



CURIEUX sous les projecteurs de GEPHI

La lettre d'actualités n°7 du DES (novembre 2019) présentait, entre autre, une approche de l'espace sémantique de l'unité lexicale « curieux » avec l'outil Gargantext.

Nous reprenons ici cette même approche à l'aide d'un autre outil : Gephi.

Gephi est un logiciel libre et gratuit d'analyse et de visualisation de réseaux qui peut être téléchargé, pour windows, Mac OS X et Linux, sur Internet: <https://gephi.org/>

On trouve sur Internet une vaste documentation sur ce logiciel, les algorithmes de spatialisation et les calculs statistiques utilisés, son mode d'emploi et de nombreux exemples d'application , en particulier : <http://www.martingrandjean.ch/gephi-introduction>

Initialement développé par des étudiants de l'université de technologie de Compiègne (UTC), Gephi a été sélectionné pour le Google Summer of Code de 2009 à 2013 et a remporté en 2010 le prix Duke's Choice d'Oracle dans la catégorie Innovative Technical Data Visualization.

Gephi analyse les réseaux de données en faisant appel à la [théorie des graphes](#) : éléments reliés entre eux par des liaisons.

Nous ne présenterons pas ici les algorithmes utilisés, d'un contenu technique statistique et informatique complexe, des documentations spécialisées peuvent être trouvés sur Internet. Nous ne présenterons pas non plus le détail des manipulations du logiciel qui sont simples, « presse-bouton ». Gephi ne demande aucune technicité informatique même s'il est parfois sensible et requiert une certaine expérience !

Cet article vise à fournir une introduction attractive à Gephi à partir d'exemples élémentaires sur le réseau des synonymes du mot « curieux ».

Analyse du réseau des synonymes de « curieux »

Le DES fournit 55 synonymes directs de « curieux » puis 933 synonymes de ces synonymes.

Le réseau de ces 988 mots est constitué par toutes les relations de synonymie qui, bien que symétriques, peuvent être considérées sous forme de liens « dirigés », les liens partant d'un terme lexical vers ses synonymes.

Importons dans Gephi le fichier texte des 2150 liens du réseau de ces mots.

Ce fichier se présente sous la forme simple d'un lien par ligne avec la source et la destination (target) du lien :

```
1 source;target
2 curieux;amateur-trice
3 curieux;amusant
4 curieux;anxieux
5 curieux;attachant
6 curieux;attentif
```

...

2144 unique;un
2145 unique;uniforme
2146 unique;univoque
2147 voyeur;curieux
2148 voyeur;indiscret
2149 voyeur;mateur
2150 voyeur;satyre
2151 voyeur;spectateur-trice
2152

Nous pouvons alors analyser ce réseau avec Gephi et en produire des représentations graphiques.

Gephi permet de réaliser trois actions différentes sur les graphes :

- les **spatialisations**, techniques de représentations visuelles des graphes à l'écran (espace plan de dimension 2). Il faut noter que les positions des nœuds dans l'espace (le plan) ne sont pas des positions signifiantes en soi ni par rapport à la thématique étudiée mais seulement un positionnement esthétique qui vise à optimiser la visualisation du graphe pour en permettre l'analyse. On peut en effet même déplacer manuellement certains points si jugé nécessaire pour un meilleur aspect.
- le **calcul de statistiques**, valeur issues de la théorie des graphes :
 - o Degré : nombre de liens entrant, sortant, total, pour chaque nœud,
 - o Modularité : détection de communautés, sous-graphes de nœuds qui sont les plus liés entre eux par rapport au reste du réseau,
 - o et bien d'autres : diamètre, densité, PageRank, composantes connexes, coefficient de clustering, centralité Eigenvector, plus courts chemins ...
- l'application de **filtres**, nombreux , pouvant être calculés et enchainés, pour réduire le graphe à une de ses sous-partie selon certains paramètres.

Les spatialisations se font selon des algorithmes liés à la structure des graphes. Les statistiques, issues également de la théorie des graphes, permettent d'affecter des valeurs numériques discriminantes aux nœuds ou au liens, valeurs qui permettront de paramétrer la visualisation d'une spatialisation. Les filtres utilisent les propriétés du graphe et/ou les statistiques pour en extraire un sous-graphe.

Deux premiers exemples sur le graphe du réseau complet des 988 mots.

Un résultat est très rapidement et simplement obtenu avec le choix d'une représentation visuelle du graphe, ici la **spatialisation** « [Force Atlas 2](#) » avec quelques paramètres standard (dissuader les Hubs, empêcher le Recouvrement, Echelle).

Force Atlas 2 est basée sur la spatialisation de force. L'algorithme applique l'idée que la densité des relations fonde les communautés d'une part et que la proximité spatiale manifeste le mieux visuellement les communautés d'autre part. Les algorithmes de force construisent une distribution spatiale à 2 dimensions qui vise à organiser des forces de répulsion et d'attraction entre nœuds . Les nœuds sont comme des particules chargées qui exercent une force de répulsion et les liens, au contraire, exercent une force d'attraction entre les nœuds. Les calculs sont dynamiquement présentés à l'écran et peuvent être arrêtés et relancés volontairement ce qui permet de « jouer » avec les paramètres de l'algorithme. Pour approfondir :

https://medialab.sciencespo.fr/publications/Jacomy_Heymann_Venturini-Force_Atlas2.pdf

Ensuite, optimisation de la visualisation avec les spatialisations « Ajustement des Labels » et « Déchevauchement » et enfin, le choix de visualisation des points et des labels des mots, ici proportionnelle au nombre de liens (synonymes) partant de chaque mot.

Gephi construit dynamiquement à l'écran cette représentation

Ce graphe peut être exporté au format pdf avec le choix de paramètres d'aperçu.

Complet11.pdf

Un autre résultat permet de mettre en évidence des communautés au sein du réseau complet.

A la spatialisation « Atlas Force2 » et aux ajustements visuels précédents on ajoute le calcul (presse-bouton) de la **statistique** « **modularity** » avec un paramètre de « résolution » égal à 1, 11 communautés sont trouvées, algorithme : Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, c'est la [méthode dite de Louvain](#).

Les sous-graphes (communautés) sont calculés à partir du nombre de liens (synonymes) de chaque nœud (mots). La modularité regroupera les unités lexicales les plus liées entre elles par rapport au reste du graphe. L'algorithme va, de manière itérative, associer les nœuds en communauté de plus en plus large tout en recherchant les communautés les plus denses.

Cette statistique est ensuite utilisée pour mettre en évidence visuelle les sous-groupes. La taille et la couleur des nœuds (mots) sont fonctions de la « modularity ».

Utiliser le zoom pour lire les mots.

Complet22.pdf

Sur les visualisations du réseau complet, on constate que de nombreux synonymes de niveaux 2 ne sont reliés qu'à un seul synonyme de niveau 1 (liens en périphérie du graphe complet22). Un premier **filtre** consistera à réduire le graphe aux seuls 55 synonymes directs de « curieux » avec leurs 502 liens entre eux.

Le filtre utilisé sera « **réseau ego-centré** » sur le nœud « curieux » avec une profondeur de 1.

On se ramène ainsi à un réseau de 56 points et de 502 liens.

On y applique une spatialisation pour avoir une représentation visuelle agréable et parlante.

On retrouve alors le graphe d'adjacence de « curieux ». La taille des nœuds est proportionnelle au nombre de liens.

Adjacence2.pdf

Ensuite, nous pouvons analyser et visualiser ce seul sous-graphe.

Spatialisation avec « Atlas Force2 » et ses compléments adaptatifs.

Un résultat obtenu où la taille des nœuds est fonction du nombre de liens (le degré) :

Spatial-forceatlas2-degre.pdf

Utilisation de l'**indice de modularité**. On se rapproche des résultats obtenus avec la notion d'espace sémantique (Sabine Ploux, Bernard Victorri, 6 Mar 2006, *Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. Traitement Automatique des Langues, ATALA, 1998, pp.161-182. halshs-00009433*) bien que la méthode statistique appliquée soit très différente.

En fonction de l'objectif, on peut déterminer le nombre de communautés en jouant sur le paramètre « résolution » du calcul de la modularité.

Tableau des synonymes dans 5 communautés

amateur-trice chercheur collectionneur-euse dilettante fouilleur fouinard fouineur fureteur indiscret inquisiteur investigateur	amusant beau drolatique drôle intéressant original piquant pittoresque plaisant truculent	Bizarre bizarroïde extraordinaire extravagant insolite inusité rare singulier unique	anxieux attachant attentif avide badaud désireux dilettante flâneur fouille-au-pot intéressé intrigué juge regardant soucieux spectateur-trice voyeur	surprenant paradoxal inouï incroyable Inconcevable incompréhensible inattendu étrange étonnant déconcertant
---	--	--	--	--

Spatialisation associée : *Modularite5classes.pdf*

Tableau des synonymes dans 10 communautés

<p>anxieux attentif avide désireux intéressé intrigué soucieux</p>	<p>amateur-trice dilettante</p>	<p>badaud flâneur regardant spectateur-trice voyeur</p>	<p>chercheur collectionneur-euse fouilleur fouinard fouineur fureteur indiscret inquisiteur investigateur</p>	<p>bizarroïde extravagant insolite singulier</p>
<p>amusant bizarre drolatique drôle original plaisant truculent</p>	<p>étonnant inconcevable incroyable inouï paradoxal surprenant</p>	<p>attachant beau intéressant piquant pittoresque</p>	<p>extraordinaire inusité rare unique</p>	<p>curieux déconcertant étrange fouille-au-pot inattendu incompréhensible juge</p>

Spatialisation associée : *Modularite10class.pdf*

Gephi permet également de réaliser un **site Web dynamique**, avec une simple fonction d'« export », grâce à « [sigmaexporter](#) » l'un de ses nombreux plugin. L'export obtenu (dossiers de fichiers de taille très réduite) doit être placé sur un serveur Web.

Ce site Web reprend la dernière spatialisation obtenue pour en faire une visualisation interactive du graphe des mots. On peut, sur cette page Web dynamique et interactive, rechercher un mot ou une communauté de mots et voir (graphiquement et littéralement) les synonymes qui lui sont liés ou lui appartiennent.

Site Web de la spatialisation en 5 communautés des synonymes de « curieux »

<https://synonymes-semantic.000webhostapp.com>