

Sémantique distributionnelle et diachronie

les intensifieurs de l'anglais

Guillaume Desagulier ¹ Ilaine Wang ²

¹MoDyCo (UMR 7114) Paris 8, CNRS, Paris Nanterre & Inst. Univ. de France

²U. d'Orléans, LIFO (EA 4022) – U. de Tours, LIFAT (EA 6300)

Séminaire LIFO
Université d'Orléans
20 mai 2019

outline

- ① introduction
- ② VNC
- ③ densité
- ④ HistWords
- ⑤ discussion

questions

comment rendre compte de l'évolution sémantique de marqueurs ?

- comment capturer le sens en corpus ?
- comment distinguer des étapes dans l'évolution ?
- avec quelle granularité temporelle ?
 - artificielle ? (1880s, 1890s, 1900s, etc.)
 - "naturelle" ? (1882–1908, 1909–1924, etc.)

deux approches



linguistique de corpus



TAL

études de cas

split infinitive (to ADV V_{inf})

*Gradually they began **to better understand** each other's points of view.*

*In order **to fully understand** the case it is necessary to explain the law.*

*They do not seem **to really know** how to go to work to do it.*

intensifieurs de l'anglais

completely, totally, perfectly, absolutely, entirely, utterly, slightly, somewhat, rather, pretty, fairly, quite, most, very, extremely, highly, awfully, terribly, almost, frightfully, jolly

Desagulier (2014)

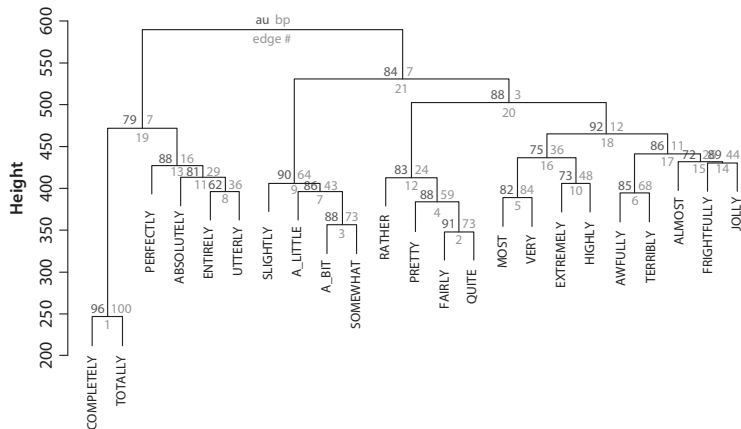


Figure 1. Cluster dendrogram of 23 degree modifiers of adjectives in English, clustered according to their adjectival collexemes (distance: Canberra; cluster method: Ward)

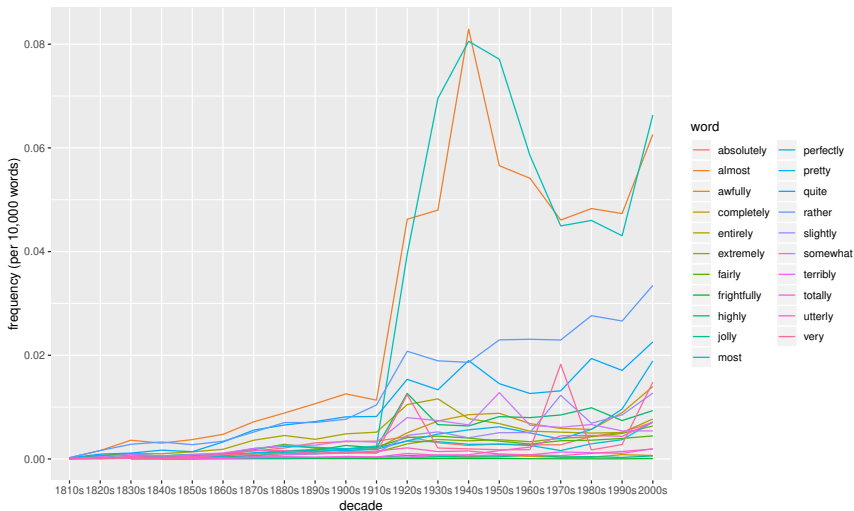
comment procède-t-on d'habitude ?

- ① extraction d'occurrences à partir d'un corpus diachronique, par ex. Davies (2010) ;
- ② classement de ces occurrences (un ou plusieurs critères) ;
- ③ faire état de la co-occurrence à plusieurs points dans l'échelle de temps.

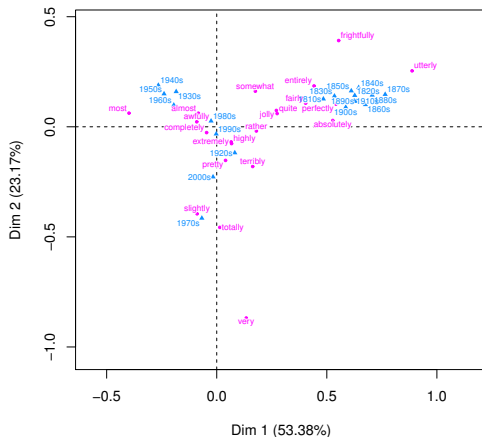
distribution des intensifieurs par décennie dans COHA

word	1810s	1820s	1830s	1840s	1850s	1860s	1870s	1880s	1890s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s
absolutely	2	9	9	16	16	34	63	86	126	136	140	175	127	106	117	111	111	178	210	287
almost	9	66	147	124	152	193	291	360	433	510	461	1878	1950	3368	2298	2199	1872	1962	1923	2543
awfully	0	0	1	1	0	3	0	0	8	6	4	10	16	18	28	21	14	13	34	27
completely	4	4	8	8	12	40	39	62	76	64	83	205	297	347	358	277	239	232	363	568
entirely	2	30	42	41	56	76	147	184	154	197	210	426	471	316	277	218	210	203	207	313
extremely	0	9	6	10	7	16	26	44	41	54	52	118	153	142	151	135	162	174	200	259
fairly	3	10	11	12	25	39	73	115	85	76	94	168	179	164	140	117	143	147	161	181
frightfully	0	0	0	1	0	0	2	2	4	0	2	3	3	3	4	2	0	1	1	3
highly	3	24	35	24	34	34	41	62	67	106	87	514	269	259	332	324	345	401	301	379
jolly	0	0	1	1	2	7	5	6	11	11	12	27	21	15	17	32	16	18	12	24
most	4	3	19	16	15	23	22	53	53	79	101	1604	2825	3273	3132	2377	1826	1869	1749	2695
perfectly	2	14	32	24	36	39	82	99	92	80	89	140	135	115	116	106	66	119	151	289
pretty	3	9	11	13	10	24	49	60	61	67	75	142	195	228	252	204	156	232	391	768
quite	8	38	46	69	57	133	226	266	293	330	333	624	542	771	590	513	534	787	694	917
rather	11	66	115	133	112	139	211	285	286	311	423	844	769	757	933	938	932	1123	1081	1360
slightly	0	4	10	6	7	7	17	32	39	47	45	185	212	166	206	207	499	280	345	516
somewhat	3	17	36	23	33	47	82	107	108	141	132	324	301	267	520	264	248	270	220	220
terribly	1	0	4	2	1	6	10	15	14	14	14	15	32	25	33	33	55	49	44	80
totally	1	5	8	6	6	5	15	18	12	20	15	43	30	33	66	93	181	185	177	289
utterly	1	6	19	27	17	42	67	65	44	45	62	87	58	63	39	32	26	44	59	78
very	0	2	7	5	11	33	41	41	43	57	49	504	86	79	72	73	742	71	113	602

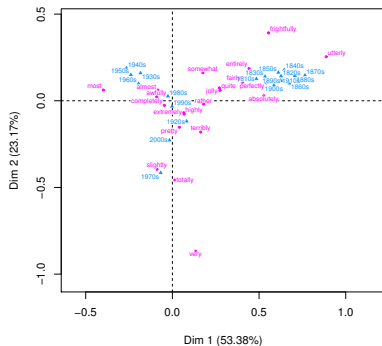
distribution des intensifieurs par décennie dans COHA



analyse des correspondances



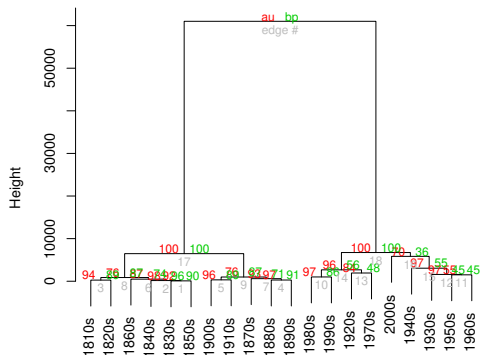
analyse des correspondances



problème

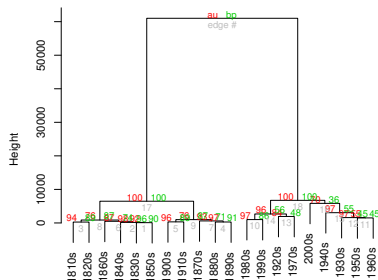
- difficile de cerner des étapes diachroniques
- difficile de cerner l'évolution du sens

classification ascendante hiérarchique



Distance: manhattan
Cluster method: ward.D

classification ascendante hiérarchique



problème

- difficile de cerner des étapes diachroniques
- pas d'adjacence entre les décennies

VNC

Hilpert et Gries (2008), Gries et Hilpert (2008), Gries et Hilpert (2010), Hilpert et Gries (2016)

Variability-based Neighbour Clustering

méthode de périodisation automatique

fondée sur la CAH

-> les périodes sont regroupées en fonction de leurs similitudes

différence : non-séparation de périodes adjacentes

VNC

algorithme

- ➊ point de départ : le corpus et ses délimitations arbitraires (en années, en décennies) ; ex. 1880s, 1890s, 1900s, 1910s, etc. ;
- ➋ les périodes adjacentes sont passées en revues par paires ; ex. 1880s–1890s, 1890s–1900s, 1900s–1910s, etc. ;
- ➌ l'indice de similarité entre les paires de périodes adjacentes est calculé en fonction d'une quantification prédéfinie (fréquences brutes/relatives de type/d'occurrence/d'hapax, mesure d'association, etc.) ;
- ➍ les périodes les plus proches sont regroupées en premier, sans déroger au principe d'adjacence ;
- ➎ calculer la valeur moyenne du regroupement ;
- ➏ répéter les étapes précédentes jusqu'à ne plus obtenir qu'un regroupement.

VNC à une variable

split vs. unsplit infinitive

Star Trek

Space : The final frontier

These are the voyages of the Starship, Enterprise

Its 5-year mission

To explore strange new worlds

To seek out new life and new civilizations

***To boldly go** where no man has gone before.*

source : <https://corpling.hypotheses.org/30>

motion chart

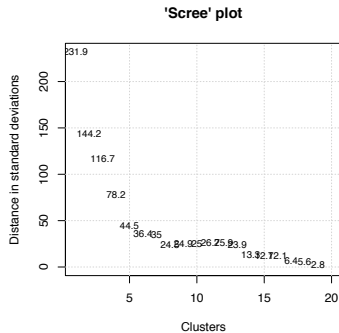
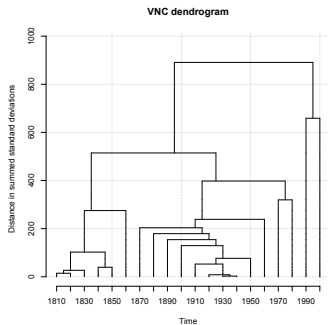
VNC à une variable

split vs. unsplit infinitive (COHA)

décennie	split infinitive	unsplit infinitive
1810	1	33
1820	0	334
1830	1	555
1840	6	662
1850	7	683
1860	19	656
1870	21	786
1880	44	833
1890	28	842
1900	56	806
1910	41	865
1920	28	799
1930	16	774
1940	15	651
1950	37	727
1960	49	656
1970	70	654
1980	80	624
1990	174	610
2000	241	604

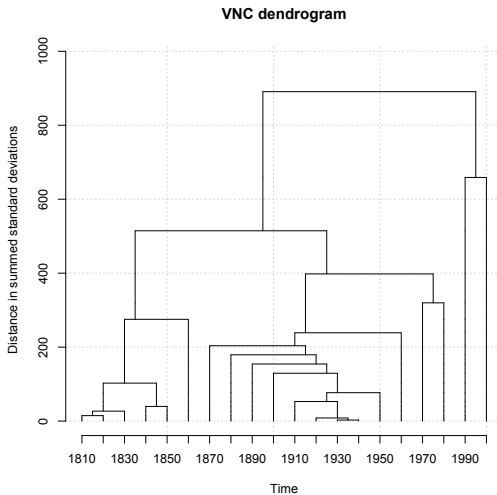
VNC à une variable

split infinitive (COHA) : *ex.to boldly go*



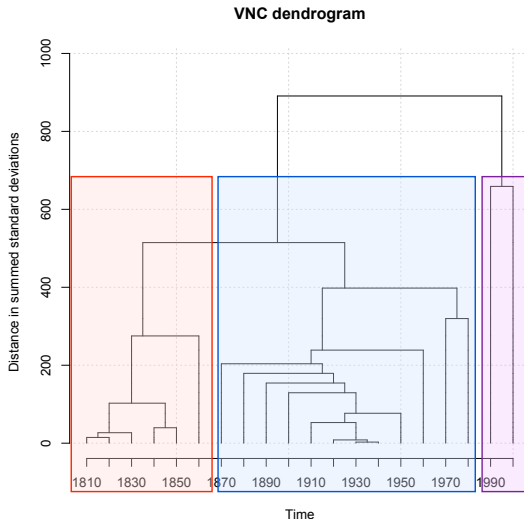
VNC à une variable

split infinitive (COHA) : *ex.to boldly go*



VNC à une variable

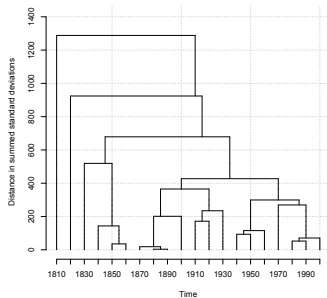
split infinitive (COHA) : *ex.to boldly go*



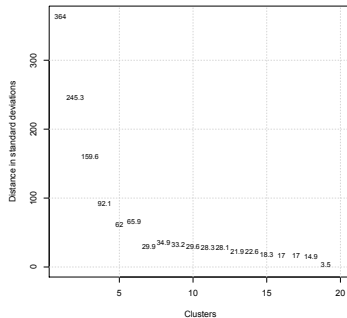
VNC à une variable

unsplit infinitive (COHA) : *ex.to go boldly*

VNC dendrogram



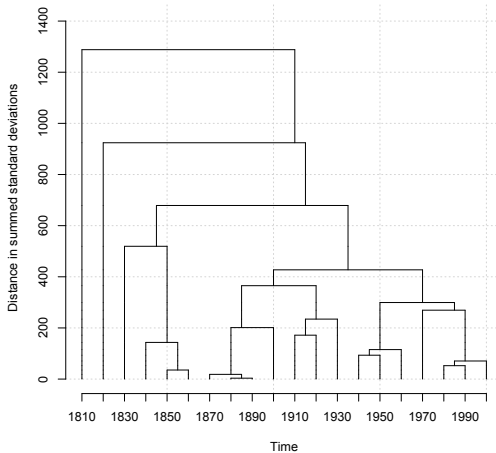
'Scree' plot



VNC à une variable

unsplit infinitive (COHA) : ex. *to go boldly*

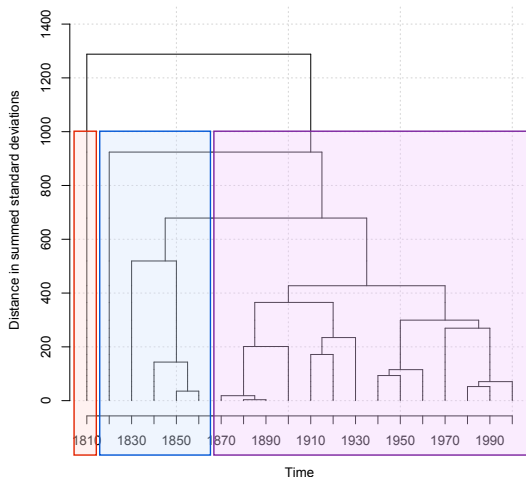
VNC dendrogram



VNC à une variable

unsplit infinitive (COHA) : ex. *to go boldly*

VNC dendrogram



VNC et MSD

Il est possible d'appliquer VNC aux **modèles de sémantique distributionnelle** (Perek et Hilpert 2017)

1810	1820	1830	1840	1850	1860	1870	1880	1890	...
0.133642342	0.133642342	0.133642342	0.247912649	0.133642342	0.150074419	0.102354937	0.102354937	0.373090579	...
0.127209096	0.127209096	0.108286251	0.166478831	0.149504469	0.119933336	0.186527314	0.119933336	0.119933336	...
0.11249435	0.11249435	0.1252518	0.1252518	0.1252518	0.134467191	0.134467191	0.134467191	0.243805699	...
0.04282996	0.072322208	0.073869258	0.060016241	0.039073274	0.053682251	0.370257283	0.045623489	0.095285637	...
0.041885443	0.041885443	0.112015072	0.041885443	0.045830093	0.045830093	0.045830093	0.118704853	0.045830093	...
0.325555253	0.065139754	0.03247229	0.043042183	0.03247229	0.04739629	0.031634915	0.095126314	0.031634915	...
...

VNC et MSD

hypothèse distributionnelle

- « You shall know a word by the company it keeps. » (Firth 1957)
- le degré de similitude sémantique entre deux unités linguistiques est fonction de la similitude des contextes dans lesquels elles apparaissent (Harris 1954 ; Miller et Charles 1991)

sémantique distributionnelle

- Lenci (2008) et Turney et Pantel (2010)
- Les MSD capturent le sens des unités d'un corpus en fonction de leur place dans le corpus.

VNC et MSD

intensifieurs de l'anglais

- COHA (406,232,024 mots, 1810–2009)
- fenêtre contextuelle : ± 2 mots
- requête POS
- stop words
- lemmatisation
- seuil de fréquence : 5

contexte gauche	intensifieur	fenêtre droite
an_at1 umbrella_nn1	pretty_rg	soon_rr go_vvgk
_np1 i_ppis1	highly_rr	pleased_jj novel_jj_nn1
i_ppis1 have_vm	rather_rr	i_ppis1 like_vv0
will_vm a_at1	jolly_rg	good_jj blow-out_nn1
old_jj man_nn1	quite_rg	exhaust_jj miller_nn2
i_ppis1 declare_vv0	quite_rg	wet_jj come_vvd
li betsy_nn1_jj_np1 betsy_nn1_np1_jj	very_rg	well_rr thank_vv0
shall_vm acknowledge_vvi	most_dat_rgt	reverend_jj mummy_nn1
...

VNC et MSD

intensifieurs de l'anglais

Pour chaque décennie :

① matrice de co-occurrences

	quite	successful	completely	expectation	moment	old	self	almost	bag	great	hand	tear	find	...
quite	590	1	0	0	1	3	0	4	0	1	3	0	8	...
successful	1	24	3	0	0	0	0	3	0	0	0	0	0	...
completely	0	3	358	1	2	3	1	10	1	0	0	0	0	...
expectation	0	0	1	3	1	1	1	1	0	0	0	0	0	...
moment	1	0	2	1	14	1	1	6	0	0	0	0	0	...
old	3	0	3	1	1	27	1	7	0	0	0	0	0	...
self	0	0	1	1	1	1	6	2	0	0	0	0	0	...
almost	4	3	10	1	6	7	2	2298	1	11	7	2	16	...
bag	0	0	1	0	0	0	0	1	2	1	1	1	0	...
great	1	0	0	0	0	0	0	11	1	26	1	1	1	...
hand	3	0	0	0	0	0	0	7	1	1	24	1	0	...
tear	0	0	0	0	0	0	0	2	1	1	1	4	0	...
find	8	0	0	0	0	0	0	16	0	1	0	0	57	...
...

VNC et MSD

intensifieurs de l'anglais

Pour chaque décennie :

- ① matrice de co-occurrences
- ② transformation PPMI (Church et Hanks 1990)

$$PMI(w, c) = \log_2 \frac{P(w, c)}{P(w)P(c)} \quad (i)$$

$$PPMI(w, c) = \max(\log_2 \frac{P(w, c)}{P(w)P(c)}, 0) \quad (ii)$$

$P(w, c)$: mesure la fréquence d'observation w et c ensemble.

$P(w)P(c)$: probabilité d'observer w dans c sachant que les deux événements sont indépendants.

PMI : mesure à quel point w et c sont observés ensemble, indépendamment du hasard.

$PPMI$: les valeurs négatives sont remplacées par des zéros.

VNC et MSD

intensifieurs de l'anglais

Pour chaque décennie :

- ① matrice de co-occurrences
- ② transformation PPMI

	quite	successful	completely	expectation	moment	old	self	almost	bag	great	hand	tear	find	...
quite	2.5884	0	0	0	0	0	0	0	0	0	0	0	0	...
successful	0	6.304	0	0	0	0	0	0	0	0	0	0	0	...
completely	0	0	3.232	0.3629	0	0	0	0	0.5556	0	0	0	0	...
expectation	0	0	0.3629	7.5626	4.2772	3.5182	5.3927	0	0	0	0	0	0	...
moment	0	0	0	4.2772	6.3841	1.8178	3.6922	0	0	0	0	0	0	...
old	0	0	0	3.5182	1.8178	5.8137	2.9332	0	0	0	0	0	0	...
self	0	0	0	5.3927	3.6922	2.9332	7.3927	0	0	0	0	0	0	...
almost	0	0	0	0	0	0	0	0.6694	0	0	0	0	0	...
bag	0	0	0.5556	0	0	0	0	0	7.3629	3.744	3.9608	6.0004	0	...
great	0	0	0	0	0	0	0	0	3.744	5.8255	1.3419	3.3814	0.1938	...
hand	0	0	0	0	0	0	0	0	3.9608	1.3419	6.1437	3.5983	0	...
tear	0	0	0	0	0	0	0	0	6.0004	3.3814	3.5983	7.6378	0	...
find	0	0	0	0	0	0	0	0	0	0.1938	0	0	5.0954	...
...

VNC et MSD

intensifieurs de l'anglais

Pour chaque décennie :

- ① matrice de co-occurrences
- ② transformation PPMI
- ③ dimensions réduites à 50 par Décomposition en Valeurs Singulières (Schutze 1992)

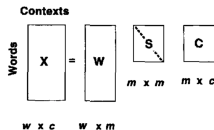


Figure A1. Schematic diagram of the singular value decomposition (SVD) of a rectangular word (w) by context (c) matrix (X). The original matrix is decomposed into three matrices: W and C , which are orthonormal, and S , a diagonal matrix. The m rows of W and the m columns of C' are linearly independent.

(Landauer et Dumais 1997, p. 238)

VNC et MSD

intensifieurs de l'anglais

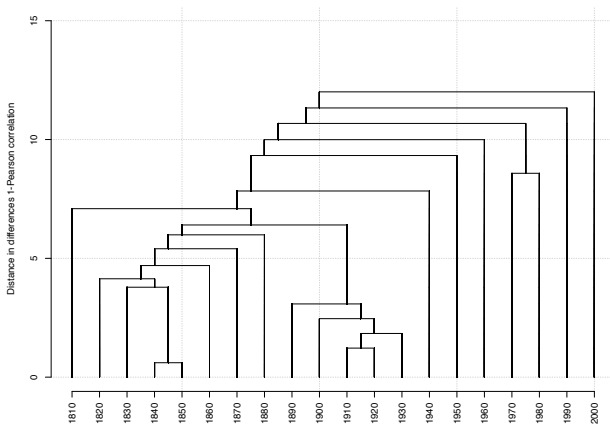
Pour chaque décennie :

- 1 matrice de co-occurrences
- 2 transformation PPMI
- 3 dimensions réduites à 50 par Décomposition en Valeurs Singulières

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	...
quite	2.588376421	2.27E-16	0	0	0	0	0	0	0	0	0	0	0	...
successful	3.81E-16	6.304034157	0	0	0	0	0	0	0	0	0	0	0	...
completely	0	0	3.23033770	0.362927846	-9.54E-16	-4.37E-16	-5.64E-16	5.82E-17	0.555572924	-2.91E-16	-1.25E-16	-2.78E-16	-3.44E-16	...
expectation	0	0	0.362927846	7.562601901	4.277197972	3.518206072	5.30267519	-4.03E-16	-1.57E-15	-6.80E-15	4.66E-15	1.53E-16	-3.75E-14	...
moment	0	0	4.42E-16	4.277197972	6.384113176	1.817766354	3.692235472	3.90E-17	-4.30E-16	-7.10E-15	4.18E-15	2.64E-16	-2.98E-14	...
old	0	0	1.93E-16	3.518206072	1.817766354	5.813661955	2.933243571	3.53E-16	1.60E-16	-3.47E-15	4.17E-15	1.28E-15	-2.46E-14	...
well	0	0	7.20E-16	5.39267519	3.692235472	2.933243571	7.30267519	1.82E-16	1.73E-16	-6.69E-15	5.28E-15	5.41E-16	-3.54E-14	...
almost	0	0	4.87E-17	-5.69E-16	-4.40E-16	1.14E-15	2.12E-16	0.669380852	2.84E-16	-4.82E-16	-3.76E-16	2.89E-16	8.33E-18	...
bag	0	0	0.555572924	4.54E-15	-7.01E-16	4.51E-15	1.22E-15	3.70E-16	7.362927846	3.744018014	3.960829403	6.000357767	1.57E-14	...
great	0	0	8.74E-16	-8.67E-17	2.83E-16	-1.01E-15	3.69E-16	2.01E-16	3.744018014	5.825547899	1.34191957	3.381447934	0.193820931	...
hard	0	0	2.22E-16	6.73E-16	-2.16E-15	-2.96E-15	-5.83E-16	6.24E-16	3.960829403	1.34191957	6.143693346	3.598259323	1.13E-14	...
tear	0	0	8.05E-16	3.05E-15	-1.63E-15	6.28E-16	-3.12E-16	-9.28E-16	6.000357767	3.381447934	3.598259323	7.637787688	1.68E-14	...
find	0	0	5.21E-16	-2.95E-16	-4.80E-16	-8.69E-16	1.08E-15	-1.19E-15	5.64E-16	0.193820931	-4.14E-16	-7.65E-16	5.095423695	...
...

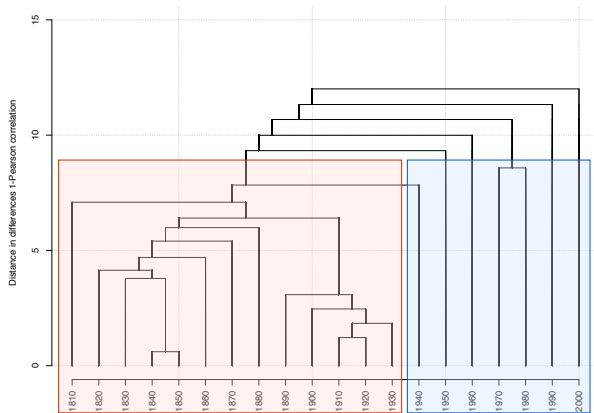
VNC et MSD

intensifieurs (merci à Florent Perek pour son code)



VNC et MSD

intensifieurs



bilan intermédiaire

- VNC permet de discerner les étapes principales dans le changement distributionnel d'un marqueur ou d'une série de marqueurs ;
- la méthode permet un retour informé sur les données ;
- elle ne permet pas d'apprécier le détail de l'évolution des profils distributionnels.

carte de densité

méthode complémentaire, inspirée de Hilpert (2016) :

- ① matrice de co-occurrences
- ② transformation PPMI
- ③ réduction par SVD
- ④ projection des vecteurs dans un espace 2D
- ⑤ tracer des courbes de densité indexées sur les fréquences par périodes

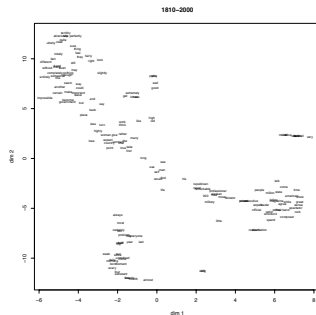
1810–1850

1860–1930

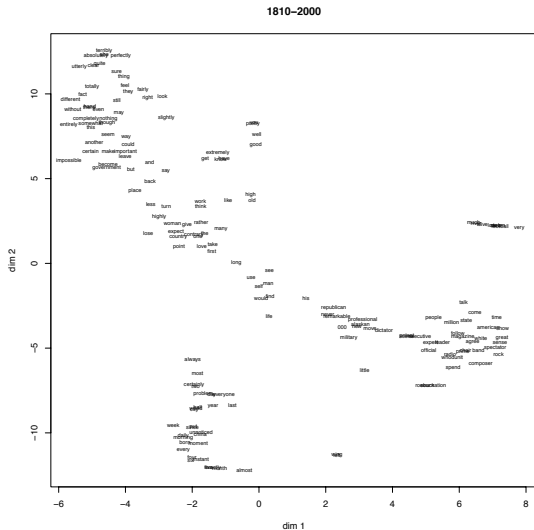
1940–2000

carte de référence

- on récupère la matrice calculée sur tout le corpus
- on projette l'ensemble sur un espace à deux dimensions avec *t*-distributed Stochastic Neighbor Embedding (van der Maaten et Hinton 2008)



carte de référence

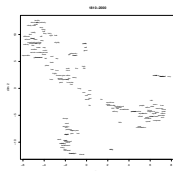


courbes de densité

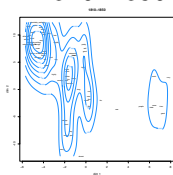
- on récupère la matrice calculée sur tout le corpus
- on projette l'ensemble sur un espace à deux dimensions avec *t*-distributed Stochastic Neighbor Embedding (van der Maaten et Hinton 2008)
- pour chacune des 3 périodes, on calcule les fréquences relatives
 - des intensifieurs
 - des mots voisins (filtrés)

carte de référence + courbes de densité

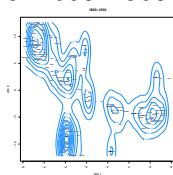
1. 1810-2000



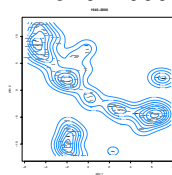
2. 1810-1850



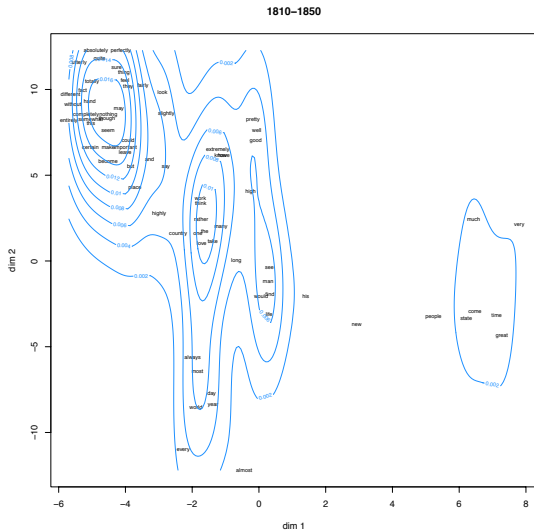
3. 1860-1930



4. 1940-2000



courbes de densité



bilan intermédiaire

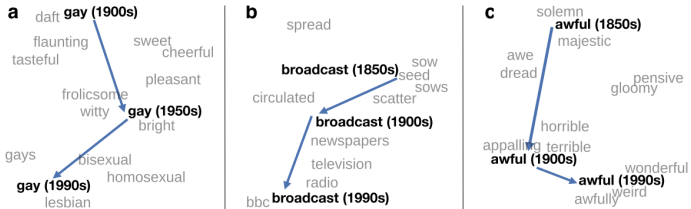
- VNC fournit une périodisation naturelle
- la projection de vecteurs peut être scindée à condition de s'appuyer sur un point de repère (carte de référence)
- cette technique issue de la linguistique de corpus a un équivalent en TAL : HistWords

HistWords (Hamilton et al. 2016)

Word Embeddings for Historical Text

Code disponible sur

<https://github.com/williamleif/histwords>



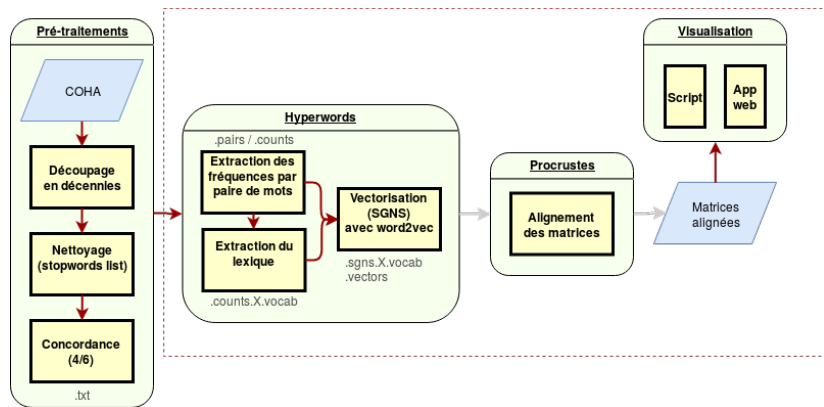
Principes

- ➊ Découpage du corpus en tranches temporelles (décennies ici)
- ➋ Construction de vecteurs pour chaque mot (word embeddings) pour chaque décennie
 - PPMI (Positive Point-wise Mutual Information)
 - SVD (Singular Value Decomposition)
 - **SGNS (Skipgram with Negative Sampling)**
ex : with the shrink he would rather have spent the hour eating
- ➌ Alignement des espaces vectoriels obtenus : méthode procustéenne (Προκρούστης dans la mythologie grecque)

Analyses possibles

- ① variations de distance entre 2 mots à travers les décennies
 - *quite* et *rather* (COHA personnalisé 1860-2000)
 - *should* et *must* (COHA 1830-2000)
- ② glissements de sens d'un mot en comparant pour chaque décennie ses plus proches voisins
 - *gay* (COHA 1830-2000)

Chaîne de traitements (expectations)



Hyperwords (Levy et al. 2015)

Pré-traitements

he would rather have spent the hour eating he had no interest in talking to the shrink about why he was n't doing anything all summer " let 's get going " john joel said

① nettoyage :

he would rather have spent the hour eating he had no interest in talking to the shrink about why he was n't doing anything all summer let s get going john joel said

② stopwords ?

he would rather have spent the hour eating he had no interest in talking to the shrink about why he was n't doing anything all summer let s get going john joel said

③ fenêtre de concordance :

rather spent hour eating interest talking

Traitements

Pairs

would	the
would	hour
rather	with
rather	the
rather	shrink
rather	he
rather	would
rather	have
rather	spent
rather	the
rather	hour
rather	eating
have	the
have	shrink

Counts

74	rather	did
29	rather	die
8	rather	died
5	rather	difference
6	rather	differences
27	rather	different
1	rather	differently
14	rather	difficult
	[...]	
21	rather	often
6	rather	oil
35	rather	old
452	rather	on
200	rather	one

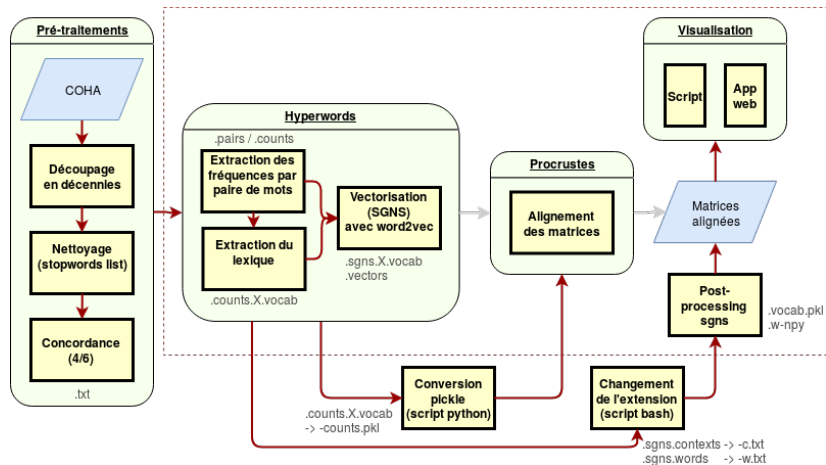
Vocab

whole	54077
change	54031
run	53983
soon	53903
brought	53825
tried	53614
hear	53277
rather	53191
south	53156
social	53072
cut	53035
military	52837
lost	52689
line	52685

Méthode procustéenne



Chaîne de traitements (reality)



Hyperwords (Levy et al. 2015)

sémantique distributionnelle et diachronie

les outils

- les outils présentés en linguistique de corpus sont complémentaires
- HistWords procède décennie par décennie, mais permet de voir les écarts entre périodes




l'étude du sens

- dans tous les cas, l'observation du changement sémantique est indirecte (sens voisins)
- et soumise à une combinaison de paramètres (heuristique)




le corpus

- il est possible de s'affranchir de la segmentation artificielle
- beaucoup plus difficile de s'affranchir des déséquilibres dans la composition du corpus





References I

-  Church, Kenneth Ward et Patrick Hanks (1990). « Word association norms, mutual information, and lexicography ». In : *Computational linguistics* 16.1, p. 22–29.
-  Davies, Mark (2010). *The Corpus of Historical American English : 400 million words, 1810–2009*. url : <https://www.english-corpora.org/coha/>.
-  Desagulier, Guillaume (2014). « Visualizing distances in a set of near synonyms : *rather, quite, fairly, and pretty* ». In : *Corpus Methods for Semantics : Quantitative Studies in Polysemy and Synonymy*. Sous la dir. de Dylan Glynn et Justyna Robinson. Amsterdam : John Benjamins, p. 145–178.





References II

-  Firth, J. R. (1957). « A synopsis of linguistic theory 1930-55. ». In : *Studies in Linguistic Analysis (special volume of the Philological Society)*. T. 1952-59. Oxford : The Philological Society, p. 1–32.
-  Gries, Stefan Th. et Martin Hilpert (2008). « The identification of stages in diachronic data : variability-based neighbour clustering ». In : *Corpora* 3.1, p. 59–81.
-  – (2010). « Modeling diachronic change in the third person singular : a multifactorial, verb-and author-specific exploratory approach ». In : *English Language & Linguistics* 14.3, p. 293–320.





References III

-  Hamilton, William L, Jure Leskovec et Dan Jurafsky (2016). « Cultural shift or linguistic drift ? comparing two computational measures of semantic change ». In : *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*. T. 2016. NIH Public Access, p. 2116.
-  Harris, Zellig S. (1954). « Distributional structure ». In : *Word* 10.2-3, p. 146–162.
-  Hilpert, Martin (2016). « Change in modal meanings ». In : *Constructions and Frames* 8.1, p. 66–85.
-  Hilpert, Martin et Stefan Th. Gries (2008). « Assessing frequency changes in multistage diachronic corpora : Applications for historical corpus linguistics and the study of language acquisition ». In : *Literary and Linguistic Computing* 24.4, p. 385–401.

References IV

-  Hilpert, Martin et Stefan Th. Gries (2016). « Quantitative approaches to diachronic corpus linguistics ». In : *The Cambridge handbook of English historical linguistics*, p. 36–53.
-  Landauer, Thomas K et Susan T Dumais (1997). « A solution to Plato's problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge. ». In : *Psychological review* 104.2, p. 211.
-  Lenci, Alessandro (2008). « Distributional semantics in linguistic and cognitive research ». In : *Italian Journal of Linguistics* 20.1, p. 1–31.
-  Miller, George A. et Walter G. Charles (1991). « Contextual correlates of semantic similarity ». In : *Language and Cognitive Processes* 6.1, p. 1–28.

References V

-  Perek, Florent et Martin Hilpert (2017). « A distributional semantic approach to the periodization of change in the productivity of constructions ». In : *International Journal of Corpus Linguistics* 22.4, p. 490–520.
-  Schutze, Hinrich (1992). « Dimensions of meaning ». In : *Supercomputing'92 : Proceedings of the 1992 ACM/IEEE Conference on Supercomputing*. IEEE, p. 787–796.
-  Turney, Peter D et Patrick Pantel (2010). « From frequency to meaning : Vector space models of semantics ». In : *Journal of Artificial Intelligence Research* 37, p. 141–188.
-  van der Maaten, Laurens et Geoffrey Hinton (2008). « Visualizing Data using t-SNE ». In : *Journal of Machine Learning Research* 9, p. 2579–2605.