

# Étude des variations sémantiques à travers plusieurs dimensions

Syrielle Montariol<sup>1,2</sup> Alexandre Allauzen<sup>3</sup>

(1) LIMSI, CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, F-91405 Orsay, France

(2) Société Générale, 17 Cours Valmy 92043 Puteaux, France

(3) ESPCI, Univ. Paris Dauphine - PSL, 75016 Paris

syrielle.montariol@limsi.fr, alexandre.allauzen@espci.psl.eu

## RÉSUMÉ

---

Au sein d'une langue, l'usage des mots varie selon deux axes : diachronique (dimension temporelle) et synchronique (variation selon l'auteur, la communauté, la zone géographique...). Dans ces travaux, nous proposons une méthode de détection et d'interprétation des variations d'usages des mots à travers ces différentes dimensions. Pour cela, nous exploitons les capacités d'une nouvelle ligne de plongements lexicaux contextualisés, en particulier le modèle BERT. Nous expérimentons sur un corpus de rapports financiers d'entreprises françaises, pour appréhender les enjeux et préoccupations propres à certaines périodes, acteurs et secteurs d'activités.

## ABSTRACT

---

### Studying semantic variations through several dimensions

In a language, word usage can vary across two axis : diachronic (time evolution), and synchronic (variation across sources, authors, communities...). In this work, we propose to leverage the capacity of contextualised embeddings models to analyse financial texts along these two axes of variation. Starting from a corpus of annual company reports spanning 20 years, we explore the ability of the language model BERT to extract interpretable variations in word usage in order to understand the stakes and concerns of specific periods, financial actors and sectors.

**MOTS-CLÉS :** Diachronie, Variation sémantique, Plongements lexicaux contextualisés, Clustering.

**KEYWORDS:** Diachrony, Semantic variation, Contextualised Embeddings, Clustering.

---

## 1 Introduction

L'usage des mots varie selon deux axes : diachronique (dimension temporelle) (Aitchison, 2001) et synchronique (en fonction de l'auteur, la communauté, la zone géographique...). Ces variations peuvent être d'envergure et d'ordre générationnel et culturel ; ou d'ampleur plus réduite, pour des variations à court terme et entre individus ou entités. Dans ce cas, elles peuvent être le résultat d'événements qui, sans altérer la signification du mot, changent ponctuellement son usage et sa connotation. Elles révèlent alors des divergences d'intérêts et de préoccupations entre les individus.

Dans cet article, nous nous concentrons sur le domaine financier et un corpus de rapports financiers d'entreprises. Dans ce cadre, modéliser les évolutions temporelles d'usage des mots peut permettre une meilleure appréhension des enjeux et préoccupations de chaque période (Matthew Purver & Pollak, 2018) tandis que les opinions, comportements et préoccupations des différents acteurs financiers peuvent transparaître à travers la façon dont ils utilisent les mots. En d'autres termes, nous cherchons

à détecter des “signaux faibles” en analysant les variations d’usage des mots. Un signal faible est une information observée à partir de données, qui peut avoir une interprétation et des conséquences ambiguës, mais peut revêtir de l’importance pour la compréhension d’événements présents ou futurs. Les variations diachroniques et synchroniques au sein d’un corpus de documents sont difficilement repérables par les analystes financiers ; mais elles peuvent révéler des informations précieuses en tant que potentiels signaux faibles de changement dans l’opinion ou la situation d’un acteur financier. Par exemple, l’évolution de la connotation du vocabulaire employé dans les communications des banques centrales (BCE et Fed) est liée à la situation économique de la période (Buechel *et al.*, 2019).

Avec la numérisation de textes historiques, les méthodes de traitement automatique des langues (TAL) pour l’analyse diachronique se sont développées rapidement ces dernières années (Tahmasebi *et al.*, 2018). De nombreux modèles reposent sur des plongements de mots statiques comme Word2Vec (Mikolov *et al.*, 2013). Ils rassemblent les différents usages d’un mot dans un unique vecteur, ce qui rend difficile une analyse des variations de contexte et d’usage. Plus récemment, les plongements de mot contextualisés sont apparus comme BERT (Devlin *et al.*, 2019) ou ELMO (Peters *et al.*, 2018). Ce type de modèles renouvellent les perspectives pour l’analyse des variations sémantiques. Dans cet article, nous utilisons le modèle BERT qui montre une nette supériorité pour la tâche de désambiguïsation sémantique par rapport à ELMO et Flair (Wiedemann *et al.*, 2019). Nous nous appuyons sur FlauBERT, une version de BERT pour le français, pour proposer une méthode de détection et d’interprétation des variations d’usages de mots dans un corpus selon plusieurs dimensions. La méthode est décrite dans la section 3 puis appliquée (section 4) sur un corpus français de rapports financiers annuels d’entreprises (section 4.1).

## 2 État de l’art

Avant que l’emploi de représentations vectorielles de mots ne se généralise, la mesure de changement sémantique reposait sur la détection de variation dans les co-occurrences de mots (Sagi *et al.*, 2009). Puis des modèles de plongements de mots diachroniques se sont développés, reposant sur l’hypothèse qu’un changement dans le contexte d’un mot reflète une évolution de sa signification et son usage. Le plus souvent, ils impliquent de diviser un corpus en strates temporelles puis d’apprendre des plongements lexicaux (Mikolov *et al.*, 2013) pour chaque mot, dans chaque strate. Pour cela, deux méthodes populaires sont l’apprentissage incrémental (Kim *et al.*, 2014) et l’alignement d’espaces vectoriels (Hamilton *et al.*, 2016). Néanmoins, ces méthodes rassemblent l’ensemble des significations et usages possibles d’un mot dans un unique vecteur, à chaque strate temporelle.

En parallèle, l’analyse de variations sémantiques dans le cas synchronique est faite en majorité à partir de méthodes de désambiguïsation sémantique. Certains auteurs utilisent des mesures de similarité entre plongements de mots pour analyser les variations de vocabulaire entre plusieurs communautés (Tredici & Fernández, 2017). Plus récemment, Schlechtweg *et al.* (2019) analysent à la fois les variations synchroniques et diachroniques dans des corpus en utilisant des plongements lexicaux et une méthode d’alignement d’espaces de représentations.

Les plongements de mots contextualisés comme BERT (Devlin *et al.*, 2019) et ELMO (Peters *et al.*, 2018) permettent à chaque occurrence d’un mot d’être représentée par un vecteur dépendant de son contexte. Pré-entraînés sur des corpus volumineux, ils améliorent l’état de l’art dans de nombreuses tâches de TAL et commencent à être utilisés pour la détection de changement sémantique. Dans un premier temps, BERT fut utilisé de façon supervisée (Hu *et al.*, 2019) pour déterminer l’évolution de

la distribution des sens d'un mot au cours du temps. Néanmoins, cette méthode implique de travailler sur un nombre réduit de mots dont l'ensemble des sens est connu l'avance. Nous explorons l'usage de méthodes de partitionnement sur l'ensemble des représentations des occurrences d'un mot pour extraire automatiquement ses sens possibles (Giulianelli *et al.*, 2019; Martinc *et al.*, 2020a).

### 3 Modélisation des variations avec FlauBERT

BERT (Bidirectional Encoder Representations from Transformers) est un modèle de représentation des mots dans leurs contextes d'utilisation (typiquement une phrase). Son utilisation repose sur l'apprentissage par transfert : pré-entraîner un modèle sur une tâche non-supervisée à partir d'un très grand volume de données, avant de le raffiner sur une nouvelle tâche. Ici, nous utilisons ce type de modèle directement pour représenter l'occurrence d'un mot dans une séquence. Notre étude portant sur des données français, nous utilisons le modèle FlauBERT (Le *et al.*, 2020) qui est une variante des modèles BERT et RoBERTa (Liu *et al.*, 2019) entraînée sur des textes français.

#### 3.1 Détection des variations

FlauBERT permet de prédire la représentation spécifique au contexte de chaque occurrence d'un mot. Afin d'en détecter les variations d'usage, nous proposons une méthode en quatre étapes.

##### Extraction des plongements contextualisés

Le corpus est d'abord segmenté en phrases, chacune d'elle étant assortie de méta-données décrivant son document d'origine, l'auteur et la date. Pour chaque mot du vocabulaire sélectionné, nous extrayons l'ensemble des plongements lexicaux contextualisés de ses occurrences dans le corpus à l'aide de FlauBERT. Le vecteur extrait correspond à la dernière couche du modèle pré-entraîné.

##### Détection des mots subissant des variations

Les travaux précédents sélectionnent manuellement une faible quantité de mots à analyser (Giulianelli *et al.*, 2019; Hu *et al.*, 2019). À l'inverse, nous souhaitons détecter automatiquement les mots qui subissent une variation significative. Pour cela, nous utilisons une approche similaire à celle de Martinc *et al.* (2020b). Pour chaque dimension, nous calculons une métrique reposant sur la moyenne des représentations vectorielles des occurrences d'un mot calculée sur l'ensemble du corpus. Prenons le cas des variations diachroniques, la moyenne des représentations observées sur tous le corpus  $u_{\text{corpus}}$  est comparée à la moyenne estimée sur chaque strate temporelle  $u_t$  en utilisant la distance cosinus. La sélection se fait par seuillage sur la moyenne des distances.

##### Partitionnement des usages

Pour chaque mot sélectionné, nous appliquons un algorithme de partitionnement sur l'ensemble des représentations vectorielles de ses occurrences. Nous retenons pour ce faire deux algorithmes en particulier : K-Means et propagation par affinité. La méthode de propagation par affinité (Frey & Dueck, 2007), moins classique que la méthode K-Means, a fait ses preuves dans la littérature de la désambiguïsation du sens des mots (Alagić *et al.*, 2018), une tâche proche de notre objectif. Elle a l'avantage de ne pas nécessiter de sélectionner manuellement le nombre de clusters.

On notera que les représentations inférées par FlauBERT contiennent des informations sémantiques mais aussi syntaxiques (Coenen *et al.*, 2019). Par construction, les clusters obtenus ne regroupent donc pas les différents sens d'un mot, mais plus largement les différentes manières dont il est utilisé.

## Matrice de distributions d'usages

Pour chaque mot, une matrice est déduite afin de représenter les distribution normalisée des clusters (donc des usages) au travers d'une dimension de variation. Par exemple, la figure 1 représente une telle matrice pour les variations temporelles et sectorielles.

### 3.2 Interprétation des variations

Les matrices de distributions d'usages de mots permettent d'extraire différentes informations :

- 1) À quel point l'usage d'un mot varie pour une dimension donnée ?
- 2) À quelle période, pour quel auteur, quelle classe, se produit la variation ?
- 3) Quel usage apparaît / disparaît ? Comment interpréter ce changement ?

Pour le premier élément, nous utilisons la divergence de Jensen-Shannon (JSD), une mesure de comparaison de deux distributions de probabilité, ainsi que sa généralisation à  $n$  distributions de probabilités  $d_1, d_2, \dots, d_n$  proposée par Ré & Azad (2014) ( $H$  est la fonction entropie) :

$$\text{JSD}(d_1, d_2, \dots, d_n) = H\left(\frac{\sum_{i=1}^n d_i}{n}\right) - \frac{\sum_{i=1}^n H(d_i)}{n} \quad (1)$$

Pour le second élément, chaque distribution des usages est comparée avec la distribution moyenne de toute la dimension. Par exemple dans le cadre diachronique, la moyenne élément-à-élément des distributions de l'ensemble des périodes est calculée. Cette distribution moyenne est comparée à celle de chaque période à l'aide de la JSD. Puis il convient de déterminer les clusters / usages concernés : ceux qui varient le plus, apparaissent ou disparaissent au cours du temps, ou sont propres à certains acteurs uniquement. Pour cela, on recherche les clusters qui se répartissent de manière inégale selon la dimension de variation.

Pour finir, une fois les clusters cibles identifiés, interpréter les usages qui y sont associés se fait de deux façons. Tout d'abord, nous faisons l'hypothèse que l'exemple (la phrase, dans notre cas) le plus proche du centroïde est représentatif du contexte des occurrences du mot dans le cluster analysé. Nous comparons donc ces phrases entre les différents clusters pour avoir une idée préliminaire de l'usage du mot dans son contexte.

Ensuite, nous mettons en place une méthode de détection de mots-clés pour caractériser les différents clusters les uns par rapport aux autres. La méthode repose sur le principe du tf-idf (Term Frequency - Inverse Document Frequency). Un cluster étant constitué d'un ensemble de phrase, nous considérons chaque cluster comme un document et l'ensemble des clusters comme un corpus. Le but est de déterminer les mots ayant de l'importance pour un document mais pas dans le reste du corpus, c'est-à-dire les mots les plus discriminants pour chaque cluster. Les mots-outils sont exclus ; nous éliminons aussi tous les mots qui apparaissent déjà dans plus de 50% des clusters. Nous calculons ensuite les scores de tf-idf des mots dans chaque cluster. Les mots ayant les scores les plus importants sont utilisés comme mots-clés pour caractériser et faciliter l'interprétation des clusters.

## 4 Expérimentation

Nous appliquons la méthode décrite à un corpus du domaine financier, pour détecter les variations d'usages de mots à travers plusieurs dimensions en plus de celle du temps. Pour nos expériences, nous

utilisons le modèle pré-entraîné et librement disponible FlauBERT-base-uncased (Le *et al.*, 2020).

## 4.1 Corpus

Les données utilisées sont issues du corpus financier CoFiF<sup>1</sup> (Daudert & Ahmadi, 2019). Il est composé des rapports financiers des 60 plus grandes entreprises françaises appartenant aux indices boursiers CAC40 et CAC Next 20. Il comporte plus de 5 millions de phrases dans 2655 rapports de différents types (rapports trimestriel, semestriel, annuel, et document de référence), de 1995 à 2018. Une particularité du corpus est la présence de tableaux de données au format brut dans le texte. Afin d'écartier ces éléments de l'analyse, lors de la division du corpus en phrases, nous excluons les phrases composées de moins de 70% de lettres (plus de 30% de chiffres, symboles et espaces). Pour finir, nous nous concentrons sur les documents de référence (DR), qui constituent presque 85% du volume de données. Ils sont publiés chaque année par les entreprises et résument leur situation financière et perspectives. Ainsi, nous aboutissons à un corpus constitué d'environ 2.7 millions de phrases. Chacune de ces phrases est associée aux méta-données du document dont elle est extraite : le nom de l'entreprise, et l'année de publication du rapport. Ce sont deux dimensions pour l'analyse de variations d'usage des mots : diachronique (par année) et synchronique (par entreprise). L'axe synchronique est étoffé en collectant des informations sur les entreprises : leur domaine d'activité (luxe, transport, chimie, ...) et leur secteur (secondaire ou tertiaire). Les deux secteurs sont équilibrés (1.4 millions de phrases pour le secteur tertiaire, et 1.3 millions pour le secondaire).

Parmi l'ensemble du vocabulaire, nous conservons les 10 000 mots les plus fréquents. Nous en excluons les mots-outils et sélectionnons les mots qui sont uniquement des noms<sup>2</sup>. En effet, le modèle BERT est fortement influencé par la catégorie grammaticale d'un mot (Coenen *et al.*, 2019); réduire ainsi l'analyse permet de limiter l'impact de la catégorie grammaticale lors du partitionnement.

## 4.2 Résultats

Certains mots ont plus de 500 000 occurrences dans le corpus ; c'est autant de phrases dont il faut extraire le plongement lexical avec FlauBERT. Pour alléger ce processus, nous échantillons 5000 phrases pour chaque mot, à partir desquelles nous mesurons la variation pour chaque dimension (par année, entreprise, domaine d'activité et secteur) en se ramenant à des plongements lexicaux statiques. Pour chaque dimension, nous conservons comme mots-cibles les 10% de mots ayant la mesure de variation la plus élevée.

Puis pour chaque mot-cible retenu, nous appliquons les algorithmes de partitionnements K-means et propagation par affinité sur l'ensemble des représentations vectorielles de ses occurrences. Afin d'évaluer la qualité du partitionnement, nous calculons le coefficient de silhouette pour chaque mot et algorithme. Puis, nous nous extrayons les distributions de probabilité pour chaque strate temporelle (analyse diachronique) et pour chaque entreprise / secteur d'activité (analyse synchronique). Enfin, pour chaque dimension, nous calculons la JSD généralisée sur l'ensemble des distributions de probabilité afin de mesurer le niveau de variation sémantique du mot dans la dimension étudiée. Les valeurs moyennes pour tous les mots-cibles du coefficient de silhouette et de la JSD par secteur et par année figurent en Table 1. Rappelons que le coefficient de silhouette se situe entre 0 et 1 (proche de

---

1. <https://github.com/CoFiF/Corpus>

2. En utilisant l'outil Wolf (Sagot & Fišer, 2008), une ressource lexicale et sémantique pour le français.

Method	S-score	JSD-secteur	JSD-année
aff-prop	0.118	1.722	1.265
kmeans3	0.094	0.137	0.075
kmeans5	0.088	0.234	0.124
kmeans7	0.071	0.230	0.167

TABLE 1: Valeurs moyenne pour tous les mots-cibles du coefficient de silhouette et de la JSD par secteur et par année

Année	Secteur
1	écologie
2	climat
3	biodiversité
4	syndicats
5	gouvernement

TABLE 2: Top 5 des mots ayant les plus fortes JSD selon les dimensions temporelle et sectorielle, avec la propagation par affinité.

zéro indique une faible qualité); et que si la JSD se situe entre 0 et 1 pour deux distribution, la version généralisée à  $n$  distributions est bornée par  $\log_2(n)$ . Pour la dimension temporelle par exemple, notre période de 18 ans mène à une borne supérieure valant  $\log_2(18) \approx 4.17$ .

Selon la table 1, l’algorithme de propagation par affinité mène au coefficient de silhouette moyen le plus élevé. Nous utilisons donc cet algorithme de partitionnement pour comparer les variations des mots-cibles entre eux. Les 5 mots-cibles ayant les plus fortes JSD selon les dimensions temporelle et sectorielle dans le corpus sont listés dans la Table 2. On note que les 3 mots les plus variables par année font tous partie du champs lexical du climat. De plus, le mot *écologie* varie fortement dans les deux dimensions; nous allons analyser et interpréter ses variations.

### Étude de cas : le mot *écologie*.

Pour ce mot, le coefficient de silhouette le plus élevé est obtenu à partir de l’algorithme K-Means avec  $k = 7$ , pour un score de 0.231. La distribution normalisée des clusters issus de ce partitionnement pour les dimensions sectorielle et temporelle est représentée sur la Figure 1. En comparant les distributions de chaque période ou secteur avec la distribution moyenne sur le corpus pour la dimension associée, nous repérons les périodes et secteurs qui se démarquent. Puis nous quantifions la variation de chaque cluster au sein d’une dimension. Cela nous mène à la dernière étape : l’interprétation des clusters. Nous extrayons les phrases les plus proches des centroïdes; puis à partir de la méthode d’extraction de mots-clés, nous associons un thème à chaque cluster (Table 3).

Par exemple, le cluster 6 connaît une forte variation temporelle, avec une proportion croissante à partir de 2007; il est associé à des problématiques de financement et de coût (Figure 1). Le cluster 2 n’est propre qu’à quelques secteurs et se concentre sur les idées de métier propre à l’écologie; il apparaît assez tardivement dans les documents. À l’inverse, les clusters 1 et 5 propres aux transports et à l’aménagement des territoires pour l’un, et à l’énergie pour l’autre, sont communs à la plupart des secteurs. Le cluster 1 est bien résumé par sa phrase-centroïde, “*ces obligations concernent pour l’essentiel l’écologie, l’aménagement du paysage et l’archéologie pour les sites de développement associés*”. Le cluster 3, plutôt associé au territoire du point de vue des ressources, est présent sur toute la période étudiée mais n’est propre qu’à quelques secteurs tels ceux du pétrole et de la chimie. Pour finir, le cluster 4 qui contient le champs lexical du risque et du danger, apparaît et prend progressivement de l’ampleur mais reste minoritaire, probablement dû au fait que les analystes financiers évitent d’employer des termes négatifs lors de la rédactions des rapports d’activité afin de ne pas inquiéter les investisseurs.

N <sup>o</sup>	Titre	Exemple de mots-clés
0	pratique	éco, concept , logement, économique, raisonné, préservant, préfabrication
1	transport	directeur, énergie, impacts, transports, aviation, initiatives, territoire, aménagement
2	métier	apprendre, structure, métiers, collaborateurs, réseau, professionnels, management
3	territoire	industrielle, sites, flux, déchets, échanges, territoriale, eaux, circulaire, ressources
4	danger	groupe, fondation, prix, intégrer, péril, polluante, excessive, concernés
5	énergie	émissions, énergie, fessenheim, industrielle, biodiversité, slovénie, co2, nucléaire
6	coût	énergie, arrêté, coût, mer, prix, stockage, économiques, milliards, aménagement

TABLE 3: Liste des clusters et interprétations pour le mot *écologie*.

Dans l'ensemble, les disparités d'usage et de connotation des mots que nous détectons sont encourageantes. La détection de variations propres à une période temporelle permettrait à un analyste de relier le résultat avec des événements de la vie réelle, tandis que les variations de connotation entre les clusters ouvrent la voie à une analyse de sentiments plus poussée.

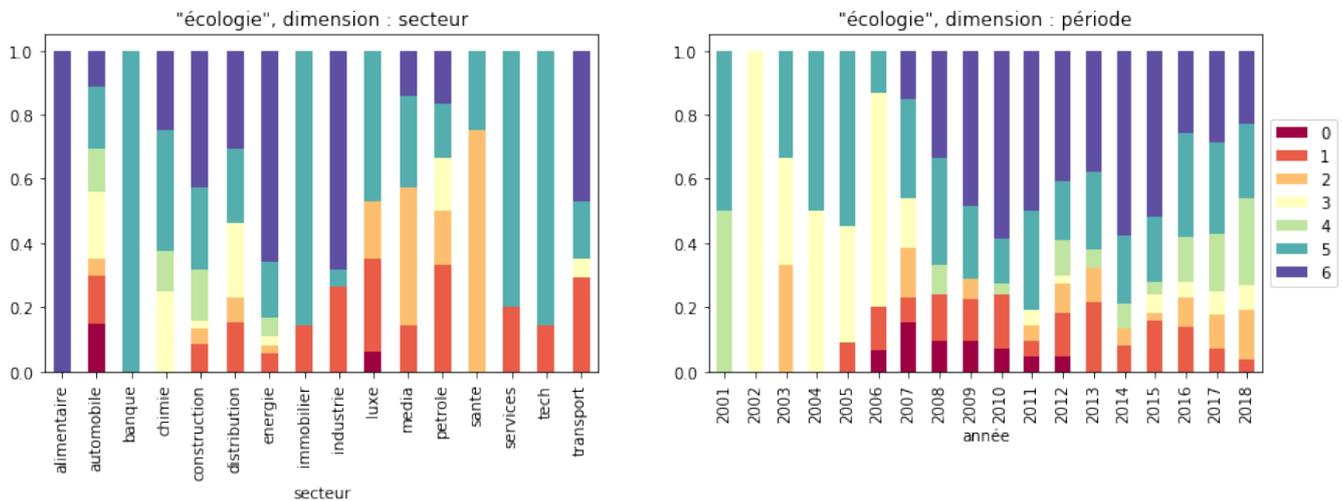


FIGURE 1: Distribution des clusters pour le mot *écologie*, par secteur d'activité (à gauche) et par année (à droite).

## 5 Conclusion

Cet article est une investigation préliminaire de la capacité des plongements lexicaux contextualisés de BERT à détecter des variations diachroniques et synchroniques d'usages de mots. Nous montrons sur une étude de cas que notre méthode permet de détecter et d'interpréter de façon fine les variations d'usage dans plusieurs dimensions.

L'étape suivante est de proposer une méthode d'évaluation de notre processus. Peu de corpus annotés de variations sémantiques étant disponibles, nous nous tournons vers la génération de variations synthétiques (Shoemark *et al.*, 2019). Nous définissons différents scénarios de variations d'usage d'un mot selon différentes dimensions, puis les simulons en générant des corpus comportant ces variations. Le but est d'évaluer les capacités de détection de notre méthode sur ces données synthétiques, pour chaque scénario.

## Références

- AITCHISON J. (2001). Language change : Progress or decay? In *Cambridge Approaches to Linguistics*. Cambridge University Press.
- ALAGIĆ D., ŠNAJDER J. & PADÓ S. (2018). Leveraging lexical substitutes for unsupervised word sense induction. In *Thirty-Second AAAI Conference on Artificial Intelligence* : [link](#).
- BUECHEL S., JUNKER S., SCHLAAK T., MICHELSEN C. & HAHN U. (2019). A time series analysis of emotional loading in central bank statements. In *Proceedings of the Second EconLP Workshop*, Hong Kong : [D19-5103](#).
- COENEN A., REIF E., YUAN A., KIM B., PEARCE A., VIÉGAS F. B. & WATTENBERG M. (2019). Visualizing and measuring the geometry of bert. In *NeurIPS* : [NIPS2019\\_9065](#).
- DAUDERT T. & AHMADI S. (2019). CoFiF : A corpus of financial reports in French language. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing*, p. 21–26, Macao, China : [W19-5504](#).
- DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). Bert : Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT* : [N19-1423](#).
- FREY B. J. & DUECK D. (2007). Clustering by passing messages between data points. *Science*, **315**(5814), 972–976. DOI : [10.1126/science.1136800](#).
- GIULIANELLI M., FERNANDEZ R. & DEL TREDICI M. (2019). Contextualised word representations for lexical semantic change analysis. In *EurNLP* : [link](#).
- HAMILTON W., LESKOVEC J. & JURAFSKY D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. p. 1489–1501. DOI : [10.18653/v1/P16-1141](#).
- HU R., LI S. & LIANG S. (2019). Diachronic sense modeling with deep contextualized word embeddings : An ecological view. p. 3899–3908. DOI : [10.18653/v1/P19-1379](#).
- KIM Y., CHIU Y.-I., HANAKI K., HEGDE D. & PETROV S. (2014). Temporal analysis of language through neural language models. DOI : [10.3115/v1/W14-2517](#).
- LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUTEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2020). Flaubert : Unsupervised language model pre-training for french. In *Proceedings of LREC 2020* : arXiv : [1912.05372](#).
- LIU Y., OTT M., GOYAL N., DU J., JOSHI M., CHEN D., LEVY O., LEWIS M., ZETTLEMOYER L. & STOYANOV V. (2019). Roberta : A robustly optimized BERT pretraining approach. *arXiv* : [1907.11692](#).
- MARTINC M., MONTARIOL S., ZOSA E. & PIVOVAROVA L. (2020a). Capturing evolution in word usage : Just add more clusters? In *Companion Proceedings of the Web Conference 2020, WWW '20*, p. 343–349, New York, NY, USA : Association for Computing Machinery. DOI : [10.1145/3366424.3382186](#).
- MARTINC M., NOVAK P. K. & POLLAK S. (2020b). Leveraging contextual embeddings for detecting diachronic semantic shift. In *Proceedings of LREC 2020* : arXiv : [1912.01072](#).
- MATTHEW PURVER, ALJOŠA VALENTINČIČ M. P. & POLLAK S. (2018). Diachronic lexical changes in company reports : An initial investigation. In M. EL-HAJ, P. RAYSON & A. MOORE, Éd., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France : [lrec2018-9\\_W27](#).

- MIKOLOV T., SUTSKEVER I., CHEN K., CORRADO G. S. & DEAN J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, p. 3111–3119. [NIPS2013\\_5021](#).
- PETERS M. E., NEUMANN M., IYYER M., GARDNER M., CLARK C., LEE K. & ZETTLEMOYER L. (2018). Deep contextualized word representations. *arXiv* : [1802.05365](#).
- RÉ M. A. & AZAD R. K. (2014). Generalization of entropy based divergence measures for symbolic sequence analysis. *PLoS ONE*, **9**. DOI : [10.1371/journal.pone.0093532](#).
- SAGI E., KAUFMANN S. & CLARK B. (2009). Semantic density analysis : Comparing word meaning across time and phonetic space. In *Proceedings of the Workshop on Geometrical Models of Natural Language Semantics*, Athens, Greece : [W09-0214](#).
- SAGOT B. & FIŠER D. (2008). Building a free French wordnet from multilingual resources. In *OntoLex*, Marrakech, Morocco. HAL : [inria-00614708](#).
- SCHLECHTWEG D., HÄTTY A., DEL TREDICI M. & SCHULTE IM WALDE S. (2019). A wind of change : Detecting and evaluating lexical semantic change across times and domains. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, p. 732–746, Florence, Italy : Association for Computational Linguistics. DOI : [10.18653/v1/P19-1072](#).
- SHOEMARK P., LIZA F. F., NGUYEN D., HALE S. & MCGILLIVRAY B. (2019). Room to Glo : A systematic comparison of semantic change detection approaches with word embeddings. In *Proceedings of the 2019 EMNLP-IJCNLP Conference*, p. 66–76, Hong Kong, China : Association for Computational Linguistics. DOI : [10.18653/v1/D19-1007](#).
- TAHMASEBI N., BORIN L. & JATOWT A. (2018). Survey of computational approaches to diachronic conceptual change. *arXiv* : [1811.06278](#).
- TREDICI M. D. & FERNÁNDEZ R. (2017). Semantic variation in online communities of practice. In *IWCS 2017 - 12th International Conference on Computational Semantics - Long papers* : [W17-6804](#).
- WIEDEMANN G., REMUS S., CHAWLA A. & BIEMANN C. (2019). Does bert make any sense ? interpretable word sense disambiguation with contextualized embeddings. In *Proceedings of KONVENS 2019*, Erlangen, Germany : [link](#).