

Des outils pour décrire certains corpus de poèmes et de chansons : les arbres additifs simultanés.

Ludovic Lebart¹

¹CNRS (R). ludovic@lebart.org

Abstract

The statistical analysis of poetic texts and songs is a methodological challenge. These texts may include choruses, repetitions or versification constraints which question the statistical significance of lexical frequencies. We must then work on "bags of words" (presence or absence of words) which evidently constitute a tiny part of the artistic aspect of the texts, but nevertheless prove worthy of interest in terms of the study of an artwork. With this coding, the dimensionality and sphericity of the cloud do not allow good visualizations in a low-dimensional space. Correspondence Analysis (CA), despite its undeniable advantage of providing a simultaneous representation of words and texts, becomes insufficient. The calculation of additive trees is essential. In this contribution, we propose a new procedure for the simultaneous representation of texts and words for additive trees. Such procedure allows us to combine the advantages of CA and clustering. We briefly illustrate this new tool with applications to corpora of lyrics of three francophone iconic singers and poets: Brassens, Aznavour and Brel.

Keywords: Additive Trees, Correspondence Analysis, Data Visualization, Poetry, Lyrics

Résumé

L'analyse statistique des textes poétiques et des chansons est un défi méthodologique. Ceux-ci peuvent comporter des refrains, des répétitions ou des contraintes de versification qui mettent en question la signification statistique des fréquences lexicales. On doit alors travailler sur des « sac de mots » (présence ou absence de vocables) qui ne constituent qu'une partie infime de l'aspect artistique des textes, mais s'avèrent cependant dignes d'intérêt au niveau de l'étude d'une œuvre. Avec ce codage, la dimensionnalité et la sphéricité du nuage ne permettent pas de bonnes visualisations dans un espace de faible dimension. L'analyse des correspondances (AC) devient insuffisante, malgré sa précieuse propriété de représentation simultanée des mots et des textes. Le calcul d'arbres additifs s'impose. Nous proposons dans cette contribution une nouvelle procédure de représentation simultanée des textes et des mots pour les arbres additifs. Cette procédure permet de cumuler les avantages de l'AC et des classifications. On illustrera très schématiquement ce nouvel outil par des applications aux corpus de trois paroliers/chanteurs et poètes francophones Brassens, Aznavour et Brel.

Mots clés : Arbres additifs, Analyse des Correspondances, Visualisations, Poésies, Chansons

1. Introduction

L'exploration statistique des recueils de poésies et de chansons se heurte à la spécificité de ces textes. Ceux-ci peuvent comporter des refrains, des répétitions volontaires, des figures de style ou des contraintes de versification qui questionnent la signification statistique des fréquences lexicales. On verra qu'il nous faudra alors travailler sur des « sacs de mots » (présence ou absence de vocables) qui ne constituent qu'une partie infime de l'aspect artistique des textes, mais s'avèrent cependant utiles au niveau de l'étude d'une œuvre dans sa temporalité. Il ne s'agit pas d'une étude approfondie ni d'une élucidation de textes poétiques, mais simplement de traces lexicales, qui s'avéreront *in fine* dignes d'intérêt. L'analyse des correspondances (AC) reste l'étape indispensable pour un premier contact avec les tableaux lexicaux, mais ne garantit

pas une vue d'ensemble. Les arbres additifs, avec leurs règles de lecture, permettent de reconstituer dans un plan des distances dans des espaces de dimensions bien supérieures à deux. Il leur manquait la représentation simultanée des lignes et des colonnes (*i.e.* : des mots et des textes) de la table lexicale de base. C'est ce que nous proposons dans cette contribution, avec trois brefs exemples d'application aux œuvres de trois chanteurs/poètes. Il faudrait au moins trois monographies de plusieurs dizaines de pages chacune pour traiter convenablement le sujet. Les brefs exemples qui suivront restent donc de simples illustrations surtout techniques et méthodologiques.

2. Insuffisance des axes principaux

Avec le codage de présence-absence de mots dans des unités textuelles, les fréquences sont en quelque sorte bridées. Les sacs de mots sont faits à partir d'une certaine taille d'unités textuelles. Ces unités seront les poèmes eux-mêmes dans le cas d'œuvres poétiques (sonnets dans le cas de Shakespeare, par exemple). Dans notre cas, pour Brassens, Aznavour et Brel, les unités textuelles seront les chansons. Les textes seront les albums (disques) ou des recueils de chansons par année de parution.

La fréquence maximale d'un mot ne peut dépasser le nombre de groupes de textes (recueils ou albums dans le cas de chansons) puisqu'il ne s'agit que d'un codage de « présence-absence » à l'intérieur de chaque unité textuelle. Dans ces conditions, la dimensionnalité et la sphéricité du nuage ne permettent que rarement de bonnes visualisations dans un espace de faible dimension. Dans le cas des méthodes en axes principaux (analyses en composantes principales ou des correspondances) les valeurs propres seront peu détachées, une cause d'instabilité pour les axes principaux. La représentation bidimensionnelle des mots et des textes, atout indéniable des méthodes en axes principaux, devient insuffisante.

3. Classification et arbres additifs

Le calcul d'arbres additifs, proposé à l'origine par Buneman (1971) a été amélioré par Sattath et Tversky (1977). Les travaux de Barthélémy et Guénoche (1988) et de Luong (1988) ont favorisé l'utilisation de ces méthodes (sous le nom d'analyses arborées) dans le champ des analyses de texte. Les arbres additifs ont été rendus aisément calculables par Saitou et Nei (1987) et vont alors s'imposer comme synthèse entre les axes principaux et les techniques de clustering (classifications hiérarchiques, partitions type k-means, etc.). Bryant (2005), puis Huson et Bryant (2006) vont ensuite justifier la méthode et rendre accessible un logiciel correspondant. Des justifications théoriques de l'efficacité de l'algorithme ont été présentées par Mihaescu *et al.* (2009).

Rappelons brièvement que le concept de hiérarchie à la base de la classification ascendante revenait à approximer les distances initiales par une distance *ultramétrique*, qui vérifie, en plus des axiomes classiques de toute distance, pour tout triplet (x, y, z) , l'inégalité :

$$d(x, y) \leq \max(d(x, z), d(y, z))$$

Les arbres additifs demandent, pour tout quadruplet (x, y, z, t) , que soit vérifiée l'inégalité plus complexe et moins intuitive, mais moins exigeante :

$$d(x, y) + d(z, t) \leq \max(\{d(x, z) + d(y, t)\}, \{d(x, t) + d(y, z)\})$$

Un arbre peut alors être dessiné avec les objets à classer comme éléments terminaux (ou « feuilles »). Plus souple que l'arbre de longueur minimale (*Minimum spanning tree*) qui dépend de $n-1$ paramètres, l'arbre additif implique $2n-3$ paramètres. Il faut donc trouver une

approximation des distances initiales qui satisfasse ces conditions, ce que permettent les travaux précités de Saitou et Nei. Le résultat important est que les distances entre nœuds (textes) du graphe (longueur du plus court chemin sur le graphe) sont des approximations des distances réelles dans tout l'espace, approximations souvent meilleures que celles fournies par un premier plan factoriel de CA.

Nous proposons donc dans cette contribution une procédure de représentation simultanée des textes et des mots pour les arbres additifs qui permet de cumuler les avantages des méthodes en axes principaux et des classifications. Plus généralement, cette procédure s'applique à la représentation simultanée des colonnes et des lignes de toute table de contingence.

4. Sur les options de traçage des arbres additifs

Il reste à tracer ces arbres. Pour une revue assez complète des visualisations de graphes généraux, on pourra consulter Di Battista *et al.* (1999), et plus particulièrement sur les méthodes utilisant les *force-directed drawings algorithms*, l'article de Kobourov (2013) qui analyse plus de 60 publications correspondant à plusieurs dizaines d'algorithmes. A l'origine, l'algorithme de tracé de graphes de Tutte (1963) est l'une des premières méthodes de tracé fondée sur des algorithmes de ce type. Puis les méthodes proposées par Eades (1984) et l'algorithme de Fruchterman et Reingold (1991) reposent toutes deux sur des forces répulsives entre tous les nœuds du graphe, mais aussi des forces attractives entre les nœuds qui sont adjacents (les arêtes sont assimilées à des ressorts, et il s'agit de trouver un équilibre entre toutes les tensions, d'où le nom de *force-directed drawings*).

Alternativement, les forces entre les nœuds peuvent être calculées sur la base de concepts de la théorie des graphes. Les distances entre nœuds sont alors les longueurs des plus courts chemins qui les joignent. Pour les arbres additifs, ces distances sont justement une approximation des distances originales (distances du chi-2 calculées sur la table lexicale originale). L'algorithme de Kamada et Kawai (1989) utilise ces « forces de ressort » proportionnelles à ces distances calculées sur le graphe. Cet algorithme de tracé est donc le plus compatible avec les propriétés des arbres additifs, et donc avec les distances lexicales de base. Expérimentalement, on constate d'ailleurs la bonne compatibilité de ces représentations avec les plans principaux issus de l'AC du tableau lexical original.

Les capacités de l'analyse des correspondances à décrire des graphes non-orientés du type grille plane ou carte géographique à partir de leurs matrices associées ont été soulignées par Benzécri (1973, chap. 10) et Lebart *et al.* (1998). On aurait pu penser utiliser de nouveau l'AC pour obtenir un tracé de l'arbre additif. Mais les arbres (graphes connexes sans cycle) sont mal visualisés par cette méthode, et réciproquement, comme le notent Sattath et Tversky (1977) "*It is interesting to note that tree and spatial models are opposing in the sense that very simple configurations of one model are incompatible with the other model. For example, a square grid in the plane cannot be adequately described by an additive tree.*"

Finalement, nous partirons donc des représentations fournies par la méthode de Kamada-Kawai la plus adaptée pour les arbres additifs, pour les enrichir par une représentation simultanée des lignes et colonnes de la table lexicale.

5. Représentation simultanée et mots caractéristiques

On peut présenter directement l'Analyse des Correspondances comme la recherche de la meilleure représentation simultanée possible des proximités entre lignes et colonnes d'une table de contingence (Lebart *et al.*, 1984). On peut en effet chercher sur un axe (pour commencer)

un positionnement simultané des textes et des mots de façon à obtenir une relation doublement barycentrique : mots au barycentre des textes, et textes au barycentre des mots (les poids étant respectivement les profils lexicaux en ligne et en colonne calculés sur la table lexicale de base). Cette double relation est impossible, car la prise de barycentre est contractante : les mots doivent être à l'intérieur de l'intervalle couvert par les textes et, simultanément, les textes à l'intérieur de l'intervalle couvert par les mots. Pour que la relation soit possible, il faut dilater ces barycentres (coefficient $\beta > 1$). La solution optimale correspond à une valeur de β la plus proche de 1 qui nous donne les positionnements des mots et des textes sur le premier axe de l'analyse des correspondances. On notera la simplicité de cette présentation de l'AC obtenue directement à partir des relations doublement barycentriques connues sous le nom de « relations de transitions ».

On ne peut donc pas espérer trouver une double représentation simultanée optimale, mais une simple représentation simultanée est suffisante (mots comme barycentres des textes, qui sont les nœuds de l'arbre). On enrichira cette représentation simultanée en faisant aussi intervenir la notion de mots caractéristiques (ou spécificités).

La procédure de représentation simultanée que nous proposons comprend les étapes suivantes : (colonnes et lignes jouent des rôles similaires et peuvent être interchangeables).

- 1) Analyse des correspondances préliminaire de contingence (table lexicale).
- 2) Choix de la dimension nx de l'espace jugé significatif (en général par *bootstrap*) (12 axes par exemple). Les distances seront calculées à partir des nx premiers axes principaux de l'AC (on verra quand nos exemples illustratifs, $nx = 12$). Cela permet une régularisation des distances initiales, procédure bien connue en analyse discriminante et dans certaines procédures de Deep Learning.
- 3) Calcul de l'arbre additif (*Neighbors-Joining method*) sur la matrice de distances entre les coordonnées des colonnes (textes) sur les nx premiers axes.
- 4) Tracé de l'arbre (procédure de Kamada-Kawai).
- 5) Positionnement barycentrique des lignes (mots - formes, lemmes) à partir des coordonnées des points-colonnes (textes) (nœuds de l'arbre) déduites de la procédure (4) et du profil textuel des lignes (mots/graphies/lemmes).
- 6) Calcul, directement à partir de la table lexicale, pour chaque colonne-texte, des lignes/mots caractéristiques (seuil probabiliste fixé) à partir des valeurs-test (pour les valeurs-test, voir par exemple : Lebart *et al.* ; 1998, 2019).
- 7) Tracés de nouvelles arêtes (couleur et épaisseur différentes de celles des arêtes de l'arbre additif) joignant sur le graphe chaque point-colonne (texte) à ses lignes (mots) caractéristiques. Dans le cadre de cette contribution, le graphique sera monochrome.

Ces sept étapes sont en fait valables pour toutes tables de contingence. Dans le cas de données textuelles, il faut ajouter une étape « 0 » de calcul de la table lexicale à partir des textes. Dans le cas des textes formés de chansons ou de poèmes, il faut encore ajouter une étape « -1 » préliminaire de conversion des textes bruts des chansons en « sacs de mots ».

On illustrera ces « arbres augmentés » (arbres additifs avec représentation simultanées des lignes et des colonnes) avec des applications aux corpus de trois musiciens et poètes prédominants dans la chanson francophone du vingtième siècle : Georges Brassens, Charles Aznavour et Jacques Brel.

Le code *python* de la chaîne complète des sept ou neuf étapes de traitement allant des textes bruts aux visualisations simultanées d'arbres additifs sera libre et disponible.

6. Le corpus Brassens

Nous évoquons ici le recueil de 194 chansons chantées et enregistrées par le musicien–poète français Georges Brassens (1921–1981) regroupées en 14 recueils correspondant à autant d’albums (disques). Cet auteur non-conformiste, qui a fréquenté les mouvements anarchistes, a cependant reçu en 1967 le prix de poésie de l’Académie Française. Traduit en plusieurs langues (anglais, italien, allemand, japonais...), il a été à l’origine de la vente de plusieurs dizaines de millions de disques.

Les textes poétiques de Brassens sont particulièrement riches en figures de style (litotes, métaphores, anaphores, euphémismes, allégories, ...) qui posent des problèmes lors de l’utilisation du mot (graphie ou lemme) comme unité statistique de base (Rochard, 2009). Brunet (2004) lors d’une étude de l’œuvre poétique d’Arthur Rimbaud, lance un avertissement qui s’applique à cette section et aux deux sections suivantes : « ... L’emploi des outils documentaires et statistiques ne va pas sans une certaine naïveté qui donne sa foi aux mots imprimés, dans leur innocence première. Or chez Rimbaud, les mots sont souvent pipés. Ce sont des leurres, des prête-noms, et la réalité qu’ils désignent et qu’ils cachent se dérobe aux plus savantes exégèses... ». De plus, on a vu que dans le cas des chansons pouvant comporter des refrains ou des répétitions intentionnelles, les fréquences lexicales n’ont plus la signification statistique qu’on leur donne dans les tableaux lexicaux. Les contraintes de versification (alexandrins par exemple, rimes) sont difficiles à intégrer dans les outils de description, mais ont aussi influencé le choix des mots et leur fréquence.

Pour remédier à cette défaillance statistique des fréquences des formes graphiques (graphies) ou des lemmes, on va donc transformer chaque texte de chanson en vocabulaire non pondéré, autrement dit, chaque élément n’apparaîtra qu’une seule fois à l’intérieur d’une chanson donnée (*words bag*). Comme souvent en analyses textuelles, on aura en fait deux jeux de données provenant d’une part du fichier brut, d’autre part du fichier lemmatisé. Le fichier lemmatisé a l’avantage de réduire la diversité des flexions et donc de permettre des seuils de fréquence minimale plus bas. Le fichier des graphies garde la diversité originale des formes ce qui est fondamental dans le cas de textes poétiques. On obtiendra donc à chaque étape deux points de vue différents et complémentaires.

Pour la plupart des analyses de type AC portant sur l’ensemble des 194 chansons, ou sur les 170 chansons dont Brassens est le seul auteur, une dimension est dominante, qu’il s’agisse de lemmes ou de graphies : elle oppose les textes anciens (quatre ou six premiers albums) aux textes plus récents. Les textes anciens, comme les poèmes externes, ont un vocabulaire que l’on peut qualifier de classique, voire galant ou précieux, pour forcer le trait. Les plus récents ont un vocabulaire plus cru, parfois argotique, provocateur, « salle de garde ».

La figure 1 nous donne le tracé de l’arbre additif « nu » avec ses 14 nœuds. La table lexicale comporte 944 lignes (formes graphiques) et 14 colonnes (14 recueils des 170 chansons entièrement écrites et composées par Brassens). La figure 2 qui reprend le tracé de l’arbre à l’identique ne comporte qu’un tout petit extrait des graphies actives. Pour chacun des 14 recueils analysés, on ne retient que les 6 graphies les plus caractéristiques au sens des valeurs-tests. Ces valeurs, calculées directement à partir de la table lexicale de base, permettent d’obtenir les mots les plus caractéristiques de chaque recueil, avant toute analyse.

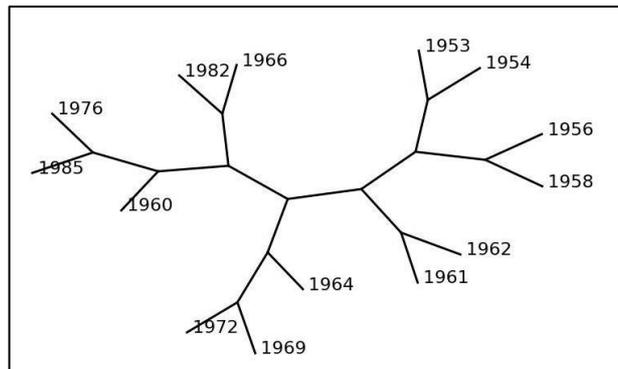


Figure 1 : Brassens. Tracé de l'arbre additif seul. Les 14 textes (albums) sont repérés par l'année de parution, de 1953 à 1985 (1985 est une publication posthume et une compilation hétérogène du point de vue des dates). La présence ou l'absence d'une graphie dans une chanson peut paraître un indicateur sommaire, et pourtant les années sur l'arbre sont loin d'être réparties au hasard... (par exemple, premières années sur la partie droite). Des « anomalies » de ce point de vue, comme la position de l'année 1960, loin de ses voisins dans le temps, méritent des commentaires de spécialistes dont le volume dépasse le cadre de cette communication technique. Si une direction horizontale correspond approximativement au premier axe de l'AC de la même table lexicale, les diverses branches de l'arbre déploient les axes suivants (les distances ont été calculées à partir des 12 premiers axes).

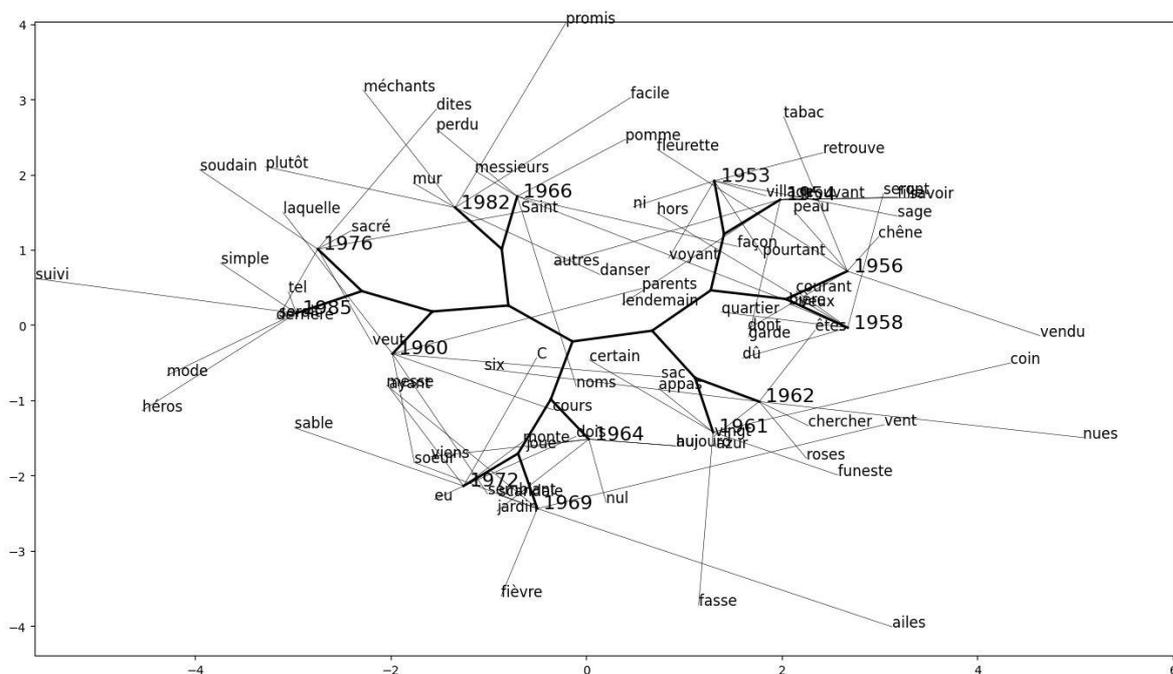


Figure 2 : Brassens. Reprise du tracé de l'arbre additif de la figure 1, avec représentation simultanée de graphies et des 14 textes. Ici, pour des raisons évidentes d'encombrement graphique, le nombre de graphies caractéristiques a été limité à 6 par texte. Le format et l'absence de couleur limitent également la portée de ce qui est un document de travail : les 14 textes sont caractérisés par les 944 formes graphiques apparaissant plus de 4 fois. Bien que les quelques graphies portées sur cette figure soient jointes par un trait fin aux textes qu'elles caractérisent le plus, il ne faut pas perdre de vue qu'elles sont des barycentres (dilatés) des textes/années, et donc qu'elle se positionnent par rapport à l'ensemble de ces textes.

7. Le corpus Aznavour

Notre corpus comprend 295 chansons reparties en 18 recueils (albums). Charles Aznavour (1924 – 2018) a commencé sa carrière musicale dans les années 1940, et a enregistré près de mille deux cents chansons interprétées en plusieurs langues. Il a écrit ou coécrit plus de mille chansons, que ce soit pour lui-même ou d'autres artistes. Elles sont publiées dans le livre : (« Chansons : L'intégral ». Aznavour, 2010). Aspect peut-être moins connu du grand public, Aznavour fut longtemps parolier pour d'autres chanteurs, avant d'être un des chanteurs français les plus reconnus en dehors du monde francophone. Il a écrit pour Édith Piaf, Eddie Constantine, Gilbert Bécaud, Juliette Gréco, Sylvie Vartan, Mireille Mathieu. Il a collaboré avec Fred Astaire, Frank Sinatra, Andrea Bocelli, Bing Crosby, Ray Charles, Liza Minnelli, Tom Jones, etc. Aznavour était un homme de scène, un grand acteur de cinéma récompensé en tant que tel, un voyageur infatigable, avec des engagements politiques parfois spectaculaires, et des décorations internationales impressionnantes... un peu aux antipodes du poète bourru et timide de la section 6.

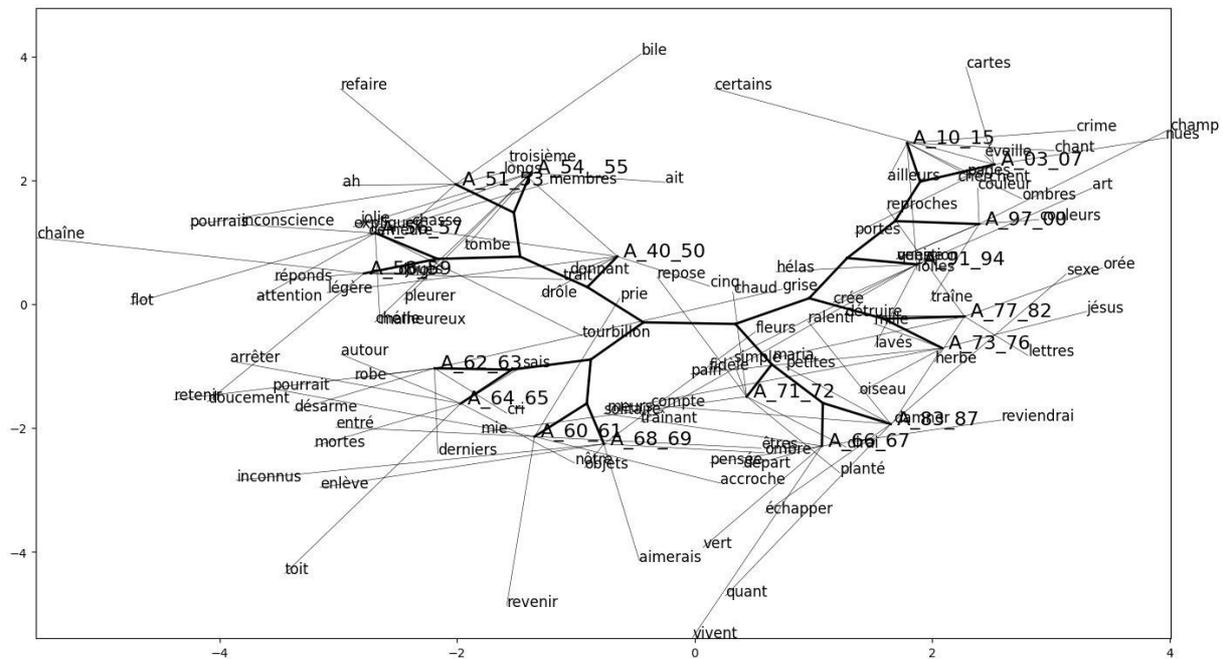


Figure 3 : Aznavour. Esquisse sommaire du tracé de l'arbre additif avec représentation simultanée des graphies et des textes. Ce sont ici les années récentes qui constituent la partie droite de l'arbre (« 10-15 » signifie « 2010-2015 »). Même limitation concernant le nombre de graphies (6 éléments caractéristiques par texte). Les textes (albums) sont repérés par l'année de parution, de 1940 à 2015 (intervalle de temps deux fois plus large que pour Brassens, et pour Brel qui va suivre). Contrairement à Brassens, le vocabulaire est simple, peu littéraire, mais parfois plus cosmopolite.

8. Le corpus Brel

Jacques Brel (1929 – 1978) est un auteur-compositeur-interprète, poète, acteur et réalisateur belge. Jacques Brel est considéré comme un des plus grands auteurs-interprètes de la chanson francophone. Au sommet de sa popularité, il abandonne pourtant les tours de chant en 1967. Avec plus de 25 millions d'albums vendus, il demeure un représentant très important de la

chanson française à l'étranger. Ses chansons sont une source d'inspiration pour bon nombre d'auteurs-interprètes anglophones comme David Bowie, Mort Shuman, Leonard Cohen. Plusieurs de ses chansons sont traduites et chantées par Ray Charles, Nina Simone, Frank Sinatra. Jacques Brel est une personnalité attachante d'une sensibilité explosive. Le petit exercice technique et méthodologique auquel nous nous livrons ici permet à peine d'effleurer les qualités de son œuvre et l'étendue de son talent.

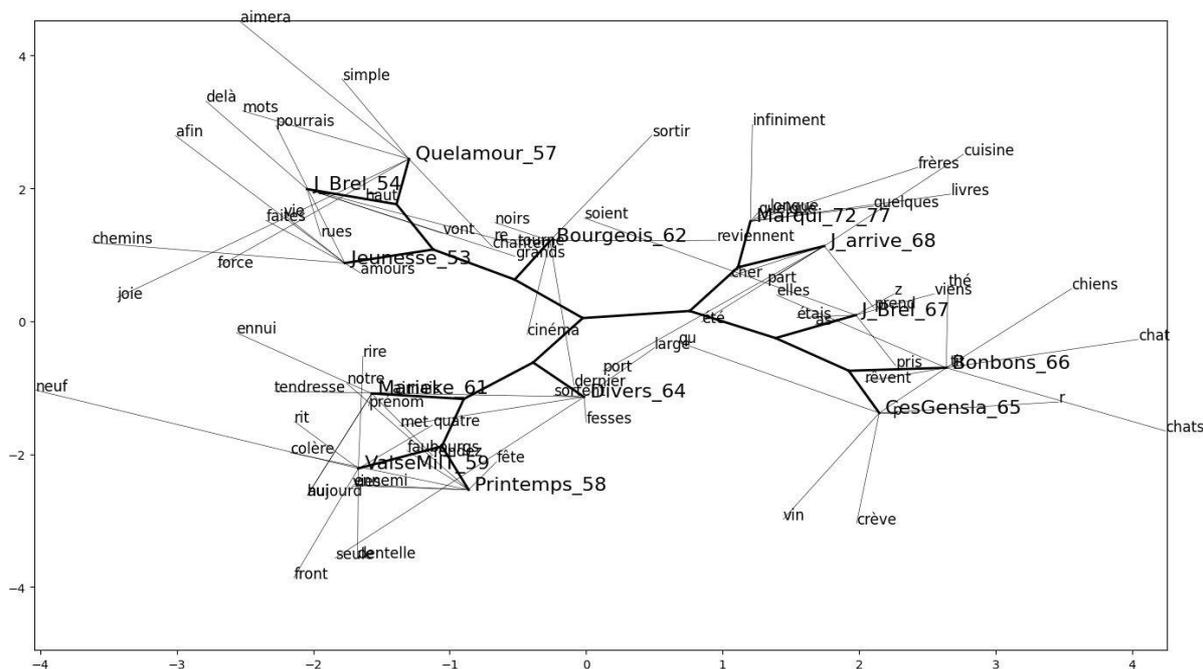


Figure 4 : Brel. Tracé de l'arbre additif avec représentation simultanée des graphies et des textes. Les textes (albums) sont repérés par un abrégé du nom de l'album (en général, titre de la chanson la plus populaire), suivi de l'année de parution, de (19)53 à (19)77 (intervalle de temps assez comparable à celui de Brassens). Comme pour les deux auteurs précédents, la simple présence ou absence d'un mot permet de déterminer des périodes temporelles relativement homogènes.

Lecture de la figure 3 : Différentes phases de la vie de Jacques Brel se retrouvent sur ce graphique, avec l'enthousiasme, la gaieté, l'émotion des premières chansons (quadrant haut gauche). Une certaine aigreur et des railleries parfois d'une grande férocité (quadrant bas droit). La position détachée à droite des formes graphiques *chat*, *chats*, *chiens*, mais aussi *thé*, *vin* caractérisent l'époque des railleries vis-à-vis de la tiédeur des modes de vie sédentaires et bourgeois, loin des *aimera*, *amours*, *simple*, *chemins* des chansons de jeunesse, à gauche.

On peut aussi positionner les titres des chansons (qui ont des profils lexicaux que l'on peut projeter en tant qu'éléments supplémentaires). En pratique, cette opération se réalise de façon interactive, car on ne peut surcharger plus des graphiques déjà à la limite de la lisibilité.

9. Conclusion

Les méthodes de visualisation de données utilisent certes des méthodes algébriques ou algorithmiques voisines de celles de l'intelligence artificielle. On a pu en effet montrer que l'AC est aussi une méthode neuronale (Lebart, 1997). Mais une visualisation n'est pas une décision à prendre, ni une tâche à exécuter. C'est presque le contraire. On ne pose pas de question pour agir, on soumet des données pour comprendre et réfléchir. L'approche est non-supervisée, une phase de travail dont le *Deep Learning* aura d'ailleurs de plus en plus besoin selon les prédictions du texte de référence de Le Cun *et al.* (2015). Comme en analyse des

correspondances avec ses représentations simultanées, le pattern observable des points-colonnes (textes, recueils) nous dit *comment* ces points s'organisent, et la présence des point-lignes (graphies) nous disent *pourquoi* ils se sont organisés de cette façon : des textes sont proches parce qu'ils utilisent souvent les mêmes mots. Et les mots habillent le squelette de l'arbre additif. Mais outre la difficulté pratique de diffusion des vraies visualisations obtenues (petits formats, absence de couleur), interviennent les difficultés inhérentes aux textes poétiques et aux chansons, qui échappent en partie, on l'a vu notamment en section 6 de cette contribution à propos de Brassens, aux exigences courantes des explorations textuelles.

Bibliographie

- Aznavour C. (2010). *Chansons : l'intégrale*. Points. Don Quichotte Editions. Paris : Seuil.
- Barthélémy J.-P. et Guénoche A. (1988). *Les arbres et les représentations de proximité*. Paris : Masson.
- Benzécri J.-P. (1973). *L'Analyse des Données*. Tome II : L'analyse des correspondances. Paris : Dunod.
- Brunet É. (2004). Statistiques Rimbaldiennes, SI@T, *Les littératures de l'Europe unie*, Cesenatico, Italie, 88-113, hal-01362731.
- Bryant D. (2005). On the uniqueness of the selection criterion in Neighbor-Joining. *Journal of Classification*, vol. (22), 1: 3-16.
- Buneman P. (1971). The recovery of trees from measurements of dissimilarity. In: Hodson F. R. D. Kendall G., and Tautu P., (Editors). *Mathematics in the archeological and historical sciences*. Edinburgh University Press, Edinburgh: 387-395.
- Di Battista, G. Eades, P., Tamassia R., et Tollis, I.G. (1999). *Graph Drawing: Algorithms for the Visualization of Graphs*, Englewood Cliffs : Prentice. Hall.
- Eades P. (1984). A heuristic for graph drawing. *Congressus Numerantium*, 42:149–160.
- Fruchterman T. et Reingold E. (1991). Graph drawing by force-directed placement. *Softw. – Pract. Exp.*, 21(11):1129–1164.
- Huson D.H. et Bryant D. (2006). Application of Phylogenetic Networks in Evolutionary Studies, *Molecular Biology and Evolution*, vol. (23), 2: 254-267.
- Kamada T. et Kawai S. (1989). An algorithm for drawing general undirected graphs. *Inform. Process. Lett.*, 31:7–15.
- Kobourov, S. G. (2013). Force-Directed Drawing Algorithms. in: *Handbook on Graph Drawing and Visualization*. Chapman and Hall/CRC.
- LeCun Y, Bengio Y, Hinton G. (2015). Deep learning. *Nature*. May 28;521(7553):436-44.
- Lebart L. (1997). Correspondence analysis, discrimination and neural networks. In: *Data Science, Classification and Related Methods*. Hayashi C., Ohsumi N., Yajima K., Tanaka Y., Bock H.- H., Baba Y. (eds). Berlin : Springer, 423-430.
- Lebart L. (2000). Contiguity Analysis and Classification. In: *Data Analysis*, (Wolfgang Gaul, Otto Opitz, Martin Schader, eds), Berlin : Springer, 233-244.
- Lebart L., Morineau A., Warwick K. (1984). *Multivariate Descriptive Statistical Analysis*. New York: Wiley and Sons.
- Lebart L., Pincemin B., Poudat C. (2019). *Analyse des Données Textuelles*. Québec : PUQ,
- Lebart L., Salem A., Berry L. (1998). *Exploring Textual Data*. Dordrecht, Boston : Kluwer Academic Publisher.
- Luong X. (1988). *Méthodes d'analyse arborée. Algorithmes, applications*. Thèse pour le doctorat ès sciences. Université Paris V.

- Mihaescu R., Levy D. et Pachter L. (2009). Why Neighbor-Joining works? *Algorithmica*, vol. (54) : 1-24.
- Reinert M. (1983). Une méthode de classification descendante hiérarchique : Application à l'analyse lexicale par contexte, *Les Cahiers de l'Analyse des Données*, 3, 187-198.
- Rochard L. (2009). *Les mots de Brassens*, Paris : Edition du Cherche Midi.
- Sattath S. et Tversky A. (1977). Additive similarity trees. *Psychometrika*, vol. (42), 3: 319-345.
- Saitou N. et Nei M. (1987). The neighbor joining method: a new method for reconstructing phylogenetic trees, *Molecular Biology and Evolution*, vol. (4), 4: 406-425.
- Tutte, W. T.(1963). How to draw a graph. *Proc. London Math. Society*, 13(52):743–768.