi:10.1037/0033-295X.102.2.211
- Faerch, C., & Kasper, G. (1980). Processes and strategies in foreign language learning and communication. *Interlanguage Studies Bulletin*, 5(1), 47–118. Retrieved from <http://files.eric.ed.gov/fulltext/ED213247.pdf>
- Fasel Lauzon, V., Pekarek Doehler, S., & Pochon-Berger, E. (2009). Identification et observabilité de la compétence d'interaction: le désaccord comme microcosme actionnel. In A. Näf, V. Fasel-Lauzon, & E. Pochon-Berger (Eds.), *bulletin vals asla* (pp. 121–241). Neuchâtel: Bulletin suisse de linguistique appliquée. Retrieved from <http://www2.unine.ch/repository/default/content/sites/codi/files/shared/documents/VALS2009.pdf>
- Fasel Lauzon, V., Pekarek Doehler, S., Pochon-Berger, E., & Steinbech Kohler, F. (2009). L'oral? L'oral! Mais comment? *Babylonia*, 2, 41–45. Retrieved from [http://www2.unine.ch/repository/default/content/sites/codi/files/shared/documents/Oral\\_Babylonia.pdf](http://www2.unine.ch/repository/default/content/sites/codi/files/shared/documents/Oral_Babylonia.pdf)
- Flumian, C., Labascoule, J., Lause, C., & Royer, C. (2011). *Nouveau Rond-point pas à pas A2*. (C. Puren, Ed.). Paris: Maison des Langues.
- Fodor, J. (1983). *The modularity of mind*. Cambridge MA: MIT Press.
- Fontan, L. (2012). *De la mesure de l'intelligibilité à l'évaluation de la compréhension de la parole pathologique en situation de communication*. Université de Toulouse.
- Forst, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123(1), 71–99. Retrieved from <http://www.haskins.yale.edu/Reprints/HL1068.pdf>
- Forst, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 104–115. Retrieved from [http://www.haskins.yale.edu/sr/sr086/sr086\\_13.pdf](http://www.haskins.yale.edu/sr/sr086/sr086_13.pdf)
- Frauenfelder, U. H., & Nguyen, N. (2000). La reconnaissance des mots parlés. In J. . Rondal & X. Seron (Eds.), *Troubles du Langage: Bases Théoriques, Diagnostic et Rééducation* (pp. 213–240). Bruxelles: Mardaga.

- Frenzel, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: a new perspective on the N400-P600 distinction. *Neuropsychologia*, 49(3), 574–9. doi:10.1016/j.neuropsychologia.2010.12.003
- Gaillard, P. (2000). *Etude de la perception des transitoires d'attaque des sons de steeldrums: particularités acoustiques, transformation par synthèse et catégorisation*. Université Toulouse le Mirail, Toulouse.
- Gaillard, P., Magnen, C., & Billières, M. (2006). Expertise phonétique: une réflexion vers la notion de "surdité phonologique." In Sensolier (Ed.), *Les expertises sensorielles: nature et acquisition* (pp. 11–14). Paris: Le Sensolier.
- Galisson, R. (1990). De la linguistique appliquée à la didactologie des langues-cultures - Vingt ans de réflexions disciplinaires. *Études de Linguistique Appliquée*, 95(79), 119–159.
- Galisson, R. (1994). Formation à la recherche en didactologie des langues-cultures. *Études de Linguistique Appliquée*, 95(JULSEP), 119–159. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=4290397>
- Gallien, C. (2001). Documents d'écoute et débutants: compte-rendu d'une étude empirique. *Mélanges CRAPEL*, 26. Retrieved from [http://atilf.fr/IMG/pdf/melanges/26\\_gallien.pdf](http://atilf.fr/IMG/pdf/melanges/26_gallien.pdf)
- Gaonac'h, D., & Larigauderie, P. (2000). *Mémoire et fonctionnement cognitif: la mémoire de travail*. Paris: Armand Colin (collection U).
- Garcia-Deban, C. (1999). Évaluer l'oral. *Pratiques*, 103/104, 193–212.
- Germain, C. (1993). *Evolution de l'enseignement des langues: 5000 ans d'histoire*. Paris: Clé Internationale.
- Germain, C., & Netten, J. (2005). Place et rôle de l'oral dans l'enseignement / apprentissage d'une L2. *Babylonia*, 2, 7–10. Retrieved from <https://www.caslt.org/pdf/16-Netten-Germain-Babylonia-2005.pdf>
- Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 16(3), 430–45. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/2140402>
- Gineste, M.-D., & Le Ny, J.-F. (2002). *Psychologie cognitive du langage: de la reconnaissance à la compréhension*. Paris: Dunod.
- Goetz, E. T., Anderson, R. C., & Schallert, D. L. (1981). The representation of sentences in memory. *Journal of Verbal Learning and Verbal Behavior*, 20, 369–385. Retrieved from [https://www.ideals.illinois.edu/bitstream/handle/2142/18025/ctrstreadtechrpev01979i00144\\_opt.pdf?sequence=1](https://www.ideals.illinois.edu/bitstream/handle/2142/18025/ctrstreadtechrpev01979i00144_opt.pdf?sequence=1)

- Gougenheim, G., Michéa, R., Rivenc, P., & Sauvageot, A. (1956). *L'élaboration du français fondamental: étude sur l'établissement d'un vocabulaire et d'une grammaire de base*. Paris: Didier.
- Gouvernement du Québec Ministère de l'Éducation. (2006). *Programme de formation de l'école québécoise*. Québec: Bibliothèque nationale du Québec.
- Grand-Clément, O. (2010). *La grammaire en dialogues - niveau grand débutant*. (M. Grandmangin-Vainseine & C. Ligonie, Eds.). Paris: Clé International.
- Grandaty, M. (2004). L'oral, objet d'enseignement? Gestion et régulation des tâches langagières en classe de maternelle. In *9ème colloque de l'AIRDF*.
- Grataloup, C., Hoen, M., Veuillet, E., Collet, L., Pellegrino, F., & Meunier, F. (2009). Speech restoration: an interactive process. *Journal of Speech, Language, and Hearing Research: JSLHR*, 52(4), 827–38. doi:10.1044/1092-4388(2008/06-0235)
- Gremmo, M.-J. (1987). à propos d'écoute... écoute. *Mélanges CRAPEL*, 55–67. Retrieved from <http://www.atilf.fr/IMG/pdf/melanges/5gremmo-2.pdf>
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66(3), 377–388.
- Griffiths, T. D., & Warren, J. D. (2004). What is an auditory object? *Nature Reviews Neuroscience*, 5(11), 887–92. doi:10.1038/nrn1538
- Guastavino, C. (2003). *Étude sémantique et acoustique de la perception des basses fréquences dans l'environnement sonore urbain*. Université Paris 6.
- Gueunier, N. (2005). Y a-t-il une linguistique de l'oral? *Le Français Aujourd'hui*, 148(1), 15. doi:10.3917/lfa.148.0015
- Guillaume, P. (1979). *La psychologie de la forme*. Paris: Flammarion.
- Guyot, F., Castellengo, M., Fabre, B. (1997). Etude de la catégorisation d'un corpus de bruits domestiques. In D. Dubois (Ed.), *Catégorisation et Cognition: de la perception au discours* (pp. pp. 41–58). Paris: Kimé.
- Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 362(1481), 801–11. doi:10.1098/rstb.2007.2089
- Halté, J. F. (2005). Intégrer l'oral : Pour une didactique de l'activité langagière. In J. F. Halté & M. Rispaïl (Eds.), *L'oral dans la classe. Compétences, enseignement, activités* (pp. 11–31). Paris: L'Harmattan.
- Holec, H. (1990). Des documents authentiques, pour quoi faire? *Mélanges Pédagogiques*. Retrieved from <http://www.atilf.fr/IMG/pdf/melanges/5holec-2.pdf>

- Houix, M. O. (2003). *Catégorisation auditive des sources sonores*. Université de Maine.
- Husserl, E. (1950). *Idées directrices pour une phénoménologie*. Paris: Gallimard.
- Ivan, M., & Avram, C. (2006). De vive voix. *Dialogos*, 14, 60–62.
- Jaworski, V. (2012). Le bruit et le droit. In A. Pecqueux (Ed.), *Les bruits de la ville* (pp. 83–94). Paris: Seuil.
- Jisa, H. (2003). L'acquisition du langage. *Terrain*, 40(40), 115–132. doi:10.4000/terrain.1562
- Kamber, A., & Skupien, C. (2009). Les documents radiophoniques dans l'enseignement de la compréhension orale. *Mélanges CRAPEL*, 31, 173–189. Retrieved from [http://web.atilf.fr/IMG/pdf/melanges/10\\_Kamber\\_Skupien.pdf](http://web.atilf.fr/IMG/pdf/melanges/10_Kamber_Skupien.pdf)
- Kekenbosch, C., & Denhière, G. (1988). L'activation et la diffusion de l'activation. *L'année Psychologique*, 88(2), 237–256. doi:10.3406/psy.1988.29268
- Kerbrat-Orecchioni, C., & Traverso, V. (2004). Types d'interactions et genres de l'oral. *Langages*, 38(153), 41–51. doi:10.3406/lgge.2004.931
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. doi:10.1037/0033-295X.95.2.163
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. New York: Cambridge University press.
- Kintsch, W., & Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5(3), 257–274. doi:10.1016/0010-0285(73)90036-4
- Kubovy, M., & Schutz, M. (2010). Audio-Visual Objects. *Review of Philosophy and Psychology*, 1(1), 41–61. doi:10.1007/s13164-009-0004-5
- Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: challenges to syntax. *Brain Research*, 1146, 23–49. doi:10.1016/j.brainres.2006.12.063
- Labascoule, J., Lause, C., & Royer, C. (2009). *Rond-Point 1*. Barcelone: Difusion FLE.
- Labascoule, J., Lause, C., & Royer, C. (2010). *Nouveau Rond-point pas à pas A2*. (C. Puren, Ed.). Paris: Maison des Langues.
- Laberge, D. (1995). *Attentional processing*. Cambridge: Harvard University Press.
- Lafontaine, L. (2000). L'enseignement de l'oral en classe de français : une passion à vivre et à découvrir. *Québec Français*, 118(4), 42–44. Retrieved from <http://www.lizannelafontaine.com/PDF/t4.pdf>

- Lafontaine, L. (2005). La place de la didactique de l'oral en formation initiale des enseignants de français langue d'enseignement au secondaire. *Nouveaux Cahiers de La Recherche En Éducation*, 8(1), 95. doi:10.7202/1018160ar
- Lafontaine, L., & Messier, G. (2009). Représentations de l'enseignement et de l'évaluation de l'oral chez des enseignants et des élèves du secondaire en français langue d'enseignement. *Revue Du Nouvel-Ontario*, 34, 119-144. Retrieved from <http://www.lizannelafontaine.com/fr/QF-2011-lafontaine.PDF>
- Lafontaine, L., & Pharand, J. (2011). Les stratégies d'écoute et leurs manifestations non verbales chez des élèves de 5<sup>e</sup> et 6<sup>e</sup> année. *AQEP Vivre Le Primaire*, 24, 45-48.
- Le Ny, J.-F. (2005). *Comment l'esprit construit du sens*. Paris: Odile Jacob.
- Leroy, C. (1985). La notation de l'oral. *Langue Française*, 65(1), 6-16. doi:10.3406/lfr.1985.6401
- Locke, J. (1972). *Essai philosophique concernant l'entendement humain*. Paris: Vrin.
- Lotze, N., Tune, S., Schlesewsky, M., & Bornkessel-Schlesewsky, I. (2011). Meaningful physical changes mediate lexical-semantic integration: top-down and form-based bottom-up information sources interact in the N400. *Neuropsychologia*, 49(13), 3573-82. doi:10.1016/j.neuropsychologia.2011.09.009
- Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1999). The role of situational continuity in narrative understanding. In H. Oostendorp Goldman, S.R. (Ed.), *The Construction of mental representations during reading* (pp. 196-218). Mahwah, N.J.: Lawrence Erlbaum Associates.
- Magnen, C. (2009). Approche dynamique de la perception de la parole: catégorisation de la substance et de la variabilité phonétique par des francophones en langue maternelle et des hispanophones en langue étrangère. Toulouse: Université Toulouse II le Mirail.
- Mandler, G. (2002). Organization: What levels of processing are levels of. In M. A. Conway (Ed.), *Levels of processing, 30 years on: Memory, special issue*. Hove: Psychology Press.
- Marslen-Wilson, W. D. (1989). Access and integration: Projecting sound onto meaning. In W. D. Marslen-Wilson (Ed.), *Lexical representation and processing* (pp. 2-24). Cambridge: MIT Press.
- Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, 101(4), 653-75. Retrieved from <http://courses.umass.edu/linguist716-kmyu/static/pages/readings/marslen-wilsonwarren1994.pdf>

- Matthen, M. (2009). On the Diversity of Auditory Objects. *Review of Philosophy and Psychology*, 1(1), 63–89. doi:10.1007/s13164-009-0018-z
- Mayer, N. D., & Tormala, Z. L. (2010). “Think” versus “feel” framing effects in persuasion. *Personality & Social Psychology Bulletin*, 36(4), 443–54. doi:10.1177/0146167210362981
- McAdams, S. (1994). La reconnaissance de sources et d'événements sonores. In S. McAdams & E. Bigand (Eds.), *Penser les sons, Psychologie cognitive de l'audition*. Paris: Presses Universitaires de France.
- Menand, R. (2006). *Taxi*. Paris: Hachette FLE.
- Merleau-Ponty, M. (2002). *Causeries, 1948*. (S. Ménasé, Ed.). Paris: Seuil.
- Mervis, C. B., & Rosch, E. (1981). Categorization of Natural Objects. *Annual Review of Psychology*, 32(1), 89–115. doi:10.1146/annurev.ps.32.020181.000513
- Metz, C. (1970). Images et pédagogie. *Communications*, 15, 162–168.
- Miller, G. A. (1962). Some psychological studies of grammar. *American Psychologist*, 17(11), 748–762. doi:10.1037/h0044708
- Miller, G. A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley.
- Ministère de l'éducation nationale, F. (1970). *Le français fondamental premier degré*. Paris: Institut national de recherche et de documentation pédagogiques.
- Ministère de l'éducation nationale, F. (1972). *Le français fondamental second degré*. Paris: Institut national de recherche et de documentation pédagogiques.
- Miquel, C. (2005). *La grammaire en dialogues - niveau débutant*. (M.-C. Couët-Lannes, Ed.) (p. 128). Paris: Clé International.
- Miquel, C. (2007). *La grammaire en dialogue - niveau intermédiaire*. Paris: Clé International.
- Niessen, M., Cance, C., & Dubois, D. (2010). Categories for soundscape: toward a hybrid classification. *Acta Acustica*, 92(6), 1–14. Retrieved from [http://www.ai.rug.nl/senscog/documents/Niessen\\_2010.pdf](http://www.ai.rug.nl/senscog/documents/Niessen_2010.pdf)
- Niessen, M. E. (2009). Understanding a soundscape through its components. *Perception*. Retrieved from <http://www.ai.rug.nl/~tjeerd/publications/niessen091.pdf>
- Niessen, M., Maanen, L. Van, & Andringa, T. C. (2008). Disambiguating Sounds through Context. *Artificial Intelligence*.

- Nonnon, E. (1999). Note de synthèse. *Revue Française de Pédagogie*, 129(1), 87–131. doi:10.3406/rfp.1999.1067
- Nonnon, É. (2004). Écouter peut-il être un objectif d'apprentissage? *Le Français Aujourd'hui*, 146(3), 75. doi:10.3917/lfa.146.0075
- Norris, D., Cutler, A., McQueen, J. M., & Butterfield, S. (2006). Phonological and conceptual activation in speech comprehension. *Cognitive Psychology*, 53(2), 146–93. doi:10.1016/j.cogpsych.2006.03.001
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(5), 1209–1228. doi:10.1037/0278-7393.21.5.1209
- Nudds, M. (2010). What Are Auditory Objects? *Review of Philosophy and Psychology*, 1(1), 105–122. Retrieved from <https://www.era.lib.ed.ac.uk/bitstream/1842/1777/1/waoa.pdf>
- Onursal, I. (2008). Le dialogue dans les manuels d'enseignement de français langue étrangère et la notion d'authenticité. *Synergies Turquie*, 1, pp. 65–75. Retrieved from <http://gerflint.fr/Base/Turquie1/onursal.pdf>
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educational Psychologist*, 38(1), 1–4. Retrieved from <http://cis.msje.edu/evoc/637/References/Pass-CognitiveLoadTheoryAndID.pdf>
- Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, 32, 1–8. Retrieved from [http://www.ucs.mun.ca/~bmann/0\\_ARTICLES/CogLoad\\_Paas04.pdf](http://www.ucs.mun.ca/~bmann/0_ARTICLES/CogLoad_Paas04.pdf)
- Parpette, C. (1997). Le discours oral: des représentations à la réalité. In *Les linguistiques appliquées et les sciences du langage, 14 -15 novembre 1997*. Université de Strasbourg. Retrieved from <http://lesla.univ-lyon2.fr/sites/lesla/IMG/pdf/doc-192.pdf>
- Parpette, C. (2008). De la compréhension orale en classe à la réception orale en situation naturelle: une relation à interroger. *Les Cahiers de l'Acedle*, 5(1). Retrieved from [http://acedle.org/IMG/pdf/Parpette\\_Cah5-1.pdf](http://acedle.org/IMG/pdf/Parpette_Cah5-1.pdf)
- Parpette, C., & Peutot, F. (2006). Les enregistrements de classes ordinaires comme support d'enseignement du français langue de scolarisation. *Les Cahiers de l'Acedle*, 2. Retrieved from [http://acedle.org/IMG/pdf/Parpette-C\\_cah2.pdf](http://acedle.org/IMG/pdf/Parpette-C_cah2.pdf)
- Pavio, A. (1971). *Imagery and verbal process*. New York: Rinehart & Winston.
- Pavio, A. (1986). *Mental representation: a dual coding approach*. New York: Oxford University Press.

- Pecqueux, A. (2012). Le son des choses; les bruits de la ville. In A. Pecqueux (Ed.), *Les bruits de la ville* (pp. 5–15). Paris: Seuil.
- Perruchet, P. (1988). *Les automatismes cognitifs*. Bruxelles: Mardaga.
- Perruchet, P., & Nicolas, S. (1998). L'apprentissage implicite: un débat théorique. *Psychologie Française*, 43(1), 13–25. Retrieved from <http://leadserv.u-bourgogne.fr/IMG/pdf/pdf2364Ko.pdf>
- Picoche, J. (1999). Dialogue autour de l'enseignement du vocabulaire. *Études de Linguistique Appliquée*, 116, 421–433. Retrieved from <http://jpicochelinguistique.free.fr/ENSEIGNEMENT/articlesdefond.html>
- Poisson-Quinton, S., Siréjols, É., & Bruley, C. (2011). *Intro*. Paris: Clé International.
- Porcher, L. (1995). *Le français langue étrangère*. Paris: Hachette éducation.
- Pottier, B. (1999). L'étude du lexique. *Bulletin Hispanique*, 101(2), 613–620. doi:10.3406/hispa.1999.5024
- Pottier, B. (2003). À propos du Dictionnaire d'analyse du discours. *Mots. Les Langages Du Politique*, 71. Retrieved from <http://mots.revues.org/8613>
- Puren, C. (1988). *Histoire des méthodologies de l'enseignement des langues*. Paris: Clé International.
- Puren, C. (1994). *La didactique des langues étrangères à la croisée des méthodes. Essai sur l'éclectisme*. (F. Lapeyre, Ed.) *Les langues modernes* (p. 206). Paris: Didier.
- Puren, C. (2006a). De l'approche communicative à la perspective actionnelle. *Le Français Dans Le Monde*, 347, 37–40. Retrieved from [http://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.christianpuren.com%2Fapp%2Fdownload%2F5523333851%2FPUREN\\_2006g\\_Configurations\\_didactiques\\_revue%252BF LM\\_n347.pdf&ei=4Q0oVNvZMdbYatqOgdAH&usg=AFQjCNHolR8egZxk77Au9SI hIOChLMRaMA&bvm=bv.76247554,d.d2s](http://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCEQFjAA&url=http%3A%2F%2Fwww.christianpuren.com%2Fapp%2Fdownload%2F5523333851%2FPUREN_2006g_Configurations_didactiques_revue%252BF LM_n347.pdf&ei=4Q0oVNvZMdbYatqOgdAH&usg=AFQjCNHolR8egZxk77Au9SI hIOChLMRaMA&bvm=bv.76247554,d.d2s)
- Puren, C. (2006b). Le cadre européen de référence et la réflexion méthodologique en didactique des langues-cultures: un chantier à reprendre. *APLV*, 16–22. Retrieved from <http://www.aplv-languesmodernes.org//spip.php?article35>
- Py, B. (2000). La construction interactive de la norme comme pratique et comme représentation. *Aile*, 12. Retrieved from <http://aile.revues.org/1464>
- Resche-Rigon. (2009). Dire ou vouloir dire. Du langage idéal au bricolage des langues. In D. Dubois (Ed.), *Le sentir et le dire* (pp. 47–74). Paris: L'Harmattan.
- Rey, A. (Ed.). (2008). *Le Grand Robert de la langue française, dictionnaire alphabétique et analogique de la langue française de Paul Robert*. Paris. Retrieved from file:///Applications/Le Grand Robert.app/aide/Aide.htm

- Rey, N. (2004). L'évolution de l'enseignement du français oral ou comment est-il devenu objet d'enseignement dans les plans d'études genevois. In *9ème colloque de l'AIRDF*.
- Rivenc, P., & Guberina, P. (1962). *Voix et Images de France (1er degré). Cours et méthode audiovisuelle d'enseignement du français aux étrangers*. Paris: CREDIF, Didier.
- Robert, J.-M. (2004). Proximité linguistique et pédagogie des langues non maternelles. *Ela. Etudes de Linguistique Appliquée*, 136(4), 499–511. Retrieved from <http://www.cairn.info/revue-ela-2004-4-page-499.htm>
- Rondal, J. A., Esperet, E., Gombert, J. E., Thibaut, J. P., & Comblain, A. (1999). Le développement du langage oral. In J. A. Rondal & E. Esperet (Eds.), *Manuel de psychologie de l'enfant* (pp. 479–564). Liège: Mardaga.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), 192–233. doi:10.1037/0096-3445.104.3.192
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. doi:10.1016/0010-0285(75)90024-9
- Rouet, J.-F., Coutelet, B., & Dinet, J. (2004). La recherche d'informations dans les documents complexes: processus cognitifs, apprentissage et développement. In *JETSIC*.
- Roulet, E. (1991). La pédagogie de l'oral en question(s). In M. Wirthner, D. Martin, & P. Perrenoud (Eds.), *Parole étouffée, parole libérée, fondements et limites d'une pédagogie de l'oral* (pp. 41–54). Neuchâtel: Delachaux et Niestlé.
- Savin, H., & Bever, T. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Verbal Behavior*, 9, 295–302. Retrieved from [http://dingo.sbs.arizona.edu/~tgb/pdfs/beverpdf\\_43.pdf](http://dingo.sbs.arizona.edu/~tgb/pdfs/beverpdf_43.pdf)
- Schaeffer, P. (1977). *Traité des objets musicaux* (Nouvelle é.). Paris: Seuil.
- Schafer, R. M. (1977). *Le paysage sonore, Le monde comme musique* (2010th ed.). Clamecy: Wildproject.
- Schaff, A. (1969). *Langage et connaissance*. Paris: Seuil, Anthropos.
- Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language*, 25(3), 279–294. doi:10.1016/0749-596X(86)90002-1

- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Böhner, M. (2010). Is It Really Robust? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6(4), 147–151. doi:10.1027/1614-2241/a000016
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84(1), 1–66. doi:10.1037/0033-295X.84.1.1
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, 34(3), 114–23. doi:10.1016/j.tins.2010.11.002
- Shamma, S. A., & Micheyl, C. (2010). Behind the scenes of auditory perception. *Current Opinion in Neurobiology*, 20(3), 361–6. doi:10.1016/j.conb.2010.03.009
- Shepard, R. N. (1981). Psychophysical complementary. In M. Kubovy & J. Pomerantz (Eds.), *Perceptual organization* (pp. 279–341). Hillsdale, NJ: Lawrence Erlbaum.
- Sheppard, D. M., Duncan, J., Shapiro, K. L., & Hillstrom, A. P. (2002). Objects and Events in the Attentional Blink. *Psychological Science*, 13(5), 410–415. doi:10.1111/1467-9280.00473
- Sieroff, E. (1992). Introduction à l'attention sélective: définitions et propriétés. *Revue de Neuropsychologie*, 2(1), pp. 3–27.
- Sivonen, P., Maess, B., & Friederici, A. D. (2006). Semantic retrieval of spoken words with an obliterated initial phoneme in a sentence context. *Neuroscience Letters*, 408(3), 220–5. doi:10.1016/j.neulet.2006.09.001
- Spanghero-Gaillard, N. (2008). *Comment l'être humain comprend? Quelques éléments de réflexion à partir de l'apprentissage d'une langue étrangère, dossier de synthèse*. Université Toulouse II le Mirail, Toulouse.
- Spanghero-Gaillard, N., & Billières, M. (2007). Actes du Deuxième colloque International de Didactique Cognitive des langues. In *Deuxième colloque International de Didactique Cognitive des langues*. Toulouse.
- Spanghero-Gaillard, N., & Gaillard, P. (2007). Du rôle des bruits environnementaux dans la compréhension des documents authentiques: est-ce que comprendre un dialogue c'est seulement comprendre la langue? *Mélanges CRAPEL*, n°31, pp. 191–204. Retrieved from <http://www.atilf.fr/spip.php?article3734>
- Spanghero-Gaillard, N., & Garnier, E. (2013). Le programme Langues en actes: apprentissages croisés (théâtre et langue espagnole). *Les Langues Modernes*, 4, 115–122.
- Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origg, G., & Wilson, D. (2010). Epistemic Vigilance. *Mind & Language*, 25(4), 359–393. doi:10.1111/j.1468-0017.2010.01394.x

- Sussman, E. S., Horv ath, J., Winkler, I., & Orr, M. (2007). The role of attention in the formation of auditory streaming. *Perception & Psychophysics*, *69*(1), 136–152.
- Sussman, E., & Steinschneider, M. (2009). Attention effects on auditory scene analysis in children. *Neuropsychologia*, *47*(3), 771–85. doi:10.1016/j.neuropsychologia.2008.12.007
- Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in natural word. In B. C. M. Moore, M. K. Tyler, & W. Marslen-Wilson (Eds.), *The perception of speech: from sound to meaning* (Vol. ?, pp. 1105–1122).
- Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011a). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(4), 1253–62. doi:10.1037/a0021925
- Thompson, S. K., Carlyon, R. P., & Cusack, R. (2011b). An objective measurement of the build-up of auditory streaming and of its modulation by attention. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(4), 1253–62. doi:10.1037/a0021925
- Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255–262. doi:10.1037/h0029997
- Tiberghien, A. (2013).  tudier les pratiques d’enseignement et d’apprentissage de la physique et contribuer   leur d veloppement. In *10 com icar*. Lyon: ENS Lyon.
- Tremblay, S., Vachon, F., & Jones, D. M. (2005). Attentional and perceptual sources of the auditory attentional blink. *Perception & Psychophysics*, *67*(2), 195–208.
- Tricot, A. (1998). Charge Cognitive et apprentissage, Une pr sentation des travaux de John Sweller. *Revue de Psychologie de L’ ducation*, *1*, pp. 37–64. Retrieved from <http://andre.tricot.pagesperso-orange.fr/tricotRPE.pdf>
- Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review*, *69*(4), 344–354. Retrieved from [http://www.alicekim.ca/PsycholRev62\\_62.pdf](http://www.alicekim.ca/PsycholRev62_62.pdf)
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organisation of Memory* (pp. 381–403). New York: Academic Press.
- Tulving, E. (1976). Echoic processes in recall and recognition. In J. Brown (Ed.), *Recall and recognition* (pp. 37–73). New York: Wiley.
- Vallat, C. (2012). *Etude de la strat gie d’ tayage dans les interactions de classe de Franais Langue  trang re, en milieu universitaire chinois*.

- Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–91. doi:10.1162/jocn.2008.20054
- Van Den Broek, P. (1994). Comprehension and Memory of narrative texts. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. pp. 539–588). San Diego: Academic Press.
- Van Den Broek, Young, M., Tzeng, Y., Linderholm, T., P. (1999). The landscape model of reading: interferences and the online construction of a memory representation. In H. Van Oostendorp Goldman, S.R. (Ed.), *The Construction of mental representations during reading* (pp. pp.71–98). Mahwah: N.J. Erlbaum.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Vanderveer, N. J. (1979). *Ecological acoustics: human perception of environmental sounds*. Cornell University.
- Vialettes, L. (2012). *Apprentissage de la lecture: étude longitudinale à partir de tâches de lecture à voix haute d'enfants anglophones scolarisés en école d'immersion française*. Université de Toulouse.
- Woloszyn, P. (2012). Du paysage sonore aux sonotopes. In A. Pecqueux (Ed.), *Les bruits de la ville* (pp. 53–62). Paris: Seuil.
- Zwaan, R. A. (1993). *Aspects of literary comprehension*. Amsterdam: John Benjamins.
- Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The Construction of Situation Models in Narrative Comprehension: An Event-Indexing Model. *Psychological Science*, 6(5), 292–297. doi:10.1111/j.1467-9280.1995.tb00513.x
- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(2), 386–397. doi:10.1037/0278-7393.21.2.386
- Zwaan, R. A., Radvansky, G. A., Hilliard, A. E., & Curiel, J. M. (1998). Constructing Multidimensional Situation Models During Reading. *Scientific Studies of Reading*, 2(3), 199–220. doi:10.1207/s1532799xssr0203\_2

# Webographie

Site de l'ONU présentant l'actualité de l'alphabétisation (dernière consultation le 27 septembre 2014) :

<http://www.un.org/apps/newsFr/storyF.asp?NewsID=17467&Cr=alphab%C3%A9tisation&Cr1=UNESCO#.Uf51woWiE7A>

Représentation schématique de l'oreille humaine (externe, moyenne, interne) (dernière consultation le 27 septembre 2014) :

[http://www.google.fr/imgres?client=firefox-a&hs=Wmf&sa=X&rls=org.mozilla:fr:official&biw=1387&bih=707&tbm=isch&tbnid=yT32qSDP\\_XLjKM:&imgrefurl=http://coccinelledemoiselle.wordpress.com/2010/02/2le-systeme-auditif/&docid=t19urmzfoU6qM&imgurl=http://coccinelledemoiselle.files.wordpress.com/2010/02/anatomie\\_oreille.jpg&w=582&h=410&ei=RDkrUqqxAclI0wXF14HACQ&zoom=1&iact=rc&dur=486&page=1&tbnh=133&tbnw=190&start=0&ndsp=27&ved=1t:429,r:11,s:0,i:115&tx=103&ty=81](http://www.google.fr/imgres?client=firefox-a&hs=Wmf&sa=X&rls=org.mozilla:fr:official&biw=1387&bih=707&tbm=isch&tbnid=yT32qSDP_XLjKM:&imgrefurl=http://coccinelledemoiselle.wordpress.com/2010/02/2le-systeme-auditif/&docid=t19urmzfoU6qM&imgurl=http://coccinelledemoiselle.files.wordpress.com/2010/02/anatomie_oreille.jpg&w=582&h=410&ei=RDkrUqqxAclI0wXF14HACQ&zoom=1&iact=rc&dur=486&page=1&tbnh=133&tbnw=190&start=0&ndsp=27&ved=1t:429,r:11,s:0,i:115&tx=103&ty=81)

Présentation de la convention ICOR (laboratoire ICAR, Lyon 2) (dernière consultation le 27 septembre 2014) :

[http://icar.univ-lyon2.fr/projets/corinte/bandeau\\_droit/convention\\_icor.htm](http://icar.univ-lyon2.fr/projets/corinte/bandeau_droit/convention_icor.htm)

Site du bar le *Wallace* où se sont déroulées des prises de son (dernière consultation le 27 septembre 2014) :

[http://www.lewallace.com/accueil\\_le\\_wallace.vdom](http://www.lewallace.com/accueil_le_wallace.vdom)

Site du bar le *Florida* où se sont déroulées des prises de son : (dernière consultation le 27 septembre 2014)

<http://www.leflorida-capitole.fr/>

Site de la Plateforme d'Études Techniques et de Recherche en Audition (dernière consultation le 27 septembre 2014) :

<http://petra.univ-tlse2.fr/>

Banque de données de sons environnementaux (dernière consultation le 27 septembre 2014) :

<http://www.sound-fishing.net/>

<http://www.universal-soundbank.com/ambiances-sonores.htm>

Site de l'observatoire des inégalités (dernière consultation le 27 septembre 2014) :

<http://www.inegalites.fr/spip.php?article245>

Portail national des professionnels de l'éducation, créé et géré par le ministère de l'éducation nationale (dernière consultation le 27 septembre 2014) :

<http://eduscol.education.fr/cid47915/liste-des-mots-classee-par-ordre-alphabetique.html>

Base de données lexicographique *Lexique* (dernière consultation le 27 septembre 2014) :

<http://www.lexique.org>

Site de l'Alliance Française de Toulouse (dernière consultation le 27 septembre 2014) :

<http://www.alliance-toulouse.org/>

Site du Département d'Enseignement du Français Langue Étrangère de l'Université de Toulouse 2 (dernière consultation le 27 septembre 2014) :

<http://defle.univ-tlse2.fr/>

Site de Cécile Colle qui a réalisé les dessins du Classe-Images et Ralph Nuhn (artistes plasticiens) : (dernière consultation le 27 septembre 2014)

<http://www.imiteme.com/>

# Table des illustrations

## Table des figures

Figure 1: Fréquences de recours aux supports de Compréhension Orale déclarées par les enseignants de FLE .....	40
Figure 2 : Exploitation du contexte sonore environnemental dans les dialogues pédagogiques .....	41
Figure 3: Schéma du modèle <i>bottom-up</i> de la perception auditive selon Houix (2003) et McAdams (1994) .....	56
Figure 4: Schéma du modèle de perception Top-down, selon Gaillard (2000).....	57
Figure 5 : Catégories d'items construites pour le lieu cuisine .....	127
Figure 6 : Catégories d'items construites pour le lieu <i>Terrasse de café</i> .....	130

## Table des tableaux

Tableau 1 : Catégories d'items construites pour le lieu <i>cuisine</i> .....	127
Tableau 2 : Catégories d'items construites pour la terrasse de café.....	130
Tableau 3 : Récapitulatif des signaux sonores pour le test <i>d'identification des lieux</i> ..	139
Tableau 4 : Catégorie de réponses au test <i>d'identification</i> pour la scène sonore la <i>préparation d'un repas</i> (global).....	142
Tableau 5 : Répartition des réponses au test <i>d'identification</i> pour <i>la préparation d'un repas</i> .....	142
Tableau 6: Catégorie de réponses au test <i>d'identification</i> pour la scène sonore <i>la vaisselle</i> (global).....	143
Tableau 7 : Répartition des réponses au test <i>d'identification</i> pour <i>la vaisselle</i> .....	143
Tableau 8 : Catégories de réponses au test <i>d'identification</i> pour la scène sonore <i>la terrasse de café</i> (global).....	145
Tableau 9 : Répartition des réponses au test <i>d'identification</i> pour <i>la terrasse de café</i> .....	145
Tableau 10 : Évaluation du niveau du dialogue <i>la préparation d'un repas</i> .....	159
Tableau 11 : Évaluation du niveau du dialogue <i>la vaisselle</i> .....	160
Tableau 12 : Évaluation du niveau du dialogue à <i>la terrasse d'un café</i> .....	161
Tableau 13: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur la construction de suites de 3 dessins cibles (VD).....	188
Tableau 14: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur la construction de suites de 3 dessins cibles (VD) pour le panel de francophones.....	189
Tableau 15: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur la construction de suites de 3 dessins cibles (VD) pour le panel d'apprenants débutants .....	190

---

Tableau 16: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire-VD) pour l'ensemble des dialogues .....	192
Tableau 17: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue A .....	193
Tableau 18: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue B .....	194
Tableau 19: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) pour le dialogue C.....	195
Tableau 20: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour l'ensemble des dialogues.....	196
Tableau 21: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue A.....	197
Tableau 22 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue B.....	198
Tableau 23 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel de francophones et pour le dialogue C.....	199
Tableau 24: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - DV) chez le panel d'apprenants débutants et pour l'ensemble des dialogues.....	200
Tableau 25 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue A.....	201
Tableau 26 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue B.....	202
Tableau 27 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les dessins sélectionnés (compréhension parcellaire - VD) chez le panel d'apprenants débutants et pour le dialogue C.....	203

Tableau 28: Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les temps de réponses (VD) enregistrés par les panels de francophones et d'apprenants débutants .....	204
Tableau 29 : Récapitulatif des effets de la variable "expertise linguistique" (VI) sur les temps de réponse (VD) enregistrés par les panels de francophones et d'apprenants débutants et associés aux suites de 3 dessins cibles.....	205
Tableau 30: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel de francophones .....	206
Tableau 31: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel de francophones et associés aux suites de 3 dessins cibles.....	207
Tableau 32: Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel d'apprenants débutants .....	207
Tableau 33 : Récapitulatif des effets de la variable "contexte sonore environnemental" (VI) sur les temps de réponse (VD) enregistrés par le panel d'apprenants débutants et associés aux suites de 3 dessins cibles.....	208
Tableau 34: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues A et B - VI) sur la construction de suites de 3 dessins cibles, pour les panels de francophones et d'apprenants débutants .....	238
Tableau 35: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues A et C - VI) sur la construction de suites de 3 dessins cibles, pour les panels francophones et apprenants débutants ..	239
Tableau 36: Récapitulatif des effets des variables "dialogue" et "contexte sonore environnemental" (Comparaison des dialogues B et C - VI) sur la construction de suites de 3 dessins cibles, pour les panels de francophones et d'apprenants débutants .....	240

---

## Table des dessins

Dessins 1 : Dessins cibles pour le dialogue <i>La préparation du repas</i> .....	170
Dessins 2 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue <i>La préparation d'un repas</i> .....	171
Dessins 3 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue <i>La préparation d'un repas</i> .....	172
Dessins 4 : Dessins avec les 2 facteurs de variation (EN-CONT) pour le dialogue <i>La préparation d'un repas</i> .....	173
Dessins 5 : Dessins cibles pour le dialogue <i>La vaisselle</i> .....	174
Dessins 6 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue <i>La vaisselle</i> .....	175
Dessins 7 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue <i>La vaisselle</i> .....	176
Dessins 8 : Dessins avec les 2 facteurs de variation (EN-CONT) pour le dialogue <i>La vaisselle</i> .....	177
Dessins 9 : Dessins cibles pour le dialogue <i>À la terrasse d'un café</i> .....	178
Dessins 10 : Dessins avec une variation de la situation d'énonciation (EN) pour le dialogue <i>À la terrasse d'un café</i> .....	179
Dessins 11 : Dessins avec une variation du contenu propositionnel (CONT) pour le dialogue <i>À la terrasse d'un café</i> .....	180
Dessins 12 : Dessins avec les 2 facteurs de variation (EN-CONT) pour le dialogue <i>À la terrasse d'un café</i> .....	181

## Table des images

Image 1 : Représentation schématique de l'oreille humaine (externe, moyenne, interne) .....	53
Image 2 : Plan du centre ville de Toulouse où apparaissent les 3 lieux de prises de sons pour la <i>terrasse de café</i> .....	132
Image 3 : Prise de sons au <i>Wallace</i> , pour la <i>terrasse de café</i> .....	133
Image 4 : Prise de sons au <i>Florida</i> pour la <i>terrasse de café</i> .....	133
Image 5 : Prise de sons rue <i>St Rome</i> , pour la <i>terrasse de café</i> .....	133
Image 6 : Prise de sons dans un domicile privé, pour la <i>cuisine</i> .....	134
Image 7 : Capture d'écran de la consigne sur le Classe-Images .....	166
Image 8 : Capture d'écran de l'"écoute" sur le Classe-Images.....	166
Image 9 : Capture d'écran de la présentation des dessins pour le dialogue <i>La préparation d'un repas</i> sur le Classe-Images .....	167
Image 10 : Capture d'écran des résultats sur le Classe-Images.....	168

# Liste des abréviations

AB	Attentional Blinding - Aveuglement attentionnel
AC	Approches Communicatives
ASA	Analyse des Scènes Auditives
BEL	Bureau d'Étude et de Liaison
BELC	Bureau d'Enseignement pour la Langue et la Civilisation
CE	Compréhension Écrite
CECRL	Cadre Européen Commun de Références pour les Langues
CO	Compréhension Orale
COD	Complément d'Objet Direct
CONT	dessins avec une variation du CONTenu propositionnel
CRÉDIF	Centre de Recherche et d'Étude pour la Diffusion du Français
CSE	Contexte Sonore Environnemental
DCS	Dialogue + Contexte Sonore environnemental
DDL	Didactique Des Langues
DEFLE	Département d'Étude du Français Langue Étrangère (Université de Toulouse 2 le Mirail)
DLE	Didactique des Langues Étrangères
DLM	Didactique des Langues Maternelles
DS	Dialogue Seul
EN	dessins avec une variation de la situation ENonciation
EN-CONT	dessins avec une variation de la situation d'ENonciation et du CONTenu propositionnel
ESTM	Echoic Short Term Memory
FF	Français Fondamental
FLE	Français Langue Étrangère

FLM	Français Langue Maternelle
IRMf	Imagerie par Résonance Magnétique fonctionnelle
LÉ	Langue(s) Étrangère(s)
LM	Langue(s) Maternelle(s)
MAO	Méthode Audio-Orale
MAV	Méthodologie Audio-Visuelle
MCT	Mémoire à Court Terme
MDT	Mémoire De Travail
MDT CT	Mémoire de Travail à Court Terme
MDT LT	Mémoire de Travail à Long Terme
MEQ	Ministère de l'Éducation au Québec
MLT	Mémoire à Long Terme
PA	Perspective Actionnelle
PE	Production Écrite
PO	Production Orale
QCM	Questionnaire à Choix Multiples
RA	Réponse Attendue
SGAV	méthodologie Structuro-Globale Audio-visuelle
VIF	Voix et Images de France
VSTM	Visual Short-Term Memory

# Présentation des annexes

Dans ce mémoire, nous avons à plusieurs reprises renvoyé le lecteur aux annexes. Celles-ci sont disponibles sur le CD joint. Elles sont organisées sous la forme de différents dossiers de fichiers qui reprennent l'organisation générale du mémoire.

Le premier dossier est intitulé *Étude sur les supports utilisés par des enseignants de FLE*. Il est composé de trois documents : le premier correspond à une présentation, sous forme de tableaux des participants à l'étude que nous avons menée auprès d'enseignants de FLE. Ils nous ont permis de savoir quel type de supports ils utilisent et auprès de quels étudiants (présentée dans le *Contexte didactique* de ce mémoire). Le deuxième document correspond à un exemplaire des questionnaires que nous leurs avons fournis. Enfin, le troisième document est un récapitulatif des réponses des participants quant à l'utilisation des documents authentiques en fonction des niveaux des apprenants.

Le deuxième dossier est intitulé *Étude sur les dialogues pédagogiques de niveau 1 en FLE*. Dans ce dossier apparaissent deux documents : le premier récapitule, sous la forme d'une bibliographie les manuels que nous avons analysés dans le cadre de cette étude (présentée dans le *Contexte didactique* de ce mémoire). Le deuxième document correspond à un tableau où figurent les différents supports de CO que nous avons analysés, avec, ou sans contexte sonore environnemental, en accompagnement des dialogues ainsi que la fonction que ce CSE peut assumer (situationnelle ou de transcodage).

Le troisième dossier est intitulé *Étude sur les attentes auditives*. Ce dossier contient différents supports qui ont tous participé à l'étude à la base de l'élaboration des scènes sonores environnementales. Le premier de ces documents est une présentation des participants à cette étude, et le deuxième correspond aux enregistrements audio des entretiens, dont une retranscription est également disponible. Cette retranscription s'est faite en suivant la norme ICOR (du laboratoire ICAR) que nous avons adaptée. La description de ces adaptations est également disponible dans ce dossier.

Le quatrième dossier est intitulé *Enregistrement des scènes sonores* et comprend les enregistrements (bruts) que nous avons réalisés à la *terrasse d'un café* ainsi que dans une *cuisine*, ainsi que les tableaux que nous avons utilisés pour l'étiquetage des pistes sonores. Dans ce dossier seront également présentées les différentes versions de chaque scène sonore que nous avons réalisées.

Dans le cinquième dossier, intitulé *Validation et sélection des scènes sonores* apparaissent des exemplaires du test d'identification et du protocole de leur sélection. Par ailleurs, la présentation des participants à ces tests est également proposée.

Dans le sixième dossier, intitulé *Dialogues et validation*, regroupe les enregistrements des dialogues (sans contexte sonore). Ceux-ci sont accompagnés d'un exemplaire des questionnaires proposés aux participants pour leur validation. Enfin, les participants au test de validation des dialogues sont également présentés.

Le septième dossier intitulé *Supports et participants du Classe-Images* présente l'ensemble des dessins réalisés pour chaque dialogue, les juges experts qui ont participé à leur élaboration, ainsi que les dialogues présentés seuls et accompagnés d'un contexte sonore environnemental. Un tableau récapitulatif des participants au Classe-Images est également disponible dans ce dossier.

Enfin, le programme du Classe-Images est accessible dans ce CD.

# Table des matières

<b>REMERCIEMENTS</b>	<b>3</b>
<b>SOMMAIRE</b>	<b>7</b>
<b>INTRODUCTION GENERALE</b>	<b>11</b>
<b>CONTEXTE DIDACTIQUE : L'ORAL EN FRANÇAIS LANGUE ETRANGERE</b>	<b>17</b>
1. Qu'est-ce que l'oral	17
1.1. L'oral comme modalité première de la langue	18
1.1.1. L'antériorité de l'oral sur l'écrit (phylogénèse et ontogénèse)	18
1.1.2. L'oral, un objet mal défini	19
1.2. L'oral en classe de langue	22
1.2.1. L'oral, objet d'apprentissage... mais pas toujours d'enseignement	22
1.2.2. L'évaluation de l'oral : une difficile distinction entre compréhension et production	24
2. La place de l'oral dans trois grandes méthodologies d'enseignement/apprentissage des langues étrangères	27
2.1. L'oral dans la méthodologie structuro-globale audio-visuelle	27
2.2. L'oral dans les approches communicatives	29
2.3. L'oral dans la perspective actionnelle	31
3. Les supports de CO en FLE	34
3.1. Documents authentiques ou dialogues fabriqués	35
3.1.1. Documents authentiques ou pédagogiques ?	35
3.1.2. Une étude sur les supports utilisés par des enseignants de FLE	39
3.2. Une étude sur les dialogues pédagogiques de niveau 1 en FLE	42
3.2.1. Les fonctions des sons environnementaux : un parallèle avec les images de la méthodologie SGAV	42
	299

3.2.2. Des dialogues avec et sans contexte sonore	44
---	----

---

**PARTIE 1 : CADRE THEORIQUE** **49**

**PREMIER CHAPITRE : DES ETUDES SUR LA PERCEPTION...** **49**

1. Qu'est-ce que percevoir ?	50
1.1. Comment perçoit-on auditivement ?	52
1.2. Les modèles de perception	54
2. L'analyse des scènes auditives (ASA)	59
2.1. La perception des sources dans un environnement bruité	63
2.1.1. L'intégration (ou fusion)	64
2.1.2. La ségrégation (ou scission)	65
2.2. L'analyse des scènes auditives : l'influence de l'attention	67
3. Le sémantisme des sons de l'environnement	72
3.1. Des nuisances sonores à la sémantique du son	72
3.2. Qu'est-ce que la représentation d'un phénomène acoustique ?	74
3.3. Sons environnementaux, langage et représentation... un lien sémantique ?	76
Ce qu'il faut retenir...	79

**DEUXIEME CHAPITRE : ... AUX ETUDES SUR LA COMPREHENSION** **81**

1. Éléments définitoires : Qu'est-ce que comprendre ?	82
2. Des études sur la compréhension orale : de la perception du signal à la (re-) construction de sens	86
2.1. Reconnaissances des unités de la chaîne parlée	89
2.2. Des unités perçues à l'émergence d'une signification	91
3. Des études sur la compréhension de textes narratifs	95
3.1. Les dialogues pédagogiques : une forme de narration ?	96
3.2. Les modèles de compréhension de textes narratifs	97
3.2.1. Le modèle de construction-intégration : l'élaboration d'un modèle de situation	97
3.2.2. Le modèle landscape : la fluctuation de l'activation	103
Ce qu'il faut retenir	104

**TROISIEME CHAPITRE : HYPOTHESES** **105**

1. Qu'est-ce qu'être débutant en langue ?	107
2. Le rôle du contexte sonore environnemental dans la compréhension d'un dialogue	111

---

**PARTIE 2 : ELABORATION DU MATERIEL EXPERIMENTAL** **119**

**PREMIER CHAPITRE : ÉLABORATION DU MATERIEL SONORE ENVIRONNEMENTAL** **121**

1. Une étude sur les attentes auditives	121
---	-----

---

1.1. Méthodologie	122
1.1.1. Objectif, cadre théorique et hypothèses	122
1.1.2. Participants et déroulement	123
1.1.3. L'encodage des données verbales	124
La norme ICOR	124
L'extraction des données	125
1.2. Résultats	126
1.2.1. La cuisine	127
1.2.2. La terrasse de café	129
2. Construction des scènes sonores	132
2.1. Objectifs	132
2.2. Conditions d'enregistrement	132
2.3. Montage et manipulations du matériau sonore	134
Les constantes	134
Les facteurs de variation	135
Remarques empiriques	136
3. Identification des scènes sonores	138
3.1. Méthodologie	138
3.1.1. Objectifs et hypothèses	138
3.1.2. Participants et déroulement	138
Consignes et déroulement	138
Les distracteurs	139
Participants	140
3.2. Résultats	140
3.2.1. La situation Cuisine	141
Étiquetage des réponses	141
Résultats pour <i>la préparation d'un repas</i>	142
Résultats pour <i>le la vaisselle</i>	143
3.2.2. La situation Terrasse de café	144
Étiquetage des réponses	144
Résultats	144
4. Sélection des scènes sonores	146
4.1. Méthodologie	146
4.1.1. Objectifs et hypothèses	146
4.1.2. Participants et déroulement	146
4.2. Résultats	147
4.2.1. La préparation d'un repas	147
4.2.2. La vaisselle	148
	301

4.2.3. La terrasse de café	148
<b>DEUXIEME CHAPITRE : ÉLABORATION DU MATERIEL LINGUISTIQUE</b>	<b>151</b>
1. Méthode	151
1.1. Rédaction et enregistrement des dialogues	151
1.1.1. Des situations stéréotypées	152
1.1.2. Le niveau de langue	153
La préparation d'un repas	155
<i>La vaisselle</i>	155
<i>À la terrasse d'un café</i>	156
1.1.3. Durée	156
1.1.4. Enregistrements	157
1.2. Évaluation des dialogues	157
1.2.1. L'évaluation du niveau de langue	158
1.2.2. Une éventuelle utilisation en classe	158
2. Résultats	158
2.1. L'évaluation du niveau de langue	158
Dialogue A : la préparation d'un repas	158
Dialogue B : la vaisselle	160
Dialogue C : À la terrasse d'un café	161
2.2. Une éventuelle exploitation en salle de classe	162
<b>PARTIE 3 : EXPERIMENTATION PRINCIPALE</b>	<b>165</b>
1. Méthode	165
1.1. Le Classe-Images	166
1.1.1. Déroulement des passations	166
1.1.2. Les données recueillies	168
1.2. Constitution du corpus d'images	169
1.2.1. La préparation d'un repas	169
1.2.2. La vaisselle	173
1.2.3. À la terrasse d'un café	177
1.2.4. Remarques	181
1.3. Participants	182
1.3.1. Le panel de participants francophones	182
1.3.2. Le panel d'apprenants de niveau A	183
1.4. Variables du test	184
1.4.1. Les variables indépendantes	184
1.4.2. Les variables dépendantes	185
2. Résultats	187

2.1. La construction de suites de trois dessins cibles : la compréhension en termes binaires	187
2.1.1. La variable « expertise linguistique »	187
2.1.2. La variable « contexte sonore environnemental »	188
2.2. Les dessins sélectionnés: évaluation des parcelles de compréhension.	190
2.2.1. La variable « expertise linguistique »	191
2.2.2. La variable « contexte sonore environnemental »	196
2.3. Compréhension et charge cognitive :	203
2.3.1. La variable « expertise linguistique »	204
2.3.2. La variable « contexte sonore environnemental »	205
<b>PARTIE 4 : DISCUSSION</b>	<b>211</b>
1. Discussion des résultats du Classe-Images	212
1.1. Synthèse des principaux résultats	212
1.1.1. Qu'est-ce qu'être apprenant débutant en langue	213
1.1.2. Le rôle du contexte sonore environnemental	217
1.2. Temps de réponse, charge cognitive et réussite	222
1.3. Profils d'erreurs et effets d'interaction	224
2. Discussion méthodologique	229
2.1. Le matériel utilisé	230
2.1.1. Le matériel linguistique	230
2.1.2. Le corpus d'images	232
2.1.3. Le matériel sonore environnemental	233
2.2. Des différences inter-dialogues	237
2.3. Les participants au test	244
3. Discussion didactique et théorique	245
3.1. Apprendre à comprendre ?	246
3.1.1. Une éventuelle utilisation du Classe-Images en classe de langue	246
3.1.2. Vers un enseignement stratégique de la compréhension	247
3.2. Comprendre et apprendre?	250
3.2.1. Qu'est-ce qu'apprendre ?	251
3.2.2. Comprendre pour apprendre ...	254
<b>CONCLUSION GENERALE</b>	<b>261</b>
<b>BIBLIOGRAPHIE</b>	<b>269</b>
<b>WEBOGRAPHIE</b>	<b>287</b>

<b>TABLE DES ILLUSTRATIONS</b>	<b>289</b>
<b>TABLE DES FIGURES</b>	<b>289</b>
<b>TABLE DES TABLEAUX</b>	<b>290</b>
<b>TABLE DES DESSINS</b>	<b>293</b>
<b>TABLE DES IMAGES</b>	<b>294</b>
<b>LISTE DES ABREVIATIONS</b>	<b>295</b>
<b>PRESENTATION DES ANNEXES</b>	<b>297</b>
<b>TABLE DES MATIERES</b>	<b>299</b>

