Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition),
Traitement Automatique des Langues Naturelles (TALN, 27e édition),
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition)
Nancy, France, 08-19 juin 2020

Une base de données de phrases en français pour l'étude du rôle conjoint des incertitudes sémantique et acoustique dans la perception de la parole.

Loriane Leprieur[1] Olivier Crouzet[1,2] Étienne Gaudrain[2,3]
(1) Laboratoire de Linguistique de Nantes - LLING / UMR6310 Université de Nantes - CNRS
chemin de la Censive et du Tertre, 44312 Nantes Cedex, France
(2) ENT Department - University Medical Center Groningen, Rijksuniversität Groningen, Pays-Bas
(3) Centre de recherche en Neurosciences de Lyon - CNRS UMR 5292- CNRL Inserm U1028
CH Le Vinatier - Bâtiment 452, 95 bd Pinel, 69675 Bron Cedex
loriane.leprieur@etu.univ-nantes.fr, olivier.crouzet@univ-nantes.fr,
etienne.gaudrain@cnrs.fr

RÉSUMÉ

Les effets de contexte dans la perception de la parole reposent aussi bien sur des sources acoustiques que sémantiques. Le contexte acoustique fournit des informations essentielles pour l'adaptation au locuteur et aux variations dialectales. En parallèle, le contexte sémantique contribue à prédire un ensemble de mots éligibles pour une interprétation licite des énoncés. Afin d'étudier plus précisément les interactions entre ces effets de contexte, nous avons créé une base de données de phrases courtes conçues pour observer ces phénomènes dans des protocoles expérimentaux. Cette base de données est constituée de 28 triplets de phrases porteuses terminées par des cibles de paires minimales de mots CV ou CVC, autour de voyelles acoustiquement proches associées à 4 contrastes vocaliques. Afin d'évaluer la validité des 3 catégories de contexte sémantique considérées, des mesures de similarité sémantique et de fréquence lexicale ont été réalisées à partir de différents corpus de langue française.

ABSTRACT

A dataset of french sentences to study the joint roles of semantic and acoustic uncertainty in speech perception.

Context effects in speech perception rely on both acoustic and semantic sources of information. On one hand, acoustic context provides information concerning speaker-specific and dialectal variation. On the other hand, semantic contextual information contributes to the selection of appropriate lexical candidates. In order to investigate the interaction between these sources of contextual information, a dataset of short sentences has been conceived that are dedicated to studying these phenomena in future research. The final database is organised around 28 carrier sentence triplets that end with minimal CV or CVC word-pairs. Target word-pairs are articulated around 4 french vowel constrasts. In order to assess the validity of the 3 corresponding semantic categories, measures of semantic similarity and lexical frequency have been computed from various language databases.

MOTS-CLÉS : perception de la parole, base de données, effets de contexte, plongements de mots, incertitude.

KEYWORDS: speech perception, database, context effects, word embeddings, entropy.

Volume 1 : Journées d'Études sur la Parole, pages 353–361.
HAL : hal-02798558.

1 Introduction

La perception du signal sonore est soumise à différentes sources d'incertitude qui trouvent leur origine dans la grande variabilité des réalisations phonétiques (Joos, 1948; Peterson, 1961; Nearey, 1989; Meunier, 2005). Les effets de contexte constituent deux sources d'information *extrinsèque* qui interviennent dans la catégorisation perceptive ((Ladefoged & Broadbent, 1957; Connine & Clifton, 1987) et interagissent avec l'analyse *intrinsèque* d'un segment.

Les sources d'incertitude acoustiques sont liés aux caractéristiques sonores des énoncés : des paramètres tels que la fréquence des formants, la fréquence fondamentale, la durée ou l'intensité, varient pour un segment donné (variation *intrinsèque*) mais s'expriment aussi par des « tendances » globales à l'intérieur d'une fenêtre plus large que le segment (variation *extrinsèque*). Ces différentes sources de variation interviennent dans la catégorisation perceptive d'un segment.

1.1 Effets de contexte acoustique

Les effets du contexte acoustique ont été initialement mis en évidence dans les travaux princeps de Ladefoged & Broadbent (1957). Les participants devaient répéter un mot cible (/bɪt/, /bɛt/, /bat/ ou /bʌt/) précédé d'une phrase de consigne *« Please say what this word is »*. Cette phrase était altérée synthétiquement sur les valeurs de fréquences de F_1 et F_2 pour créer six contextes acoustiques distincts correspondant à des manipulations phonétiques du contexte : (1) F_1 et F_2 originaux, (2) F_1 bas, (3) F_1 haut, (4) F_2 bas, (5) F_2 haut, (6) F_1 bas et F_2 haut. Les réponses perceptives des participants étaient toujours réalisées sur une cible acoustique non-modidiée. Néanmoins, les auteurs montrent que les réponses perceptives sont systématiquement influencées par le contexte acoustique de la phrase : /bɪt/ est perçu /bɪt/ à 87.5 % dans le contexte d'origine mais comme /bɛt/ à 90 % quand F_1 est abaissé. Inversement /bɛt/ est perçu /bɛt/ par 77 % à 95 % des participants dans les contextes où F_1 est intact ou abaissé, mais comme /bɪt/ à 97 % quand F_1 est rehaussé. /bat/ est moins bien reconnu dans son contexte d'origine (à 58 % contre 42 % pour /bɛt/) mais une hausse de F_1 fait basculer la perception vers /bɛt/ à 80 %. Enfin, /bʌt/ est reconnu /bʌt/ à 82 % dans son contexte d'origine, mais une baisse de F_2 fait diminuer la performance d'identification à 38 % au profit de /bat/ (60 %). Ces changements de catégorisation perceptive sont interprétés comme des effets de compensation : si la cible ne change pas, les variations de son environnement modifient sa perception.

Ces effets ont également été confirmés plus récemment par Sjerps (Sjerps *et al.*, 2013). Les expériences présentées reproduisent les effets de compensation observés tout en donnant des pistes de réponse sur le niveau auquel ces variations sont traitées dans la catégorisation des phonèmes. Les auteurs utilisent pour cela un stimulus au format [Vpapu] où la voyelle cible (V) est un des dix intervalles d'un continuum entre [i] et [ɛ]. Le mot porteur [papu] est également modifié pour avoir deux contextes $F_1 + 200$ Hz et $F_1 - 200$ Hz sur les deux voyelles. En tâche de catégorisation les participants doivent d'abord identifier comme [i] ou [ɛ] les dix intervalles cibles dans les deux contextes porteurs. Les résultats montrent que le basculement de la perception de la voyelle de [i] vers [ɛ] sur les paliers médians est affecté par le contexte : si les paliers 5 et 6 sont reconnus comme [ɛ] dans environ 40 % des essais lorsque F_1 est abaissé, le taux de reconnaissance de [ɛ] augmente à 65 % lorsque la fréquence de F_1 est élevée. Cette expérience suggère aussi que ces effets de compensation prennent place y compris en présence d'une quantité temporelle d'information très limitée. La seconde expérience est une tâche de discrimination en 4I-oddity : un objet déviant parmi un ensemble de 4. Les participants devaient trouver la voyelle déviante dans un ensemble exclusivement [ipapu] ou [ɛpapu] sur les dix variantes

du continuum de cibles et dans les deux contextes porteurs. Les participants ont montré davantage de difficultés à reconnaître [ɛpapu] dans un contexte de F_1 bas (environ 80 % indépendamment de l'écart entre le déviant et les standards), et inversement pour la reconnaissance de [i] en contexte de F_1 élevé (environ 60 % pour un palier déviant proche du standard, environ 80 % lorsque le palier déviant est éloigné). Ce type de tâche demande une concentration portée davantage sur les indices acoustiques du signal, et non sur ses propriétés phonologiques. Ces résultats soutiennent l'idée que les effets de compensation prennent place à une étape très précoce des processus de catégorisation.

1.2 Effets de contexte sémantique

La catégorisation phonétique est aussi influencée par des informations provenant du contexte sémantique. Ces phénomènes ont été initialement mis en évidence par (Connine & Clifton, 1987) avec un matériel constitué de paires minimales de mots CVC où la variation portait sur le caractère voisé / non-voisé de la consonne initiale (p. ex. : ang. *dime* / *time*). Ces paires avaient subi un traitement synthétique pour créer un continuum de dix paliers, dont 5 allant vers le voisement et 5 allant vers le dévoisement de la consonne. Ces cibles étaient précédées de deux phrases introductrices possibles, chacune tendant vers le sens d'un seul mot de la paire. Les participants identifiaient le mot cible (ce qui donnait une indication sur la reconnaissance de la consonne initiale variable) et devaient dire si la phrase finale formée faisait alors sens ou non. Les résultats ont montré que les cibles phonétiques sont perçues correctement indépendamment du contexte sur les paliers extrêmes du continuum, mais que la perception varie selon le contexte sur les seuils proches de la frontière catégorielle entre les deux consonnes : le contexte correspondant au mot voisé exerce une attraction de la variable vers la consonne voisée, et inversement. L'expérience montre également un temps de réaction plus lent lorsque la cible n'est pas sémantiquement liée à la phrase. Selon (Connine & Clifton, 1987), le contexte sémantique apporterait une information destinée à faciliter la perception mais il aurait un rôle tardif : l'ambiguité générée lors de la phase perceptive initiale serait résolue à un niveau post-perceptif (décisionnel) par l'information contextuelle.

Ces deux sources d'informations –acoustique intrinsèque d'une part, liée au contexte sémantique d'autre part– exercent donc des influences combinées sur la catégorisation du signal. Gaskell & Marslen-Wilson (2001) ont cherché à mieux comprendre ces mécanismes à partir de deux approches concernant l'impact potentiel du contexte sémantique dans les modèles de reconnaissance des mots. Selon certains modèles, l'information sémantique peut bloquer très précocément l'émergence d'un candidat lexical alors que d'autres modèles attribuent un rôle primordial aux informations ascendantes et n'intègrent les informations sémantiques que très tardivement dans le processus d'identification. Cette opposition correspond à des divergences observées dans la littérature. Par exemple, Tabossi (1988), utilisant du matériel fondé sur la distinction sémantique entre homophones (p. ex. : *bank* désigne en anglais à la fois une institution financière et le bord d'une rivière), observe des effets du contexte sémantique. À l'inverse, Connine *et al.* (1994), qui avaient eu recours à des versions ambiguës intermédiaires entre deux mots phonologiquement distincts (p. ex : ang. *dip / tip*), concluaient au caractère tardif du contexte sémantique.

Parmi les hypothèses expliquant les divergences des résultats, Gaskell & Marslen-Wilson (2001) explorent la possibilité qu'il n'existe pas de mécanisme d'analyse sémantique résolvant l'ambiguité entre deux mots phonologiquement distincts alors que le contexte pourrait par contre résoudre l'ambiguité lexicale entre homophones. Pour ce faire, ils étudient le comportement de mots subissant une altération liée à la coarticulation (p. ex. : ang. *ru**n** / ru**m** – **d**oes / **p**icks*) dans une tâche d'amorçage

multimodal de répétition : les participants doivent réaliser une tâche de décision lexicale sur une présentation visuelle du mot cible (*rum* ou *run*) juste après la présentation auditive du mot en contexte.

Les auteurs observent que en situation où le contexte phonétique influence l'interprétation vers une cible coronale (« rum does », réalisé comme un équivalent phonétique de « run does ») la présence du contexte sémantique facilite l'accès vers l'interprétation de la cible coronale (« run ») sans bloquer la cible non-coronale (« rum »). Les deux représentations sont accessibles. Ces expériences suggèrent donc que l'information acoustique ascendante demeure prioritaire. Le contexte sémantique n'apporte une information que lorsque les indices phonétiques et phonologiques sont insuffisants pour résoudre l'ambiguité entre deux représentations actives.

La base de données que nous présentons ici a pour objectif d'étudier plus spécifiquement les interactions entre les effets du contexte sémantique (Connine & Clifton, 1987; Gaskell & Marslen-Wilson, 2001) et ceux de l'adaptation au contexte acoustique (Ladefoged & Broadbent, 1957; Sjerps *et al.*, 2013) en proposant un ensemble de phrases contrôlées du point de vue de paramètres objectifs fondés sur des méthodes de plongements de mots (Mikolov *et al.*, 2013).

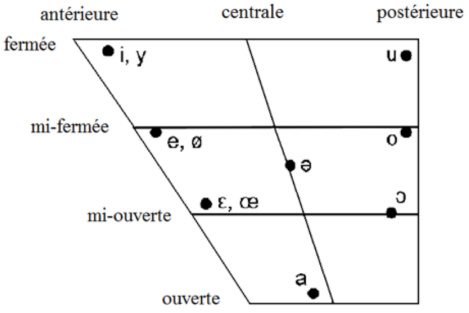
2 Méthode

La construction de la base de données s'est faite en deux étapes :

1. Une première phase de pré-sélection de mots-cible ;
2. Une seconde phase de construction de phrases à partir de mesures de similarité lexicale entre mots (Mikolov *et al.*, 2013; Fauconnier, 2015).

Les relations de similarité ont ensuite été évaluées pour les phrases générées et celles qui ne remplissaient pas les critères de relation établis entre les 3 catégories de lien sémantique ont été supprimées de la base de données. Des mesures de fréquence des mots-cible sélectionnés ont été réalisées afin de vérifier l'absence de déséquilibre entre les deux groupes.

2.1 Sélection des mots-cible



Voyelle	F_1	F_2	F_3
a	788	1503	2737
ɛ	660	2080	2954
a	788	1503	2737
ɔ	634	1180	2690
y	305	2046	2535
ø	469	1605	2581
o	461	855	2756
u	311	804	2485

FIGURE 1 – Trapèze vocalique du français, avec les valeurs des fréquences des trois premiers formants caractéristiques des voyelles cibles pour chaque paire vocalique de l'ensemble de données. Ces valeurs de fréquence des formants sont extraites de Calliope (Coll.) (1989).

Les mots-cible sont des mots français correspondant à des paires distinctives de type CV ou CVC selon les paires sélectionnées (p. ex. : « gare » /gaʁ/ et « guerre » /gɛʁ/) et qui se distinguent par leur voyelle. À partir d'une recherche de mots monosyllabiques du français dans la base de données BRULEX (Content *et al.*, 1990), nous avons sélectionné des paires de mots associées à 4 couples de voyelles du français. Ces couples de voyelles correspondent à des catégories dont les valeurs de fréquence des formants relevées sont proches dans un espace vocalique $F_1 \sim F_2 \sim F_3$ (Calliope (Coll.), 1989). Les mots cibles sont tous des noms communs. Pour une paire donnée, ils sont de même genre ou ont la même forme au pluriel afin que le déterminant soit identique dans les différents contextes sémantiques. La base de données constituée porte ainsi sur les paires [a / ɛ], [a / ɔ], [y / ø] et [o / u]. Nous avons initialement identifié 46 paires de mots qui pouvaient potentiellement servir de base à la construction des phrases.

2.2 Construction des phrases

Pour chacune des 46 paires de mots identifiées, nous avons procédé à la conception de phrases porteuses (Sujet-Verbe) correspondant à 3 catégories sémantiques :

Contexte 1 Phrase porteuse dont la signification est associée au mot 1 de la paire (exemple : « balle », /bal/) mais n'est pas liée au mot 2 (exemple : « belle », /bɛl/) ;

Contexte 2 Phrase porteuse dont la signification est associée au mot 2 de la paire (exemple : « belle », /bɛl/) mais n'est pas liée au mot 1 (exemple : « balle », /bal/) ;

Contexte 0 Phrase porteuse dont la signification n'est associée à aucun des deux mots de la paire ;

Des exemples de ces 3 contextes sont donnés dans le tableau 1.

TABLE 1 – Exemple de combinaison entre une paire de mots et les phrases forgées correspondant aux 3 contextes sémantiques possibles. Le contexte 1 favorise le mot 1, le contexte 2 favorise le mot 2, le contexte 0 ne favorise aucun des deux mots.

Mot de la paire	Contexte	Phrase finale
(1) balle	(1) Le joueur a dévié la	Le joueur a dévié la balle.
	(2) Le prince a charmé la	Le prince a charmé la balle.
	(0) La salade a raccourci la	La salade a raccourci la balle.
(2) belle	(1) Le joueur a dévié la	Le joueur a dévié la belle.
	(2) Le prince a charmé la	Le prince a charmé la belle.
	(0) La salade a raccourci la	La salade a raccourci la belle.

Les phrases ont été constituées à l'aide de mesures de plongements de mots (Word2vec, Mikolov *et al.*, 2013) issues d'un modèle du corpus Wikipedia en français (Fauconnier, 2015). Ces mesures de plongements de mots (ang. *word embeddings*) fournissent une estimation de la similarité sémantique entre deux mots exprimée par un vecteur allant de 0 (aucune proximité) à 1 (proximité maximale). Sur la base d'essais progressifs, nous avons fixé des valeurs de similarité types comme seuils d'acceptabilité minimal ou maximal pour les mots composant les phrases de chaque contexte. Dans un premier temps, nous avons d'abord cherché à sélectionner les verbes afin d'obtenir une valeur de similarité supérieure à 0.250 avec la cible de leur contexte et inférieure à 0.150 avec la cible opposée

ou présentant un différentiel de similarité d'au-moins 0.2 entre les deux contextes. Par exemple, pour la paire « balle » / « belle » le contexte 1 correspond au verbe « dévier » dont la proximité avec « balle » est de 0.334 alors que sa similarité avec le mot « belle » n'est que de 0.096 (différence = 0.238). Pour le contexte 2 lié à « belle », le verbe « charmer » a une similarité de 0.343 avec la cible mais de seulement 0.044 avec le mot « balle » (différence = 0.299).

Les noms-sujet de chaque phrase ont ensuite été choisis de manière à respecter l'une des deux conditions suivantes :

— soit répondre aux mêmes seuils d'acceptabilité que pour le verbe (au-moins 0.250 pour le contexte relié et au-plus 0.150 pour le contexte non-relié),
— soit, si cette condition n'était pas possible, présenter une différence *positive* en faveur du contexte supposé relié.

Ainsi, pour la même paire « balle » / « belle », le contexte 1 a pour sujet « joueur » dont la proximité avec la cible « balle » est de 0.374 alors que sa similarité avec la cible « belle » est de 0.095. Le contexte 2 a pour sujet « prince » qui présente une valeur de similarité de 0.186 avec la cible, et de seulement 0.054 avec « balle ». On voit que le seuil de 0.250 entre le sujet et la cible n'est pas respecté. Par contre, la différence entre les deux contextes est bien positive.

Pour la construction des phrases correspondant au contexte 0 (aucun lien avec l'un des deux membres de la paire), nous avons sélectionné un sujet et un verbe dont les valeurs de similarité sont systématiquement inférieures à 0.100 avec les deux cibles. Toujours pour la même paire, le sujet « salade » correspond à une proximité de 0.051 avec « balle » et de 0.099 avec « belle » ; le verbe « raccourcir » présente des valeurs de similarité de 0.075 avec « balle » et de 0.052 avec « belle ».

Sur les 46 paires de mots initiales, 18 n'ont pas permis de trouver des combinaisons de 3 phrases porteuses respectant ces conditions. Au final, nous obtenons un ensemble de 28 triplets de phrases porteuses associés à 28 paires de mots. Le biais sémantique induit par la phrase porteuse peut être vu comme la combinaison des valeurs de proximité du sujet et du verbe avec le mot-cible.

Sur la base de ces éléments, les phrases sont composées de manière à ce que, pour une paire de mots, les 3 contextes soient exprimés avec un accord en genre et en nombre. Les verbes sont conjugués au même temps. L'uniformité des déterminants / genre / nombre / temps verbal au sein des trois contextes d'une même paire permet d'interchanger ces contextes devant les cibles pour opérer des manipulations du contexte sémantique. Les phrases sont relativement courtes, de 6 à 12 syllabes, ce qui permet une interprétation rapide de la structure tout en conservant une quantité suffisante de matériel acoustique et sémantique pour générer les effets de contexte prédits.

2.3 Enregistrement audio

L'enregistrement de l'intégralité des phrases a été effectué en chambre sourde par deux locuteurs natifs du français (français urbain de l'ouest de la France), une femme de 20 ans et un homme de 22 ans étudiants à l'Université de Nantes. Les phrases étaient présentées dans un ordre aléatoire et répétées trois fois en imposant un débit relativement soutenu à travers l'interface de présentation visuelle des phrases.

Pour chaque phrase enregistrée, nous avons ensuite déterminé deux informations temporelles : le moment de transition entre la phrase porteuse et la cible, ainsi que la position correspondant au milieu de la voyelle. Le point de transition a été déterminé en deux phases. Dans un premier temps, nous avons positionné ce point à partir de l'étude acoustique (spectrogramme et forme d'onde) des signaux.

Dans un second temps, un script permettait d'écouter les séquences en isolant la phrase porteuse du mot cible afin de déterminer dans quelle mesure ce point temporel permettait de séparer les deux parties de la phrase de manière satisfaisante. Cette position temporelle pouvait alors être modifiée de manière à améliorer le découpage. Un script Python a été conçu afin de séparer les phrases porteuses des cibles avant de les recombiner par *cross-splicing*, en suivant un *design* de carré-latin : une cible enregistrée à l'origine en contexte 0 est rattachée à la porteuse de contexte 1, une cible enregistrée en contexte 1 est rattachée à la porteuse de contexte 2 et une cible enregistrée en contexte 2 est rattachée à la porteuse de contexte 0. Cette méthode permet de neutraliser complètement les effets de coarticulation (courte et longue distance) entre la phrase porteuse et le mot-cible.

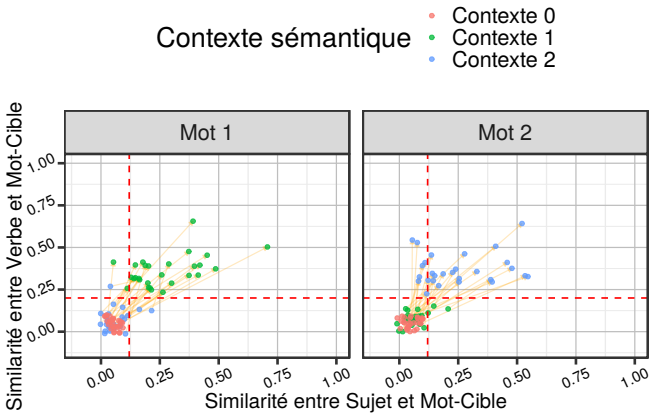
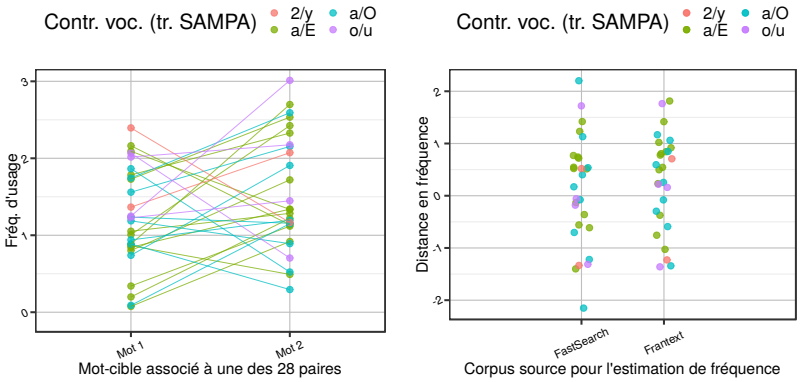


FIGURE 2 – Valeurs de similarité entre le mot-cible et respectivement le sujet (abcisse) / le verbe (ordonnée). Chaque point correspond à un mot d'une paire (Mot-cible 1 dans le graphique de gauche, mot-cible 2 dans le graphique de droite). Les points verts correspondent aux phrases porteuses associées au mot 1 (contexte 1), les points bleus correspondent aux phrases porteuses associées au mot 2 (contexte 2), les points rouges correspondent aux phrases porteuses de contexte 0. Les traits rouges en pointillés sont des repères visuels pour distinguer les nuages de points.

3 Résultats

Le corpus final est constitué de 28 paires de mots qui sont réparties de manière déséquilibrée entre les 4 contrastes vocaliques : 13 paires [a / ɛ], 9 paires [a / ɔ], 2 paires [y / ø], et 4 paires [o / u]. Les mesures de similarité caractérisant les relations entre mot-cible, verbe et nom-sujet en fonction du type de contexte sont présentées dans la figure 2. Ces données permettent de vérifier le caractère opérant des 3 catégories de contexte considérées : les valeurs de similarité du sujet et du verbe sont regroupées en position haute et / ou droite du graphique pour le contexte sémantiquement relié (contexte 1 / mot 1, contexte 2 / mot 2) alors que les valeurs sont regroupées en position basse et / ou gauche pour le contexte sémantiquement non-relié (contexte 1 / mot 2, contexte 2 / mot 1). Les valeurs associées au contexte 0 sont dans tous les cas localisées en bas / à gauche.

Pour chaque paire de mots, nous avons également recueilli les fréquences d'occurrence issues de la base de données Lexique (New *et al.*, 2001). Ces mesures sont représentées dans la figure 3. La différence de fréquence d'usage entre membres d'une paire est non-significative aussi bien sur le corpus *Frantext* (textes écrits, $\overline{x} = 0.307, sd = 0.898$; $t_{27} = 1.81, p = 0.082$), que sur le corpus *FastSearch* (pages web, $\overline{x} = 0.109, sd = 1.03, t_{27} = 0.559, p = 0.58$).



(a) Valeurs individuelles de fréq. (b) Distance en fréq.

FIGURE 3 – Fréquences d'occurrence ($\log_{10}$ sur 1 million) des mots-cible relevées dans la base de données Lexique (New *et al.*, 2001). (3a) valeurs individuelles comparées entre mots d'une même paire pour la fréquence d'occurrence sur des textes écrits (Frantext). (3b) distance de fréquence entre le mot 1 et le mot 2 issues de pages web (Fastsearch) et de textes écrits (Frantext).

4 Discussion

L'objectif de cette base de données est de mettre à disposition un matériel expérimental permettant de confronter les informations liées aux sources d'incertitude acoustique et sémantique dans les mécanismes de perception de la parole en fournissant des mesures objectives de relations sémantiques fondées sur des modèles de plongements de mots (Mikolov *et al.*, 2013). Ce travail fournit une liste de $3 \times 28 \times 2 = 168$ phrases distinctes correspondant à la combinaison de 28 paires de mots acoustiquement proches avec 3 contextes sémantiques distincts. Ces phrases ont été enregistrées par deux locuteurs francophones et manipulées afin de produire des croisements entre phrase porteuse et mot-cible qui permettent de supprimer la contribution des effets de coarticulation de la phrase porteuse vers le mot-cible.

Des mesures supplémentaires sont en cours sur un modèle alternatif du corpus Wikipedia en français (Gaudrain & Crouzet, 2019) afin de vérifier les valeurs issues du modèle de Fauconnier (2015). Nous avons également soumis l'ensemble des phrases possibles à un échantillon de locuteurs natifs du français et recueilli leurs jugements d'interprétabilité de 1 (non-interprétable) à 5 (totalement interprétable) afin de comparer nos mesures avec des estimateurs issus de réponses fournies par des locuteurs. Ces données sont en cours de récolte.

Cette base de données, grâce aux méta-informations fournies (valeurs de similarité, fréquences des mots, position temporelle du milieu de la voyelle du mot-cible…) et aux manipulations réalisées (*cross-splicing*), est multidisciplinaire et d'autres applications et expériences pouvant utiliser ce type de données peuvent être envisagées, notamment autour de questions d'acquisition et de *machine learning*.

L'intégralité des enregistrements (originaux, segmentés et réassemblés par *cross-splicing*) ainsi que la liste des phrases et les mesures réalisées sont mis à disposition sur un dépôt Zenodo (https://doi.org/10.5281/zenodo.3818582).

Remerciements

Ce travail a reçu le soutien de la « Mission pour les Initiatives Transverses et l'Interdisciplinarité » (MITI, CNRS, FR) et du programme Marie Skłodowska–Curie (PRESTIGE–2017–2–0044, UE).

Références

CALLIOPE (COLL.) (1989). *La Parole et son traitement automatique*. Paris : Masson.

CONNINE C. M., BLASKO D. G. & WANG J. (1994). Vertical similarity in spoken word recognition : Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics*, **56**(6), 624–636. DOI : 10.3758/bf03208356.

CONNINE C. M. & CLIFTON C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology : Human Perception and Performance*, **13**(2), 291–299. DOI : 10.1037/0096-1523.13.2.291.

CONTENT A., MOUSTY P. & RADEAU M. (1990). Brulex. une base de données lexicales informatisée pour le français écrit et parlé. *L'année psychologique*, **90**(4), 551–566. DOI : 10.3406/psy.1990.29428.

FAUCONNIER J.-P. (2015). French word embeddings. http://fauconnier.github.io.

GASKELL G. & MARSLEN-WILSON W. D. (2001). Lexical ambiguity resolution and spoken word recognition : Bridging the gap. *Journal of Memory and Language*, **44**(3), 325–349. DOI : 10.1006/jmla.2000.2741.

GAUDRAIN E. & CROUZET O. (2019). word2vec model trained on lemmatized French Wikipedia 2018. type : dataset, DOI : 10.5281/zenodo.3241447.

JOOS M. (1948). *Acoustic Phonetics*. Language monographs. Linguistic Society of America.

LADEFOGED P. & BROADBENT D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, **29**(1), 98–104. DOI : 10.1121/1.1908694.

MEUNIER C. (2005). Invariants et variabilité en phonétique. In N. NGUYEN, S. WAUQUIER-GRAVELINES & J. DURAND, Éds., *Phonologie et Phonétique : Forme et Substance*, chapitre 13, p. 349–374. Paris : Lavoisier.

MIKOLOV T., CHEN K., CORRADO G. S. & DEAN J. (2013). Efficient estimation of word representations in vector space. http://arxiv.org/abs/1301.3781v3.

NEAREY T. M. (1989). Static, dynamic, and relational properties in vowel perception. *The Journal of the Acoustical Society of America*, **85**(5), 2088–2113. DOI : 10.1121/1.397861.

NEW B., PALLIER C., FERRAND L. & MATOS R. (2001). Une base de données lexicales du français contemporain sur internet : LEXIQUE. *L'année psychologique*, **101**(3), 447–462. DOI : 10.3406/psy.2001.1341.

PETERSON G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research*, **4**(1), 10–29. DOI : 10.1044/jshr.0401.10.

SJERPS M. J., MCQUEEN J. M. & MITTERER H. (2013). Evidence for precategorical extrinsic vowel normalization. *Attention, Perception, & Psychophysics*, **75**(3), 576–587. DOI : 10.3758/s13414-012-0408-7.

TABOSSI P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, **27**(3), 324–340. DOI : 10.1016/0749-596x(88)90058-7.