
Institut National des Langues et Civilisations Orientales

**Variations Phonétiques des Accents de la
Langue Italienne**

MASTER

TRAITEMENT AUTOMATIQUE DES LANGUES

Parcours :

Ingénierie Multilingue

par

Giovanna FAVIA

Directeur de mémoire :

Cyril Grouin

Encadrant :

Ioana Vasilescu et Martine Adda-Decker

Année universitaire 2017/2018

TABLE DES MATIÈRES

Liste des figures	5
Liste des tableaux	5
Résumé	7
Remerciements	9
Introduction	11
I État de l'Art	15
1 Variation des accents dans les langues	17
1.1 Variantes régionales en italien	17
1.2 Autres phénomènes de variation : natifs vs non natifs	19
1.3 Différences inter-locuteurs : le genre	20
II Corpus et Méthode	21
2 Corpus	23
2.1 Pré-traitements	23
2.2 Annotation de l'origine géographique des locuteurs	23
2.3 Reconnaissance vocale	24
3 Méthode	27
3.1 Annotation des enregistrements par origine géographique	27
3.2 Extraction des phonèmes, durées et positions	27
3.3 Détection automatique de l'origine des locuteurs	28
3.4 Influence du genre sur l'identification de l'origine	29
III Analyses acoustiques et variantes régionales	31
4 Motifs de variation régionale du système consonantique	33
4.1 Renforcement phono-syntaxique	35
4.2 Redoublement Intervocalique	37
4.3 Consonnes, voyelles et géminées	38
5 Le système vocalique	41
5.1 Espace vocalique en italien	41
5.2 Espace vocalique et particularités régionales	43

5.3	Détection automatique des parlers régionaux avec WEKA	47
6	Autres facteurs de variation : genre du locuteur	51
6.1	Présentation	51
6.2	Classification des accents selon le sexe du locuteur	52
	Conclusion et perspectives	53
	Bibliographie	57

LISTE DES FIGURES

0.1	Carte des variantes dialectales de l'italien	13
2.1	Extrait du fichier tabulaire avec les informations nécessaires à la segmentation automatique	24
2.2	Extrait du fichier audio segmenté automatiquement	25
3.1	Extrait du fichier tabulaire employé pour effectuer les analyses acoustiques sur les phonèmes consonantiques	28
3.2	Extrait du fichier employé pour la détection de l'origine des locuteurs	28
4.1	Représentation des occurrences des consonnes géminées dans le corpus	34
4.2	Extrait du script PRAAT pour extraire la durée de chaque phonème	36
5.1	Triangle vocalique de la langue italienne	41
5.2	Représentation des occurrences des voyelles dans le corpus	42
5.3	Triangle vocalique issu des prononciations des voyelles des locuteurs d'origine sarde	43
5.4	Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlers du Nord (pointillés = femmes, lignes = hommes)	44
5.5	Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlers du Centre (pointillés = femmes, lignes = hommes)	45
5.6	Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlers du Sud (pointillés = femmes, lignes = hommes)	45
5.7	Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlers de Sardaigne (pointillés = femmes, lignes = hommes)	46
5.8	Comparaison des timbres des voyelles moyennes entre tous les parlers présents dans le corpus, y compris ceux des locuteurs non natifs	47
6.1	Spectrogrammes correspondant aux phonèmes /tʃ/ et /ʃ/ prononcés par un locuteur avec un italien standard	54
6.2	Spectrogramme correspondant au phonème /tʃ/ prononcé par un locuteur de la capitale	55

LISTE DES TABLEAUX

2.1	Description quantitative du corpus GRR	25
4.1	Table des phonèmes de la langue italienne	33
4.2	Moyenne des durées (ms) des phonèmes géminés par groupes de consonnes	35
4.3	Moyenne des durées (ms) des phonèmes consonantiques simples par groupes de consonnes	35
4.4	Moyenne des durées des phonèmes en position de renforcement phonosyntaxique	37
4.5	Moyenne des durées (ms) des phonèmes en position intervocalique	38

4.6	Moyenne des durées (ms) des phonèmes occlusifs géminés	38
4.7	Moyenne des durées (ms) des phonèmes occlusifs simples	38
4.8	Moyenne des durées (ms) des voyelles en présence de consonnes géminées .	39
5.1	Tableau des taux d'erreur produit par l'extraction automatique des for- mats des voyelles	44
5.2	Valeurs moyennes des formants	46
5.3	Résultat de la détection des trois classes par l'algorithme Ibk	48
5.4	Résultat de la détection des trois classes par l'algorithme Naïve Bayes . . .	48
5.5	Résultat de la détection des trois classes par l'algorithme J48	48
5.6	Résultats de la détection de la classe correspondante aux traits acous- tiques d'un italien standard par l'algorithme Ibk	49
5.7	Résultats de la détection de la classe correspondante aux traits acous- tiques d'un italien standard par l'algorithme Naïve Bayes	49
5.8	Résultats de la détection de la classe correspondante aux traits acous- tiques d'un italien standard par l'algorithme de tree J48	49
6.1	Pourcentage de perception d'un accent régionale chez les deux sexes	51
6.2	Résultat de la détection des trois classes chez les femmes	52
6.3	Résultat de la détection des trois classes chez les hommes	52

RÉSUMÉ

Ce mémoire de recherche vise à analyser les variations phonétiques des accents de l'italien contemporain, divisés en trois grandes zones (Nord, Centre et Sud) au sein d'un grand corpus de parole continue et semi-spontanée de journaux radiodiffusés. Les analyses acoustiques sont fondées sur la durée des phonèmes consonantiques d'une part, et sur les valeurs des formants des voyelles d'autre part. Pour ce qui concerne la durée des phonèmes consonantiques, l'attention est portée sur les phénomènes de renforcement phono-syntaxique et de redoublement intervocalique qui se produisent sur toutes les classes de consonnes, à l'exception des fricatives, chez les locuteurs des trois zones considérées. Pour les phonèmes vocaliques, une analyse des triangles vocaliques est d'abord menée, suivie d'une classification automatique des accents des locuteurs. Une dernière étude porte sur l'importance du genre du locuteur dans un contexte de détection automatique de l'origine.

REMERCIEMENTS

Je tiens à remercier toute l'équipe du PluriTal de l'Inalco, de Paris III, ainsi que de ParisX pour leurs enseignements et disponibilité au cours de ces deux années. Un remerciement particulier à Cyril Grouin pour avoir accepté de diriger mon mémoire avec professionnalité. Je remercie également tous les membres du laboratoire LIMSI, en particulier ceux du groupe du "Traitement du langage parlé" pour leur gentillesse et disponibilité, en particulier Ioana Vasilescu, pour ses enseignements et encadrement tout au long de mon stage. Parmi mes collègues de master, un remerciement particulier à Gabriele Chignoli pour ses conseils au sujet de cette étude de recherche.

INTRODUCTION

Ce mémoire de recherche porte sur l'analyse acoustique de l'italien contemporain. Le travail s'appuie sur l'analyse automatisée d'un corpus de journaux radiodiffusés en italien de différentes régions de la Péninsule et au-delà (Île de Sardaigne, accents étrangers). Ce travail a été mené dans le laboratoire LIMSI-CNRS et dans le groupe "Traitement du Langage Parlé" (TLP) dont l'un des principaux domaines de recherche est la transcription automatique de la parole. La linguistique outillée, qui fait appel à des grands corpus et utilise des méthodes d'analyse issues des technologies vocales et de la statistique a changé le paysage de l'analyse phonétique de l'oral. L'attention des chercheurs s'est portée de plus en plus sur des données hétérogènes, enregistrées dans des conditions écologiques¹, qui illustrent mieux - par rapport aux corpus de laboratoire - les processus actifs dans les langues [Vasilescu, 2016]. L'évolution de la phonétique vers le statut de "science de la parole" est en rapport avec les technologies vocales et notamment la transcription automatique de la parole. L'interaction avec ce domaine a eu deux effets bénéfiques : d'une part, la phonétique a pu s'appuyer sur les corpus acquis pour entraîner des systèmes automatiques. Ces systèmes totalisent des dizaines, voir des centaines d'heures et sont souvent acquis dans des conditions moins contrôlées comme les journaux radiodiffusés. Le caractère plus spontané des données permet de saisir des processus qui échappent lorsque l'objet d'étude est représenté par des données contrôlées, de laboratoire. Il est notamment possible de saisir des motifs de la variation phonétique qui renseignent de différentes caractéristiques d'une langue comme ses particularités régionales, les spécificités de prononciation des locuteurs selon le statut social, ou encore les changements phonétiques. D'autre part, la phonétique a pu s'appuyer sur ces systèmes pour les analyses proprement dites. Notamment les systèmes de transcription automatique de la parole servent au prétraitement des données dans la mesure où ils permettent un alignement fin entre les sons et le texte, représentant la transcription orthographique de ce qui a été dit.

Notre travail prend en compte l'italien contemporain. La langue italienne compte 60 millions de locuteurs dans la péninsule italienne, environ 660 000 locuteurs en Suisse. Elle compte encore quelques locuteurs en Istrie (Croatie) et elle est la troisième langue à Malte, après le maltais et l'anglais. L'italien est aussi toujours parlé avec différents niveaux de maîtrise, par des communautés d'immigrés dans plusieurs pays du monde. Par opposition à d'autres langues européennes, comme l'anglais et le français, l'italien ne présente pas une variété définie standard. Lorsqu'on parle d'italien standard on fait plutôt référence à un modèle abstrait, valable pour la langue écrite [Berruto, 2010]. Pour des raisons culturelles et historiques, qui n'ont vu l'Italie comme un pays unique que récemment et par les grands écrivains d'origine florentine, la variante régionale de la Toscane, la plus proche du latin, est communément admise comme la variante de l'italien la plus semblable à l'italien normalisé

1. enregistrements de parole spontanée

[Berruto, 2010], cependant, l'influence de l'accent de la capitale a récemment modifié cette dénomination [Bertinetto and Loporcaro, 2005].

Variantes dialectales Dans le cadre d'une étude basée sur les variantes phonétiques de l'italien, on aurait tendance à parler de variante régionale, ce qui n'est pas tout à fait correct. Les accents de la langue italienne ne varient pas par rapport à la région, mais plutôt par rapport à la zone [Bonomi et al., 2010, Berruto, 2010]. On peut donc diviser les variantes de la langue en 5 zones, comme l'illustre la carte 0.1² :

- la variante septentrionale (Piémont, Ligurie, Lombardie, Emilia-Romagna, Veneto, Trentino, Friuli).
- la variante centrale (Toscane, Lazio, Umbria, Marches).
- la variante méridionale (Campanie, Abruzzes, Molise, Pouilles septentrionales settentrionale, lucana, Calabre septentrionale).
- la variante méridionale extrême (Pouilles méridionales, Calabre méridionale et Sicile).
- la variante de Sardaigne (considérée par beaucoup de linguiste comme une langue à part entière).

Phonèmes La langue italienne possède 30 phonèmes consonantiques, dont toutes les consonnes, sauf /z/, peuvent être phonétiquement géminées. Les phonèmes vocaliques, sont 9, dont deux semi-voyelles (/j/ et /w/, les graphèmes *i* et *u* lorsqu'ils sont suivi d'une voyelle).

L'analyse de la variation phonétique en italien, en lien avec les particularités régionales, a fait l'objet de nombreuses études dédiées aux différents niveaux possibles d'analyse, variation vocalique [E.L.Renwick and Ladd, 2016] et consonantique [Elinor M.Payne, 2005, Mattei and Benedetto, 2008, Bafle, 2014, Turco and Braun, 2016]. Cependant peu d'études s'intéressent à cette variation sur de grands corpus traités automatiquement.

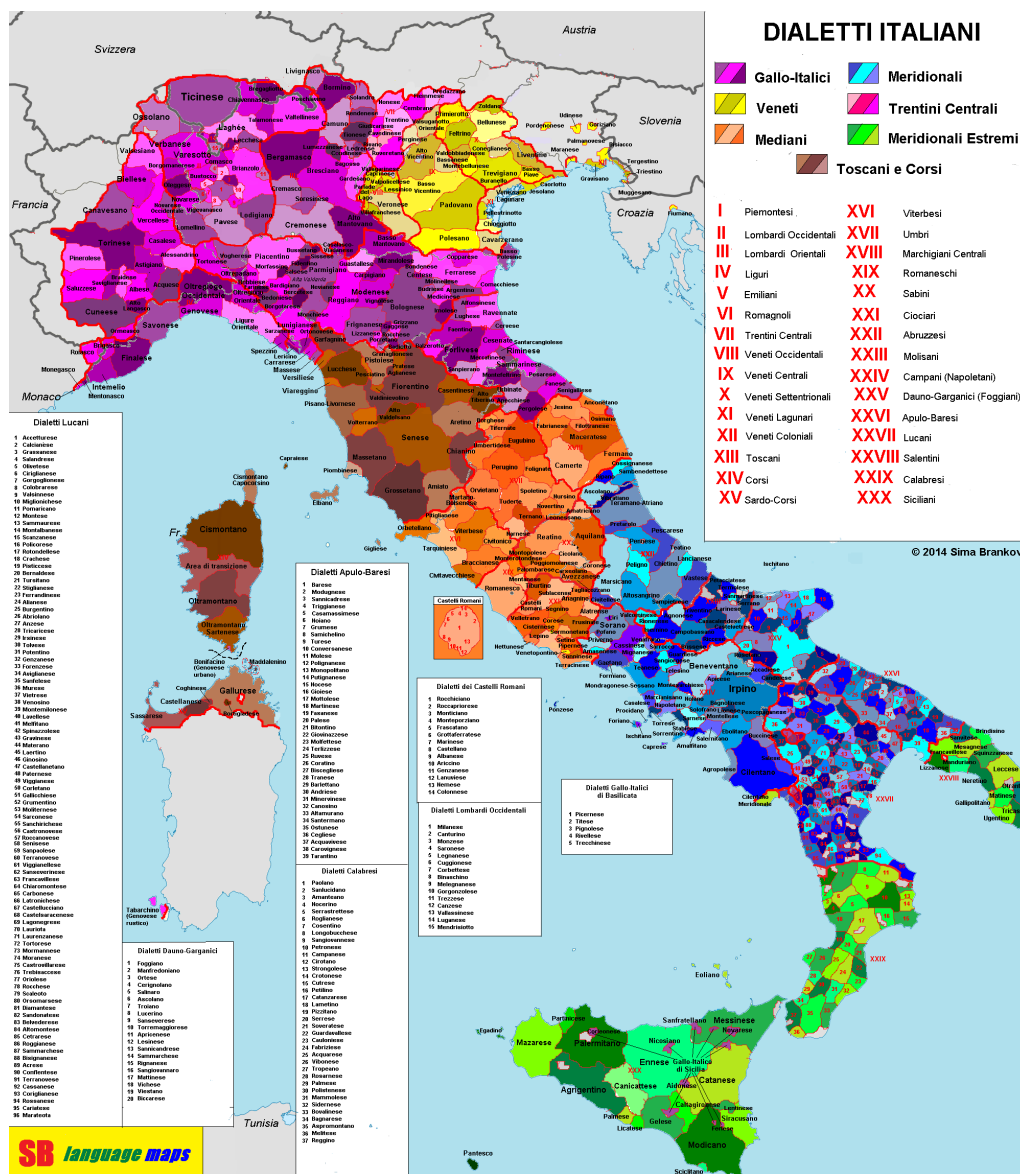
Notre étude prend en compte la phonétique de l'italien contemporain et se focalise sur quelques particularités régionales qui caractérisent les consonnes et les voyelles de la langue à partir d'un corpus oral de 5 heures environ ayant bénéficié d'une ségmentation automatique.

Ce mémoire est organisé comme suit. Après un chapitre introductif dans lequel nous décrivons le corpus et la méthodologie, **une première partie** sera dédiée aux particularités du système consonantique et notamment à un phénomène caractéristique de l'italien appelé le *redoublement consonantique*. Après avoir introduit la question des consonnes géminées phonologiques, et analysé leur spécificité de durée dans notre corpus, nous allons définir les zones et les conditions d'occurrence des consonnes redoublées, et tenter d'expliquer la fréquence observée, ainsi que leur interaction avec le facteur origine géographique. Nous analyserons également l'interaction consonnes/voyelles. Nous estimerons ainsi la variabilité de la durée des voyelles au contexte consonantique et en lien avec la région géographique.

Une **deuxième partie** sera dédiée au système vocalique. Nous décrivons les espaces vocaliques des voyelles selon les régions géographiques suite à l'extraction automatique de paramètres acoustiques dont les deux premiers formants F1 et F2. Une

2. L'image présentée représente une répartition plus détaillée des accents de la péninsule, par une séparation de la région de Vénétie et de la Toscane.

FIGURE 0.1 – Carte des variantes dialectales de l'italien



sous-partie sera aussi dédiée à la détection automatique de l'origine géographique des enregistrements en se fondant sur les formants des voyelles et avec une méthode d'apprentissage supervisé à l'aide du logiciel WEKA.

Enfin, une dernière partie sera aussi consacrée au facteur genre, également responsable de patterns de la variation phonétique. Nous estimerons dans quelle mesure il y a une interaction entre le sexe du locuteur et son origine géographique et mènerons des expériences de détection automatique des accents en interaction avec ce facteur.

Enfin dans la conclusion, nous proposerons un début d'analyse spectrale de deux motifs de la variation de plus en plus rencontrés dans l'italien parlé, à savoir la production de la consonne affriquée /ʃ/ à la place de la fricative /tʃ/. Ces productions montrent que l'analyse de l'oral faite sur des données ayant un caractère peu contrôlé et en quantité importante peut pointer sur des changements phonétiques naissants.

Dans le futur, il serait intéressant de prendre en compte à la fois les patterns acoustiques des segments se prêtant à ces confusions et la fréquence des phénomènes en lien avec le caractère standard ou non standard de la langue.

Une conclusion et des perspectives clôtureront ce mémoire.

Première partie

État de l'Art

VARIATION DES ACCENTS DANS LES LANGUES

La langue varie selon plusieurs facteurs, qu'il s'agisse de sa variante écrite ou parlée. Dans sa dimension orale elle est, de manière encore plus marquée, soumise à des variations continues, qui peuvent être définies comme intrapersonnelles ou interpersonnelles. Elles peuvent être causées par différents facteurs, comme le temps (variation diachronique), l'espace (variation diatopique) ou les caractéristiques sociales des locuteurs (variation diastratique) [Ledegen and Léglise, 2013]. Ces trois facteurs interagissent et s'influencent les uns avec les autres. Notre étude se focalise cependant sur les variations phonétiques due aux facteurs diatopiques, donc sur les différences régionales qui se reflètent dans des particularités phonétiques pertinentes d'un point de vue acoustique et perceptif, tout en étudiant également les autres facteurs qui influencent l'accent présents dans nos données, tels que le statut de langue native/non native (l'accent étranger) et les particularités socio-culturelles des locuteurs (l'influence du genre).

1.1 Variantes régionales en italien

Les études basées sur des corpus illustrant des parlars de différentes régions sont bien représentées dans la phonétique de l'italien. Ces études prennent en compte les différentes particularités au niveau consonantique ou vocalique et mettent en évidence la variabilité de réalisation selon l'origine géographique du locuteur. Parmi les aspects consonantiques, nous observons que l'opposition consonne simple vs. consonne géminée, ou double est la plus étudiée. Cette diversité est fondée sur une plus longue durée des consonnes géminées, qui peut varier selon la nature du phonème ainsi que sa position à l'intérieur du mot, et dont la production est anticipée par une petite pause, comme décrit par [Elinor M. Payne, 2005].

L'étude conduite par [Turco and Braun, 2016] prend en compte un corpus assez restreint, produit en laboratoire, correspondant à 15 minutes d'enregistrement, basé sur une série de phrases, lues par neuf locuteurs de sexe différent, provenant de six villes d'Italie situées au nord, au centre et au sud de la péninsule. Elle a pour but d'analyser, d'une part la durée des premières consonnes d'un mot ne contenant aucune géminée, en la comparant à la durée de la même consonne, prononcée dans la paire minimale, contenant une géminée. C'est le cas des paires minimales *mese* prononcé /mese/ (mois) et *messe* prononcé /mes :e/ (messes). D'autre part, la même analyse est menée sur des mots contenant une géminée, en opposition aux mots contenant deux consonnes consécutives différentes (clusters). Ensuite, l'étude s'étend à la

variabilité de la durée de la première voyelle qui précède la géminée et celle qui précède le cluster, ainsi qu'à la comparaison entre la longueur de la consonne géminée comparée à la même consonne à l'intérieur du cluster.

L'analyse montre que, sans distinction, les neuf locuteurs ont produit un premier phonème plus court dans un mot ne contenant pas de géminée ni de cluster, par rapport aux premiers phonèmes des mots contenant une double consonne et un cluster. Au contraire, la durée de la première voyelle aurait tendance à être réduite lorsqu'elle se trouve en présence d'une géminée, ce qui n'est pas le cas pour les voyelles qui précèdent une consonne simple. Enfin, l'analyse de la durée des géminées vs clusters, montre que la durée d'une consonne à l'intérieur du cluster semble avoir une durée bien plus longue que celle d'une géminée, quelle que soit sa nature [Turco and Braun, 2016].

Les analyses de M. Mattei et M. G. Di Benedetto présentent des traits communs avec l'étude de Turco et Braun. Dans le cadre du projet *GEMMA* (un projet qui se focalise sur la gémination, conçu par l'université de Rome, La Sapienza), l'étude se fonde sur la relation entre consonne simple vs géminée, mais avec une attention exclusive dédiée aux consonnes nasales /n/ et /m/ et aux voyelles cardinales de l'italien /a/, /i/ et /u/. Avec un corpus contrôlé, constitué de 12 mots, contenant une séquence de type VCV et sa contrepartie géminée VCCV, prononcée hors contexte trois fois par six locuteurs natifs de l'italien ayant un parler standard, l'étude a pour but l'analyse des durées des voyelles et des consonnes contenues dans ces mots. En plus des consonnes, l'étude s'intéresse aux voyelles qui les entourent. La durée des consonnes géminées est, en moyenne, deux fois plus longue que celle des non géminées ; en revanche, les voyelles qui précèdent la géminée ont une durée inférieure par rapport aux autres, tandis que, encore un fois, on observe une longueur réduite pour les voyelles qui suivent la géminée [Mattei and Benedetto, 2008].

Parmi les études dédiées à la variation régionale liée aux voyelles, citons les travaux de thèse de doctorat de [Renwick, 2012], qui mène d'abord des analyses acoustiques sur la réalisation phonémique des timbres mi-ouverts en italien. Elle montre que le contraste phonémique entre /e/ et /ɛ/, et entre /o/ et /ɔ/, est marginal, c'est-à-dire peu de paires minimales permettent d'opposer vraiment les timbres. Par conséquent, la réalisation acoustique des 4 timbres varie beaucoup avec les locuteurs et les régions et il est difficile d'établir une règle. Dans une étude plus récente, elle s'appuie sur la différence phonétique qui existe entre les locuteurs de différentes zones de l'Italie [E.L. Renwick and Ladd, 2016] s'interrogent sur l'habileté des locuteurs natifs de l'italien — originaires de régions différentes participants à l'étude — de pouvoir distinguer l'aperture, et sur l'existence d'un phénomène commun qui puisse établir une règle sur l'ouverture de ces voyelles. L'étude montre l'impossibilité de définir une règle de base qui permettrait d'appuyer les jugements perceptifs des locuteurs. Une exception est faite pour les locuteurs de Sardaigne, qui semblent ne pas avoir l'opposition phonologique d'aperture des voyelles moyennes, et qui ont une tendance à une prononciation toujours fermée.

La question de la différenciation des dialectes au sein d'une même langue se pose pour la transcription automatique. Si la question se pose moins pour l'italien, il est crucial pour des langues latines comme l'espagnol de pouvoir transcrire automatiquement des variétés dialectales qui montrent des particularités phonétiques. Par exemple, dans les travaux de [Caballero et al., 2009] dédiés à la transcription automatique de l'espagnol, la question est posée de l'entraînement des systèmes sur des données qui illustrent les parlers de l'Espagne vs de l'Amérique Latine.

1.2 Autres phénomènes de variation : natifs vs non natifs

L'augmentation du multiculturalisme et le nombre toujours plus grand de personnes bilingues ou qui interagissent avec une langue autre que leur langue maternelle, a introduit un nouveau défi en recherche sur la nécessité de savoir reconnaître une langue ou un dialecte, qu'il s'agisse de sa version écrite ou orale. De plus, la plupart des systèmes existants de reconnaissance de la parole, ne sont entraînés que sur des corpus de langue autochtone ou sur un dialecte particulier, alors qu'il est nécessaire de répondre aux besoins de différentes variantes, surtout lorsqu'il s'agit de corpus en parole spontanée. Par conséquent, l'identification automatique de l'accent non natif est devenu un sujet porteur dans un contexte de communication internationale et multilingue.

De nombreuses études aussi bien en linguistique (notamment en phonétique) qu'en traitement automatique du langage parlé, s'interrogent sur les facteurs linguistiques qui caractérisent et influencent les accents non-natifs. Ces domaines sont pertinents pour notre étude car parmi les accents identifiés dans le corpus, il y a des accents non natifs. Les études linguistiques s'intéressent aux différents niveaux qui interviennent lorsqu'on parle d'accent non natif, cependant ceux qui s'intéressent au niveau segmental sont nombreux. L'étude menée par [Vieru-Dimulescu and Boula de Mareuil, 2006] vise à caractériser différents accents étrangers en langue française à partir d'un corpus de parole spontanée, prononcé par 36 locuteurs de 6 langues étrangères (arabe, anglais, allemand, espagnol, italien et portugais) qui ont librement parlé pendant environ 5 minutes au sujet d'un texte issu du projet PFC (*Phonology of Contemporary French* [Delas-Roussarie and Durand, 2003]). L'étude est divisée en deux parties, dont la première prévoit la reconnaissance de l'accent d'origine du locuteur par des locuteurs natifs et la deuxième une analyse des traits de l'accent de chaque langue en se basant sur les formants des voyelles et sur le rythme. Les résultats montrent que, d'après les locuteurs natifs, l'accent dont on reconnaît mieux l'origine est l'arabe, tandis que le plus difficile à reconnaître est le portugais, probablement à cause de la présence de sons semblables à ceux du français en langue portugaise. Ensuite, les quatre autres accents se divisent en deux groupes : anglais-allemand et espagnol-italien. Les locuteurs français ont souvent mélangé ces deux groupes d'accents. Par la suite, l'étude s'intéresse aux différences acoustiques (le timbre des voyelles) et prosodiques (le rythme), se basant sur l'hypothèse formulée dans des études précédentes selon laquelle la qualité des voyelles prononcées par un locuteur non natif, permettrait une meilleure reconnaissance de sa langue d'origine, par rapport à la prosodie [Laurent, 1998] et [Boula de Mareuil et al., 2004]. Par l'étude du triangle vocalique produit pour chaque accent, le phonème vocalique français /y/ semble être la plus parlante en termes de caractéristiques acoustiques des non natifs. Pour ce qui concerne la prosodie, les résultats montrent qu'il existe une claire différence de rythme seulement entre les locuteurs arabes et italiens.

Une étude menée par [Bahari et al., 2018] a pour but la reconnaissance automatique de l'accent de locuteurs provenant de Russie, Inde, États-Unis, Thaïlande, Vietnam et Chine, parlant anglais dans une conversation téléphonique, pour un total de 22 heures d'enregistrement. Étant donné le faible nombre d'extraits de certains locuteurs et la nature des deux langues, l'accent vietnamien et chinois ont été associés, pour un total de 5 origines différentes. En se basant sur les résultats de trois diffé-

rents algorithmes d'apprentissage supervisé, ils montrent avoir rapporté le meilleur score de reconnaissance chez les locuteurs des États-Unis, suivis des locuteurs d'origine indienne, alors que le plus faible appartient aux locuteurs d'origine russe.

1.3 Différences inter-locuteurs : le genre

Enfin, parmi les facteurs qui influencent la variation phonétique, les caractéristiques socio-culturelles jouent un rôle également important. Le genre, entre autre, est associé aux spécificités de prononciations, responsables aussi de la performance des systèmes de reconnaissance vocale. Dans les premiers systèmes de reconnaissance de la parole, une voix féminine était plus difficile à reconnaître par les machines, parce qu'elles étaient toujours entraînées sur des voix masculines. Ensuite, les avancées technologiques ont mené à la production de systèmes plus performants tels qu'il n'existe plus de différence significative entre les taux de reconnaissance d'enregistrements produits par des hommes ou des femmes.

[Adda-Decker and Lamel, 2005] ont étudié les motivations qui mènent à un taux d'erreur différent dans la reconnaissance de la parole à partir d'enregistrements produits par des voix d'hommes et des femmes, en langue anglaise et française. Avec un corpus de 20 heures d'enregistrements issus de broadcast news radio et télévisés, ainsi que de conversions téléphoniques, elles ont pu calculer une meilleure performance dans la reconnaissance de la parole prononcée par des femmes et ceci dans les deux langues. En ce qui concerne le corpus de broadcast news, elles ont justifié ce résultat par le fait qu'il y a plus d'hommes interviewés, ce qui implique une quantité plus importante de données difficiles à transcrire. Cependant, malgré les défis acoustiques dans la reconnaissance vocale appliquée sur des données produites par une voix féminine (difficultés liées aux timbres ; à la haute fréquence fondamentale et à une plus basse fréquence des formants), à partir du deuxième corpus, issu de conversations téléphoniques, elles ont calculé des résultats semblables, en remarquant une plus grande tendance à l'effacement ou à la prononciation incomplète de certains phonèmes, ainsi qu'à une plus haute vitesse dans la production du discours chez les hommes. Ces différences se traduisent par la réduction temporelle des phonèmes et par une mauvaise détection de ces phonèmes par le système.

La tendance des femmes vers une prononciation plus claire et standard serait probablement due à des causes socio-linguistiques, en raison du rôle joué par les femmes dans l'éducation et dans la transmission de la langue à l'enfant [Adda-Decker and Lamel, 2005].

Deuxième partie

Corpus et Méthode

CORPUS

2.1 Pré-traitements

Présentation Notre corpus se compose de 150 enregistrements audio d'une durée comprise entre 1 et 30 minutes, de journaux radiodiffusés (broadcast news) provenant de la chaîne radio italienne RAI, entre les années 1992 et 1998 pour un total de 5 heures environ. Le corpus a été collecté et segmenté automatiquement par le laboratoire LIMSI. La segmentation automatique a été effectuée avec le système automatique décrit dans [Matassoni et al., 2012]. Le corpus est issu d'un projet antérieur mené en collaboration avec l'IRST. L'alignement forcé est le processus pendant lequel un système de reconnaissance vocale réalise la correspondance entre un signal entrant et sa transcription en posant des frontières entre les phonèmes et en leur associant le symbole phonétique approprié. Cette démarche suppose une transcription orthographique préalable réalisée par un humain [Vasilescu, 2018].

2.2 Annotation de l'origine géographique des locuteurs

Nous avons annoté semi-manuellement la totalité du corpus et avons déterminé l'origine géographique probable des locuteurs selon la répartition classique en Nord, Centre, Sud, Standard, Île de Sardaigne, accents non natifs.

En tant que journal radio les locuteurs couvrent différentes régions de l'Italie. Par contre, s'agissant d'une radio publique et nationale, il n'est pas étonnant d'entendre un grand nombre de locuteurs dont l'accent est peu marqué ou pas perceptible du tout, comme le montrent les chiffres ci-dessous :

- 58,3% de locuteurs définis comme standard ou dont l'accent n'a pas pu être reconnu
- 18,0% de locuteurs provenant du centre de l'Italie
- 16,9% de locuteurs provenant du nord de l'Italie
- 6,79% de locuteurs provenant du sud de l'Italie
- 1,61% de locuteurs d'origine étrangère
- 0,76% de locuteurs d'origine sarde

Nous constatons qu'au-delà des locuteurs que nous pouvons définir comme parlant un italien standard, les locuteurs du nord et du centre sont prépondérants. Nous émettons l'hypothèse que la très faible présence de locuteurs du sud est due à la localisation géographique de la capitale, Rome, dont le parler est considéré comme un parler du centre, et à l'importance économique des villes du nord de la péninsule.

2.3 Reconnaissance vocale

Le système de reconnaissance vocale se fonde sur un jeu de phonèmes correspondant globalement au système phonémique de la langue tel que décrit par les linguistes. Le jeu de phonèmes peut être néanmoins simplifié par rapport aux descriptions théoriques, notamment lorsque certaines oppositions ne sont pas très fonctionnelles. C'est le cas par exemple de l'opposition d'aperture des voyelles mi-ouvertes en italien /e/ et /o/. Si la description du système phonémique de l'italien [Berruto, 2010] prévoit une opposition phonémique entre /e/ et /ɛ/ d'une part, et entre /o/ et /ɔ/ d'autre part, dans la pratique langagière, l'opposition est peu productive et les paires minimales peu nombreuses. En raison de ce risque réduit de confusion lexicale, il n'est pas nécessaire d'en tenir compte pour la reconnaissance vocale. Des incohérences peuvent venir des mots étrangers, dont la phonétisation automatisée produite pour construire le dictionnaire nécessaire pour la reconnaissance vocale, se base sur les règles propres à l'italien. Ainsi le mot français *forfait*, utilisé aussi en italien pour indiquer un prix approximatif à payer, est bien prononcé avec un /ɛ/ final, mais il présente les trois dernières lettres *a*, *i* et *t* en tant que phonèmes dans sa segmentation, alors qu'elles ne sont pas prononcées selon les règles de prononciation de l'italien. Cependant, l'équivalence quasi-systématique graphème/phonème de l'italien, dont l'orthographe est phonétique (par opposition au français qui est étymologique), rend ce type de segmentation assez fiable.

Les éléments de travail à disposition au début de l'étude étaient donc : une série d'enregistrements audio et une liste de fichiers tabulaires, dont un extrait est fourni sur la figure 2.1.

FIGURE 2.1 – Extrait du fichier tabulaire avec les informations nécessaires à la segmentation automatique

```

3.878 0.12 g GRuno .[g]i GRR920828_1900 f0 male
3.998 0.08 i GRuno g[i]e GRR920828_1900 f0 male
4.078 0.14 e GRuno i[e]R GRR920828_1900 f0 male
4.218 0.08 R GRuno e[R]e GRR920828_1900 f0 male
4.298 0.03 e GRuno R[e]u GRR920828_1900 f0 male
4.328 0.19 u GRuno e[u]n GRR920828_1900 f0 male
4.518 0.03 n GRuno u[n]o GRR920828_1900 f0 male
4.548 0.21 o GRuno n[o]s GRR920828_1900 f0 male
4.758 0.26 s sera o[s]e GRR920828_1900 f0 male
5.018 0.16 e sera s[e]r GRR920828_1900 f0 male
5.178 0.03 r sera e[r]a GRR920828_1900 f0 male
5.208 0.14 a sera r[a]. GRR920828_1900 f0 male
5.348 0.37 . [silence] a[.]. GRR920828_1900 f0 male
9.49 0.11 i i .[i]t GRR920828_1900 f0 male
9.60 0.03 t titoli i[t]i GRR920828_1900 f0 male
9.63 0.07 i titoli t[i]t GRR920828_1900 f0 male
9.70 0.07 t titoli i[t]o GRR920828_1900 f0 male
9.77 0.05 o titoli t[o]l GRR920828_1900 f0 male
9.82 0.04 l titoli o[l]i GRR920828_1900 f0 male
9.86 0.08 i titoli l[i]. GRR920828_1900 f0 male
9.94 0.25 . [silence] i[.]. GRR920828_1900 f0 male

```


Ces fichiers se composent de plusieurs colonnes, spécifiant : le temps de début du phonème (le time code); sa durée; le phonème lui-même; le mot dans lequel le phonème apparaissait; le phonème accompagné par le phonème précédent et le suivant; le nom du fichier dans lequel ils ont été prononcés; ainsi que le sexe du locuteur.

En combinant les fichiers audio, convertis en format *.wav*, avec les tabulaires précédents, à l'aide d'un script PRAAT, il a été possible d'obtenir de manière automatique, pour chaque enregistrement, un *TextGrid* contenant la représentation spectrale de l'audio suivi par quatre bandes (tiers) utilisées pour l'étiquetage automatique, spécifiant respectivement : le phonème prononcé; le mot dans lequel le phonème est prononcé; le sexe du locuteur et une toute dernière couche, la quatrième, était destinée à l'annotation semi-manuelle du lieu de provenance du locuteur qui a été rajoutée a posteriori. L'image 2.2 en montre un extrait.

FIGURE 2.2 – Extrait du fichier audio segmenté automatiquement

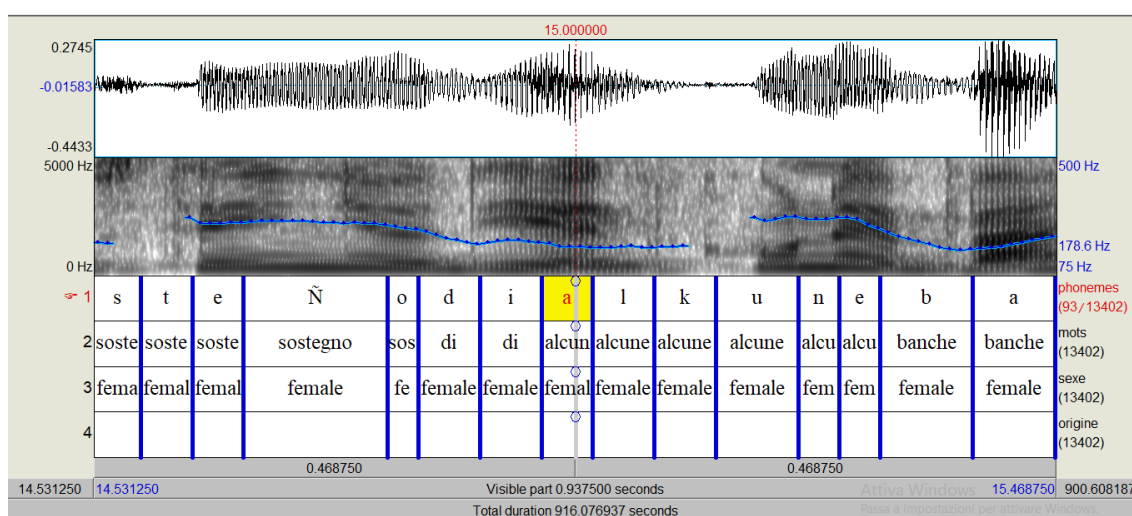


TABLE 2.1 – Description quantitative du corpus GRR

Mots	Mots Distincts	Audio	Consonnes	Voyelles
256 008	21 564	5h environ	73.5% dont 6% géminées	26.5%

MÉTHODE

3.1 Annotation des enregistrements par origine géographique

À partir du corpus oral et de sa transcription manuelle phonétisée, nous avons réalisé une annotation semi-manuelle pour spécifier l'origine du locuteur sur chacun des phonèmes prononcés. Une fois identifiée l'origine, chaque premier phonème prononcé par chaque nouveau locuteur a été annoté selon la zone géographique d'origine et, de manière automatique, la même annotation a été reportée aux autres phonèmes du même locuteur. Notre travail d'annotation repose sur les classes suivantes :

- *NI* pour le nord
- *CI* pour le centre
- *SudI* pour le sud
- *SI* pour l'italien standard
- *Sardegna* pour la Sardaigne
- et *EI* pour les locuteurs non natifs

3.2 Extraction des phonèmes, durées et positions

Aucune information concernant la région d'origine du locuteur n'était fournie, c'est pourquoi la distinction entre variante méridionale et variante méridionale extrême, dont on a l'habitude de parler au sujet des accents, n'apparaît pas ici. De plus, il n'a pas été possible de prendre en compte toutes les sources régionales identifiées à l'intérieur du corpus. Comme on a pu l'observer, les locuteurs de Sardaigne et les non natifs n'étaient pas assez nombreux, ils n'ont été pris en compte que de manière très marginale dans nos analyses. Comme précisé, notre étude se fonde sur des analyses acoustiques prenant en compte la nature des consonnes et des voyelles de l'italien. Nous avons utilisé le logiciel PRAAT pour extraire automatiquement tous les phonèmes ainsi que leur durée. La durée des mots et la position du phonème à l'intérieur du mot ont aussi été exploitées. La figure 3.1 présente un extrait du fichier tabulaire de sortie.

Nous avons complété les informations disponibles grâce au logiciel de statistique RStudio, pour calculer les moyennes des durées. Pour analyser et détecter automatiquement la source régionale d'un extrait, nous avons repris les données et la méthode décrite ci-dessus. Nous nous sommes à nouveau basés sur le calcul de la durée comme mesure permettant la caractérisation des voyelles du corpus par rapport à leur position dans le mot.

FIGURE 3.1 – Extrait du fichier tabulaire employé pour effectuer les analyses acoustiques sur les phonèmes consonantiques

fichier	timing	phoneme_precedent	phoneme	phoneme_successif	mot	longueur_mot	posit
GRR920828_1900.TextGrid	3.878	_	g i	GRuno 5	debut	120 120 SI	male
GRR920828_1900.TextGrid	3.998	g	i e	GRuno 5	central	80 200 SI	male
GRR920828_1900.TextGrid	4.078	i e	R	GRuno 5	central	140 340 SI	male
GRR920828_1900.TextGrid	4.218	e R	e	GRuno 5	central	80 420 SI	male
GRR920828_1900.TextGrid	4.298	R e	u	GRuno 5	central	30 450 SI	male
GRR920828_1900.TextGrid	4.328	e u	n	GRuno 5	central	190 640 SI	male
GRR920828_1900.TextGrid	4.518	u n	o	GRuno 5	central	30 670 SI	male
GRR920828_1900.TextGrid	4.548	n o	s	GRuno 5	final	210 880 SI	male
GRR920828_1900.TextGrid	4.758	o s	e	sera 4	debut	260 260 SI	male
GRR920828_1900.TextGrid	5.018	s e	r	sera 4	central	160 420 SI	male
GRR920828_1900.TextGrid	5.178	e r	a	sera 4	central	30 450 SI	male
GRR920828_1900.TextGrid	5.208	r a	_	sera 4	final	140 590 SI	male
GRR920828_1900.TextGrid	5.348	a _	i	[silence]	9 debut	4142 0 SI	male
GRR920828_1900.TextGrid	9.490	_	i t	i l	debut	110 0 SI	male
GRR920828_1900.TextGrid	9.600	i t	i	titoli 6	debut	30 30 SI	male
GRR920828_1900.TextGrid	9.630	t i	t	titoli 6	central	70 100 SI	male
GRR920828_1900.TextGrid	9.700	i t	o	titoli 6	central	70 170 SI	male
GRR920828_1900.TextGrid	9.770	t o	l	titoli 6	central	50 220 SI	male
GRR920828_1900.TextGrid	9.820	o l	i	titoli 6	central	40 260 SI	male
GRR920828_1900.TextGrid	9.860	l i	_	titoli 6	final	80 340 SI	male

3.3 Détection automatique de l'origine des locuteurs

En raison de l'absence de différenciation entre phonèmes ouverts et fermés dans la segmentation forcée, nous nous sommes appuyés sur les formants pour détecter automatiquement l'origine du locuteur. La figure 3.2 montre un extrait du fichier employé pour entraîner les algorithmes d'apprentissage dans WEKA.

FIGURE 3.2 – Extrait du fichier employé pour la détection de l'origine des locuteurs

```
@relation accent
@attribute voyelle string
@attribute mot string
@attribute position string
@attribute duree numeric
@attribute f1 numeric
@attribute f2 numeric
@attribute f3 numeric
@attribute f4 numeric
@attribute val_pitch numeric
@attribute min_pitch numeric
@attribute max_pitch numeric
@attribute harmonicity numeric
@attribute harmonicity_mean numeric
@attribute class {NI,CI,SudI}
@data
i,il,debut,60,387.1144502504,2629.62063187461,3175.52573337427,4265.70336262902,213.129784870848,213.129784
e,tema,central,70,568.726067406489,1706.26213721723,3080.31718650371,4283.64942221465,270.465765650676,265.
a,tema,final,60,745.212150754812,1728.20643158687,2967.18747266574,3947.1373138155,250.510642425951,250.323
e,del,central,50,477.123305451407,2010.57249793165,3038.8301878239,4420.90135040378,223.690293469082,204.94
o,convegno,central,50,498.86437472448,1185.81472832853,2728.11003002557,3984.77841231841,209.911821978709,2
e,convegno,central,110,485.354553727289,2245.46952740967,3062.27881001242,4247.63436805118,200.236296211806
o,convegno,final,30,611.981388530173,1553.77527849548,2833.48720214522,3960.17124927593,306.614031027949,20
a,Alvaro,debut,170,619.29083159668,2029.56545868208,2880.34178820743,4110.26607680994,236.647040661087,226.
a,Alvaro,central,110,799.742519481631,1643.40381024765,2827.33859057534,3811.90307580608,201.833841170191,1
o,Alvaro,final,30,567.714444981391,1676.93250487897,2676.92852462921,3798.65330922623,200.692427944406,197.
e,e,debut,30,573.084104743709,1700.34210943991,2718.70784841926,3905.04435578596,197.657855461079,196.22802
e,l,debut,30,570.149358266651,1780.27033320801,2710.55607889266,3900.98969151114,194.641350329626,192.01593
e,l,final,30,617.250240747095,1834.68144255274,3121.3562780491,4481.582064276,184.100333350752,181.48465656
```

3.4 Influence du genre sur l'identification de l'origine

Notre dernière étude, dédiée à la détection de l'accent compte tenu du sexe du locuteur, a été possible en se basant, d'un côté, sur une estimation des annotations des origines géographiques des locuteurs des deux sexes, et d'un autre côté sur la même méthode employée pour les systèmes de détection avec le logiciel WEKA en divisant le corpus en deux : uniquement des locuteurs de sexe masculin et uniquement des locutrices de sexe féminin.

Troisième partie

**Analyses acoustiques et
variantes régionales**

MOTIFS DE VARIATION RÉGIONALE DU SYSTÈME CONSONANTIQUE

Présentation Le système consonantique de l'italien standard se compose de 21 phonèmes consonantiques. En termes de mode d'articulation, l'italien a 6 classes (occlusive; fricative; affriquée; nasale; vibrante et latérale) et deux glides, ou semi-voyelles (/j/ comme dans le mot "piano" et /w/, dans le mot "buono"). En termes de lieu d'articulation, l'italien possède 6 lieux d'articulation (bilabiale; labio-dentale; dentale; alvéolaire; palatale et vélaire). Sur les 21 phonèmes consonantiques, 20 présentent des géménées. Par géménéation, on entend la prononciation prolongée d'une consonne, anticipée par une pause, dont la durée peut varier par rapport à la nature de la consonne [Elinor M. Payne, 2005].

Dans le tableau 4.1 les géménées de chaque consonne sont représentées à droite du symbole pour la consonne non géménéée.

TABLE 4.1 – Table des phonèmes de la langue italienne

	Bilabial	Labiodent.	Dentale	Alvéolaire	Palatale	Vélaire
Occlusive	p/p : b/b :		t/t : d/d :			k/k : g/g :
Affriquée				ts/ts : dz/dz :	tʃ/tʃ : tʒ/tʒ :	
Nasale	m/m :		n/n :		ɲ/ɲ :	
Vibrante				r/r :		
Fricative		f/f : v/v :		s/s : z	ʃ/ʃ :	
Latérale				l/l :	ʎ/ʎ :	

Selon les régions, la prononciation des consonnes peut varier. Ainsi, en ce qui concerne la zone septentrionale, les particularités suivantes ont été observées [Bertinetto and Loporcaro, 2005] : sonorisation de la fricative alvéolaire intervocalique, manque du redoublement des palatales, et absence du redoublement syntaxique¹.

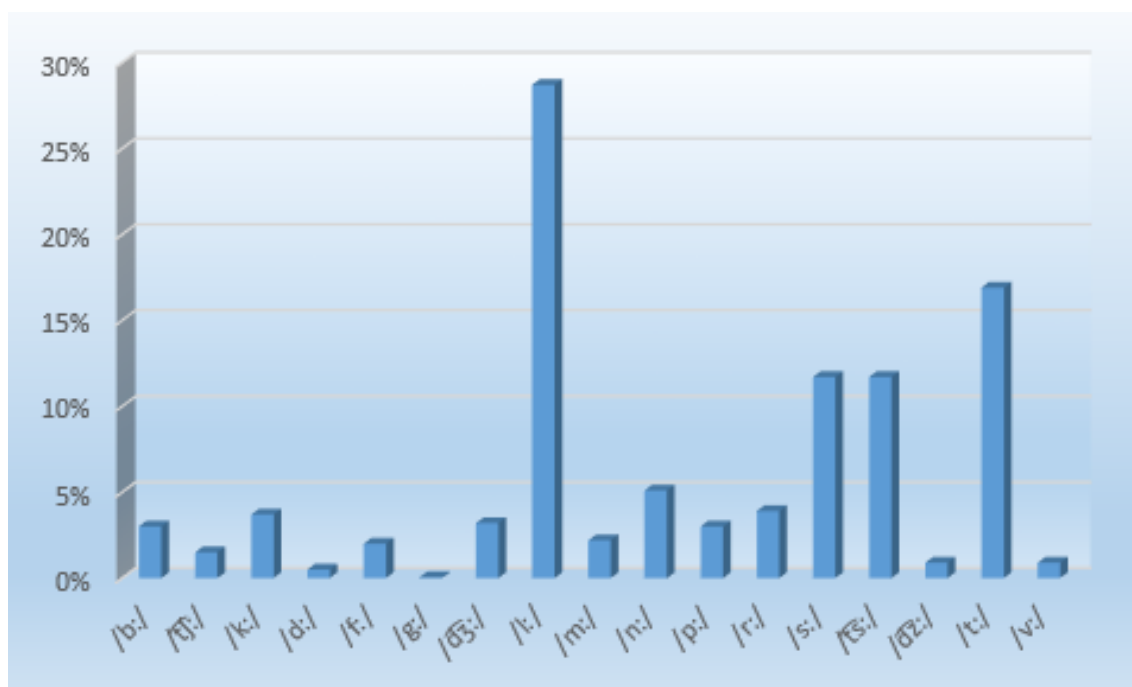
Dans le cas du centre, on assiste à un affaiblissement du /r/ géméné, l'affrication de la fricative /s/ qui devient /ts/, le renforcement du /b/ intervocalique et encore une fois la perte de la composante palatale dans le phonème latéral /ʎ/ qui est prononcé comme /j/.

1. L'absence de redoublement syntaxique renvoie à la prononciation allongée de la première consonne d'un mot, lorsqu'il est précédé par une voyelle à la fin du mot précédent, ce qui est presque toujours le cas en italien, étant donnée la grande majorité de mots qui se terminent par une voyelle.

Dans la zone méridionale, on observe la gémination de certaines consonnes, en particulier des occlusives, en position intervocalique ; la sonorisation des occlusives sourdes post-nasales et une prononciation rétroflexe des syllabes *tr*, *dr* et *ll* comme déjà observé par [Bonomi et al., 2010].

Étude des géminées Comme indiqué précédemment, la quasi totalité des phonèmes consonantiques de l'italien peut avoir une réalisation phonologique géminée, comme dans le cas des mots *fato* ("le destin") vs *fatto* ("un fait") ou *pala* ("pelle") vs *palla* ("ballon"). La figure 4.1 présente les pourcentages de consonnes géminées que nous avons identifiées dans notre corpus.

FIGURE 4.1 – Représentation des occurrences des consonnes géminées dans le corpus



Nous nous sommes également intéressés à l'occurrence des consonnes géminées dans le corpus et à leur durée par rapport aux consonnes seules. Nous pouvons noter que les consonnes géminées représentent 6% des données. Parmi les consonnes, la latérale /l/ est la plus fréquente suivie par l'occlusive dentale /t/, comme on le voit sur la figure 4.1. Enfin en termes de durée, nos données confirment des tendances déjà observées à partir de corpus plus restreints, à savoir que les consonnes simples sont plus courtes que les consonnes géminées, et ce pour toutes les consonnes indépendamment de la région géographique (tableaux 4.2 et 4.3).

Si la gémination a un statut phonologique, même si l'on enregistre des motifs de variation régionale, d'autres motifs acoustiques touchant la réalisation des consonnes ne relèvent pas d'oppositions phonémiques mais sont très caractéristiques des différences géographiques. C'est notamment le cas du redoublement consonantique (en italien, *radoppiamento fonsintattico*).

Dans la suite de ce chapitre, nous nous intéressons aux phénomènes qui comportent une gémination impropre et non phonologique, en l'occurrence, le *renforcement phono-syntaxique* et le *redoublement intervocalique*.

TABLE 4.2 – Moyenne des durées (ms) des phonèmes géminés par groupes de consonnes

Zone	occlusives	fricatives	affriquées	nasales	latérales	vibrantes
NI	93.3	118.39	108.5	99.2	65.1	85.25
CI	96.4	109.6	120.2	95.9	62.94	85.74
SudI	95.5	112.8	110.4	100.6	65.69	83.52
SI	92.3	108.5	110.3	91	65.1	84.46

TABLE 4.3 – Moyenne des durées (ms) des phonèmes consonantiques simples par groupes de consonnes

Zone	occlusives	fricatives	affriquées	nasales	latérales	vibrantes
NI	64	69.1	91.3	60.3	79.7	44.5
CI	63.8	70.9	95	59.7	43.5	44.3
SudI	63.8	70.7	95.9	61	71.7	44.4
SI	63.4	72.8	96	58.5	49.2	44.6
Moyenne	63.7	70.8	94.55	59.8	61	44.4

4.1 Renforcement phono-syntaxique

4.1.1 Présentation

Par *renforcement phono-syntaxique* on entend un allongement se manifestant par une durée plus longue du premier phonème consonantique d'un mot, lorsqu'il est précédé par une voyelle, à la fin du mot précédent. Il est souvent perceptible à l'occurrence de prépositions ou pronoms monosyllabiques suivies par un nom débutant par une consonne, comme “*a*”, “*va*”, “*te*”, etc. [Bertinetto and Loporcaro, 2005]. Ce phénomène est à l'origine de mots et de syntagmes homophones. C'est le cas, par exemple, du verbe *accoppiare* (“coupler”) et de l'expression *a coppia* (“par paires”) ou du verbe *alettare* (“aliter”) et l'expression *a letto* (“au lit”), entre autres.

Une explication concernant le phénomène de redoublement serait la chute dans la graphie de l'italien, de la dernière consonne dans les prépositions latines comme *ad* ou *et*, mais qui serait restée dans la prononciation et causerait ainsi le renforcement phono-syntaxique.

Ce phénomène concerne plutôt les régions centrales et méridionales et moins celles du nord [Bonomi et al., 2010]. Cette différence serait justifiée par l'absence de consonnes géminées dans les dialectes de l'Italie du nord, en contact avec l'espagnol et le français [Bertinetto and Loporcaro, 2005]. Cela empêcherait probablement une plus longue et marquée prononciation des consonnes en début de mot comme c'est le cas dans certaines géminées phonologiques.

4.1.2 Analyse acoustique

Cette première analyse s'est focalisée sur la durée des consonnes susceptibles d'être affectées par le phénomène de redoublement à l'initiale de mot. À cet effet, nous avons utilisé un fichier tabulaire contenant des informations telles que la durée,

la position (définie comme *début*, *central* ou *final*), le phonème précédent, le phonème suivant et l'origine. Ce fichier résulte de la segmentation automatique qui contient les bornes temporelles pour chaque phonème et il a été enrichi afin d'être exploité à l'aide d'un script PRAAT. En ce qui concerne l'extraction des durées, elles sont calculées en millisecondes, les chiffres résultantes ont été multipliés par 1000, de manière à pouvoir rendre visibles les très petits nombres produits par le calcul.

FIGURE 4.2 – Extrait du script PRAAT pour extraire la durée de chaque phonème

```
debut = Get starting point: 1, y
fin = Get end point: 1, y
duree = round((fin - debut)*1000)
if (mot$ != mot_precedent$)
  dureeMot = 0
  position$= "debut"
elseif (mot$ != mot_successif$)
  position$= "final"
else
  position$= "central"
endif

if (mot$ == mot_precedent$ or mot$ == mot_successif$)
  dureeMot = dureeMot + duree
endif
```

Ensuite, ces informations ont été exploitées à l'aide du logiciel de statistique *RStudio*. Le logiciel est doté d'une interface graphique sur laquelle il est possible de charger des données de grandes dimensions sous forme d'objets, ainsi que d'écrire et d'exécuter des scripts R. Suite au chargement du fichier résultant du script PRAAT, avec toutes les informations nécessaires pour l'analyse des consonnes du corpus étudiée, des objets booléens ont été créés correspondant à chaque entrée, de manière à savoir quels phonèmes étaient bien des voyelles ou des consonnes et lesquels des phonèmes consonantiques étaient des géminées.

Le tableau 4.4 synthétise les résultats sous forme de moyennes de durée. De manière générale, on observe l'existence du renforcement phono-syntaxique pour tous les locuteurs, peu importe leur origine géographique. Nos résultats montrent que ce phénomène décrit dans la littérature d'un point de vue syntaxique et à partir de corpus de laboratoire [Elinor M. Payne, 2005] caractérise la quasi-totalité des variantes de l'italien. De l'analyse de la base de données que nous avons constituée, il ressort que la longueur d'un phonème consonantique placé en début de mot et précédé par une voyelle à la fin du mot précédent, est de 98 ms ; tandis qu'une consonne se trouvant en milieu ou en fin de mot aurait une durée moyenne de 64 ms, indépendamment de l'origine géographique du locuteur.

TABLE 4.4 – Moyenne des durées des phonèmes en position de renforcement phono-syntaxique

Zone	Occlusives	Affriquées	Fricatives	Nasales	Latérales	Vibrantes
NI	71.9	101.2	80.8	69.8	55.7	50.3
CI	72.4	95.9	75.5	66.6	63.2	52.8
SudI	71.5	94.3	76.5	67.3	58.1	52
SI	72.8	100.6	80.4	67.6	63.1	52.4
Moyenne	72.15	98	78.3	67.8	60	51.8

Nous observons que le redoublement concerne moins les régions du nord, mais que les données de ces régions sont aussi influencées par cette gémination impropre, concernant cependant des consonnes de nature différente par rapport aux locuteurs du centre et du sud. On relève ainsi un renforcement plus prononcé des occlusives pour les locuteurs du centre et un plus faible pour ceux du nord, mais l'envers pour les affriquées et les nasales.

La moyenne de durée élevée pour les locuteurs du centre semble être due à certaines occlusives qui sont particulièrement touchées par le phénomène et notamment les labiales, comme décrit dans le travail de [Bonomi et al., 2010].

4.2 Redoublement Intervocalique

4.2.1 Présentation

Nous avons étendu cette analyse à un autre contexte comparable, susceptible d'être à l'origine d'une gémination non phonologique. Il s'agit du contexte intervocalique. Ainsi, dans le premier cas, le phénomène d'allongement consonantique se produit au-delà de la frontière du mot, tandis que dans le cas traité dans cette section le phénomène se produit en contexte intra-lexical. La gémination phonologique est très fréquente en langue italienne et elle est presque toujours observable en position intervocalique.

Comme pour le renforcement phono-syntaxique, probablement pour les mêmes raisons, ce phénomène semble également être plus souvent répandu parmi les locuteurs du centre et du sud que dans les régions du nord. De plus, il a plutôt tendance à se produire lorsqu'il s'agit de consonnes occlusives, en particulier le /b/ et il est possible de l'entendre lorsque le /b/ est précédé par la liquide /l/ ou la vibrante /r/, comme dans le mot *libro*, qui peut donc être prononcé /lib :ro/ [Bertinetto and Loporcaro, 2005].

La raison de ce redoublement, en particulier celui du /b/, s'explique par deux causes. La première cause est d'ordre acoustique, le phonème /b/ est difficilement prononcé sans redoublement dans les paroles des zones du sud. La deuxième cause renvoie à la volonté d'insister sur le mot contenant ce phonème [Argenziano and de Filippis, 2010].

Pour l'analyse du redoublement intervocalique, nous avons fait appel au même format de données, en prenant en compte la position intra-lexicale des consonnes en contexte post et pré-vocalique. Nous nous sommes focalisés sur les occlusives, étant le type de phonème sur lequel il a plus tendance à se produire.

4.2.2 Analyse acoustique

Le tableau 4.5 montre les moyennes de durée des consonnes en position intralexicale intra vocalique.

TABLE 4.5 – Moyenne des durées (ms) des phonèmes en position intervocalique

Zone	/b/	/p/	/k/	/g/	/t/	/d/
NI	67.26	89.97	73.86	69.83	69.1	53.77
CI	78.50	76.03	63.90	64.72	67.90	53.42
SudI	81.61	85.55	68.1	64.65	62.32	51.75
SI	72.21	82.16	71.91	67.49	65.59	55.56

TABLE 4.6 – Moyenne des durées (ms) des phonèmes occlusifs géminés

Zone	/b :/	/p :/	/k :/	/g :/	/t :/	/d :/
NI	74.90	101.61	105.3	110	95.48	72.75
CI	90.05	101.75	99.5	100	96.78	90.66
SudI	90.9	106.66	109.4	80	102.53	84.16
SI	85.01	94.93	99.6	90	94.99	89.89

TABLE 4.7 – Moyenne des durées (ms) des phonèmes occlusifs simples

Zone	/b/	/p/	/k/	/g/	/t/	/d/
NI	62.7	69.4	72	61.7	66.9	51.4
CI	69.2	67.9	71.3	59.1	64.4	51.3
SudI	70.5	66	70.8	60.1	65.3	50.1
SI	65.4	65.5	70.8	60.8	65.7	52.4

Les résultats montrent que dans le cas de l'occlusive labiale /b/, ce sont bien les locuteurs du centre et du sud qui montrent un redoublement. Cela se traduit dans une durée plus importante du phonème. Les locuteurs du nord semblent produire des durées plus longues dans le cas des dentales et vélaires occlusives. La raison pourrait peut-être se trouver, encore un fois, dans le dialecte d'origine des zones prises en considération. En effet, les occlusives /d/ et /g/, entre autres types de phonèmes, subissent un affaiblissement consonantique ou *lénition*, pour utiliser la dénomination de Laura Bafle, dans les parlers du sud et du centre sud [Bafle, 2014].

4.3 Consonnes, voyelles et géminées

4.3.1 Présentation

Cette troisième expérience joue un rôle de liaison, pour l'introduction du troisième chapitre de cette étude, dédiée à l'analyse des voyelles en langue italienne et leur analyse à travers ce corpus. Dans cette expérience, nous prenons en compte aussi bien les consonnes géminées que les voyelles, ainsi que la longueur des voyelles en

présence des géminées. Dans cette expérience, nous ne faisons pas de distinction entre les différents accents de l'italien, comme on l'a fait jusqu'ici et comme on fera par la suite. Cette recherche est inspirée de l'étude de G.Turco et B.Braun présentée dans l'état de l'art.

En raison d'un faible nombre de paires minimales présentes dans le corpus, nous ne prendrons pas en compte la longueur de la première consonne, mais uniquement la longueur d'une voyelle en position centrale, en présence ou pas de consonnes géminées.

4.3.2 Analyse acoustique

TABLE 4.8 – Moyenne des durées (ms) des voyelles en présence de consonnes géminées

Voyelles	cvc/vvc/cvv	Gv	#vG/vG	GvG
	66.7	58.3	61.6	66.9

Les résultats confirment les tendances observées par G.Turco et B. Braun, qui ont observé une durée réduite de la voyelle en position centrale, lorsqu'elle est suivie d'une géminée (61.6 *millisecondes*), tandis qu'elle est plus allongée lorsqu'elle est suivie d'un phonème autre qu'une double consonne (63.1). Les voyelles en contexte post et pré-géminée ont la durée la plus importante.

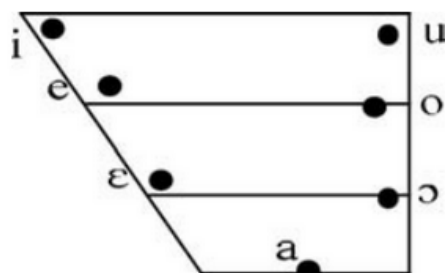
En présence non seulement de deux consonne, mais de deux géminées, le locuteur aurait une plus forte tendance à prolonger le son vocalique du milieu. Tandis que les valeurs de durée les plus réduites sont observées lorsqu'un phonème vocalique est prononcé en début de mot et est suivi par une géminée. En effet, dans de la parole spontanée, le locuteur pourrait avoir tendance à s'appuyer d'avantage sur la dernière voyelle précédant le mot, sachant que la plupart des mots en langue italienne se terminent par une voyelle, et gommer ainsi la première voyelle du mot par la présence d'une géminée.

LE SYSTÈME VOCALIQUE

5.1 Espace vocalique en italien

Le système vocalique de l'italien possède 7 voyelles phonologiques (et 5 graphèmes) : /i/, /e/, /ɛ/, /a/, /ɔ/, /o/, /u/. La figure 5.1 présente le triangle vocalique de l'italien, issu de [Bertinetto and Loporcaro, 2005].

FIGURE 5.1 – Triangle vocalique de la langue italienne



Ces timbres sont phonologiques car ils s'opposent dans des paires minimales. On appelle paires minimales les couples de mots avec signification différente, ayant la même graphie, mais une prononciation différente. C'est le cas des mots *pesca* ("pêche", l'action de pêcher) prononcé /peska/ ou *pesca* ("pêche", le fruit) prononcé /peska/ ou bien *botte* ("tonneau") prononcé /bot ɛ/ ou *botte* ("coups") prononcé /bot ɛ/. Les timbres d'aperture moyenne antérieurs (/e/ /ɛ/) et postérieurs (/o/ /ɔ/) sont définis par certains auteurs comme illustrant un contraste phonologique marginal, dans la mesure où il y a peu de paires minimales qui les opposent [E.L.Renwick and Ladd, 2016]. La figure 5.2 présente les occurrences des voyelles dans notre corpus.

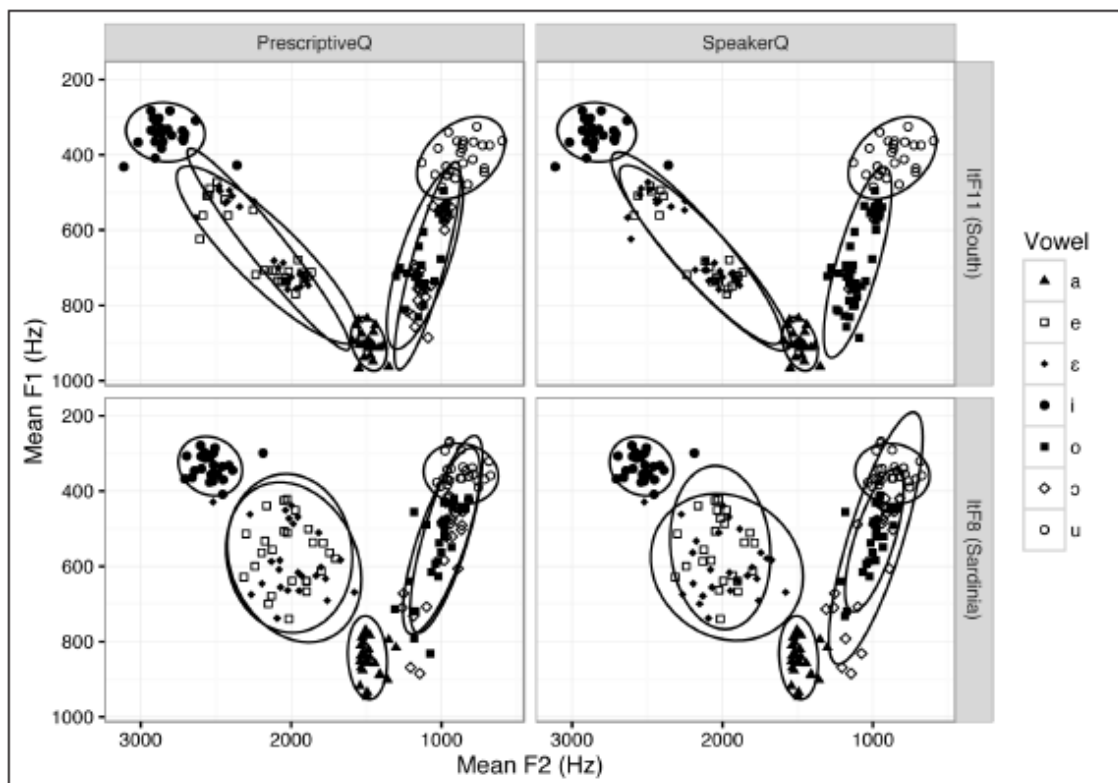
FIGURE 5.2 – Représentation des occurrences des voyelles dans le corpus



La phonétique de l'italien s'est penchée plus en détail sur le statut des voyelles d'aperture moyenne dont le nombre de paires minimales est réduit, à savoir /e/, /ɛ/ d'une part, et /o/, /ɔ/ d'autre part. Il a été noté ainsi qu'une variation dans la réalisation des voyelles /e/, /o/ avec un timbre plus ouvert ou plus fermé est dépendant de l'origine géographique du locuteur. On avance ainsi que dans les parlers du nord il y a une plus faible préférence à distinguer le timbre ouvert vs fermé, privilégiant plutôt la variante ouverte. Cependant on note une forte variabilité interlocuteur, parfois même ceux d'une même ville, et ce aussi bien au niveau acoustique que perceptif [Renwick, 2018].

Nous constatons également que le contraste /e/ vs /ɛ/ a une valeur phonémique moins importante que celui du /o/, /ɔ/. Ce qui signifie que certains mots de la langue peuvent être prononcés de manière complètement aléatoire pour ce qui est de la sélection de /e/ vs /ɛ/, mais ce ne sera pas le cas pour les réalisations /o/, /ɔ/. Les locuteurs originaires de Sardaigne par contre semblent ne pas avoir une complète distinction entre les deux voyelles d'aperture moyenne, comme le montre la figure 5.3, où les réalisations des locuteurs de Sardaigne sont comparées avec celles des locuteurs du sud [E.L.Renwick and Ladd, 2016].

FIGURE 5.3 – Triangle vocalique issu des prononciations des voyelles des locuteurs d'origine sarde



5.2 Espace vocalique et particularités régionales

Dans cette section nous étudions les particularités acoustiques des voyelles de l'italien considérées en fonction de la variante régionale. Nous avons utilisé un script PRAAT pour extraire les paramètres acoustiques caractérisant les 5 voyelles telles que encodées par le système de reconnaissance vocale (pour rappel, les oppositions /o/ vs /ɔ/ et /e/ vs /ε/ ne sont pas prises en compte dans le jeu de phonèmes mis à disposition). Les paramètres acoustiques extraits sont les suivants :

- les formants F1, F2, F3 et F4
- les valeurs moyenne, minimale et maximale de la fréquence fondamentale
- l'harmonie au milieu du phonème et la moyenne de l'harmonicité à l'intérieur du phonème

Au moyen de l'outil statistique RStudio, nous avons calculé l'écart-type des formants F1 et F2 de chaque voyelle, ainsi que leur moyenne. Sachant que dans l'annotation automatique les voyelles moyennes /o/ vs /ɔ/, et /e/ vs /ε/ ne sont pas distinguées (nous avons un seul timbre encodé comme timbre fermé /e/ et /o/), ce calcul a été effectué 10 fois (écart-type et moyenne pour chacune des 5 voyelles).

$$F1 > x(F1) + 3 \times \sigma(F1) \quad (5.1)$$

$$F2 > x(F2) + 3 \times \sigma(F2) \quad (5.2)$$

Les données dont la F1 et la F2 ne correspondaient pas à cette équation, ont été écartées de l'ensemble en diminuant les données de 1,6%. Grâce à ce processus, un taux d'erreur, dû à l'extraction des traits acoustiques, correspondant à chaque voyelle et à chaque type de parler, a pu être calculé. Les résultats sont présentés dans le tableau 5.1.

TABLE 5.1 – Tableau des taux d'erreur produit par l'extraction automatique des formats des voyelles

Zone	/e/		/o/		/a/		/i/		/u/	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
NI	2.14%	9.04%	1.8%	7.4%	1.8%	16.6%	0.8%	18.8%	1.7%	1.1%
CI	1.7%	12.3%	1.4%	4%	1.5%	20.2%	0.6%	22%	1.5%	1.3%
SudI	4.3%	13.1%	1.6%	3.7%	2.6%	27.3%	0.9%	27.1%	1.8%	0.8%

Une fois obtenues des valeurs pertinentes, nous avons procédé à la représentation des espaces vocaliques sous forme de triangles ayant pour axes les deux premiers formants. Les phonèmes reliés par des pointillés correspondent aux locutrices, tandis que les phonèmes reliés par des lignes correspondent aux locuteurs.

FIGURE 5.4 – Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlers du Nord (pointillés = femmes, lignes = hommes)

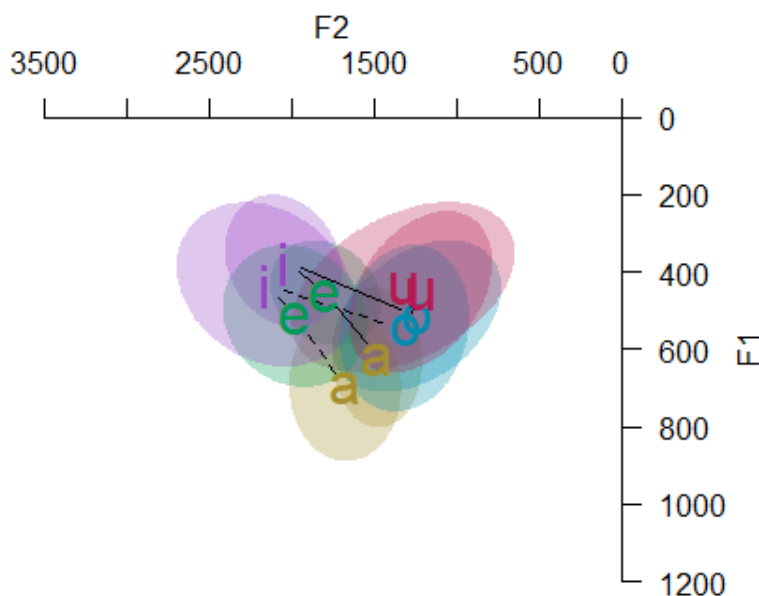


FIGURE 5.5 – Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlars du Centre (pointillés = femmes, lignes = hommes)

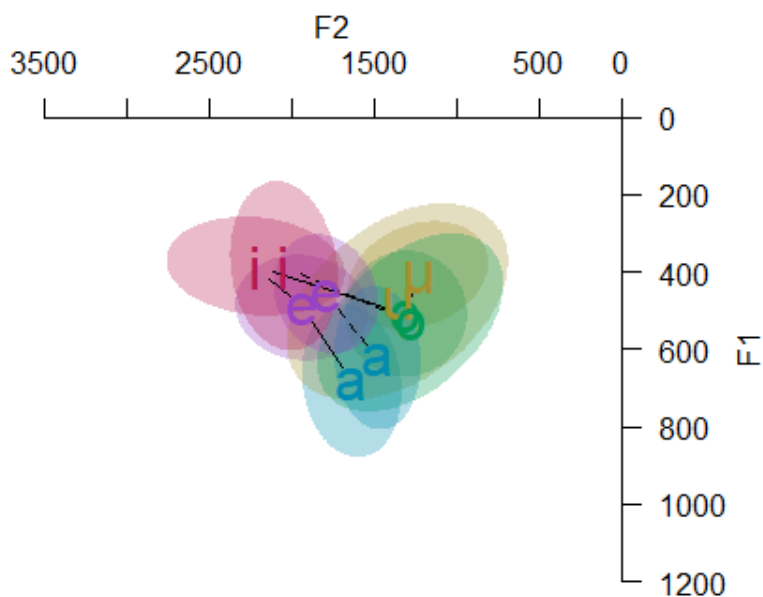


FIGURE 5.6 – Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlars du Sud (pointillés = femmes, lignes = hommes)

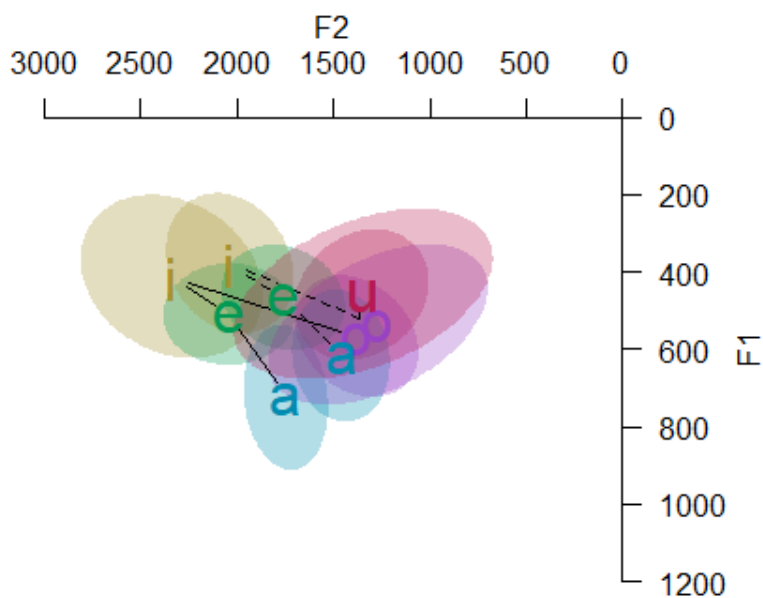
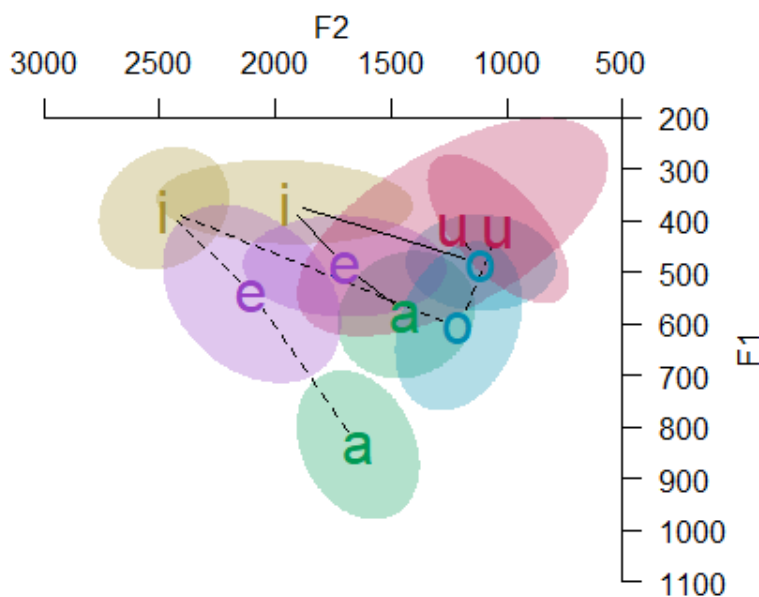


FIGURE 5.7 – Triangle vocalique produit à partir des formants F1 et F2 des voyelles des parlars de Sardaigne (pointillés = femmes, lignes = hommes)



Les résultats obtenus montrent beaucoup de variabilité dans la réalisation des timbres vocaliques et ce, pour toutes les régions géographiques considérées. Cela se manifeste à travers un recouvrement partiel des ellipses correspondant à chaque voyelle. Le tableau qui synthétise les moyennes des formants montre également des valeurs similaires pour toutes les régions.

TABLE 5.2 – Valeurs moyennes des formants

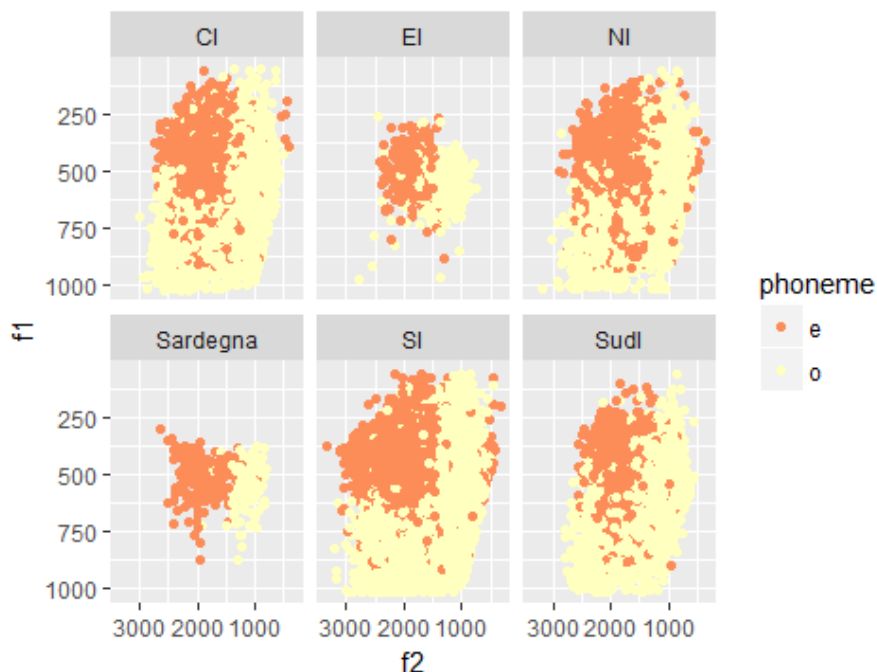
Zone	/e/		/o/		/a/		/i/		/u/	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
NI	510.8	91695.4	511.5	1695.8	511.5	11699	512.6	1703.6	506	1700.9
CI	515.9	1705.1	516	1701.7	515.5	1704.1	516.7	1699.5	513.4	1705.9
SudI	502.5	1686.9	504.7	1698.1	503.3	1687.7	501.9	1694.9	504.7	1639.3
Sardaigne	514.9	1675.7	489.2	1674.9	474.8	1673.4	504.7	1674.8	459.1	1711.2

À la lumière de ces résultats, nous ne pouvons pas nous prononcer sur la possibilité de typer les régions à travers la réalisation globale des timbres vocaliques. Ces voyelles sont produites en contexte et dans un flux de parole semi-préparée ou spontanée, et la coarticulation inhérente à une parole non contrôlée est responsable de ces superpositions des espaces.

Notons toutefois la différence entre les réalisations vocaliques des locuteurs sardes et celles du reste de l'Italie. Les ellipses sont plus distinctes et nous rencontrons moins de recouvrement des espaces dédiés à chaque timbre. Pour beaucoup de linguistes, la Sardaigne constitue un isolat linguistique, voire une langue à part, et

cette individualité semble se traduire par une utilisation différente de l'espace vocalique.

FIGURE 5.8 – Comparaison des timbres des voyelles moyennes entre tous les parlers présents dans le corpus, y compris ceux des locuteurs non natifs



La figure 5.8 montre la disposition pour les voyelles d'aperture moyenne /e/ et /o/. Encore une fois nous pouvons noter la forte variabilité qui concerne toutes les régions à l'exception de la Sardaigne. Sauf pour la Sardaigne nous pouvons faire l'hypothèse de réalisation à la fois ouverte et fermée pour les *e* et *o* graphémiques.

5.3 Détection automatique des parlers régionaux avec WEKA

Par la suite nous nous sommes penchés sur la question de la classification automatique des variantes régionales à partir des traits acoustiques. À cet effet, toutes les caractéristiques extraites ont servi à alimenter les trois algorithmes de classification choisis, disponibles dans WEKA, que nous avons utilisés afin de déterminer l'origine géographique des sources présentes dans le corpus. Les algorithmes testés ont été choisis en se basant sur leur capacité à prévoir plusieurs classes, étant donnée la nature de l'étude, basée sur la détection des trois origines régionales possibles.

Pour que les résultats soient pertinents, les classes du corpus ont été divisées en sous-ensembles égaux, pour un corpus de départ qui consistait en 40,7% de locuteurs du nord, 43,05% du centre et 16,23% du sud. Nous n'avons pas considéré les locuteurs de Sardaigne.

Expériences d'identification automatique des variantes dialectales : est-il possible d'identifier automatiquement toutes les variantes ? Les algorithmes

choisis pour la classification sont : *Ibk*¹, basé sur la proximité des valeurs ; *Naïves Bayes*, à l'image du choix fait par [Bahari et al., 2018] et les arbres de décision *J48*, qui permet d'analyser les résultats par ordre d'importance dans le graphe. Ces algorithmes ont été entraînés sur un corpus d'entraînement, et vérifiés sur un corpus de test, correspondant respectivement à 70 % et 30 % du total. Les résultats ci-dessous sont présentés sous forme d'une matrice de confusion, dans laquelle en colonne sont représentées les classes a priori (les trois régions géographiques) et en ligne les prédictions.

TABLE 5.3 – Résultat de la détection des trois classes par l'algorithme *Ibk*

Zone	NI	CI	SudI
NI	15.82%	8.94%	8.56%
CI	8.56%	14.9%	9.83%
SudI	11.65%	10%	11.64%

TABLE 5.4 – Résultat de la détection des trois classes par l'algorithme Naïve Bayes

Zone	NI	CI	SudI
NI	5.9%	18.7%	8.7%
CI	8.5%	12.9%	11.78%
SudI	5.1%	10.8%	11.78%

TABLE 5.5 – Résultat de la détection des trois classes par l'algorithme *J48*

Zone	NI	CI	SudI
NI	15.8%	11.6%	5.8%
CI	9.6%	15.1%	8.6%
SudI	12.2%	9.5%	11.6%

Les résultats montrent que les particularités acoustiques permettent d'attribuer correctement les fichiers de test à la bonne variété et ce pour tous les algorithmes testés, à l'exception de Naïve Bayes, qui prévoit avec une fréquence trois fois plus grande les locuteurs du Nord comme étant du Centre.

Nous observons également que les parlers du centre sont plus faciles à prédire, suivis de ceux du nord, tandis que SudI reste la classe avec les plus bas pourcentages de réussite.

Nous relevons également que les traits acoustiques des locuteurs du Sud sont souvent attribués à ceux du Nord, en particulier pour ce qui est du premier et dernier des algorithmes testés. Une explication pourrait en être la prononciation fermée de la voyelle /e/ dans certains mots, tels que les adverbes se terminant en *-mente*, par rapport à la prononciation ouverte, fréquente auprès des locuteurs du centre. Ou bien une manque d'attributs pour classer les classes correctement.

1. correspondent au NKK plus proches voisins dans WEKA

Expériences de caractérisation du parler standard : de quel dialecte est-il le plus proche ? Ensuite, les mêmes algorithmes ont été employés pour une détection basée sur le modèle produit par le corpus de train, mais avec un corpus de test constitué exclusivement de formants issus des productions vocaliques des locuteurs avec un parler standard. Cette analyse a pour but de vérifier si le parler standard peut être qualifié de proche d’une des trois zones étudiées. Les résultats sont rapportés dans les tableaux 5.6, 5.7 et 5.8.

TABLE 5.6 – Résultats de la détection de la classe correspondante aux traits acoustiques d’un italien standard par l’algorithme Ibk

Zone	NI	CI	SudI
SI	31.9%	40.4%	27.26%

TABLE 5.7 – Résultats de la détection de la classe correspondante aux traits acoustiques d’un italien standard par l’algorithme Naïve Bayes

Zone	NI	CI	SudI
SI	25.21%	44.57%	30.21%

TABLE 5.8 – Résultats de la détection de la classe correspondante aux traits acoustiques d’un italien standard par l’algorithme de tree J48

Zone	NI	CI	SudI
SI	33.57%	43.12%	23.30%

Les résultats des trois algorithmes ont en commun un fort rapprochement des parlers standard, qui ne sont pas marqués par un accent, comme étant similaires au parler du centre. Cela pourrait être expliqué par l’ampleur de la région dite “centrale”, qui est plus importante que les autres. De nombreux traits acoustiques seraient ainsi caractéristiques du Centre. De plus, comme mentionné dans l’introduction, le fait que l’italien ne possède pas officiellement de langue standard, a comme conséquence que la variante de Toscane soit facilement adoptée comme parler standard. Le choix a été fait sur la base de considérations historiques et culturelles, il a comme conséquence que le parler de Rome et de Toscane, régions du Centre, sont très représentés dans les données.

AUTRES FACTEURS DE VARIATION : GENRE DU LOCUTEUR

6.1 Présentation

Dans ce chapitre nous menons une réflexion sociolinguistique concernant la perception de l'accent d'origine chez les femmes, par rapport aux hommes, ainsi qu'une analyse des résultats issus des algorithmes de classification de WEKA appliqués à deux sous-corpus divisés selon le genre des locuteurs.

Le nombre de locuteurs dans le corpus étudié est important. Il n'a pas été possible de repérer le nombre exact d'hommes et de femmes intervenants dans ces enregistrements de radio couvrant différentes régions. Cependant, comme mentionné dans la section consacrée au corpus et à la méthode, le sexe du locuteur est précisé en correspondance de chaque phonème prononcé. À partir de cette indication on peut estimer une beaucoup plus grande présence d'hommes que de femmes.

De manière plus détaillée, on compte 77,1 % des phonèmes prononcés par des hommes et seulement 22,9 % prononcés par des femmes. De plus, en présence d'un locuteur féminin il a souvent été impossible de repérer l'accent d'origine, comme le montre le tableau ci-dessous, les femmes ont été identifiées comme ayant un accent standard, deux fois plus que les hommes.

TABLE 6.1 – Pourcentage de perception d'un accent régionale chez les deux sexes

SI	79.5%	32.3%
Autre	20.5%	67.7%

Une explication à ce phénomène résiderait probablement dans des raisons sociolinguistiques, comme le montre l'étude de [Nelson et al., 2014] qui révèle que les femmes se trouveraient en position plus susceptible d'être jugées comme moins compétentes par rapport aux hommes, en particulier lorsqu'elles laissent percevoir la présence d'un accent dans leur façon de parler, ou bien, comme l'expliquent [Adda-Decker and Lamel, 2005], par le rôle de la femme dans l'éducation et la transmission de l'accent des parents aux enfants, d'où la tendance à adopter une prononciation standard d'autant plus que le cadre des informations radio-diffusés suppose un besoin d'accessibilité au plus grand nombre.

6.2 Classification des accents selon le sexe du locuteur

Les traits acoustiques extraits automatiquement pour caractériser les voyelles ont été divisés en deux sous-corpus : un corpus dont les voyelles n'ont été prononcées que par des femmes et un corpus où les phonèmes vocaliques n'ont été prononcés que par des hommes.

Pour cette dernière analyse, les résultats ont été tirés par une validation croisée en se basant sur un seul algorithme de classification : le J48, ayant produit le meilleur résultat dans l'étude précédente.

TABLE 6.2 – Résultat de la détection des trois classes chez les femmes

Zone	NI	CI	SudI
NI	63.7%	15.8%	20.46%
CI	14.2%	74.9%	10.9%
SudI	17.1%	16.5%	66.3%

TABLE 6.3 – Résultat de la détection des trois classes chez les hommes

Zone	NI	CI	SudI
NI	72%	18.9%	9.1%
CI	17.4%	68.6%	13.9%
SudI	10.7%	15.1%	74.1%

Le tableau des résultats montre une meilleure détection de l'accent chez les hommes, qui de ce fait montrent un accent non-standard plus marqué se traduisant par des traits acoustiques des voyelles spécifiques aux régions géographiques retenues. Cependant il est important de tenir compte de la grande quantité de données correspondantes aux hommes, ce qui permet également un meilleur entraînement de l'algorithme sur les données correspondant aux hommes.

CONCLUSION ET PERSPECTIVES

Dans ce mémoire, nous avons étudié la variation phonétique de la langue italienne. La perspective adoptée a été celle de la variation diatopique telle que reflétée par des phénomènes spécifiques à la phonétique de l'italien concernant le système consonantique et vocalique. Notre travail repose sur un corpus de 5 heures de données radiodiffusées de la radio-télévision italienne RAI, transcrites manuellement et segmentées automatiquement avec un système de reconnaissance vocale pour l'italien développé au LIMSI lors de projets précédents. À partir des données segmentées automatiquement, nous avons procédé à une annotation des sources en grandes régions géographiques (Nord, Centre, Sud, Sardaigne et non natif) et nous avons également étudié la possibilité d'identifier les locuteurs selon le genre. Cette dernière étude nous a permis de considérer les phénomènes de variation en rapport avec le locuteur, homme ou femme.

L'analyse de la variation consonantique s'est focalisée sur un phénomène spécifique à l'italien, la gémination. Nous avons commencé par estimer les différences de durée entre les consonnes simples et les consonnes géminées. Par la suite, nous nous sommes intéressés au renforcement phono-syntaxique et au redoublement intervocalique qui sont des phénomènes similaires à la gémination, mais non phonologiques. Ces phénomènes sont effectivement observables aussi bien dans les parlers du centre et du sud, que dans ceux du nord. Cependant, de manière globale, il est possible d'associer des spécificités de réalisation de ces phénomènes à une région.

Dans une deuxième phase, nous nous sommes intéressés au contexte vocalique. Nous avons notamment mené une analyse de la dispersion des formants F1 et F2 selon les zones géographiques. À l'exception de la Sardaigne, la dispersion des voyelles est similaire d'une région à l'autre et les ellipses sont superposées, ce qui est spécifique de la parole continue, où la coarticulation est présente. Cette difficulté d'isoler des zones dialectales est aussi observable dans les résultats des algorithmes d'apprentissage, dont les pourcentages montrent une différenciation, mais sans être très performants.

Enfin, la détection des origines par rapport au sexe du locuteur a fourni des résultats préliminaires montrant qu'il est plus facile d'associer une région dialectale aux locuteurs masculins. Cette analyse nécessite cependant de considérer des corpus plus équilibrés, car dans notre étude il y avait plus d'hommes que de femmes (ce qui a favorisé un sur-apprentissage des caractéristiques des locuteurs masculins).

Dans un futur travail, nous envisageons d'étudier les différences d'accents entre affriquées et fricatives. Cet intérêt est en lien avec la tendance, presque-systématique d'un point de vue perceptif, des locuteurs du centre, en particulier, à prononcer la consonne affriquée /tʃ/, comme la fricative /ʃ/. En raison de la difficulté de trou-

ver des paramètres spectraux identifiables automatiquement et dans les temps par rapport aux contraintes du stage, nous nous sommes arrêtés à une analyse comparative des spectres des deux types de segments. À titre d'illustration, nous présentons dans les figures 6.1 et 6.2 les sorties du spectrogramme pour les phonèmes /tʃ/ et /ʃ/. Dans la figure 6.1, nous pouvons comparer l'affriquée /tʃ/ et la fricative /ʃ/, prononcées par un locuteur d'italien standard, dans la figure 6.2, le même phonème affriquée, prononcé par un locuteur du centre. Nous observons que, dans le deuxième cas, le spectre présente des ressemblances avec celui de la figure 6.1 correspondant à la fricative /ʃ/, ce qui suggère que l'affriquée est prononcée comme une fricative dans le parler de la capitale.

Finalement, pour conclure, ce travail s'inscrit dans la linguistique outillée de grands corpus permettant d'ancrer les théories linguistiques de la réalité langagière. Ce travail peut servir aussi à la reconnaissance vocale des solutions l'amélioration via des variantes de prononciation qui intègrent les phénomènes observés comme la gémination non phonologique.

FIGURE 6.1 – Spectrogrammes correspondant aux phonèmes /tʃ/ et /ʃ/ prononcés par un locuteur avec un italien standard

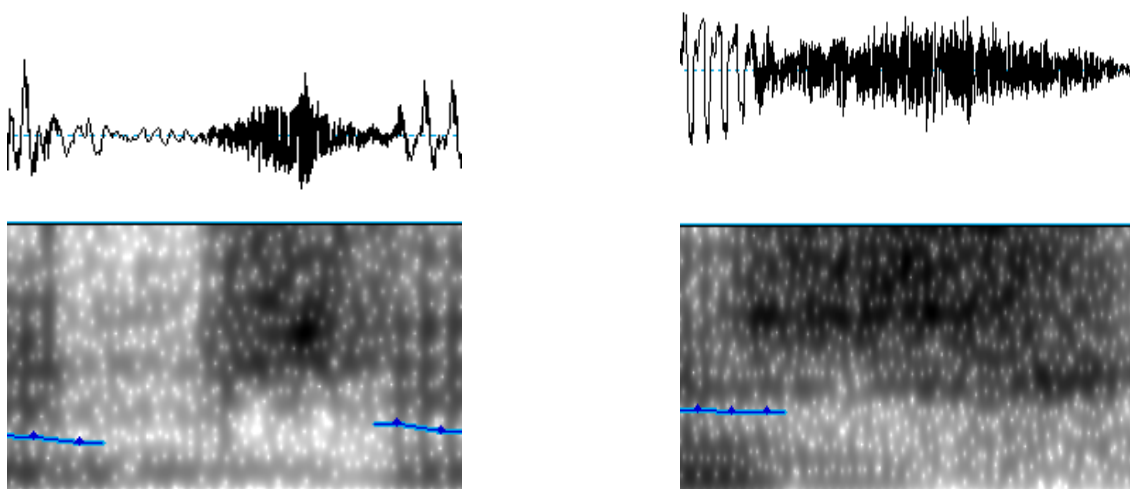
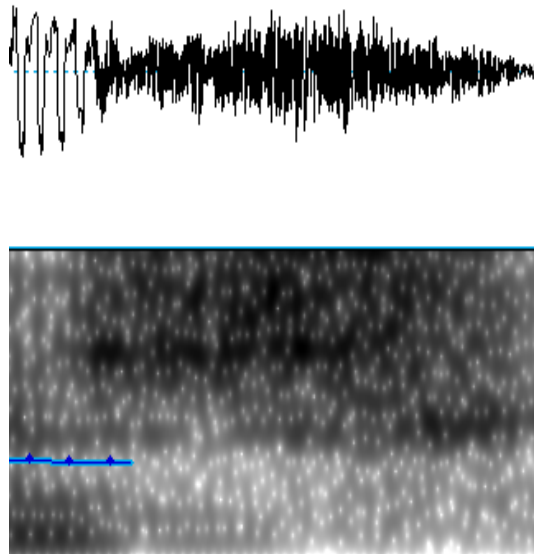


FIGURE 6.2 – Spectrogramme correspondant au phonème /tʃ/ prononcé par un locuteur de la capitale



BIBLIOGRAPHIE

- [Adda-Decker and Lamel, 2005] Adda-Decker, M. and Lamel, L. (2005). Do speech recognizer prefer female speakers? In *Proc. of Interspeech*. – Cité pages 20 et 51.
- [Argenziano and de Filippis, 2010] Argenziano, S. and de Filippis, G. (2010). Dialetto napoletano- vocalismo e consonantismo. In Editore, N. R., editor, *A lenga turrese*. – Cité page 37.
- [Bafile, 2014] Bafile, L. (2014). Betacismo e altre alternanze in dialetti italiani meridionali. il rafforzamento è l'opposto della lenizione? *Rivista di Linguistica Letteratura Cinema Teatro Arte*. – Cité pages 12 et 38.
- [Bahari et al., 2018] Bahari, M. H., Saeidiy, R., hamme, H. V., and Leeuweny, D. V. (2018). Accent recognition using i-vector, gaussian mean supervector and gaussian posterior probability supervector for spontaneous telephone speech. *Radboud Repository of the Radboud University Nijmegen*, pages 7344–7348. – Cité pages 19 et 48.
- [Berruto, 2010] Berruto, G. (2010). *Treccani*. – Cité pages 11, 12 et 24.
- [Bertinetto and Loporcaro, 2005] Bertinetto, P. M. and Loporcaro, M. (2005). The sound pattern of standard italian, as compared with varieties spoken in florence, milan and rome. *Journal of the International Phonetic Association*. – Cité pages 12, 33, 35, 37 et 41.
- [Bonomi et al., 2010] Bonomi, I., Masini, A., Morgana, S., and Piotti, M. (2010). *Elementi di linguistica italiana*. Carocci. – Cité pages 12, 34, 35 et 37.
- [Boula de Mareüil et al., 2004] Boula de Mareüil, P., Brahimy, B., and Gendrot, C. (2004). Role of segmental and supersegmental cues in the perception of magrebian-accented French. In *Proc of INTERSPEECH*. – Cité page 19.
- [Caballero et al., 2009] Caballero, M., Marío, J. B., and Moreno, A. (2009). Multidialectal spanish modeling for asr. *Speech Communication*, 51(3):217–229. – Cité page 18.
- [Delas-Roussarie and Durand, 2003] Delas-Roussarie, E. and Durand, J. (2003). *Corpus et variation en phonologie du français, Méthodes et analyses*. Presses Universitaires du Mirail Toulouse. – Cité page 19.
- [Elinor M.Payne, 2005] Elinor M.Payne (2005). Phonetic variation in italian consonant gemination. *Journal of the International Phonetic Association*. – Cité pages 12, 17, 33 et 36.
- [E.L.Renwick and Ladd, 2016] E.L.Renwick, M. and Ladd, D. R. (2016). Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Journal of the Association for Laboratory Phonology*, (19):1–29. – Cité pages 12, 18, 41 et 42.

- [Laurent, 1998] Laurent, B. (1998). *Aspects de phonétique expérimentale contrastive: "l'accent" anglo-américain en français*. PhD thesis, Université Paris III. – Cité page 19.
- [Ledegen and Léglise, 2013] Ledegen, G. and Léglise, I. (2013). Variation et changement linguistique. In Edition, E., editor, *Sociolinguistique des langues en contact*. – Cité page 17.
- [Matassoni et al., 2012] Matassoni, M., Brugnara, F., and Gretter, R. (2012). Evalita 2011: Automatic speech recognition large vocabulary transcription. In *Evaluation of Natural Language and Speech Tools for Italian*. – Cité page 23.
- [Mattei and Benedetto, 2008] Mattei, M. and Benedetto, M. G. D. (2008). Acoustic analysis of singleton and geminate nasals in italian. *The European Student Journal of Language and Speech*. – Cité pages 12 et 18.
- [Nelson et al., 2014] Nelson, R. J., Signorella, M., and Botti, K. (2014). Accent, gender, and perceived competence. *SAGE journal*. – Cité page 51.
- [Renwick, 2012] Renwick, M. (2012). *Vowels of Romanian: Historical, Phonological and Phonetic Studies*. PhD thesis, Cornell University. – Cité page 18.
- [Renwick, 2018] Renwick, M. (2018). Phonetic implementation of mid vowel contrasts across Italian varieties. In *Proc of British Association of Academic Phoneticians*. – Cité page 42.
- [Turco and Braun, 2016] Turco, G. and Braun, B. (2016). An acoustic study on non-local anticipatory effects of italian length contrast. *The Journal of the Acoustic Society of America*, (4):2247–2256. – Cité pages 12, 17 et 18.
- [Vasilescu, 2016] Vasilescu, I. (2016). *De la variation dans la production et la perception de la parole : analyses linguistiques de grand corpus en lien avec le traitement automatique*. PhD thesis, Université Paris-Saclay. Habilitation à diriger des recherches. – Cité page 11.
- [Vasilescu, 2018] Vasilescu, I. (2018). Studying variation in romanian: deletion of the definite article -l in continuous speech. *Linguistic Vanguard*. – Cité page 23.
- [Vieru-Dimulescu and Boula de Mareüil, 2006] Vieru-Dimulescu, B. and Boula de Mareüil, P. (2006). Perceptual identification and phonetic analysis of 6 foreign accents in French. In *Proc. of Interspeech*. – Cité page 19.