

# Calcul des proximités entre les entrées du Dictionnaire de la participation

Gérald SÉDRATI-DINET

16/12/2013

## 1 Présentation et définitions

Le Groupement d'intérêts scientifiques (GIS) sur la participation du public aux processus décisionnels et à la démocratie participative a entrepris la réalisation d'un dictionnaire critique et interdisciplinaire de la participation. Une particularité de ce dictionnaire en ligne réside dans le mode de navigation entre les notices qui le composent : contrairement à un dictionnaire classique offrant une navigation suivant l'ordre alphabétique des entrées, le dictionnaire de la participation se caractérise par une navigation selon leur proximité. Le présent document détaille le calcul permettant d'établir une mesure de la proximité entre les entrées du dictionnaire.

Dans la suite de ce document, on utilisera les définitions suivantes :

- une *notice* du dictionnaire est une page du dictionnaire définissant une expression donnée par le titre de la notice à travers une courte définition et une explication détaillée dans le corps de la notice ;
- une *entrée* du dictionnaire est une expression, composée d'un ou de plusieurs mots, définie dans une ou plusieurs notices du dictionnaire ; plusieurs entrées peuvent être définies conjointement dans la même notice lorsqu'il s'agit de *termes en opposition* ;

- une *famille* est un regroupement d’entrées selon l’angle dans lequel ces termes sont considérés ; cinq familles ont ainsi été identifiées : la famille des *acteurs*, concernant les sujets appréhendés non dans le sens grammatical du terme, mais comme protagonistes de la participation ; la famille des *notions*, regroupant les termes qui font le plus débat, dont les usages sont multiples et problématiques ; la famille des *théories*, rassemblant les concepts forts (et faibles) de la participation et autour desquels de véritables théories et/ou des courants de pensée se sont formés au fil des années ; la famille des *dispositifs*, réunissant les termes par lesquels on met en pratique une démarche participative ; et la famille des *termes en opposition*, se rapportant aux termes donnant à voir les tensions dégagées entre les expressions lorsqu’elles sont utilisées en régime de participation ;
- un *axe* regroupe les entrées les plus proches de celle(s) de la notice consultée qui sont de la même famille, permettant ainsi un parcours du dictionnaire selon un angle de vue particulier ;
- une *orbite* regroupe les entrées les plus proches de celle(s) de la notice consultée qui appartiennent à une autre famille que celle(s) de la notice consultée, permettant ainsi un parcours transversal du dictionnaire ;
- un *mot étranger* est un mot proche d’une entrée du dictionnaire sans lui-même en être une.

## 2 Critères d’évaluation de la proximité entre deux entrées du dictionnaire

La proximité entre deux entrées du dictionnaire est évaluée selon la règle canonique suivante : *deux entrées sont d’autant plus proches qu’elles sont souvent employées ensemble dans les notices du dictionnaire.*

Afin de calculer la proximité entre deux entrées du dictionnaire, il est donc nécessaire d’évaluer, dans le corpus constitué par l’ensemble des notices du dic-

tionnaire, les critères permettant d'établir à quel degré ces deux entrées apparaissent ensemble dans le corpus étudié. Pour satisfaire de tels critères, il conviendrait ainsi d'évaluer :

1. si ces deux entrées sont ou non présentes dans les mêmes notices ;
2. le nombre de notices dans lesquelles elles sont toutes deux présentes ;
3. le nombre d'occurrences de chacune de ces deux entrées dans chaque notice où elles sont toutes deux présentes ;
4. leur degré de présence simultanée au sein d'une notice, c'est-à-dire si les occurrences des deux entrées apparaissent simultanément :
  - (a) dans la courte définition et/ou dans le corps de la notice ;
  - (b) en n'étant séparées que par un nombre minimal d'autres mots (voire de manière adjacente si les occurrences des deux entrées ne sont séparées par aucun autre mot) ;
  - (c) dans le même paragraphe ;
  - (d) dans des paragraphes adjacents ;
  - (e) dans la même phrase ;
  - (f) dans des phrases adjacentes ;
  - (g) dans le même groupe syntaxique ;
  - (h) dans des groupes syntaxiques adjacents, etc.

L'évaluation de ces critères, dans l'ordre indiqué ci-dessus, permettrait ainsi d'obtenir une mesure de plus en plus précise de la proximité entre deux entrées. Toutefois, le problème posé par le dictionnaire de la participation n'est pas tant de calculer la mesure exacte de la proximité entre deux entrées, que celui d'obtenir les entrées les plus proches de celle(s) de chaque notice consultée. Il n'est donc pas nécessaire d'évaluer tous les critères précédemment listés pour chaque paire d'entrées. Une évaluation des premiers d'entre eux peut s'avérer suffisante pour obtenir un classement par ordre de proximité décroissant, qui soit le plus exact possible, des entrées proches de celle(s) de chaque notice consultée.

Il a été ainsi estimé qu'il était acceptable pour chaque notice du dictionnaire d'évaluer le nombre de mots séparant deux entrées (critère 4b) afin de calculer une proximité suffisamment précise pour en déduire avec assez de justesse les entrées les plus proches de celle(s) de chaque notice consultée. En effet, en appliquant ce critère, les trois premiers listés ci-dessus sont également implicitement pris en compte, étant donné que l'on recherche dans chaque notice où les deux entrées sont présentes le nombre de mots séparant chaque occurrence de chacune des entrées entre lesquelles on calcule la proximité. Si les résultats obtenus n'étaient toutefois pas satisfaisant en s'arrêtant à ce critère, le calcul pourrait être itérativement affiné en appliquant successivement les suivants.

### 3 Identification des entrées du dictionnaire

Une entrée du dictionnaire est déterminée par le titre de la ou des notices dans laquelle elle est définie. Cependant, certaines notices peuvent définir simultanément plusieurs entrées, lorsqu'il s'agit de *termes en opposition*. Dans ce cas les différentes entrées sont séparées par une barre oblique dans le titre de la notice, qui doit par conséquent être décomposé.

De plus, le titre de certaines notices précisent le contexte dans lequel une entrée est définie en accolant à cette entrée une expression entre parenthèses. Il convient de ne pas tenir compte de ces précisions entre parenthèses lors de l'identification des entrées du dictionnaire.

Les entrées du dictionnaire peuvent apparaître dans les notices en tant que telles c'est-à-dire avec l'orthographe exacte employée pour définir les entrées, mais elles peuvent aussi potentiellement être présentes sous une forme déclinée par exemple, au pluriel. Le calcul de proximité doit ainsi prendre en compte pour chaque entrée les occurrences de toutes ses déclinaisons.

Par ailleurs, la présence d'une entrée, ou de l'une de ses déclinaisons, dans les notices du dictionnaire doit bien entendu être indépendante de la casse avec laquelle elle est typographiée.

La proximité entre les entrées du dictionnaire devant permettre une navigation selon l'axe ou les orbites de chacune d'entre elles, il est en outre nécessaire de considérer sa ou ses familles d'appartenance lors de ce calcul.

Enfin, comme nous le verrons dans la section suivante *Indexation des notices du dictionnaire*, il s'agit non seulement de calculer la proximité entre les entrées du dictionnaire mais également entre chacune d'entre elles et les mots étrangers présents dans tout le corpus des notices. Pour ces derniers, afin de ne conserver que la proximité entre les entrées du dictionnaire et les mots étrangers significatifs, il est en outre nécessaire de définir de manière configurable une liste de mots à ignorer : conjonctions, articles, auxiliaires, etc.

Pour tenir compte techniquement de ces contraintes, les entrées du dictionnaire et leurs déclinaisons sont représentées dans une table `ppddp_dico_entries` en base de données ayant la structure suivante :

<b>colonne</b>	<b>description</b>	<b>format</b>
<i>id</i>	identifiant unique de l'entrée ou du mot étranger	entier positif non nul auto-incrémenté
<i>eid</i>	identifiant unique de l'entrée ou du mot étranger	entier non nul, positif égal à l' <i>id</i> pour une entrée, ou négatif opposé à l' <i>id</i> pour un mot étranger
<i>name</i>	nom canonique de l'entrée ou du mot étranger	chaîne de caractères en minuscules, sans précision entre parenthèses
<i>inflex</i>	identifiant de l'entrée dont celle-ci est une déclinaison.	entier positif égal à l' <i>eid</i> pour une déclinaison, zéro pour une entrée canonique et nul pour un mot étranger ignoré

<b>colonne</b>	<b>description</b>	<b>format</b>
<i>family_1</i>	appartenance à la première famille ( <i>notions</i> )	booléen à vrai si l'entrée appartient à la première famille
<i>family_2</i>	appartenance à la deuxième famille ( <i>acteurs</i> )	booléen à vrai si l'entrée appartient à la deuxième famille
<i>family_3</i>	appartenance à la troisième famille ( <i>théories</i> )	booléen à vrai si l'entrée appartient à la troisième famille
<i>family_4</i>	appartenance à la quatrième famille ( <i>dispositif</i> )	booléen à vrai si l'entrée appartient à la quatrième famille

## 4 Indexation des notices du dictionnaire

Puisque le calcul de la proximité entre deux entrées revient à déterminer dans chaque notice le nombre de mots séparant chaque occurrence de chacune des entrées, ce calcul porte sur une projection unidimensionnelle des notices, chaque notice étant ainsi représentée sous forme de la liste ordonnées des mots qu'elle contient. Une telle projection est une opération courante dans le domaine des sciences de l'information désignée par le terme *indexation*.

## Extrait de la notice Wikipédia sur l'indexation automatique de documents

Pour un texte, un index très simple à établir automatiquement est la liste ordonnée de tous les mots apparaissant dans les documents avec la localisation exacte de chacune de leurs occurrences ; mais un tel index est volumineux et surtout peu exploitable.

L'indexation automatique tend donc plutôt à rechercher les mots qui correspondent au mieux au contenu informationnel d'un document. On admet généralement qu'un mot qui apparaît souvent dans un texte représente un concept important. Ainsi, la première approche consiste à déterminer les mots représentatifs par leur fréquence. Cependant, on s'aperçoit que les mots les plus fréquents sont des mots fonctionnels (ou mots outils, mots vides). En français, les mots *à*, *de*, *le*, *un*, *et*, *les*, etc. sont les plus fréquents. En anglais, ce sont *of*, *the*, etc.

Il est évident que l'on ne peut pas garder ces mots à haute fréquence mais peu porteur de sens en terme. C'est pourquoi on introduit dans les moteurs de recherche des opérations de filtrage de ces mots. Ces listes de mots sont appelées anti-lexiques ou plus fréquemment *stoplist*.

Une autre opération est ensuite couramment appliquée lors de l'indexation. Elle consiste à effacer les terminaisons (flexions de nombre, genre, conjugaison, déclinaison) afin de retrouver les racines des mots. Cette opération est appelée *stemming* (une autre solution voisine appelée lemmatisation conduit globalement au même résultat). Ce procédé permet de relever les fréquences en cumulant les nombres d'occurrence des variations des mêmes mots.

Chaque unité documentaire (chaque document ou chaque passage de document) peut alors faire l'objet d'une représentation vectorielle : les coordonnées représentent les fréquences des mots non vides. Lorsque l'on effectue cette opération pour un corpus de documents ou de pages web on obtient une matrice dont les colonnes représentent un document et les coordonnées la fréquence des termes.

Les moteurs de recherche de première génération s'appuient sur des formules de pondération, généralement pour affecter un poids élevé aux termes non-distribués uniformément au sein du corpus. Il existe un grand nombre de formules de pondération dont le but est de distribuer le poids pour contribuer à la différenciation informationnelle des documents. Certaines formules de pondération harmonisent les poids en fonction de la longueur des documents où la fréquence des termes est globalement plus élevée, d'autres formules s'appuient sur la fréquence maximale des termes afin de concilier l'aspect multi-thématique d'un document avec des documents mono thématiques. Les formules de pondération les plus connues sont TF-IDF (term frequency . inverse document frequency).

Les moteurs de seconde génération s'appuient non seulement sur la fréquence des termes pour indexer les pages web mais aussi sur la popularité des sources. En naviguant de lien en lien, les robots indexent les termes utilisés par une page web pour décrire une autre page web. À chaque fois qu'un utilisateur suit ce lien, il lui vote la pertinence des descripteurs utilisés. Le *page-ranking* est ensuite calculé selon la popularité des descripteurs et un coefficient de similarité issu de la modélisation vectorielle.

L'indexation des notices du dictionnaire peut ainsi être rapprochée de celle opérée par un moteur de recherche, à la différence près qu'il s'agit non seulement, pour chaque entrée du dictionnaire, de consigner les notices dans lesquelles cette entrée est présente, mais également, pour chacune de ces notices, le rang de chaque occurrence de l'entrée dans l'ordre des mots de la notice considérée. En outre, afin de calculer la proximité entre les entrées du dictionnaire et les *mots étrangers*, ces derniers doivent aussi être indexés. Si l'on ne calculera pas cf. la section suivante *Calcul de la proximité entre deux entrées du dictionnaire* de proximité avec les mots étrangers non significatifs définis dans une liste configurable, cf. la section précédente *Identification des entrées du dictionnaire* ceux-ci doivent néanmoins être indexés afin de conserver le rang exact des entrées et mots étrangers significatifs permettant de calculer avec justesse la distance les séparant.

Ainsi, l'indexation d'une notice du dictionnaire consiste à convertir celle-ci en une suite ordonnée des entrées et mots étrangers qui la composent.

Les notices étant enregistrées au format *HTML*, il est nécessaire en première instance d'en expurger les balises, afin de ne conserver que le texte de chaque notice, constitué par la concaténation de son titre dans le cas où celui-ci comporte une précision contextuelle entre parenthèses, cette dernière expose en effet des mots proches de l'entrée considérée de sa courte définition et de son explication détaillée. En second lieu, chaque déclinaison connue d'une entrée doit être remplacée par celle-ci dans le texte à indexer. Les entrées, puisqu'étant éventuellement composées de plusieurs mots, sont ensuite délimitées en étant encadrées par des guillemets.

À ce stade, la décomposition d'une notice en une suite ordonnée d'entrées et de mots étrangers peut se baser sur l'expression régulière suivante :



### Expression régulière inspirée du module Perl Text : :ParseWords

```
$regexp = '/^
(?:
  # entrée encadrée par des guillemets
  ("                # $quote
  ((?>[^\\""]*(?:\\\\". [^\\""]*)*))" # $quoted
  | # --OR--
  # mot non encadré de guillemets
  (                # $unquoted
  (?:\\\. | [^\\""])*?
  )
  # suivi par
  (                # $delim
  \Z(?:\n)        # fin de ligne
  | # --OR--
  (?:[\'. , ; : ? ! \~ \^ \_ \{ \} \[ \] \\\ \d \\ [ \\\ \\/ ] +) # délimiteur de mots
  | # --OR--
  (?!\^)(?=["])  # guillemet
  )
)/sux';
```

Cette dernière permet de capturer successivement les entrées du dictionnaire définies comme des chaînes de caractères encadrées par des guillemets capturées dans la variable `$quoted` ainsi que les mots étrangers dans la variable `$unquoted` si la variable `$quote` n'a rien capturé selon la manière dont ceux-ci sont séparés fin de ligne, délimiteurs de mots ou guillemet, capturé dans la variable `$delim`.

Enfin, dans le but de limiter le nombre de mots étrangers indexés, ceux de trois lettres ou moins ne sont pas pris en compte.

Cette décomposition permet ainsi de construire un index que l'on peut sto-

cker dans une table `ppddp_dico_index` en base de données ayant la structure suivante :

colonne	description	format
<i>eid</i>	identifiant unique d'une entrée ou d'un mot étranger pointant sur une ligne de la table <code>ppddp_dico_entries</code>	entier non nul, positif pour une entrée, ou négatif pour un mot étranger
<i>nid</i>	identifiant unique d'une notice	entier positif non nul, défini par le système de gestion de contenu ( <i>Drupal</i> )
<i>offset</i>	rang dans la suite ordonnée des entrées et mots étrangers composant la notice	entier positif non nul

## 5 Calcul de la proximité entre deux entrées du dictionnaire

À partir de l'index construit comme il a été décrit dans la section précédente *Indexation des notices du dictionnaire* et des critères d'évaluation de la proximité retenus dans la section *Critères d'évaluation de la proximité entre deux entrées du dictionnaire*, il est assez immédiat de calculer les proximités entre les entrées du dictionnaires et avec les mots étrangers.

Si l'on considère les variables suivantes :

$N$	= nombre total de notices dans le dictionnaire
$i$	= identifiant d'une notice
$n_i$	= nombre de mots indexés de la notice $i$
$m_i$	= nombre d'occurrences de l'origine dans la notice $i$
$j$	= identifiant d'une occurrence de l'origine dans la notice $i$
$l_i$	= nombre d'occurrences de la destination dans la notice $i$
$k$	= identifiant d'une occurrence de la destination dans la notice $i$
$O_{ij}$	= rang de l'occurrence $j$ de l'origine dans la notice $i$
$T_{ik}$	= rang de l'occurrence $k$ de la destination dans la notice $i$
$P_{OT}$	= proximité entre l'origine $O$ et la destination $T$

La proximité  $P_{OT}$  entre une origine  $O$  et une destination  $T$  peut être évaluée en additionnant, pour toutes les notices  $i$  et pour toutes les occurrences  $j$  de  $O$  ainsi que pour toutes les occurrences  $k$  de  $T$ , la distance entre  $O_{ij}$  et  $T_{ik}$ . Cette dernière, devant être minimale entre les premier et dernier mots indexés de la notice  $i$ , peut être obtenue en soustrayant de 1 la valeur absolue de la différence entre  $T_{ik}$  et  $O_{ij}$  rapportée au nombre total de mots indexés  $n_i$  dans la notice  $i$ .

La formule de calcul de la proximité  $P_{OT}$  est donc la suivante :

$$P_{OT} = \sum_{i=1}^N \sum_{j=1}^{m_i} \sum_{k=1}^{l_i} 1 - \frac{|T_{ik} - O_{ij}|}{n_i}$$

La proximité étant une relation symétrique  $P_{OT} = P_{TO}$  le calcul des proximités peut être optimisé en choisissant toujours une destination dont l'identifiant dans la table `ppddp_dico_entries` est supérieur à celui de l'origine considérée.

De plus, la proximité entre deux mots étrangers ne présentant pas d'intérêt,  $O$  peut être toujours choisi parmi les seules entrées du dictionnaire alors que  $T$  peut l'être parmi les entrées et les mots étrangers significatifs.

Enfin, dans le but de limiter les calculs de proximité avec les mots étrangers, potentiellement très nombreux, il est possible d'obtenir une approximation satis-

faisante en ne choisissant que les notices  $i$  pour lesquelles le nombre d'occurrences  $l_i$  du mot étranger  $T_i$  est supérieur à un seuil configurable  $Th_{foreign}$ .

De même, il est possible de ne considérer les notices d'une part dont le nombre d'occurrences  $m_i$  de l'entrée d'origine  $O_i$  est supérieur à un certain seuil configurable  $Th_{entry}$  et, d'autre part celles dont le nombre d'occurrences  $l_i$  de l'entrée ou du mot étranger de destination  $T_i$  est supérieur à un autre seuil configurable  $Th_{other\_entry}$ .

Les expérimentations réalisées cf. la section Résultats expérimentaux permettent d'obtenir une vitesse de calcul satisfaisante pour  $Th_{entry} = Th_{other\_entry} = 0$  c'est-à-dire que toutes les notices sont prises en compte pour le calcul des proximités entre les seules entrées du dictionnaire et  $Th_{foreign} = 10$  c'est-à-dire que seules sont prises en compte les notices dans lesquelles le mot étranger considéré apparaît plus de dix fois.

Les proximités ainsi calculées peuvent être stockées dans une table `ppddp_dico_proximities` en base de données ayant la structure suivante :

<b>colonne</b>	<b>description</b>	<b>format</b>
<i>eid_src</i>	identifiant unique d'une entrée d'origine pointant sur une ligne de la table <code>ppddp_dico_entries</code>	entier positif non nul
<i>eid_dst</i>	identifiant unique d'une entrée ou d'un mot étranger de destination pointant sur une ligne de la table <code>ppddp_dico_entries</code>	entier non nul, positif pour une entrée, ou négatif pour un mot étranger
<i>proximity</i>	valeur de la proximité entre l'origine et la destination	nombre décimal positif

## 6 Recherche des mots proches pour une notice du dictionnaire

L'objectif du calcul de proximités entre les entrées du dictionnaire ainsi qu'avec les mots étrangers est d'offrir à partir de chaque notice une navigation selon l'axe et les orbites de l'entrée ou des entrées lorsqu'il s'agit de *termes en opposition* correspondante(s), permettant donc, *pour chaque famille*, de consulter les notices associées aux entrées les plus proches de celle(s) considérée(s), ainsi que la visualisation des mots étrangers significatifs les plus proches.

Il s'en suit que pour la ou les entrées considérées, il est nécessaire de rechercher dans la table `ppddp_dico_proximities`, décrite dans la section précédente *Calcul de la proximité entre deux entrées du dictionnaire*, les entrées ayant les plus grandes valeurs de proximité avec celle(s) considérée(s). Cette recherche doit en outre ne pas être effectuée globalement c'est-à-dire sur l'ensemble de la table mais être répétée *pour chaque famille*. En effet, il s'agit de rechercher *pour chaque famille*, selon que celle-ci correspond à un axe ou une orbite, ainsi que pour les mots étrangers, un certain nombre d'entrées les plus proches de celle(s) considérée(s). Si l'on veut retourner le nombre désiré de mots proches *pour chaque famille*, ainsi que pour les mots étrangers, une recherche indépendante de la famille des mots proches devrait porter sur l'ensemble de la table `ppddp_dico_proximities` et nécessiterait une ventilation *a posteriori* par famille. Il est beaucoup plus pertinent et performant de s'appuyer sur les optimisations du moteur de base de données en effectuant *a priori* une décomposition des recherches *par famille*, par le biais d'une jointure avec la table `ppddp_dico_entries`.

La recherche de mots proches pour une notice se rapportant à plusieurs entrées c'est-à-dire pour la famille des *termes en opposition* pose un problème supplémentaire. En effet, il faut dans ce cas rechercher les mots proches *de chacune* des entrées définies dans la notice, en excluant pour chacune de ces recherches les autres termes définis dans la même notice, puis fusionner les

résultats en préservant un ordre décroissant des valeurs de proximité. Cette fusion consiste à ne conserver, pour une même entrée ou un même mot étranger proche de plusieurs des entrées définies par la notice considérée, que la valeur de proximité la plus élevée. Ces recherches doivent donc chacune renvoyer un nombre de résultats suffisant pour que leur fusion respecte un ordre décroissant de proximité.

De même, puisque la navigation doit permettre de pointer vers des notices et non vers des entrées, les résultats de la recherche des entrées de la famille des *termes en opposition* proches de l'entrée ou des entrées considérées doivent être fusionnés pour ne conserver que la valeur de proximité la plus élevée pour les termes définis dans la même notice. Il faut donc là aussi que la recherche renvoie un nombre de résultats suffisant pour conserver le nombre de résultats désiré.

Les notices définissant au maximum trois entrées de la famille des *termes en opposition*, il est nécessaire de tripler le nombre de résultats renvoyés par chaque recherche, pour au final obtenir le nombre désiré de mots proches. Celui-ci a été défini à 14 pour les entrées sur l'axe de celle(s) considérée(s), à 11 pour celles des orbites et à 4 pour les mots étrangers. Ces chiffres permettent une vue détaillée des mots proches pour chaque notice. Il est néanmoins possible d'en offrir également une vue rapide en ne sélectionnant dans un premier temps qu'un certain nombre d'entre eux.

## 7 Résultats expérimentaux

Une indexation et un calcul de proximité, dont les statistiques sont données dans la figure suivante, ont été réalisés sur un corpus de 253 notices :

## DICTIONNAIRE

Administrer · Configuration du site

[Statistiques](#) [Configuration](#) [Calcul des proximités](#)

Nombre de notices	253
Nombre d'entrées	211
Nombre d'entrées de la famille <i>notions</i>	56
Nombre d'entrées de la famille <i>acteurs</i>	9
Nombre d'entrées de la famille <i>théories</i>	4
Nombre d'entrées de la famille <i>dispositifs</i>	23
Nombre d'entrées de la famille <i>termes en opposition</i>	33
Nombre de mots étrangers	16709
Mots étrangers les plus courants	participation (156), sont (138), comme (131), entre (121), cette (117), elle (109), leur (104), politique (102), avec (101), mais (96), processus (91), action (87), acteurs (83), aussi (76), peut (73), urbaine (72), années (72), elles (70), être (69), droit (68)
Nombre d'entrées indexées	210
Entrées non indexées	sig

[↑ HAUT DE PAGE](#)

La liste de mots étrangers non significatifs (cf. la section *Identification des entrées du dictionnaire*) et les divers seuils (cf. la section *Calcul de la proximité entre deux entrées du dictionnaire*) ont été configurés comme indiqué par la figure suivante :

**DICTIONNAIRE**

Administrer · Configuration du site · Dictionnaire

Statistiques Configuration **Calcul des proximités**

**Mots étrangers ignorés :**

pour  
plus  
dans  
sont

Entrez ligne par ligne les mots étrangers à ignorer. Ceux-ci continueront à être indexés afin de préserver la justesse du calcul des proximités, mais ils n'apparaîtront plus en mots étrangers proches d'une entrée du dictionnaire.

CKEditor : l'ID pour exclure ou inclure cet élément est `ppddp:admin/settings/dico/config.edit-ppddp-dico-foreigns-skip`.

**Seuil pour un mot étranger :**

10

Si un mot étranger apparaît dans une notice un nombre de fois inférieur à ce seuil, il ne sera fait aucun calcul de proximité entre les entrées du dictionnaire et ce mot étranger pour la notice en question. Si ce seuil vaut 0, on calculera la proximité entre chaque entrée du dictionnaire et tous les mots étrangers pour toutes les notices, sinon, uniquement pour les notices où ce seuil est dépassé.

**Seuil pour une entrée dans une notice :**

0

Si une entrée du dictionnaire apparaît dans une notice un nombre de fois inférieur à ce seuil, il ne sera fait aucun calcul de proximité entre cette entrée du dictionnaire et les autres entrées ou les mots étrangers pour la notice en question. Si ce seuil vaut 0, on calculera la proximité de toutes les entrées du dictionnaire avec toutes les autres et avec tous les mots étrangers dans toutes les notices, sinon, uniquement pour les notices où ce seuil est dépassé.

**Seuil pour une autre entrée dans une notice :**

0

Si une entrée du dictionnaire apparaît dans une notice un nombre de fois inférieur à ce seuil, il ne sera fait aucun calcul de proximité entre les autres entrées du dictionnaire et celle-ci pour la notice en question. Si ce seuil vaut 0, on calculera la proximité entre chaque entrée du dictionnaire et toutes les autres pour toutes les notices, sinon, uniquement pour les notices où ce seuil est dépassé.

Enregistrer

↑ HAUT DE PAGE

Dans ce contexte avec des seuils pour les entrées à 0 et un seuil pour les mots étrangers à 10, la phase d'indexation a pris 30 minutes et la phase de calcul des proximités 33 minutes. Avec tous les seuils fixés à 0 c'est-à-dire en prenant en compte toutes les notices que ce soit pour les entrées du dictionnaire ou pour les mots étrangers, le calcul de proximités prend un peu moins de 18 heures et 38 minutes. Avec des seuils pour les entrées à 3 c'est-à-dire en ne prenant que les notices où l'une et l'autre des entrées apparaît plus de 3 fois et un seuil pour les mots étrangers à 10, le calcul de proximités prend un peu moins de 12 minutes. Avec un seuil pour l'entrée d'origine à 0, un seuil pour l'entrée de destination à 3 et un seuil pour les mots étrangers à 10, le calcul de proximités prend un peu moins de 17 minutes.



**DICTIONNAIRE**

Administrer · Configuration du site · Dictionnaire

Statistiques Configuration Calcul des proximités

Pour chaque notice du dictionnaire, calculer la position de chaque entrée ainsi que des mots étrangers. Cette opération prend environ 30 minutes.

Réindexer les notices du dictionnaire

Calculer la proximité entre les entrées du dictionnaire ainsi qu'avec les mots étrangers. Cette opération dépend des seuils fixés dans les paramètres et peut prendre entre quelques dizaines de minutes et plusieurs heures (si aucun seuil n'est fixé pour les mots étrangers).

Recalculer les proximités entre les entrées du dictionnaire

↑ HAUT DE PAGE

La figure suivante montre les résultats obtenus pour une notice définissant une seule entrée :

**NAVIGATION DANS LE DICTIONNAIRE**

Axe dispositifs  
débat public, tirage au sort, conseil de développement, conseil de quartier, évaluation participative des technologies, concertation, jury citoyen, évaluation participative des politiques publiques

Orbite notions  
conflit, engagement, démocratie, dispositif, délibération, risque, développement durable, débat, intérêt général, controverse, légitimité

Orbite acteurs  
professionnel, expert, peuple, riverain, usager

Orbite théories  
engagement, démocratie directe, démocratie délibérative

Orbite termes en opposition  
citoyen/élu/technicien, engagé/concerné, dispositif/milieu, intérêt général/intérêt particulier, usager/client, local/global, agrégation/délibération, démocratie représentative/démocratie participative, inclusion/exclusion, proximité/distanc e, empowerment/capacitation

Mots étrangers  
mais, question, participation, sont

---

**NOTICE DE DICTIONNAIRE**

## Conférence de citoyens

Voir Modifier Devel

28 juin 2013 | Comité éditorial | Notice de Dictionnaire

**Famille**  
Dispositifs

**Définition de l'entrée**

Une « conférence de citoyens » est une procédure qui met en scène la réunion studieuse, la délibération et la recommandation de citoyens, à propos d'une décision à prendre par une entité, publique ou privée, appelée commanditaire. L'objectif annoncé est donc bien une « participation » à la décision du commanditaire. Entre la question que celui-ci formule et la recommandation que doit émettre le groupe de citoyens, s'interpose une procédure assez formalisée impliquant des étapes précises : énoncé de la question ; choix d'un responsable de l'organisation de la conférence, désignation d'un comité de pilotage, sélection d'un groupe de citoyens, choix d'un animateur, formation sur les sujets dont il s'agit de débattre, débat et recommandation. Mais chacune de ces étapes peut être différemment conçue. Or on ne dispose guère d'analyse comparative qui permettrait de choisir de manière rationnelle un dispositif plutôt qu'un autre.

> **Le commanditaire et la question posée**

Ce sont des autorités publiques qui sont les commanditaires les plus fréquents. Mais le terme recouvre des réalités très diverses. En France, citons l'Office parlementaire des choix scientifiques et techniques, la commission française du [développement durable](#) (CFDD) et de multiples autorités régionales. Mais cela peut être un institut de recherche pour réfléchir à sa politique scientifique. L'impact de la recommandation dépend en partie de la [compétence](#) du commanditaire pour prendre la [décision](#) : ainsi, la CFDD n'ayant aucun [pouvoir](#) de [décision](#) propre, la recommandation n'a eu aucun impact car elle n'a pas intéressé le ministère de tutelle.

Des entreprises privées se sont déjà intéressées à cette procédure pour anticiper les réactions de la clientèle ou du [public](#) : les assureurs anglais ont ainsi testé la possibilité d'effectuer des tests génétiques pour déterminer si la personne est assurable et à quel prix.

Si les [conférences de citoyens](#) ont été mobilisées essentiellement sur des questions scientifiques et techniques d'[intérêt général](#), rien n'empêche de poser une question plus étroite, comme l'organisation du tri sélectif dans une ville. La question peut avoir une portée générale, comme les ondes émises par les antennes relais, mais être posée dans un cadre [local](#) comme ce fut le cas de la conférence organisée sur ce sujet par la Ville de Paris.

La figure suivante montre les résultats obtenus pour une notice définissant plusieurs entrées de la famille des *termes en opposition* :

r

## NAVIGATION DANS LE DICTIONNAIRE

Axe termes en opposition  
dispositif/milieu,  
démocratie  
représentative/démocratie  
participative,  
agrégation/délibération,  
engagé/concerné,  
local/global, intérêt  
général/intérêt  
particulier,  
consensus/dissensus,  
usager/client,  
proximité/distance,  
inclusion/exclusion,  
virtuel/réel,  
empowerment/capacitati  
on, participation  
politique/participation  
civique, blanc/noir

Orbite notions  
démocratie, débat,  
dispositif, information,  
engagement, démocratie  
participative,  
délibération, risque,  
légitimité, conflit,  
intérêt général

Orbite acteurs  
expert, professionnel,  
usager, peuple, riverain,  
stakeholder, indigné, sans  
part, lanceur d'alerte

Orbite théories  
engagement, démocratie  
directe, démocratie  
délibérative, droit de  
l'environnement et  
participation

Orbite dispositifs  
débat public,  
concertation, conférence  
de citoyens, tirage au  
sort, référendum, conseil  
de quartier, community  
organizing, jury citoyen,  
assemblée générale,  
conseil de  
développement, atelier  
du futur

Mots étrangers  
participation, comme,  
sont, science

## NOTICE DE DICTIONNAIRE

### Citoyen/Élu/Technicien

Voir Modifier Révisions Devel

28 juin 2013 Comité éditorial Notice de Dictionnaire

#### Famille

Termes en opposition

#### Auteur(s)

Alice MAZEAUD

#### Définition de l'entrée

Ce tryptique organise la répartition des rôles entre les acteurs des processus participatif.

#### Mots Proches

Savoirs / Décision. Professionnels/citoyens

Le tryptique [élus/citoyens/techniciens](#) constitue l'un des schèmes d'analyse les plus partagés par ceux qui, acteurs ou chercheurs, travaillent sur la participation à l'échelle locale. Apparu dans la littérature en tant que système d'acteurs en tension, il s'est ensuite imposé comme outil descriptif et normatif dans les guides de bonnes pratiques, et autres chartes de la participation, développés par les professionnels de la démocratie locale. Aujourd'hui, alors que du côté des praticiens ce tryptique organise la répartition des rôles entre les acteurs des processus participatifs, du côté des chercheurs, il tend à être dépassé par une sociologie des acteurs du fait participatif.

#### > De la découverte d'un système en tension à la fixation des rôles

La triade [élus/citoyens/techniciens](#) est évoquée pour la première fois par Maurice Blanc (1999) dans ce qui fut le premier ouvrage de synthèse sur la démocratie locale. L'enjeu porte alors essentiellement sur les modalités d'organisation de la complémentarité entre [démocratie représentative](#) et participative. Dressant un bilan de plusieurs années de recherche sur la participation des habitants dans la politique de la ville, il mobilise le concept de « transaction sociale » (Rémy J., et al. 1991) qui constitue selon lui « un outil analytique fécond pour analyser le processus d'élaboration des compromis pratiques dans les situations de coopération conflictuelle, [et donc] très éclairant pour analyser la [démocratie participative](#) locale », et ses limites.

Dans la lignée de la grille d'Arnstein et selon le schème d'analyse alors dominant, la qualité des procédures participatives est appréciée au regard du [pouvoir](#), octroyé ou conquis, des [citoyens](#). Ainsi, dans les premiers travaux, la participation des [habitants](#) (ou [citoyens](#)) est considérée comme un trouble dans la distribution ordinaire du [pouvoir local](#) entre [élus](#) et [techniciens](#). La participation des [citoyens](#) mettrait en tension la [légitimité](#) électorale des [élus](#) et l'[expertise](#) technique des [techniciens](#). Cette irruption des [profanes](#) dans la production des politiques motiverait des stratégies du couple [élu/technicien](#), [engagé](#) dans un rapport de concurrence-collusion pour le [pouvoir](#), visant à bloquer la [parole](#) des [citoyens](#). Dans cette perspective, le [projet](#) de démocratisation du [pouvoir local](#) incarné par une participation accrue des [citoyens](#) est conditionné au déploiement de processus d'[apprentissage](#) réciproques permettant à chacun de trouver sa place dans une dynamique de construction collective de l'[intérêt général](#). Au cours du processus, les [citoyens](#) feraient l'[apprentissage](#) d'une citoyenneté active (s'exprimer en [public](#) et orienter son discours vers la construction de l'[intérêt général](#)) leur permettant d'affirmer leur [légitimité](#) face aux [légitimités](#) électives et expertes des [élus](#) et des [techniciens](#), les [techniciens](#) apprendraient à prendre en compte les savoirs d'usage en complément de leur [expertise](#) technique et enfin, les [élus](#) feraient l'expérience que la participation des [citoyens](#) ne les déposséderait pas de leur [pouvoir](#), mais au contraire contribuerait à les relégitimer.