



HAL
open science

Les doubles modaux en anglais américain : nouvelles perspectives sémantiques, sociolinguistiques et historiques à partir d'un corpus de réseaux sociaux

Cameron Morin

► To cite this version:

Cameron Morin. Les doubles modaux en anglais américain : nouvelles perspectives sémantiques, sociolinguistiques et historiques à partir d'un corpus de réseaux sociaux. 2024, <https://cle.ens-lyon.fr/anglais/langue/linguistique/les-doubles-modaux-en-anglais-americain>. hal-04730920

HAL Id: hal-04730920

<https://hal.science/hal-04730920v1>

Submitted on 10 Oct 2024

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

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

Les doubles modaux en anglais américain : nouvelles perspectives sémantiques, sociolinguistiques et historiques à partir d'un corpus de réseaux sociaux

Cameron Morin

La Clé des Langues ENS Lyon (ISSN : 2107-7029)

[Article] Cet article est une version abrégée et adaptée pour un public angliciste francophone d'un article scientifique récemment paru dans la revue *PLOS One* (Morin & Grieve 2024). Dans cet article, nous analysons l'emploi des doubles modaux en anglais américain (par exemple *might could, would might...*) à partir d'un corpus de plusieurs milliards de mots provenant de publications géolocalisées sur le réseau social Twitter entre 2013 et 2014. Nous identifions et cartographions 76 types de doubles modaux distincts pour un total de 5 349 exemples, révélant une structure et un usage plus complexes qu'on ne le pensait auparavant. Nous abordons trois débats théoriques : tout d'abord, nous montrons qu'il n'y a pas de contraintes absolues sur la formation des doubles modaux, car ceux-ci sont utilisés de manière répandue à travers les États-Unis. Deuxièmement, nous constatons une variation dialectale considérable dans leur usage, qui est le plus souvent associé à la langue afro-américaine dans la région du *Deep South*. Troisièmement, nos résultats suggèrent que les doubles modaux pourraient être une innovation de la langue afro-américaine plutôt qu'un import par des colons écossais et irlandais dans la région de l'*Upper South* des États-Unis.

Introduction

Les doubles modaux (DMs), séquences de deux verbes modaux (par exemple, *I might can go to work ; He used to could run every morning ; You might should help her out*), ont longtemps intrigué les linguistes anglicistes. Les DMs apparaissent dans certains dialectes de l'anglais américain (principalement dans le Sud), en scots, y compris le *Borders Scots* et l'*Ulster Scots* (Morin et al. 2020 ; Morin 2021a), ainsi que dans les créoles fondés sur l'anglais comme le gullah, parlé par des locuteurs afro-américains de la côte Est dont les ancêtres esclaves cultivaient le riz sur les zones côtières, et le jamaïcain (Kortmann et al. 2020). La nature grammaticale des DMs, envisagés tantôt comme combinaisons d'un adverbe et d'un modal, tantôt comme éléments lexicaux uniques, demeure à ce jour non résolue (Nagle 2003). La recherche sur les DMs examine leur statut syntaxique (Elsman & Dubinsky 2009), leurs contraintes sémantiques (Nagle & Holmes 2000 ; Morin 2021b), leurs distributions régionales et sociales (Coats 2022 ; Morin & Coats 2023 ; Coats & Morin à paraître) et leurs origines historiques (Zullo et al. 2021).

En plus de leur structure inhabituelle et de leur signification sociale marquée, les DMs sont l'une des constructions grammaticales les plus rares de la langue anglaise. Leur rareté limite toute analyse empirique, car même les corpus linguistiques standard de plusieurs millions de mots sont trop restreints pour les observer de manière fiable, y compris dans les dialectes où ils sont les plus courants. Par conséquent, l'étude des DMs repose sur des méthodes de recueil de données auprès des locuteurs eux-mêmes, telles que les jugements d'acceptabilité de faits de langue, les enquêtes dialectales et les entretiens sociolinguistiques. Ces méthodes ne sont pas observationnelles, ce qui rend difficile à la fois l'identification d'inventaires de DMs à étudier et la mise au point des meilleures approches pour les recueillir auprès d'informateurs.

Notre compréhension limitée de l'emploi des DMs conduit à deux lacunes majeures. Premièrement, nous ne connaissons pas l'inventaire complet des DMs, ni les combinaisons modales les plus courantes. Une telle description nécessiterait l'analyse de très grands corpus de langue naturelle,

et l'absence de tels corpus entrave aussi l'évaluation d'hypothèses concernant les contraintes sémantiques sur leur formation. Deuxièmement, nous savons peu de choses sur la variation dialectale dans l'emploi des DMs. Les données déjà collectées auprès d'un petit nombre d'informateurs, principalement dans le Sud des États-Unis (Hasty 2012), sont insuffisantes pour proposer une cartographie détaillée ou une analyse quantitative. Cette lacune rend difficile la compréhension des facteurs qui conditionnent la distribution régionale et sociale des DMs, ainsi que de leurs origines historiques.

Cependant, l'usage accru du langage en ligne permet pour la première fois l'étude des DMs à grande échelle (Morin et al. 2024), et ouvre de nouvelles perspectives de recherche descriptive et théorique inimaginables il y a 50 ans (Butters 1973). Notre objectif dans cette étude est de réaliser une description plus complète de l'emploi des DMs en anglais américain à l'aide d'un grand corpus de données issues de Twitter, et de considérer les manières dont ces résultats éclairent les débats théoriques actuels. Plus précisément, cet article aborde cinq questions de recherche principales, dont les deux premières sont descriptives et les trois dernières sont théoriques.

Tout d'abord, nous répondons à deux questions de recherche descriptives par une analyse quantitative d'un corpus de 8,9 milliards de mots, composé de publications géolocalisées sur Twitter dans l'ensemble des États-Unis contigus (c'est-à-dire tous les États sauf l'Alaska et Hawaï) entre 2013 et 2014 :

(1) Quel est l'inventaire et la distribution quantitative des DMs sur ce réseau social américain ?

(2) Quelle est la distribution régionale des DMs, individuellement et globalement ?

Ensuite, en nous fondant sur nos résultats empiriques, nous considérons trois questions de recherche théoriques particulièrement importantes :

(3) Quels sont les facteurs sémantiques qui sous-tendent la formation des DMs ?

(4) Quels sont les facteurs sociaux et régionaux qui expliquent la variation dans l'emploi des DMs ?

(5) Quelles sont les origines historiques des DMs en anglais américain ?

Bien que notre étude soit fondée sur des données modernes issues d'un réseau social, le grand nombre de catégories et d'occurrences de DMs que nous observons et cartographions en détail pour la première fois dans les comtés des États-Unis nous permet de remettre en question des présomptions de longue date sur ces constructions en anglais américain. Notre étude nous permet en outre de contribuer à une meilleure compréhension théorique des DMs anglais en général.

Le reste de cet article est structuré comme suit : en Section 2, nous passons en revue les recherches antérieures sur les DMs, en particulier leurs aspects syntaxiques, sémantiques, sociolinguistiques et historiques. En Section 3, nous présentons notre corpus, ainsi que ses forces et ses faiblesses méthodologiques. En Section 4, nous présentons notre analyse descriptive. En Section 5, nous examinons nos résultats empiriques à la lumière de nos questions théoriques. Enfin, en Section 6, nous résumons nos principales conclusions et nous considérons leur pertinence pour de futures études sur les DMs en anglais.

1. Recherches antérieures

1.1 Structure et signification des doubles modaux

Les DMs, constructions grammaticales rares contenant deux verbes modaux dans un seul syntagme verbal conjugué, sont un type de 'modal multiple' (Huang 2011). Parmi les exemples les plus courants, on retrouve des phrases telles que *I might could go* et *You might oughta go*. Bien qu'ils soient non-standard et peu connus pour nombre de locuteurs de l'anglais, leurs significations peuvent être généralement inférées, *might* et *may* étant souvent apparentés à *maybe*, et *could* et *ought to* à *be able to* et *have to*.

La sémantique des DMs est généralement contrainte par un ordre 'épistémique-radical' : le premier modal indique tend à indiquer le degré de certitude du locuteur vis-à-vis de la possibilité de l'évènement, et le second tend à refléter l'attitude du locuteur quant à cette possibilité (Nagle et Holmes 2000). Par ailleurs, les DMs servent souvent des fonctions pragmatiques telles que la négociation interpersonnelle et l'atténuation (Mishoe & Montgomery 1994). Les modaux *might* et *may* sont les plus acceptables en première position, et certains auteurs soutiennent que le DM le plus commun est *might oughta* (Hasty 2012).

La structure syntaxique des DMs a longtemps été débattue, et certains vont même jusqu'à remettre en question leur existence en tant que combinaison de deux 'vrais' modaux (Nagle 2003). Divers auteurs proposent de traiter les DMs comme des constituants verbaux parallèles, des éléments lexicaux simples, des séquences d'un adverbe et un modal, ou des prédicats bi-propositionnels (Nagle 2003). Des études génératives récentes proposent de les analyser comme des combinaisons de modaux comportant des caractéristiques syntaxiques distinctes (Elsman et Dubinsky 2009). Ces études visent à réconcilier le comportement des modaux individuels avec la contrainte générale d'un seul verbe conjugué par syntagme verbal en grammaire anglaise. Cependant, des problèmes théoriques subsistent quels que soient le cadre et l'analyse adoptés (Nagle 2003).

1.2 Inventaire et variation des doubles modaux en anglais américain

Dans cette section, nous examinons les inventaires des DMs disponibles dans la littérature, en mettant en lumière les limites empiriques des recherches précédentes. Les DMs sont attestés dans divers dialectes de l'anglais, y compris le *Borders Scots* (Morin 2021) et l'anglais du Tyneside (Beal 2004), ainsi qu'en *Ulster Scots* en Irlande du Nord (Corrigan 2011). Les DMs les plus courants dans les îles britanniques incluent *might can*, *will can*, *used to could* et *must can*, avec d'autres combinaisons moins fréquentes telles que *might could*, *should can* et *would could* (Morin 2023).

Cependant, les DMs sont principalement attestés dans les dialectes de l'anglais américain, notamment dans le Sud, comme on le voit dans les enquêtes dialectales traditionnelles du *Linguistic Atlas of the Middle and South Atlantic States* (McDavid et O'Cain 1980) et du *Linguistic Atlas of the Gulf States* (Pederson et al. 1986), ainsi que dans de nombreuses études dialectologiques en Alabama (Feagin 1979), dans les Carolines (Mishoe et Montgomery 1994), le Tennessee (Hasty 2012), au Texas (DiPaolo 1989) et dans les Appalaches (Wolfram et Christian 1976).

Il existe par ailleurs une grande diversité régionale dans l'emploi des DMs, dont la majorité des formes attestées est inventoriée dans la base de données MultiMo (Reed et Montgomery 2012). Celle-ci contient 2 030 exemplaires de DMs américains répartis sur 30 types, combinant 9 modaux centraux et 2 semi-modaux. Les combinaisons les plus courantes incluent *might could*, *might can*, et *used to could*. Cependant, ces fréquences doivent être interprétées avec prudence, en raison de la variété des méthodes de collecte des données dans ces études.

D'une manière générale, seul un nombre limité de combinaisons de modaux a été attesté en anglais américain, bien que ces combinaisons permettent 110 DMs potentiels. En effet, seuls 30 DMs ont été documentés, avec des désaccords sur les combinaisons les plus courantes, hormis l'accord sur la prédominance de *might could* (DiPaolo 1989, Hasty 2012). Cette variation pourrait découler de

la taille des échantillons, des méthodes de collecte des données, ou de variations dialectales régionales et sociales. Hasty (2012) note que l'acceptabilité des DMs varie à travers le Sud, tandis que les données du LAGS et du LAMSAS suggèrent qu'il y a une distribution homogène des DMs dans les États du Golfe et dans les États de l'Atlantique Sud.

De plus, les DMs ont reçu peu d'attention dans la recherche sur la langue afro-américaine. Fennell et Butters (1996) notent que les DMs sont utilisés dans le Sud par les locuteurs blancs et afro-américains, mais ailleurs surtout par les Afro-Américains. Les données du LAGS (Mishoe et Montgomery 1994) suggèrent aussi que les DMs sont légèrement plus acceptables pour les Afro-Américains, notamment *might could* et *might would*, et sont également plus acceptables pour les locuteurs de classes sociales défavorisées, indépendamment de l'ethnicité.

Enfin, dans une étude récente, Coats (2022) a utilisé un corpus de 1,2 milliard de mots de transcriptions de vidéos YouTube pour étudier les DMs en Amérique du Nord. Après vérification manuelle, il a identifié 1 035 occurrences de DMs, représentant 67 types, bien plus que ceux attestés auparavant. Coats a également cartographié ces DMs au niveau de l'État, montrant une concentration dans le sud-est mais aussi une répartition plus large dans le pays. Ses résultats révèlent en outre une surprenante absence d'occurrences dans les Appalaches et les régions du *Deep South* à forte population afro-américaine, où l'on s'attendrait à en trouver. Coats conclut que les DMs ne sont pas restreints à des dialectes régionaux spécifiques, mais sont une ressource disponible pour la plupart des locuteurs de l'anglais nord-américain.

1.3 Origines diachroniques des doubles modaux

La recherche sur les origines des DMs en anglais américain est spéculative, en raison du manque de données historiques et de notre compréhension limitée des DMs contemporains. La théorie la plus acceptée est celle selon laquelle les DMs ont été importés par les Écossais-Irlandais (*Scots-Irish*) dans les Appalaches aux 18^{ème} et 19^{ème} siècles (Crozier 1984), puis diffusés dans le Sud (Montgomery et Nagle 1993). Cette théorie est soutenue par l'attestation traditionnelle des DMs dans les zones d'immigration écossaise-irlandaise (Wolfram et Christian 1976). Cependant, les preuves empiriques sont limitées, et les inventaires de DMs dans les dialectes du scots et de l'anglais américain semblent différents (Zullo et al. 2021). Montgomery et Nagle (1993) suggèrent que le système des DMs américains se serait différencié du système britannique au cours du temps, par le biais d'une plus grande auxiliarisation de *can* et le développement d'une plus large gamme de combinaisons.

Les DMs ne sont pas courants dans l'*Ulster Scots* moderne (Corrigan 2011), mais plutôt davantage dans le *Borders Scots* (Morin 2023). Leur émergence pourrait être un développement interne ou un emprunt du scandinave (Thráinsson et Vikner 1995). Une théorie secondaire propose que les DMs seraient arrivés aux États-Unis via les Caraïbes, restructurés par un phénomène de créolisation (Cunningham 1970), bien que Fennell et Butters (1996) rejettent cette hypothèse. Zullo et al. (2021) postulent quant à eux que les modaux multiples font partie d'une 'ceinture pan-atlantique' (*pan-atlantic belt*), dont l'origine serait écossaise et qui aurait été restructurée en anglais américain suite à des phénomènes de contact avec les créoles caribéens.

Dans l'ensemble, les origines diachroniques des DMs sont incertaines, malgré un consensus qu'ils ont probablement été importés par les Écossais-Irlandais dans le Sud des États-Unis. L'idée que les DMs sont une innovation interne en anglais américain est généralement rejetée (Fennell et Butters 1996). Toutefois, en raison du manque d'observations des DMs en langue naturelle, ces conclusions sont à nos yeux prématurées, et nécessitent davantage de recherches empiriques.

2. Méthodes

2.1 Corpus

Malgré de nombreuses recherches existantes sur les DMs, il existe encore des lacunes dans notre compréhension de leur inventaire, leur distribution régionale et sémantique, et leurs origines historiques, en raison de leur rareté dans la langue naturelle. Les corpus traditionnels sont trop restreints pour observer ces formes rares à grande échelle. Cependant, l'essor des réseaux sociaux, en particulier Twitter/X, offre de nouvelles possibilités pour l'analyse des DMs. Twitter/X, en tant que réseau social très informel, encourage l'utilisation des DMs plus que d'autres variétés, y compris l'anglais parlé.

Pour cette étude, nous avons analysé un corpus de 8,9 milliards de mots provenant de 980 millions de tweets géolocalisés. Ce corpus a été collecté entre octobre 2013 et novembre 2014 aux États-Unis (Huang et al. 2016). Il est l'un des plus grands corpus de réseaux sociaux existants, et est particulièrement utile pour cartographier la variation linguistique régionale. La majorité des comtés américains sont représentés, avec une moyenne de 2 millions de mots chacun, permettant ainsi une analyse spatiale détaillée.

2.2 Collecte et nettoyage des données

Pour identifier les DMs, nous avons extrait automatiquement tous les tweets contenant des séquences de deux modaux anglais, en excluant les combinaisons dupliquées, les semi-modaux, et les DMs entrecoupés d'un adverbe ou d'un pronom. Après cette extraction, nous avons obtenu 10137 DMs potentiels, mais une analyse manuelle a révélé que beaucoup n'étaient pas de véritables occurrences. Nous avons ainsi éliminé les « faux » DMs, notamment les séquences contenant des mots qui ne sont pas des modaux (par exemple *may* pour le mois de mai), des erreurs de frappe, ou encore des modaux adjacents dans des propositions différentes.

Après avoir supprimé ces erreurs, nous avons retenu un jeu de données de 5 439 occurrences de DMs, représentant 54 % des combinaisons possibles. Toutes ces occurrences contiennent deux modaux consécutifs dans une seule proposition conjuguée. Nous avons rendu ce jeu de données public pour toute analyse ultérieure par d'autres chercheurs.

2.3 Avantages et inconvénients du corpus

Avant de présenter nos résultats, il est crucial de discuter des avantages et des limites de l'approche sur corpus de réseaux sociaux, notamment Twitter, devenue populaire en sociolinguistique computationnelle (Grieve et al. 2023). Travailler avec ces corpus permet d'accéder à de grandes quantités de données, essentielles pour observer des constructions rares comme les DMs à grande échelle. Twitter, en tant que plateforme informelle utilisée par une population diversifiée (21% des Américains en 2013, Duggan et Brenner 2013), augmente les chances d'observer des DMs, et l'accès à des métadonnées spatiales permet une cartographie précise. De plus, les corpus de langue naturelle évitent de fait le 'paradoxe de l'observateur' qui s'impose souvent en sociolinguistique, c'est-à-dire la tendance qu'ont les locuteurs de changer leurs comportements sociaux et linguistiques en présence d'un observateur tiers (Labov 1972).

Cependant, les études fondées sur les corpus présentent des limites, notamment la difficulté de généraliser les résultats à d'autres variétés linguistiques. Twitter ne reflète qu'une partie de la langue vernaculaire, avec une surreprésentation des citoyens jeunes, afro-américains, et urbains (Duggan et Brenner 2013). Bien que cette approche puisse donc compenser certaines lacunes des méthodes classiques de recueil de données auprès des locuteurs, elle limite l'accès à d'autres

communautés, comme les personnes âgées ou les habitants de zones rurales. De plus, les corpus rendent plus difficile l'analyse des nuances de signification et d'acceptabilité des constructions linguistiques : les enquêtes sociolinguistiques demeureront donc toujours une approche complémentaire importante (Morin & Basile 2022).

Malgré ces limites, Twitter reste la seule variété linguistique accessible à l'échelle que nous nécessitons et qui comporte les métadonnées requises pour ce type d'analyse. Des recherches sur des corpus comparables pour l'anglais britannique ont montré que les cartes dialectales issues de Twitter sont fidèles à celles issues des enquêtes d'élicitation (Grieve et al. 2019). Nous soutenons donc que ce corpus offre des perspectives importantes pour l'étude des DMs, tant sur Twitter qu'en anglais américain en général, en attendant un accès accru à davantage de données linguistiques naturelles et informelles.

3. Résultats

3.1 Inventaire et fréquences des doubles modaux

À partir des 5 439 occurrences de DMs observés, nous avons identifié 76 types distincts, représentant 69 % des 110 types possibles. Contrairement aux recherches précédentes, *might could* n'est pas le DM le plus fréquent dans notre corpus : il présente en effet deux fois moins d'occurrences que la première catégorie de la liste, *might can*. Cette disparité pourrait refléter des différences sociales, de registre, ou bien concernant un changement historique. En tout cas, elle remet en question les recherches antérieures qui sous-estimaient l'importance de *might can* par rapport à *might could*.

D'autres résultats importants incluent la fréquence des DMs avec *will*, *would*, *can*, et *could* en première position, qui sont rares dans MultiMo mais relativement courants dans notre corpus, tandis que *used to would* et *should oughta* sont moins fréquents que prévu. La diversité des DMs dans notre corpus Twitter est également plus grande que dans les études précédentes, avec 76 combinaisons uniques, soit plus du double de celles trouvées dans MultiMo. Ces résultats font en partie écho aux recherches de Coats (2022), qui a également identifié des combinaisons absentes de MultiMo. La Figure 1 ci-dessous représente le degré d'intersection des inventaires des DMs entre MultiMo et notre corpus.



Figure 1 : Diagramme de Venn comparant la distribution des DMs dans la base de données MultiMo et notre corpus issu de Twitter. <https://doi.org/10.1371/journal.pone.0295799.q001>

3.2 Distribution régionale des doubles modaux

Pour approfondir notre analyse des DMs, nous avons cartographié leur usage à travers les États-Unis en utilisant les informations de géolocalisation de notre corpus. La Figure 2 illustre deux résultats majeurs. Premièrement, les DMs traditionnels sont les plus courants dans le sud-est des États-Unis, ce qui confirme les recherches précédentes, bien qu'ils apparaissent également dans nombre d'autres régions, comme la Pennsylvanie et l'Ohio. Deuxièmement, et contrairement aux attentes, un certain nombre de DMs non-traditionnels, comme *should would* et *would could*, n'affichent aucun motif régional clair, ce qui remet en question l'idée que les DMs sont exclusivement des formes du Sud.

Frequency Maps for Double Modals on American Twitter (2013–2014)

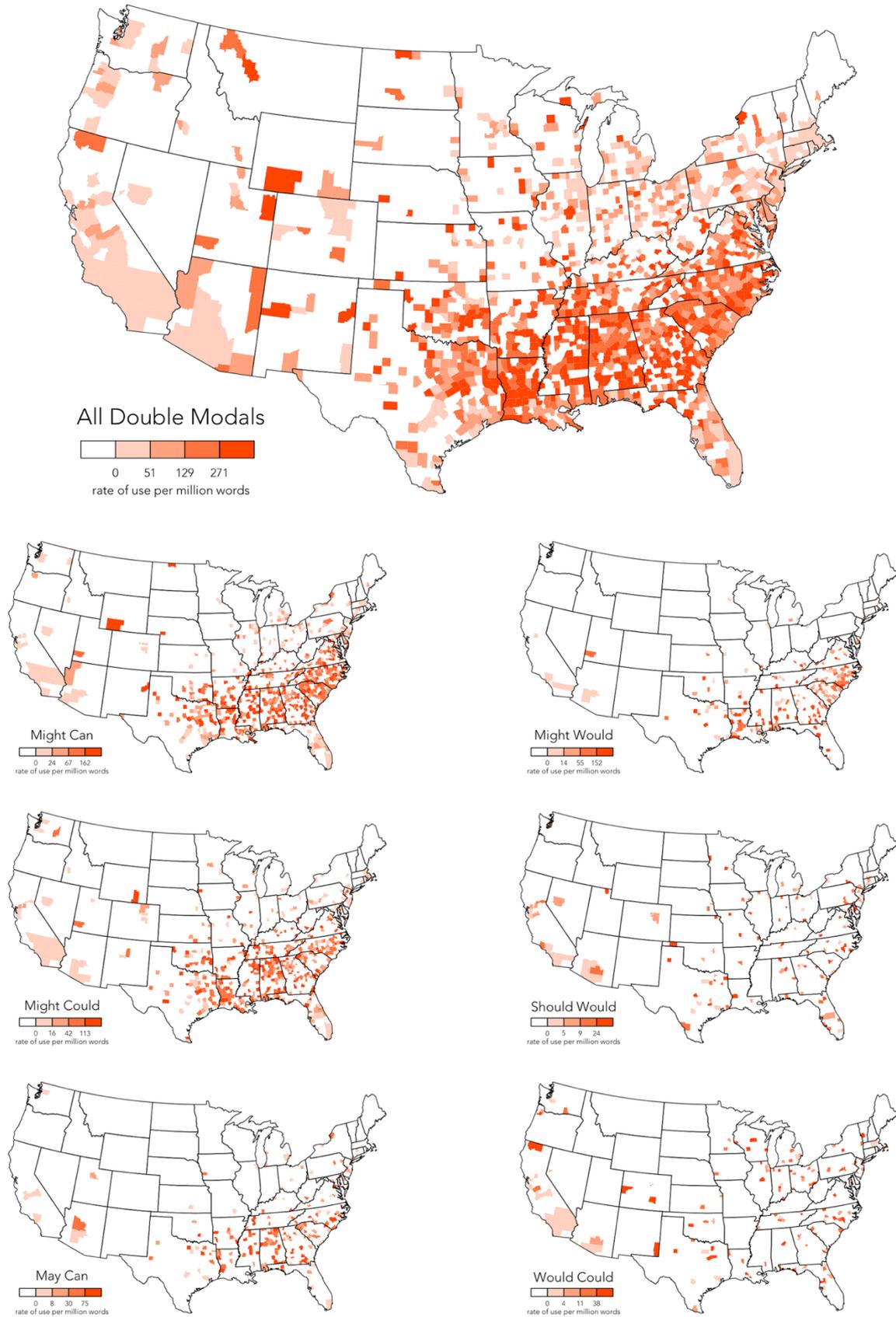


Figure 2 : Cartes de fréquences des DMs sur Twitter aux États-Unis. <https://doi.org/10.1371/journal.pone.0295799.g003>

Pour confirmer ce deuxième résultat inattendu, nous avons vérifié la validité des combinaisons non-attestées par le passé (cf. le volet de gauche de la Figure 1). Bien qu'il ne soit pas impossible qu'il y demeure un petit nombre d'erreurs typographiques difficiles à détecter, ces cas sont certainement mineurs. En effet, chaque occurrence a été analysée manuellement pour s'assurer qu'il s'agissait d'un emploi de deux verbes modaux dans une seule proposition finie. Par ailleurs, la structure de la distribution des DMs dans notre corpus, comportant un petit groupe de DMs fréquents et un grand nombre de DMs à faible fréquence (voir la section suivante), étaye la validité de l'inventaire dans son ensemble. En conclusion, ces DMs non traditionnels semblent être une caractéristique rare mais réelle de l'anglais américain, qui n'est pas facilement expliquée par les théories actuelles (Coats 2022).

4. Discussion théorique

4.1 Contraintes sémantiques sur la formation des doubles modaux

En raison du manque de connaissances sur les DMs et leurs fréquences, il a été difficile dans les études antérieures d'établir une combinatoire précise de ces constructions en anglais américain. Par contraste, nos résultats fournissent une base importante et utile pour évaluer des hypothèses au sujet de leurs contraintes sémantiques. Comme nous l'avons dit plus haut, bien que l'inventaire des DMs semble dominé par un petit nombre de formes fréquentes, il en existe une grande diversité globale. Ceci est illustré par les graphes de la Figure 3, qui montrent les combinaisons modales les plus courantes, telles que *might can* et *might could*.

Double Modal Usage on American Twitter (2013–2014)

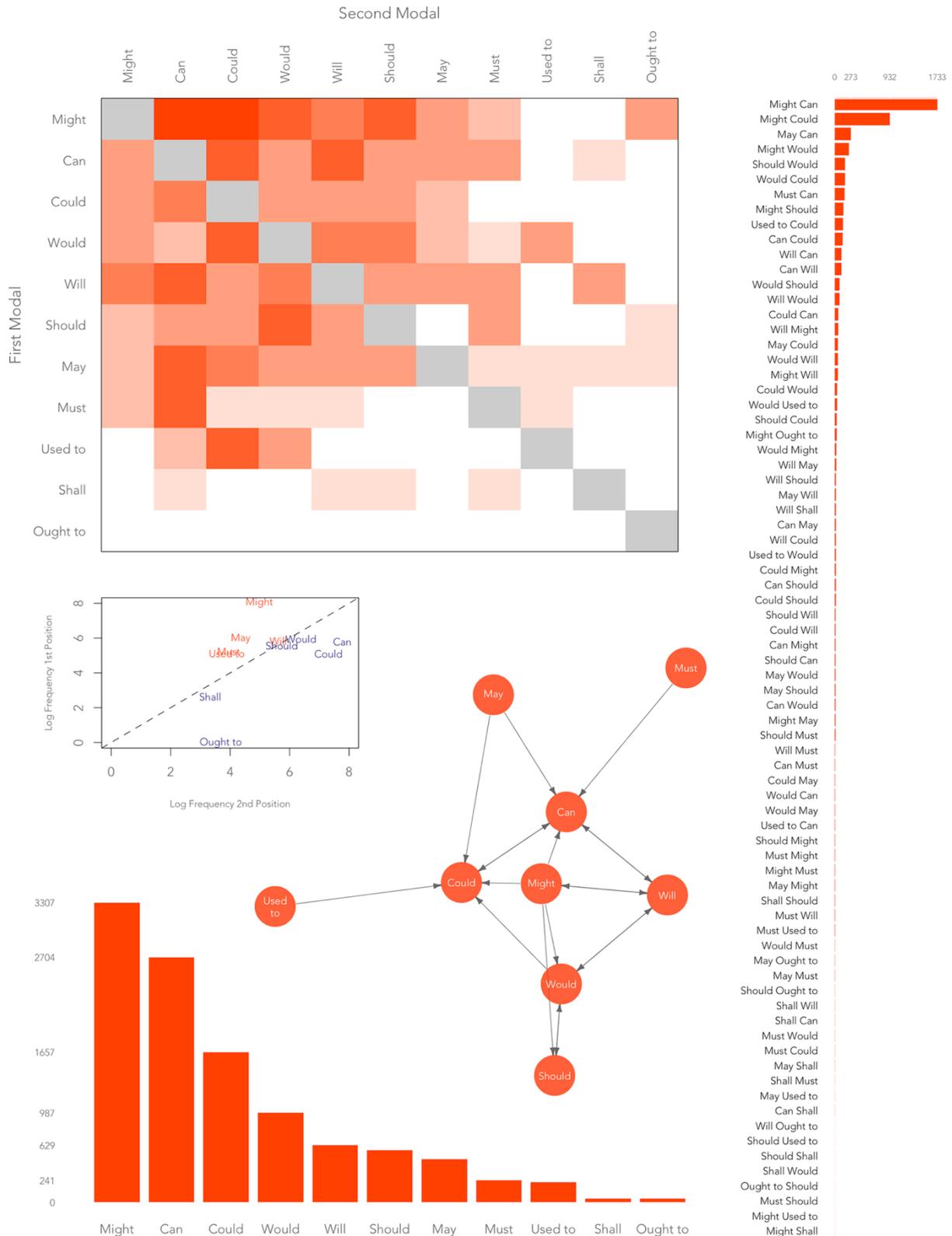


Figure 3 : Visualisations des DMs dans le corpus Twitter (dans le sens horaire à partir du haut) : graphique de densité par position modale, histogramme des DMs, schéma en réseaux par position modale, histogramme des modaux individuels au sein des DMs, et diagramme de dispersion des DMs par position modale. <https://doi.org/10.1371/journal.pone.0295799.g004>

Les graphes révèlent la prédominance de certains modaux en première position (*might, may, will*) et en deuxième position (*can, could*), tandis que *used to* et *ought to* sont généralement rares. Par ailleurs, les représentations des relations entre les modaux montrent que *might* se combine fréquemment avec *can* et *could*, alors que *will* se combine souvent avec *can, might, et would*.

Dans l'ensemble, nos résultats suggèrent qu'il n'y a pas de contraintes absolues sur la formation des DMs en anglais américain, mais plutôt des contraintes probabilistes. Ils remettent aussi en question les théories précédentes, qui postulaient un système figé contenant un petit nombre de DMs, et suggèrent même qu'avec un corpus suffisamment large, la totalité des 110 combinaisons possibles aurait été observée.

Pour mieux comprendre quelles pourraient être les contraintes sémantiques probabilistes qui influencent la formation des DMs, nous avons utilisé un cadre analytique inspiré de Coates (1983) et Palmer (1990). Ce cadre distingue deux types de significations : d'un côté, le degré de certitude du locuteur à propos d'un événement (modalité épistémique), et de l'autre, l'attitude du locuteur envers la possibilité de cet événement (modalité radicale).

Nous avons analysé manuellement un échantillon aléatoire des 10 DMs les plus courants pour voir comment leur sens change selon le contexte des tweets. Nos résultats montrent que l'ordre où la certitude vient avant l'attitude (Épistémique > Radical) est le plus fréquent (57 %), comme le suggèrent des études antérieures. Cependant, il est aussi courant de trouver deux marqueurs de certitude (28 %), et, bien que plus rare, l'usage de deux marqueurs d'attitude (11 %) est aussi présente. L'ordre où l'attitude vient avant la certitude est le moins fréquent (4 %), mais il existe tout de même.

Nous notons par ailleurs deux autres motifs sémantiques majeurs. Premièrement, les modaux de possibilité (tels que *may, might, can, could*) sont les plus courants dans les DMs, ce qui suggère qu'ils remplissent bien une fonction pragmatique d'atténuation (Mishoe & Montgomery 1994). Deuxièmement, conformément à la littérature, les DMs semblent présenter une concordance de temps (les deux modaux doivent être ou bien au présent ou bien au prétérit). Cependant, cette tendance est encore une fois probabiliste plutôt que catégorique. En effet, on retrouve un certain nombre de DMs à 'temps mixte', dont *might can*, la combinaison la plus fréquente dans tout le corpus. En général, pour les DMs les moins courants (5 occurrences ou moins), les contraintes que nous trouvons sont moins respectées, au point que ces combinaisons semblent être essentiellement libres de contraintes.

En somme, nos résultats montrent que la combinatoire des DMs est beaucoup moins rigide qu'on ne l'avait supposé, et que celle-ci refléterait simplement les pressions de l'usage linguistique, ce qui fait écho aux études récentes sur les DMs en anglais parlé (Coats 2022 ; Coats 2023).

4.2 Sources de variation dialectale dans l'emploi des doubles modaux

Comme décrit précédemment, nous avons identifié deux grands motifs régionaux dans l'utilisation des DMs : les DMs traditionnels sont principalement utilisés dans le Sud, tandis que de nombreux DMs 'non-traditionnels' sont dispersés à travers le pays sans concentration régionale apparente. Dans cette section, nous nous concentrons plus particulièrement sur le Sud des États-Unis pour évaluer la variation régionale et sociale des DMs traditionnels, en particulier les 10 DMs les plus fréquents, tels que *might can* et *might could*. Nous avons utilisé une analyse d'autocorrélation locale spatiale pour cartographier leur utilisation. Notre travail de cartographie révèle que les DMs sont plus fréquents dans le *Deep South* que dans les Appalaches, contrairement à ce que disent les études précédentes (Wolfram et Christian 1976).

Cette concentration d'emploi dans le *Deep South*, région qui présente une importante population afro-américaine, concorde en revanche avec les résultats du LAGS, qui suggèrent que les DMs sont davantage employés par les locuteurs afro-américains (à l'exception de *might could*). Afin d'explorer ces tendances plus en détail, nous avons comparé les cartes d'emploi des DMs à des données démographiques de population afro-américaine par comté dans cette région. Nos résultats sont illustrés en Figure 4. Ils révèlent une corrélation positive pour 9 des 10 DMs les plus fréquents, en particulier pour ceux comportant le modal *can* en deuxième position. Ainsi, nos résultats suggèrent qu'il y a une forte association entre les DMs et la langue afro-américaine.

Double Modal Hotspot Compared to African American Population

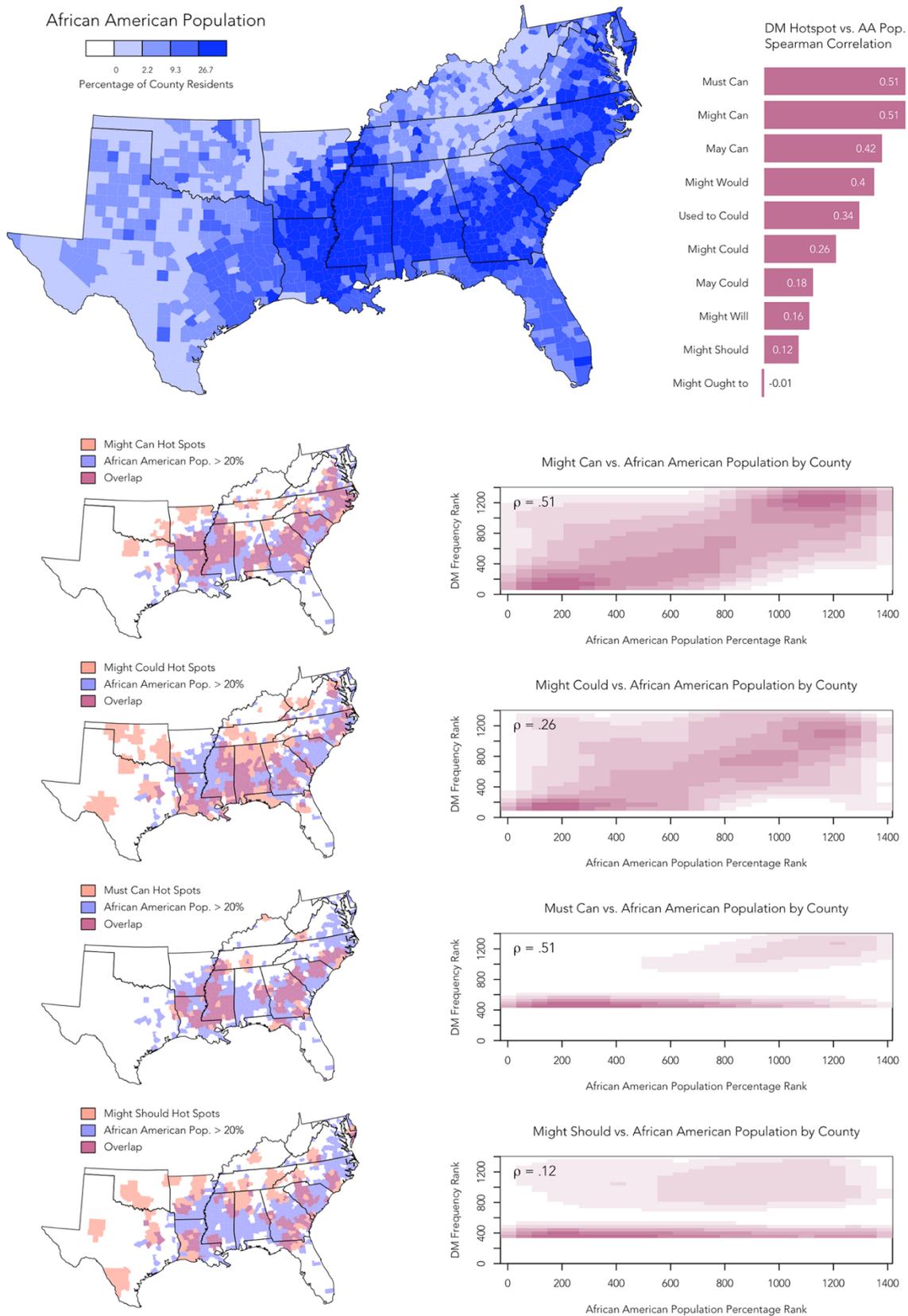


Figure 4 : cartes d'autocorrélation spatiale comparant les motifs d'emplois des DMs aux proportions démographiques afro-américaines au sud-est des États-Unis. <https://doi.org/10.1371/journal.pone.0295799.g006>

Nous constatons en effet que *might can*, le DM le plus fréquent dans notre corpus, est fortement corrélé aux proportions de population afro-américaine du Sud (coefficient de corrélation $\rho = 0,51$), tandis que *might could*, qui est traditionnellement considéré comme étant le plus courant, montre une corrélation beaucoup plus faible ($\rho = 0,26$). Cette différence est visible dans les cartes qui recourent ces deux distributions : *might can* est employé principalement dans les zones à forte population afro-américaine, tandis que *might could* est plus fréquent dans des régions à plus faible population afro-américaine (par exemple dans la Géorgie et le Mississippi).

Ce résultat révèle une variation sociale et régionale significative dans l'utilisation des DMs entre les locuteurs blancs et afro-américains dans le Sud, et suggère que ces deux communautés favorisent l'emploi de combinaisons modales distinctes. Les DMs comme *might can* indexeraient donc une identité afro-américaine plus marquée, tandis que d'autres DMs, comme *might could*, seraient plus neutres et associés à l'anglais parlé par les locuteurs blancs dans le Sud. Cette hypothèse pourrait expliquer pourquoi *might can* est plus fréquent que *might could* dans notre corpus : les populations afro-américaines n'ont peut-être pas été suffisamment représentées dans les recherches précédentes, ce qui aurait conduit à une sous-estimation de la variation dans l'emploi des DMs.

4.3 Une origine afro-américaine des doubles modaux ?

Enfin, nos résultats pourraient avoir des implications importantes pour la compréhension des origines diachroniques des DMs en anglais américain. La théorie dominante suppose que les DMs ont été introduits aux États-Unis par des immigrants écossais-irlandais, principalement dans les Appalaches. Cependant, nous trouvons que les DMs traditionnels sont plus courants dans le *Deep South*, en particulier les régions à forte population afro-américaine, ce qui remet en question cette théorie. Par exemple, le DM *used to could*, typique en scots, est plus commun dans le *Deep South* que dans les Appalaches, et d'autres DMs fréquents dans notre corpus (par exemple *may can* et *must would*) ne sont pas bien attestés dans les îles britanniques.

Nos résultats suggèrent donc que le système traditionnel des DMs pourrait être une innovation américaine, spécifiquement de la langue afro-américaine, plutôt qu'une importation écossaise-irlandaise. Bien que cette théorie afro-américaine ait été envisagée à une reprise dans la littérature (Fennell & Butters 1996), elle a été univoquement rejetée depuis. Cependant, nos données issues de Twitter tendent au contraire à soutenir cette théorie, en suggérant que les DMs pourraient avoir été diffusés par les locuteurs afro-américains aux locuteurs blancs dans le Sud. Par la suite, certains DMs auraient perdu leur signification sociale première au fil du temps. Ce ne serait pas la première fois qu'un fait de langue afro-américain ait été adopté en anglais américain général (Grieve et al. 2018).

Ceci étant dit, nous voulons souligner que nos résultats ne prouvent pas catégoriquement la théorie de l'origine afro-américaine des DMs. Nous soutenons simplement qu'elle mérite d'être sérieusement envisagée, contrairement à ses traitements dans les études passées. Si les DMs proviennent effectivement de la langue afro-américaine, il reste à établir comment et pourquoi ce système serait apparu. Par exemple, il serait nécessaire d'enquêter plus spécifiquement sur les liens entre le système des DMs américains et la créolisation, étant donné que ces faits de langue sont fréquents dans de nombreux créoles fondés sur l'anglais, notamment le gullah parlé aux États-Unis.

En d'autres termes, bien que ces théories diachroniques demeurent spéculatives et nécessitent plus de recherches, les résultats de notre étude, qui représente l'analyse empirique la plus complète des DMs à partir d'un échantillon d'occurrences naturelles, remettent en question les présupposés traditionnels sur leurs origines, et suggèrent de nouvelles pistes de recherche, notamment au sujet de leurs possibles origines afro-américaines.

Conclusion

Dans cet article, nous avons analysé l'emploi des DMs dans un corpus massif de données géolocalisées issues d'un réseau social américain, apportant trois nouveaux éclairages sur leur structure, leur utilisation et leurs origines en anglais américain. Premièrement, nous avons montré que la diversité des DMs est bien plus grande que ce qui était supposé auparavant, en identifiant 76 catégories de DMs sur les 110 possibles. Cette diversité suggère qu'il n'y a pas de contraintes absolues sur la formation des DMs, mais plutôt des pressions probabilistes d'usage linguistique.

Deuxièmement, nos résultats indiquent que les facteurs sociaux et régionaux influençant la variation dialectale des DMs sont également plus complexes qu'on ne le pensait. Les DMs sont plus fréquents dans les régions à forte population afro-américaine, particulièrement dans le *Deep South*. Nos résultats révèlent une association étroite avec la langue afro-américaine, et remettent en question l'idée que les DMs sont principalement utilisés par les Blancs dans les Appalaches.

Troisièmement, nos découvertes apportent des éléments importants pour le débat théorique sur les origines des DMs, en proposant que ces formes pourraient être une innovation afro-américaine, plutôt qu'une importation écossaise-irlandaise. Bien que nous ne puissions pas confirmer de manière définitive ces origines, une telle théorie mérite d'être sérieusement envisagée à l'avenir.

Enfin, nos résultats jettent les bases pour de futures recherches sur la syntaxe, la sémantique, et la signification sociolinguistique des DMs, en reconsidérant comment ces formes complexes sont employées pour exprimer des identités sociales distinctes aux États-Unis.

Bibliographie

- BEAL, Joan. 2004. « English Dialects in the North of England: Morphology and Syntax », in Bernd Kortmann et Edgar W. Schneider (éds.), *A Handbook of Varieties of English, Volume 2: Morphology and Syntax*. Berlin : Mouton de Gruyter, pp.114–141.
- BUTTERS, Ronald. 1973. « Acceptability Judgments for Double Modals in Southern Dialects », in Charles Bailey et Roger Shuy (éds.), *New Ways of Analyzing Variation in English*. Washington DC : Georgetown University Press, pp.276–286.
- COATES, Jennifer. 1983. *The Semantics of Modal Auxiliaries*. London, Canberra : Croom Helm.
- COATS, Steven. 2024. « Naturalistic Double Modals in North America », *American Speech*, volume 99, n°1, pp.47-77.
- . 2023. « Double Modals in Contemporary British and Irish Speech », *English Language and Linguistics*, volume 27, n°4, pp.693-718.
- COATS, Steven et MORIN, Cameron. À paraître. « Double Modals Beyond the Atlantic: New Evidence from Computational Sociolinguistics », *English Today*.
- CORRIGAN, Karen. 2011. « Grammatical Variation in Irish English », *English Today*, volume 27, n°2, pp.39-46.
- CROZIER, Alan. 1984. « The Scotch-Irish Influence on American English », *American Speech*, volume 59, n°4, pp.310-331.
- CUNNINGHAM, Irma. 1970. *A Syntactic Analysis of Sea Island Creole ('Gullah')* [PhD thesis]. University of Michigan.
- DIPAOLLO, Marianna. 1989. « Double Modals as Single Lexical Items », *American Speech*, volume 64, n°3, pp.195-224.
- DUGGAN, Maeve et BRENNER, Joanna. 2013. *The Demographics of Social Media Users—2012*.

Pew Research Center. Disponible sur : <https://www.pewresearch.org/internet/2013/02/14/the-demographics-of-social-media-users-2012>.

- ELSMAN, Minta et DUBINSKY, Stanley. 2009. « Double Modal Syntactic Patterns as Single Modal Interactions », *University of Pennsylvania Working Papers in Linguistics*, volume 5, n°1, pp.75-82.
- FEAGIN, Crawford. 1979. *Variation and Change in Alabama English: A Sociolinguistic Study of the White Community*. Washington DC : Georgetown University Press.
- FENNELL, Barbara et BUTTERS, Ronald. 1996. « Historical and Contemporary Distribution of Double Modals in English », in Edgar W. Schneider (éd.), *Focus on the USA*. Amsterdam/Philadelphia : John Benjamins, pp. 265–288.
- GRIEVE, Jack, HOVY, Dirk, JURGENS, David, KENDALL, Tyler S, NGUYEN, Dong, STANFORD, James N. et SUMNER, Meghan (éds.). 2023. *Computational Sociolinguistics*. Lausanne : Frontiers Media SA.
- GRIEVE, Jack, MONTGOMERY, Chris, NINI, Andrea, MURAKAMI, Akira et GUO, Diansheng. 2019. « Mapping Lexical Dialect Variation in British English Using Twitter », *Journal of English Linguistics*, volume 2, p.1-11.
- GRIEVE, Jack, NINI, Andrea et GUO, Diansheng. 2018. « Mapping Lexical Innovation on American Social Media », *Journal of English Linguistics*, volume 46, n°4, pp.293-319.
- HASTY, Daniel. 2012. *This Might Could Help Us Better Understand Syntactic Variation: The Double Modal Construction in Tennessee English*. Michigan State University.
- HUANG, Nick. 2011. *Multiple Modals*. Yale Grammatical Diversity Project: English in North America. Disponible sur : <https://ygdp.yale.edu/phenomena/multiple-modals>.
- HUANG, Yuan, GUO, Diansheng, KASAKOFF, Alice et GRIEVE, Jack. 2016. « Understanding US Regional Linguistic Variation with Twitter Data Analysis », *Computers, Environment and Urban Systems*, volume 59, pp.244-255.
- KORTMANN, Bernd, LUKENHEIMER, Kirsten et EHRET, Katharina (éds.). 2020. *The Electronic World Atlas of Varieties of English*. Zenodo.
- LABOV, William. 1972. *Language in the Inner City*. Philadelphia : University of Philadelphia Press.
- MCDAVID, Raven et O'CAIN, Raymond. 1980. *Linguistic Atlas of the Middle and South Atlantic States*. Chicago : Chicago University Press.
- MISHOE, Margaret et MONTGOMERY, Michael. 1994. « The Pragmatics of Multiple Modal Variation in North and South Carolina », *American Speech*, volume 69, n°1, pp.3-29.
- MONTGOMERY, Michael et NAGLE, Stephen. 1993. « Double Modals in Scotland and the United States: Trans-atlantic Inheritance or Independent Development? », *Folia Linguistica Historica*, volume 14, n°1, pp.91-107.
- MORIN, Cameron. 2021a. « Double Modals in Scots: a Speaker's Choice Hypothesis », in Laura Baranzini et Louis de Saussure (éds.), *Aspects of Tenses, Modality, and Evidentiality*. Leiden : Brill, pp.165–182.
- . 2021b. « Construire une Triangulation Méthodologique en Variation Syntaxiques Restreinte », *Échanges Linguistiques en Sorbonne 7*, pp. 48–71.
- . 2023. *Social Meaning in Construction Grammar: Double Modals in Dialects of English* [PhD thesis]. Université Paris-Cité.
- MORIN, Cameron, DESAGULIER, Guillaume et GRIEVE, Jack. 2020. « Dialect Syntax in

- Construction Grammar: Theoretical Benefits of a Constructionist Approach to Double Modals in English », *Belgian Journal of Linguistics*, volume 34, pp.252-262.
- MORIN, Cameron et COATS, Steven. 2023. « Double Modals in Australian and New Zealand English », *World Englishes*, pp.1-24.
- MORIN, Cameron et GRIEVE, Jack. 2024. « The Semantics, Sociolinguistics, and Origins of Double Modals in American English: New Insights from Social Media », *PLOS One*, volume 19, n°1.
- MORIN, Cameron, DESAGULIER, Guillaume et GRIEVE, Jack. 2024. « A Social Turn for Construction Grammar: Double Modals on British Twitter », *English Language and Linguistics*, pp.1-29.
- NAGLE, Stephen. 2003. « Double Modals in the Southern United States: Syntactic Structure or Syntactic Structures? », in Roberta Facchinetti, Frank Palmer et Manfred Krug (éds.), *Modality in Contemporary English*. Berlin, Boston : Mouton de Gruyter, pp.359–372.
- NAGLE, Stephen et HOLMES, Patsy. 2000. « On the Semantics of English Double Modals », in Christiane Dalton-Puffer et Nikolaus Ritt (éds.), *Words: Meaning, Structure, Function: A Festschrift for Dieter Katovsky*. Berlin & New York : Mouton de Gruyter, pp.217–234.
- PALMER, Frank. 2001 (1986). *Mood and Modality: Second Edition*. Cambridge : Cambridge University Press.
- PEDERSON, Lee, MCDANIEL, Susan, BAILEY, Guy et BASSETT, Marvin (éds.). 1986. *Linguistic Atlas of the Gulf States*. Athens : University of Georgia Press.
- REED, Paul et MONTGOMERY, Michael. 2012. *The MultiMo Database*. Disponible sur : <http://artsandsciences.sc.edu/multimo.welcome>.
- THRÁINSSON, Höskuldur et VIKNER, Sten. 1995. « Modals and Double Modals in the Scandinavian Languages », *Lund University Working Papers in Scandinavian Syntax*, volume 55, pp.51-88.
- WOLFRAM, Walt et CHRISTIAN, Donna. 1976. *Appalachian Speech*. Arlington, VA : Center for Applied Linguistics.
- ZULLO, David, PFENNINGER, Simone et SCHREIER, Daniel. 2021. « A Pan-Atlantic "Multiple Modal Belt"? », *American Speech*, volume 96, n°1, pp.1-46.

Pour citer cette ressource :

Cameron MORIN, "Les doubles modaux en anglais américain : nouvelles perspectives sémantiques, sociolinguistiques et historiques à partir d'un corpus de réseaux sociaux", *La Clé des Langues* [en ligne], Lyon, ENS de LYON/DGESCO (ISSN 2107-7029), octobre 2024. Consulté le 10/10/2024. URL: <https://cle.ens-lyon.fr/anglais/langue/linguistique/les-doubles-modaux-en-anglais-americain>