



HAL
open science

Analyse Diachronique de Corpus : le cas du poker

Gaël Lejeune, Lichao Zhu

► **To cite this version:**

Gaël Lejeune, Lichao Zhu. Analyse Diachronique de Corpus : le cas du poker. Journées d'Analyse statistique des Données Textuelles (JADT) 2018, Jun 2018, Rome, Italie. hal-02467662

HAL Id: hal-02467662

<https://hal.science/hal-02467662v1>

Submitted on 5 Feb 2020

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.

Analyse Diachronique de Corpus : le cas du poker

Gaël Lejeune¹, Lichao Zhu²

¹ STIH, Sorbonne Université – gael.lejeune@sorbonne-universite.fr

² LLSHS, Université Paris XIII – lichao.zhu@univ-paris13.fr

Abstract

In this paper we will investigate a diachronic corpus. We want to highlight how people's mentalities evolve regarding the gambling especially the poker game and how the evolution is correlated with the way that the game is considered in press articles. We study plain or metaphorical meanings of the terms in question by using clustering and statistical methods in order to detect changes of meanings in a relatively large period of time.

Résumé

Dans cet article nous nous intéressons à l'étude diachronique de corpus de presse dans le but d'illustrer des évolutions dans la vision de la société sur les jeux d'argent et de hasard ainsi que sur les joueurs. Nous utilisons des méthodes de statistique textuelles et de *clustering* pour détecter les grandes tendances visibles sur notre échelle de temps en nous focalisant sur le poker. Nous montrons que si le regain de popularité du jeu de poker se traduit par un traitement médiatique plus important, les métaphores exploitant la notion de poker restent très fréquentes.

Keywords: analyse diachronique, corpus, jeux d'argent et de hasard

1. Introduction

L'analyse diachronique de corpus opère sur un champ assez large. Nous pouvons en juger par exemple en observant les nombreux travaux sur l'évolution des langues, travaux qui passionnent aussi bien la communauté scientifique (Dediu & de Boer 2016) que les médias si l'on se fie par exemple à l'intérêt renouvelé porté par ceux-ci sur l'évolution des dictionnaires. Dans le champ purement scientifique, les intérêts dans le domaine embrassent tous les niveaux de l'analyse linguistique même si la morphologie (Macaulay 2017) et le lexique (la néologie par exemple chez Gérard et al. 2014). La sémantique est un autre aspect des études diachroniques notamment pour étudier les représentations mentales des locuteurs (Hamilton et al. 2016). Le travail présenté ici s'intéresse à une autre catégorie de représentations mentales qui est l'image que certaines activités ludiques peuvent prendre au cours du temps. Nous nous intéressons ici à un jeu d'argent et de hasard qui

a connu une sorte de nouvelle jeunesse ces dernières années : le jeu de poker. Dans ce travail, nous nous inspirons de l'analyse de l'usage du lexique dans (Hamilton et al. 2016), nous souhaitons examiner l'évolution de l'usage d'un mot, d'un terme particulier au cours du temps. Ce travail, même si notre ambition est moins large, peut se rattacher aux études sur la néologie sémantique (Sablayrolles 2002) ou néosémie (Rastier et Valette 2009). Pour illustrer l'intérêt que représente le poker en tant que phénomène de société, nous pouvons considérer le retentissement autour du *MoneyMaker Effect*¹ ou encore cette citation du journal Le Monde daté du 22 janvier 2007 qui illustre le changement d'image de ce jeu: « *Considéré il y a encore peu de temps comme un jeu sulfureux se jouant dans les arrière-salles de bars louches ou dans des appartements huppés à l'abri des regards indiscrets, le poker fait une entrée en force à la télévision* ». En particulier, dans sa variante à la mode Texas Hold'Em, le poker est redevenu un jeu dont on parle et dont on parle plutôt positivement. Notre objectif est d'une part de mesurer à quel point ce regain d'attention a pu se traduire par une amélioration de l'image du jeu de poker en général. D'autre part, il s'agit de voir dans quelle mesure les usages métaphoriques du terme poker, plutôt connotés "négativement" (poker menteur, coup de poker²...) ont pu évoluer conjointement à cette plus grande popularité du jeu lui-même. Dans la section 2 nous présenterons le corpus que nous avons constitué pour cette étude. Puis, nous proposerons dans les deux sections suivantes une analyse statistiques des prédicats puis une analyse sous forme de *clustering*. Enfin, nous présenterons nos conclusions et perspectives.

2. Présentation de notre corpus d'étude

De manière à pouvoir s'affranchir des variations de choix éditoriaux entre journaux, nous nous avons souhaité nous concentrer sur une seule publication. Nous avons choisi le Monde ce qui nous permettait d'exploiter des articles dont la publication s'étalent sur 30 ans : 1988-2017. Pour la partie 1988-2005 nous avons utilisé le corpus du monde distribué par ELRA³, nous avons restreint aux textes contenant le terme poker. Pour les années 2006 à 2017 nous avons extrait d'Europresse⁴ les articles qui comportait le terme poker. Dans les deux cas nous avons considéré toutes les variantes possibles dans la casse. Nous avons ainsi obtenu 3528 textes dont la répartition dans le

¹ Par exemple : http://www.slate.com/articles/news_and_politics/explainer/2011/06/the_moneymaker_effect.html

² Dans le sport par exemple, on remarque des contextes de « tentative désespérée », « dernière chance » ...

³ http://catalog.elra.info/product_info.php?products_id=438&language=fr

⁴ <http://www.europresse.com/fr/>

temps est présentée Figure 1. Nous pouvons observer que le nombre d'articles a connu une chute entre 2005 et 2006. Ceci semble être dû au fait que nous passons à ce moment précis d'une étude du corpus complet du monde tel qu'existant auprès d'ELRA à une étude fondée sur la base Europresse. De fait, sur nos critères de recherche, la base Europresse ne totalise que 47 articles pour 2003 (contre 129 dans le corpus ELRA), 62 pour 2004 (contre 117) et 67 articles pour 2005 (contre 117). Les contraintes respectives d'utilisation de ces deux sources de données nous ont interdit de pouvoir disposer d'un corpus dont la constitution soit constante. Nous nous sommes efforcés de s'affranchir de ce biais en adaptant notre méthodologie (notamment le *clustering*).

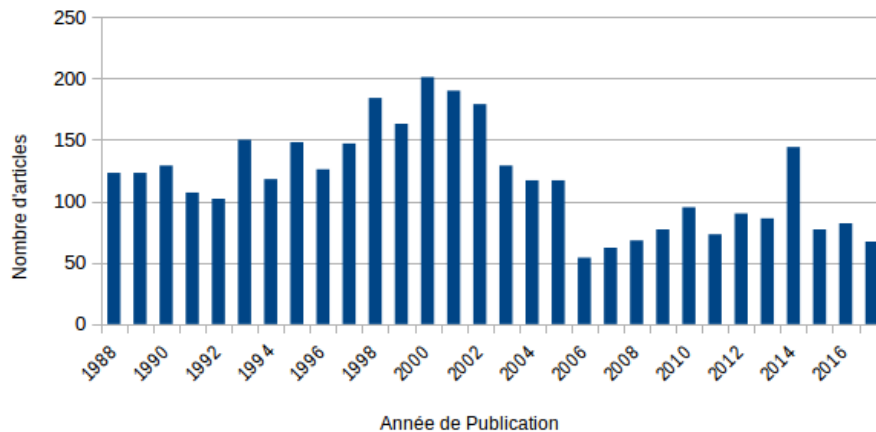


Figure 1 : Répartition du nombre d'articles par année

Nous avons 4353 occurrences du terme recherché, leur répartition est instructive (Figure 2) : la très grande majorité des articles (2834/3528 soit 80,33%) ne comporte qu'une seule occurrence. Nous pensons que ceci est le reflet de deux tendances. D'une part le sujet de l'article est rarement le poker pour lui même, il est question d'un personnage qui par ailleurs joue au poker par exemple. D'autre part, cette rareté de la répétition révèle un usage massivement métaphorique, en effet comme l'a montré (Lejeune 2013) une métaphore perd de sa force en étant répétée. Si un terme est répété, il est très probable qu'il soit employé dans son sens plein. Si cette observation était faite sur des noms de maladies infectieuses, il nous semble que ceci est avant tout lié au genre de texte et que cela s'applique également ici. Si nous allons un peu plus loin, nous pouvons faire l'hypothèse que la métaphore peut être filée, mais qu'elle est rare dans les articles expositifs. D'autre part, dans le cas peu probable d'une métaphore filée, les conventions stylistiques impliquent de changer le terme employé, le journaliste utilisera plutôt des termes du même champ lexical.

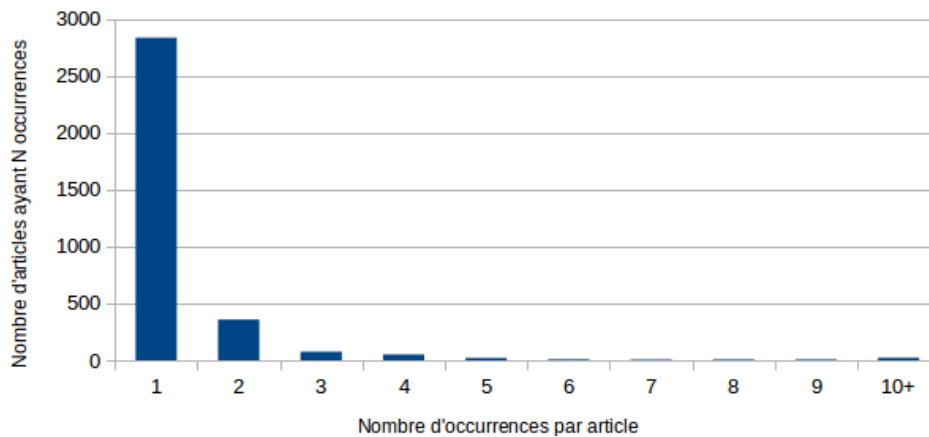


Figure 2 : Répartition des d'articles selon le nombre d'occurrences du terme « poker »

La répartition des articles entre ceux qui comportent une et une seule occurrence et ceux qui en comportent plusieurs montre des variations importantes dans le temps (Figure 3). Si l'on observe des périodes de 5 ans, on peut se rendre compte que le nombre d'articles comprenant plusieurs occurrences de "poker" représente 15% des articles sélectionnés sur la période 1988-1992, se pourcentage descend à 10% jusqu'en en 2003 puis remonte progressivement pour finalement rester au-dessus de 20% à partir de 2004-2008 avec une pointe à 30% pour les périodes 2007-2011 à 2009-2013.

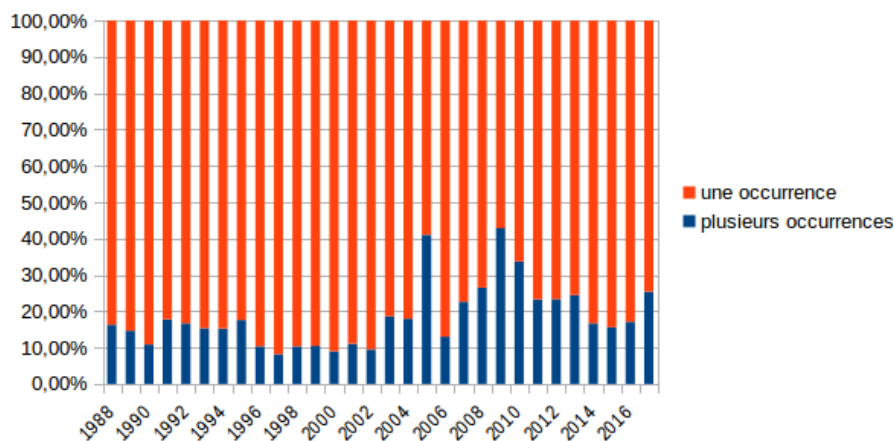


Figure 3 : Répartition par année des articles selon le nombre d'occurrences

3. Prédicats et séquences figées

Dans la théorie linguistique lexique-grammaire de M. Gross (1975) et de G. Gross (2012), les prédicats sont considérés comme les noyaux d'une phrase capables de disposer d'arguments, grâce à leurs propriétés transformationnelles et distributionnelles. Parmi les apports de cette théorie figurent le « schéma d'arguments » et les « prédicats appropriés ». Nous relevons dans notre corpus les contextes gauches et droits des séquences

figées « partie de poker » et « coup de poker » afin de distinguer leurs emplois métaphoriques et non métaphoriques. Ce travail est fait en étudiant le premier verbe précédant ou suivant l'expression (sans remonter au-delà d'une phrase). Nous montrons dans les tableaux 1 et 2 les 20 verbes les plus fréquents pour chaque contexte se trouvent le plus fréquemment dans ces contextes (20 dans les contextes gauches, 20 dans les contextes droits).

Tableau 1 : Effectif des verbes dans le contexte gauche de “[partie|coup] de poker”

être (76)	jouer (62)	faire (15)	tenter (14)	gagner (11)
avoir (11)	ressembler (10)	prendre (9)	tenir (8)	lancer (8)
perdre (7)	voir (6)	partir (6)	engager (6)	agir (5)
réussir (4)	livrer (4)	remporter (3)	organiser (3)	mener(2)

Tableau 2 : Effectif des verbes dans le contexte droit de “[partie|coup] de poker”

être (98)	avoir (75)	jouer (16)	pouvoir (13)	devoir (8)
gagner (7)	engager (7)	venir (6)	livrer (6)	faire (5)
vouloir (4)	voir (4)	tenter (4)	tenir (4)	réussir (4)
prendre (4)	monter (4)	bluffer (4)	aller (4)	retrouver (3)

Hormis les verbes « être » et « avoir » qui sont susceptibles d'être des verbes auxiliaires ou semi-auxiliaires, pour les autres verbes on peut se trouver dans trois cas de figure :

- h) Verbe support
- i) Prédicat approprié : le sens littéral de l'expression peut être activé
- j) Prédicat non approprié : le sens métaphorique de l'expression est activé

Le cas des verbes support n'est pas pertinent pour notre étude. Pour le second cas, nous observons que le verbe jouer, prédicat approprié pour les deux séquences décrites, est très souvent lié à un usage métaphorique. Dans le troisième cas, de loin le plus fréquent. Les verbes « tenter », « s'engager », « réussir », « mener », « lancer » voire « remporter » ne sont pas tout à fait congruents avec le sens premier de la séquence, c'est-à-dire qu'ils ne sont pas des prédicats appropriés au sens propre du jeu de poker. Des occurrences de ces verbes dans le corpus confirment cette intuition :

*Il leur fallait **lancer** la partie de poker que Bonn et Paris s'apprêtent à jouer sur le GATT (1993)*

*les enjeux de la partie de poker qui **s'engagera** mercredi à la mi-journée lorsque l'ambassadeur[...] (2017)*

*[ils] avaient pu croire un moment que leur coup de poker allait **réussir**. (1989)*

[Celui qui] est davantage connu pour ses coups de poker financiers continue à **mener** sa stratégie (2015)

Elle venait de **remporter** la partie de poker menteur qui constitue l'essentiel des premiers hectomètres. (1995)

4. Étude des champs lexicaux par *clustering*

Si les séquences « partie de poker » et « coup de poker » sont ambiguës dans le sens où elles figurent dans des champs lexicaux différents, on peut se demander ce qu'il en est des champs lexicaux du terme « poker » en général. Pour étudier cette question, nous avons réalisé un *clustering* de notre corpus. Nous avons utilisé l'implantation des k-moyennes (*K-means*) de la bibliothèque Python *scikit-learn*. Nous avons fixé le nombre de *clusters* *K* à 10^5 et le nombre maximal d'itérations à 400, la mesure des poids est le *tf-idf*. Nous avons extrait tous les n-grammes de mots avec *n* allant de 1 à 3 puis seulement nous avons utilisé une stop-list. De sorte que, par exemple, « de » n'était pas gardé en tant que tel mais que nous le retrouvions dans « coup de poker » ou « loi de Robien ». Nous avons tout d'abord travaillé sur le corpus lemmatisé puis nous avons observé que les résultats étaient semblables sans lemmatisation, nous avons donc supprimé ce pré-traitement. Nous allons maintenant décrire chaque *cluster* en donnant la proportion du corpus qu'il couvre ainsi que les 10 termes les plus significatifs.

Cluster 0, « sport et poker 1 » : 3,1 % (club, football, équipe, Ligue, France, championnat, saison, joueurs, OM, Marseille) Ce *cluster* comporte deux volets : l'un sur les « coups de poker » dans les championnats de football et l'autre où il est question des championnats de poker eux mêmes.

Cluster 1, « politique » : 18,79 % (ministre, président, politique, gouvernement, pays, État, premier ministre, premier, États, faire). Un *cluster* autour de l'action politique, notamment au niveau européen. Un exemple intéressant de métaphore (filée) ici : « M. Erdogan remet tout en jeu, comme un joueur de **poker** fait tapis »

Cluster 2, « fourre-tout » : 38,01 % (être, bien, film, vie, entre, Jean, monde, France, temps, homme) Le seul de nos *clusters* qui n'ait pas d'unité ni de tendance thématique, ici les expressions contenant *poker* sont pour moitié métaphoriques.

Cluster 3, « culture_1 » : 5,13 % (film, Booker Prize, roman, prix, livres, livre, littéraire, base, prix littéraire, attribué). Ce *cluster* rassemble les livres ayant trait au *poker*, les expressions liées sont prises dans leur sens littéral

⁵ Selon la méthode du coude (*elbow method*), la valeur optimale se situait entre 9 et 12.

(l'expression « coup de poker » y est quasi absente).

Cluster 4 « finance »: 4,2 % (Vivendi, marché, groupe, Bourse, marches, actionnaires, titres, taux, millions, fonds, terme, milliards, prix) Il se caractérise uniquement par des thématiques associées au domaine de finance et notamment aux coups de poker boursiers.

Cluster 5 « sport et poker 2 »: 5,04 % (Coupe, match, équipe, joueurs, France, club, football, finale, francs, PSG). Nous avons ici un *cluster* sur le sport où environ la moitié des articles concernent toutefois le poker lui même.

Cluster 6 « industrie du poker »: 12,96 % (jeux, paris, ligne, marché, milliards, euros, millions, Internet, dollars, Bourse) Ici nous avons tout ce qui est lié à l'industrie du poker et notamment à l'essor des jeux d'argent sur Internet (dont le poker a été un fer de lance).

Cluster 7 « sport »: 3,26 % (Tour, numéros, France, coureur, étape, peloton, course, équipe, Tour de France, maillot) Nous avons ici des usages, massivement métaphoriques, dans le domaine du sport (principalement le cyclisme). Un exemple avec le terme spécialisé *flop* : « [P.A.Bosse] avait trouvé cette image [...] : Si on compare le 1500m au poker, il a un *flop* d'avance. »

Cluster 8 « culture_2 » : 7,14 % (blues, musique, CD, rock, John Lee Hooker, jazz, album, guitare, musiciens, scène) Un usage métaphorique dans le domaine de la musique avec des expressions telles que « poker face », « poker perdant »...

Cluster 9 « culture_3 » : 2,38 % (Dracula, Bram Stoker, vampire, roman, film, fantastique, Christie, Coppola, comte, Frankenstein) Le *cluster* 3 était centré sur le domaine littéraire, ici il est question de cinéma et particulièrement des personnalités liées au poker. L'usage y est surtout littéral.

Pour ce qui est de la répartition temporelle, il est très intéressant de noter que le cluster 6 (l'industrie du poker) devient le second plus important derrière le cluster 2 (à partir de 2005 (popularisation des jeux d'argent sur Internet) et plus encore à partir de 2010 (légalisation des paris en ligne). Le *cluster* 0 (sport et poker) devient plus important à partir de 2004 d'autant qu'en son sein la thématique poker y est alors largement majoritaire.

5. Conclusion

Nous avons proposé dans cet article une étude diachronique d'articles de presse contenant le mot « poker ». Notre hypothèse initiale était que ce terme était souvent employé dans des expressions métaphoriques et que le regain de popularité de ce jeu depuis quelques années avait du amener une plus grande proportion d'usage littéral. Nous avons observé que dans plus de 80 % des cas, le terme poker n'apparaissait qu'une fois dans les textes. Nous

avons montré que ceci était dû à un usage principalement métaphorique, on ne répète pas une métaphore, mais aussi au fait que le poker est rarement le sujet central de l'article. Cette tendance change quelque peu à partir de 2005, le poker devenant lié à des championnats et des retransmissions télévisuelles plutôt qu'à des tripots et des casinos. Enfin, nous avons montré que les usages métaphoriques relevaient très majoritairement de 3 domaines : la finance, la politique et le sport.

References

- Dediu D. and de Boer B. (2016)., Language evolution needs its own journal , Journal of Language Evolution, Volume 1, Issue 1, 1 January 2016, Pages 1–6
- Gérard C., Falk I., and Bernhard D. (2014). Traitement automatisé de la néologie : pourquoi et comment intégrer l'analyse thématique ? Actes du 4e Congrès mondial de linguistique française (CMLF 2014), Berlin, Pages 2627-2646
- Gross, M. (1975). Méthodes en syntaxe: régime des constructions complétives, Hermann.
- Gross, G. (2012). Manuel d'analyse linguistique: Approche sémantico-syntaxique du lexique, Presses Universitaires du Septentrion.
- Hamilton W.L., Leskovec J., and Jurafsky D. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In Proc. Of the Association for Computational Linguistics Conference (ACL) 2016
- Lejeune G. (2013) Veille épidémiologique multilingue : une approche parcimonieuse au grain caractère fondée sur le genre textuel, Thèse de doctorat en Informatique de l'Université de Caen
- Macaulay, M. & Salmons. (2017). Synchrony and diachrony in Menominee derivational morphology, J. Morphology 27: 179
- Rastier, F., Valette, M. (2009) « De la polysémie à la néosémie », Le français moderne, S. Mejri, éd., La problématique du mot, 77, 97-116.
- Sablayrolles, F. (2002). « Fondements théoriques des difficultés pratiques du traitement des néologismes », Revue française de linguistique appliquée, VII-1, pp. 97-111.