

SÉMINAIRE DELICORTAL

Sommes-nous sortis du bois ou de l'auberge ? Dépistage semi-automatique de la variation diatopique en langue spécialisée

25 octobre 2016

Patrick Drouin

Observatoire de linguistique Sens-Texte (OLST)

Université de Montréal

1. Plan
2. Qui suis-je ?
3. Variation terminologique
4. Problématique
5. Étude
6. Méthodologie
7. Résultats
8. Conclusion

Enseignement

- Terminologie

Enseignement

- Terminologie
- Localisation

Enseignement

- Terminologie
- Localisation
- Traduction et terminologie assistée par ordinateur

Enseignement

- Terminologie
- Localisation
- Traduction et terminologie assistée par ordinateur

Recherche

Directeur de l'Observatoire de linguistique Sens-Texte depuis 2015.

- Terminologie

Enseignement

- Terminologie
- Localisation
- Traduction et terminologie assistée par ordinateur

Recherche

Directeur de l'Observatoire de linguistique Sens-Texte depuis 2015.

- Terminologie
- Traitement automatique de la langue

Enseignement

- Terminologie
- Localisation
- Traduction et terminologie assistée par ordinateur

Recherche

Directeur de l'Observatoire de linguistique Sens-Texte depuis 2015.

- Terminologie
- Traitement automatique de la langue
- Linguistique de corpus

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Axes de recherche

- I Axe lexicologie et lexicographie

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Axes de recherche

- I Axe lexicologie et lexicographie
- II Axe terminographie

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Axes de recherche

- I Axe lexicologie et lexicographie
- II Axe terminographie
- III Axe traitement de corpus

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Axes de recherche

- I Axe lexicologie et lexicographie
- II Axe terminographie
- III Axe traitement de corpus
- IV Axe didactique des langues

Observatoire de linguistique Sens-Texte (OLST)

- Groupe de recherche multidisciplinaire
- Fondé en 1997 par Igor Mel'čuk
- Liens bien établis entre l'OLST et le LIDILEM

Axes de recherche

- I Axe lexicologie et lexicographie
- II Axe terminographie
- III Axe traitement de corpus
- IV Axe didactique des langues
- V Axe applications documentaires

Linguistique et traduction

- Lidjia Iordanskaja (Université de Montréal)
- Anna Joan Casademont (TÉLUQ)
- *Aline Francoeur (Université Laval)
- **François Lareau (Université de Montréal)**
- **Marie-Claude L'Homme (Université de Montréal)**
- Sébastien Marengo (Université de Sherbrooke)
- Elizabeth Marshman (Université d'Ottawa)
- Igor Mel'čuk (Université de Montréal)
- Jasmina Milićević (Dalhousie University)
- *Aurélien Picton (Université de Genève)
- Janine Pimentel (Universidade Federal do Rio de Janeiro)
- *Alain Polguère (ATILF, CNRS, Université de Lorraine)

Didactique

- Dominic Anctil (Université de Montréal)
- Pascale Lefrançois (Université de Montréal)
- Ophélie Tremblay (UQAM)

Didactique

- Dominic Anctil (Université de Montréal)
- Pascale Lefrançois (Université de Montréal)
- Ophélie Tremblay (UQAM)

Sciences de l'information

- Lyne Da Sylva (Sciences de l'information, Université de Montréal)
- Dominic Forest (Sciences de l'information, Université de Montréal)

VARIATION TERMINOLOGIQUE

Pourquoi ?

- Travaux de Wüster (1931)

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie
- Mise en évidence lente, mais inévitable des faiblesses du postulat de base de Wüster

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie
- Mise en évidence lente, mais inévitable des faiblesses du postulat de base de Wüster
- Approche aménagiste et la socioterminologie (variation diatopique/régionale, diastratique/sociale, diaphasique/stylistique)

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie
- Mise en évidence lente, mais inévitable des faiblesses du postulat de base de Wüster
- Approche aménagiste et la socioterminologie (variation diatopique/régionale, diastratique/sociale, diaphasique/stylistique)
- Traitement automatique de la langue (variation textuelle)

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie
- Mise en évidence lente, mais inévitable des faiblesses du postulat de base de Wüster
- Approche aménagiste et la socioterminologie (variation diatopique/régionale, diastratique/sociale, diaphasique/stylistique)
- Traitement automatique de la langue (variation textuelle)
- Évolution accélérée des connaissances (variation diachronique)

Pourquoi ?

- Travaux de Wüster (1931)
- Intérêt récent pour la variation en terminologie
- Mise en évidence lente, mais inévitable des faiblesses du postulat de base de Wüster
- Approche aménagiste et la socioterminologie (variation diatopique/régionale, diastratique/sociale, diaphasique/stylistique)
- Traitement automatique de la langue (variation textuelle)
- Évolution accélérée des connaissances (variation diachronique)
- Disponibilité de corpus représentatifs de la variation

PROBLÉMATIQUE

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique
- Mécanismes de description sont relativement bien établis dans les dictionnaires

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique
- Mécanismes de description sont relativement bien établis dans les dictionnaires
- Dépistage de la variation, même dans cette tradition plusieurs fois centenaire, laisse à désirer

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique
- Mécanismes de description sont relativement bien établis dans les dictionnaires
- Dépistage de la variation, même dans cette tradition plusieurs fois centenaire, laisse à désirer
- Lexicographe demeure le point fort de la démarche : connaissances et culture générale

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique
- Mécanismes de description sont relativement bien établis dans les dictionnaires
- Dépistage de la variation, même dans cette tradition plusieurs fois centenaire, laisse à désirer
- Lexicographe demeure le point fort de la démarche : connaissances et culture générale
- Le problème, c'est la mort.

Peut-on apprendre de la lexicographie ?

- Variation bien prise en compte dans le cadre de la description lexicographique
- Mécanismes de description sont relativement bien établis dans les dictionnaires
- Dépistage de la variation, même dans cette tradition plusieurs fois centenaire, laisse à désirer
- Lexicographe demeure le point fort de la démarche : connaissances et culture générale
- Le problème, c'est la mort.
- Connaissances disparaissent avec le lexicographe...

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage
- Exemples
 - pastèque, melon d'eau (liée au territoire)

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage
- Exemples
 - pastèque, melon d'eau (liée au territoire)
 - mitaine, mouffle

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage
- Exemples
 - pastèque, melon d'eau (liée au territoire)
 - mitaine, mouffle
 - céphalée, mal de tête (liée au niveau de langue)

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage
- Exemples
 - pastèque, melon d'eau (liée au territoire)
 - mitaine, mouffle
 - céphalée, mal de tête (liée au niveau de langue)
 - maladie du cœur, maladie cardiaque (reformulation en discours)

De quoi est-il question précisément ?

- Principalement de ce qui est touché par les marques d'usage
- Exemples
 - pastèque, melon d'eau (liée au territoire)
 - mitaine, mouffle
 - céphalée, mal de tête (liée au niveau de langue)
 - maladie du cœur, maladie cardiaque (reformulation en discours)
 - écologie, sciences de l'environnement (liée à l'époque)

Hypothèses

- La variation, comme les termes, est observable en corpus

Hypothèses

- La variation, comme les termes, est observable en corpus
- L'humain n'est pas toujours à même de dépister le phénomène (connaissances, culture, immersion)

Hypothèses

- La variation, comme les termes, est observable en corpus
- L'humain n'est pas toujours à même de dépister le phénomène (connaissances, culture, immersion)
- La variation peut ou pourra être dégagée à l'aide d'outils statistiques

Hypothèses

- La variation, comme les termes, est observable en corpus
- L'humain n'est pas toujours à même de dépister le phénomène (connaissances, culture, immersion)
- La variation peut ou pourra être dégagée à l'aide d'outils statistiques
- Les outils et techniques déjà disponibles en TAL permettent de prendre en charge un premier dépistage

Hypothèses

- La variation, comme les termes, est observable en corpus
- L'humain n'est pas toujours à même de dépister le phénomène (connaissances, culture, immersion)
- La variation peut ou pourra être dégagée à l'aide d'outils statistiques
- Les outils et techniques déjà disponibles en TAL permettent de prendre en charge un premier dépistage
- Ce dépistage peut appuyer ou alimenter l'intuition du linguiste

Objectifs

- Dépister les formes qui semblent les plus symptomatiques d'une variation potentielle (en général)

Objectifs

- Dépister les formes qui semblent les plus symptomatiques d'une variation potentielle (en général)
- Identifier automatiquement des traces de variation (diatopique pour cette présentation)

Objectifs

- Dépister les formes qui semblent les plus symptomatiques d'une variation potentielle (en général)
- Identifier automatiquement des traces de variation (diatopique pour cette présentation)
- Proposer au terminologue des pistes de réflexion pour compléter son intuition ou aiguiller sa recherche

Objectifs

- Dépister les formes qui semblent les plus symptomatiques d'une variation potentielle (en général)
- Identifier automatiquement des traces de variation (diatopique pour cette présentation)
- Proposer au terminologue des pistes de réflexion pour compléter son intuition ou aiguiller sa recherche
- Mettre de l'avant une méthodologie qui se prête autant au travail lexicographique que terminographique

MÉTHODOLOGIE

Les corpus utilisés

Constructeur	Canada	France
Audi	130 463	819 195
Bentley	15 677	46 214
BMW	342 835	583 282
Chevrolet	392 419	163 434
Ford	674 228	765 728
Honda	1 526 859	58 805
Hyundai	118 264	398 242
Infiniti	99 373	58 414
KIA	344 153	56,335
Landrover	78 337	83 938
Lexus	93 114	70 609
McLaren	8 494	12 696
Nisan	123 139	475 201
Porsche	1 118, 20	1 433 450
Smart	38 227	51 359
Volkswagen	463 092	468 979

Les corpus utilisés

Pays	Canada	France
Formes	22,827	28,407
Occurrences	4,147,354	5,240,698

Préparation des corpus

- Tous les corpus ont été aspirés du Web avec wget (disponible sur plusieurs plateformes)

Préparation des corpus

- Tous les corpus ont été aspirés du Web avec wget (disponible sur plusieurs plateformes)
- Les corpus ont été convertis en format texte à l'aide de scripts Python

Préparation des corpus

- Tous les corpus ont été aspirés du Web avec wget (disponible sur plusieurs plateformes)
- Les corpus ont été convertis en format texte à l'aide de scripts Python
- Certains sites comportaient d'importantes sections en anglais, elles ont été exclues (scripts Python)

Préparation des corpus

- Tous les corpus ont été aspirés du Web avec wget (disponible sur plusieurs plateformes)
- Les corpus ont été convertis en format texte à l'aide de scripts Python
- Certains sites comportaient d'importantes sections en anglais, elles ont été exclues (scripts Python)

Fiabilité des corpus

- Les sites Web sont-ils réellement localisés ?

Préparation des corpus

- Tous les corpus ont été aspirés du Web avec wget (disponible sur plusieurs plateformes)
- Les corpus ont été convertis en format texte à l'aide de scripts Python
- Certains sites comportaient d'importantes sections en anglais, elles ont été exclues (scripts Python)

Fiabilité des corpus

- Les sites Web sont-ils réellement localisés ?
- Le contenu des sites des constructeurs couvre différents aspects (loi, marketing, véhicules, etc.)

Les outils

Deux outils ont été utilisés pour l'analyse des données.

Les outils

Deux outils ont été utilisés pour l'analyse des données.

IRaMuTeQ

Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires

Outil conçu par Pierre Ratinaud de Toulouse qui permet de procéder à diverses analyses statistiques sur des corpus assez volumineux.

Les outils

Deux outils ont été utilisés pour l'analyse des données.

IRaMuTeQ

Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires

Outil conçu par Pierre Ratinaud de Toulouse qui permet de procéder à diverses analyses statistiques sur des corpus assez volumineux.

Diatopix

Outil conçu à l'OLST qui permet de visualiser de façon graphique la répartition de mots, de termes, d'expressions, etc. dans un espace Web délimité par une langue.

Démonstration (?) : <http://olst.ling.umontreal.ca/diatopix>

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre
- Le **calcul des spécificités** (Lafon 1984) permet d'identifier les formes dont la fréquence dévie de *la norme*

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre
- Le **calcul des spécificités** (Lafon 1984) permet d'identifier les formes dont la fréquence dévie de *la norme*
- Trois types de résultats : les spécificités négatives, les formes banales et les **spécificités positives**

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre
- Le **calcul des spécificités** (Lafon 1984) permet d'identifier les formes dont la fréquence dévie de *la norme*
- Trois types de résultats : les spécificités négatives, les formes banales et les **spécificités positives**
- L'outil calcule les fréquences sur les lemmes

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre
- Le **calcul des spécificités** (Lafon 1984) permet d'identifier les formes dont la fréquence dévie de *la norme*
- Trois types de résultats : les spécificités négatives, les formes banales et les **spécificités positives**
- L'outil calcule les fréquences sur les lemmes
- En plus du calcul des spécificités, une mise en opposition présence/absence des corpus est utilisée

Pourquoi IRaMuTeQ ?

- La méthodologie est basée sur une intuition qu'il est possible statistiquement d'aller cibler les formes qui se comportent différemment d'un corpus à l'autre
- Le **calcul des spécificités** (Lafon 1984) permet d'identifier les formes dont la fréquence dévie de *la norme*
- Trois types de résultats : les spécificités négatives, les formes banales et les **spécificités positives**
- L'outil calcule les fréquences sur les lemmes
- En plus du calcul des spécificités, une mise en opposition présence/absence des corpus est utilisée
- Pourquoi ne pas utiliser TermoStat ?
(<http://termostat.ling.umontreal.ca>)

Pourquoi Diatopix ?

- Validation des données obtenues à partir d'IRaMuTeQ

Pourquoi Diatopix ?

- Validation des données obtenues à partir d'IRaMuTeQ
- Vérification externe des données obtenues sur un corpus assez limité

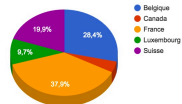
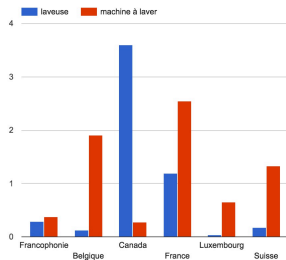
Pourquoi Diatopix ?

- Validation des données obtenues à partir d'IRaMuTeQ
- Vérification externe des données obtenues sur un corpus assez limité
- Visualisation assez utile des répartitions des formes dans les 2 corpus

Pourquoi Diatopix ?

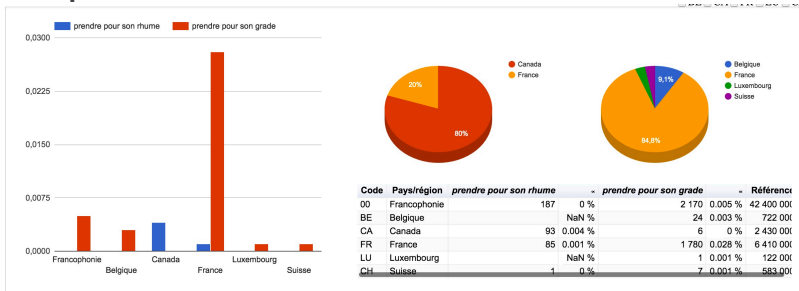
- Validation des données obtenues à partir d'IRaMuTeQ
- Vérification externe des données obtenues sur un corpus assez limité
- Visualisation assez utile des répartitions des formes dans les 2 corpus
- Obtention de données chiffrées plus importantes sur les répartitions

Diatopix



Code	Pays/région	laveuse	%	machine à laver	%	Référence
00	Francophonie	120 000	0.287 %	158 000	0.378 %	41 800 000
BE	Belgique	876	0.121 %	13 800	1.911 %	722 000
CA	Canada	87 400	3.597 %	6 770	0.279 %	2 430 000
FR	France	76 400	1.192 %	163 000	2.543 %	6 410 000
LU	Luxembourg	40	0.034 %	775	0.651 %	119 000
CH	Suisse	996	0.171 %	7 780	1.334 %	583 000

Diatopix



Les limites de Diatopix

- L'outil ne peut exploiter que les noms de domaine sur le Web

Les limites de Diatopix

- L'outil ne peut exploiter que les noms de domaine sur le Web
- La localisation est parfois trompeuse

Les limites de Diatopix

- L'outil ne peut exploiter que les noms de domaine sur le Web
- La localisation est parfois trompeuse
- Les résultats ne contiennent que ce qui est écrit et sur le Web

Les limites de Diatopix

- L'outil ne peut exploiter que les noms de domaine sur le Web
- La localisation est parfois trompeuse
- Les résultats ne contiennent que ce qui est écrit et sur le Web
- La polysémie peut jouer des drôles de tours

Validation des données

- Validation manuelle des données

Validation des données

- Validation manuelle des données
- Exploitation des 4 principales sources :
 - *IATE*
 - *Le grand dictionnaire terminologique*
 - *TermiumPlus*
 - *Usito*

Validation des données

- Validation manuelle des données
- Exploitation des 4 principales sources :
 - *IATE*
 - *Le grand dictionnaire terminologique*
 - *TermiumPlus*
 - *Usito*
- Sources suffisantes dans la majorité des cas, parfois appel à experts

Validation des données

- Validation manuelle des données
- Exploitation des 4 principales sources :
 - *IATE*
 - *Le grand dictionnaire terminologique*
 - *TermiumPlus*
 - *Usito*
- Sources suffisantes dans la majorité des cas, parfois appel à experts
- Forme est jugée pertinente quand son sens est relié au domaine de l'industrie automobile

Présence/absence

- Formes présentes dans le corpus Canada uniquement : 1343 (f. 10-3299)

Présence/absence

- Formes présentes dans le corpus Canada uniquement : 1343 (f. 10-3299)
- Formes présentes dans le corpus France uniquement : 1993 (f. 10-3452)

Présence/absence

- Formes présentes dans le corpus Canada uniquement : 1343 (f. 10-3299)
- Formes présentes dans le corpus France uniquement : 1993 (f. 10-3452)
- Analyse des 200 premières formes dans chacune des listes

Présence/absence

- Formes présentes dans le corpus Canada uniquement : 1343 (f. 10-3299)
- Formes présentes dans le corpus France uniquement : 1993 (f. 10-3452)
- Analyse des 200 premières formes dans chacune des listes
- Filtrage des noms de lieux, de modèles et des abréviations

Présence/absence

- Formes présentes dans le corpus Canada uniquement : 1343 (f. 10-3299)
- Formes présentes dans le corpus France uniquement : 1993 (f. 10-3452)
- Analyse des 200 premières formes dans chacune des listes
- Filtrage des noms de lieux, de modèles et des abréviations
- Listes résultantes : 60 pour le Canada et 20 pour la France

Canada

Rank	Form	Relevant	Regional	Canada	France	Ratio
1	sedan	Yes	Yes	3.33%	2.86%	1.16
2	<i>livrable (adj.)</i>	No	No	0.06%	0.57%	0.11
3	<i>magasiner</i>	No	Yes	16.06%	1.88%	8.54
4	<i>courriel</i>	No	Yes	27.56%	14.20%	1.94
5	<i>informationnel</i>	No	No	0.22%	1.00%	0.22
6	télédéverrouillage	Yes	Yes	1.63%	0.02%	81.5
7	<i>desservir</i>	No	No	0.84%	3.47%	0.24
8	<i>radiocommunication</i>	Yes *	Yes	0.25%	2.13%	0.11
9	<i>galvanisation</i>	No	No	0.15%	1.41%	0.11
10	<i>ultrasonique</i>	Yes *	No †	Sparse	Sparse	Unknown
11	antirouille	Yes *	Yes	0.33%	0.11%	3.00
12	<i>cueillette</i>	No	No	7.40%	3.05%	2.42
13	infoloisirs	Yes *	Yes †	Sparse	None	+
14	hydro-électrique	Yes *	Yes †	Sparse	None	+
15	sous-marinage	Yes *	Yes †	Sparse	None	+
16	<i>gradabilité</i>	No	No	Sparse	Sparse	Unknown
17	<i>approuvé</i>	No	No	7.94%	5.62%	1.41
18	multiplace	Yes *	Yes †	0.15%	Sparse	+
19	écoénergétique	Yes *	Yes	0.80%	0.02%	40.00
20	<i>déversement</i>	No	No	0.89%	0.31%	2.87

Canada - Fréquence

Forme

sedan

télédéverrouillage

antirouille

infoloisirs

hydro-électrique

sous-marinage

multiplace

écoénergétique

Complexe

traitement antirouille

système infoloisirs

servodirection hydro-électrique

rampe de sous-marinage

cabine multiplace

moteur écoénergétique

France

Rank	Form	Relevant	Regional	France	Canada	Ratio
1	<i>surteintées</i>	Yes *	Yes	0.11%	Sparse	+
2	<i>labélisés</i>	Yes *	Yes †	1.05%	0.01%	105.00
3	<i>monospace</i>	Yes	Yes	10.98%	0.09%	122.00
4	<i>radiocommande</i>	Yes *	Yes	4.13%	0.01%	413.00
5	<i>contracter</i>	No	Yes	7.57%	1.91%	3.96
6	<i>déconnectable</i>	Yes *	Yes	1.90%	Sparse	+
7	<i>infotainment</i>	Yes *	Yes	0.48%	0.04%	12.00
8	<i>camping-car</i>	Yes	Yes	4.71 %	0.83%	5.68
9	<i>non-fumeur</i>	No	No	8.94%	5.64%	1.59
10	<i>candy</i>	Yes	Yes	7.74%	3.03%	2.55
11	<i>croquis</i>	No	No	9.03%	2.07%	4.36
12	<i>athermique</i>	Yes *	Yes	0.26%	Sparse	+
13	<i>allume-cigare</i>	Yes	Yes	2.56%	0.12%	21.33
14	<i>présignalisation</i>	Yes *	Yes	0.19%	Sparse	+
15	<i>répétiteur</i>	Yes *	Yes	0.03%	Sparse	Unknown
16	<i>showroom</i>	No	Yes	10.18%	3.41%	2.99
17	<i>majoré</i>	Yes *	Yes	1.85%	0.49%	3.78
18	<i>emprunteur</i>	No	No	4.09%	0.42%	9.74
19	<i>anodisées</i>	Yes	No	2.29%	0.33%	6.94
20	<i>tweeters</i>	Yes	Yes	1.58%	0.07%	22.57

France - Fréquence

Forme

surteintées

labélisés

monospace

radiocommande

déconnectable

infotainment

camping-car

candy

athermique

allume-cigare

présignalisation

répétiteur

majoré

tweeters

Complexe

lunette surteintée

concessionnaire labélisé

clé pliante de radiocommande

airbag déconnectable

parebrise athermique

triangle de présignalisation

répétiteur de clignotant

loyer majoré

Spécificités

- Le calcul est effectué sur l'ensemble des formes du corpus

Spécificités

- Le calcul est effectué sur l'ensemble des formes du corpus
- Seules les 20 premières spécificités positives ont été analysées

Canada

Rank	Form	Relevant	Regional	Canada	France	Ratio
1	servodirection	Yes	Yes	0.27%	0.04%	6.75
2	bouton	No	No	7.83%	27.86%	0.28
3	cote	Yes *	Yes	0.12%	Sparse	+
4	retenue	Yes *	No	0.12%	0.25%	0.48
5	louer	No	No	11.35%	16.09%	0.71
6	période	No *	No	12.43%	31.75%	0.39
7	code	No *	No	23.03%	21.47%	1.07
8	catalytique	Yes *	No	0.08%	0.09%	0.88
9	caractéristique	Yes *	No	0.01%	0.01%	1.00
10	location	Yes	No	21.77%	25.63%	0.85
11	variable	No	No	14.29%	10.10%	1.42
12	arbre	Yes *	No	9.52%	24.67%	0.39
13	frein	Yes	Yes	0.72%	Sparse	+
14	stabilité	Yes *	No	9.66%	13.64%	0.39
15	contrôle	Yes *	No	15.46%	31.46%	0.49
16	propriétaire	No	No	15.55%	22.52%	0.69
17	frais	No *	No	18.85%	31.75%	0.59
18	fonction	No	No	19.27%	28.01%	0.69
19	latéral	Yes *	No	1.51%	7.04%	0.21
20	roue	Yes *	No	11.35%	37.09%	0.31

Canada - Spécificités

Forme

servodirection

cote

frein

Complexe

servodirection hydro-électrique

cote de consommation

freins antiblocage ABS

France

Rank	Form	Relevant	Regional	France	Canada	Ratio
1	gérer	No	No	8.08%	5.06%	1.60
2	branche	Yes *	No	21.84%	6.24%	3.5
3	sellerie	Yes	Yes	4.12%	0.22%	0.66
4	adapter	No	No	24.63%	6.32%	4.00
5	proposer	No	No	19.23%	20.50%	0.94
6	loyer	Yes	Yes	3.48%	7.00%	0.50
7	réinitialiser	No	No	46.32%	4.23%	10.9
8	valeur	No	No	38.89%	16.24%	2.39
9	intérêt	No	No	24.35%	19.31%	1.26
10	business	Yes *	Yes	57.97%	24.26%	2.39
11	département	No	No	27.59%	6.40%	4.31
12	côte	Yes *	No†	1.82%	0.15%	12.13
13	rendement	Yes *	No	14.15%	5.15%	2.75
14	concept	Yes *	Yes†	1.75%	0.04%	43.75
15	poursuivre	No	No	5.85%	8.75%	0.67
16	assurance	No	No	6.00%	21.73%	0.28
17	carburant	Yes	Yes	6.10%	4.31%	1.42
18	couler	Yes *	No†	7.23%	1.11%	6.51
19	extérieur	No	No	16.89%	13.42%	1.26
20	teinte	No	No	2.03%	1.91%	1.06

France - Spécificités

Forme

sellerie

loyer

business

concept

carburant

Complexe

gamme business

concept car

Conclusion

- Il est possible d'appuyer le lexicographe ou le terminologue dans le dépistage des variantes diatopiques

Conclusion

- Il est possible d'appuyer le lexicographe ou le terminologue dans le dépistage des variantes diatopiques
- Les résultats sont bruités

Conclusion

- Il est possible d'appuyer le lexicographe ou le terminologue dans le dépistage des variantes diatopiques
- Les résultats sont bruités
- La simple technique de comparaison présence/absence donne des résultats surprenants (8/20 et 14/20)

Conclusion

- Il est possible d'appuyer le lexicographe ou le terminologue dans le dépistage des variantes diatopiques
- Les résultats sont bruités
- La simple technique de comparaison présence/absence donne des résultats surprenants (8/20 et 14/20)
- Les résultats du calcul des spécificités sont plus denses d'un point de vue terminologique

Conclusion

- Il est possible d'appuyer le lexicographe ou le terminologue dans le dépistage des variantes diatopiques
- Les résultats sont bruités
- La simple technique de comparaison présence/absence donne des résultats surprenants (8/20 et 14/20)
- Les résultats du calcul des spécificités sont plus denses d'un point de vue terminologique
- La performance de la spécificité n'est pas aussi intéressante pour la variation diatopique (3/20 et 5/20)