

## Contenu des dictionnaires Lingway



# Table des matières

<b>TABLE DES MATIÈRES .....</b>	<b>2</b>
<b>• MODE DE CONSTITUTION DES DICTIONNAIRES LINGWAY .....</b>	<b>3</b>
1 ORGANISATION DES DONNÉES.....	3
1.1. <i>Modèle</i> .....	3
1.2. <i>Compilation d'un dictionnaire</i> .....	4
2 CONSTITUTION DE BRIQUES .....	4
2.1. <i>Données externes au dictionnaire</i> .....	4
2.2. <i>Données internes au dictionnaire</i> .....	5
3 RESSOURCES DISPONIBLES .....	5
3.1. <i>Dictionnaire de la langue générale</i> .....	5
3.2. <i>Briques de spécialité</i> .....	6
<b>• DICTIONNAIRES DISPONIBLES.....</b>	<b>10</b>
4 DICTIONNAIRE « PRESSE GÉNÉRALISTE ».....	10
4.1. <i>Description</i> .....	10
4.2. <i>Contenu</i> .....	10
5 DICTIONNAIRE « DOCUMENTATION TECHNIQUE » .....	11
5.1. <i>Description</i> .....	11
5.2. <i>Contenu</i> .....	11
6 DICTIONNAIRE « ERECRUTEMENT » .....	12
6.1. <i>Description</i> .....	12
6.2. <i>Contenu</i> .....	12
7 DICTIONNAIRE « JURIDIQUE ET FISCAL » .....	12
7.1. <i>Description</i> .....	12
7.2. <i>Contenu</i> .....	12
8 DICTIONNAIRE « ESERVICES » .....	13
8.1. <i>Description</i> .....	13
8.2. <i>Contenu</i> .....	13
9 DICTIONNAIRE « ECOMMERCE » .....	14
9.1. <i>Description</i> .....	14
9.2. <i>Contenu</i> .....	14
10 DICTIONNAIRE « MÉDICAL » .....	15
10.1. <i>Description</i> .....	15
10.2. <i>Contenu</i> .....	15
11 DICTIONNAIRE « CUSTOM » .....	16

# > Mode de constitution des dictionnaires Lingway

## 1 Organisation des données

### 1.1. Modèle

L'ensemble des ressources dictionnaires disponibles à Lingway est regroupé dans une base de données unique. On peut distinguer deux types d'informations présentes dans la base :

- les informations générales liées au modèle (traits morphologiques, catégories syntaxiques, traits sémantiques)
- les informations liées à l'appartenance à une ressource terminologique spécifique (dites « briques »).

#### 1.1.1 Modèle de la base

Le dictionnaire Lingway est constitué selon un modèle à trois niveaux

- Niveau morpho-syntaxique : niveau où sont codées les entrées du dictionnaire, sous la forme lemme + catégorie syntaxique + langue. Les entrées sont marquées par des attributs morphologiques : mode de flexion (singulier/pluriel, masculin/feminin, conjugaison des verbes, etc.)
- Niveau sémantique : niveau où les mots sont rattachés à leur différents sens possibles.
- Niveau conceptuel : niveau où les sens des mots sont reliés à des concepts. Les concepts sont reliés entre eux par un réseau de liens sémantiques. Ils sont marqués par des attributs sémantiques : classe (le concept de *jonquille* est marqué par la classe « fleur ») et domaine (le concept de *jonquille* est marqué des domaines « horticulture » et « botanique »).
  - Le modèle Lingway comprend 22 domaines principaux, hiérarchisés en 411 sous-domaines. Les 22 domaines principaux sont les suivants :

Agriculture, Pêche et Forêts	Médecine et Santé
Alimentation	Militaire
Art	Politique et Gouvernement
Droit	Religion
Enseignement	Sans domaine
Gestion de l'entreprise	Sciences de la Terre
Industrie - Artisanat	Sciences de la vie
Industrie - Construction	Sciences humaines
Industrie - Services	Sciences Physiques et Mathématiques
Industries manufacturières	Sports et Loisirs
Maison et jardin	Transport

- Il existe 477 classes sémantiques de type ANIMAL, HUMAIN, VEGETAL, PLANTE, ARBRE, LOGICIEL, FONCTION, ARTEFACT, etc. Ces classes permettent de regrouper les données pour des travaux terminologiques particuliers.

### 1.1.2 Ressources terminologiques ou « briques »

A chaque niveau du modèle, les informations<sup>1</sup> peuvent être incluses dans une brique. Ainsi dans une brique donnée, on pourra spécifier :

- des mots à sélectionner ou supprimer
  - le nom propre *Java* (nom de l'île) est supprimé de la brique informatique
  - le nom commun *lotus* (plante) est supprimé de la brique informatique
  - le nom commun *paramétreur* ou le nom propre *Lotus* sont définis dans une brique informatique. Ils n'existent pas dans le dictionnaire général
- des pertinences de mots à spécifier
  - par exemple il existe une brique de pertinences sur les noms de fonction. Ceci a notamment une influence sur le scoring dans LKM.
- des sens de mots à sélectionner ou supprimer
  - le sens « devin » du mot *oracle* est supprimé dans la brique informatique
  - le sens « base de données » du mot *oracle* est ajouté dans la brique informatique
- des relations entre concepts à sélectionner ou supprimer
  - par exemple ajout de liens entre SAP et ERP dans la brique informatique
  - suppression du lien entre *nettoyage* et *maintenance* dans une brique RH
- des informations sémantiques
  - Ajout d'une information de domaine à un concept, par exemple domaine « Logiciels » affecté au concept de l'entrée Lotus Notes créée dans la brique informatique

Chaque travail terminologique effectué à Lingway est importé dans la base avec un nom de brique spécifique.

## 1.2. Compilation d'un dictionnaire

La constitution d'un dictionnaire opérationnel s'effectue en plusieurs étapes :

- On définit l'ensemble des briques nécessaires à l'application visée
- On compile dans la base de données un point de vue « terminal » qui agrège l'ensemble des briques qui le constituent
- A partir du point de vue ainsi créé, on compile un dictionnaire opérationnel qui sera chargé dans les applications LKM

## 2 Constitution de briques

La constitution de briques dictionnaires peut prendre pour base deux sources de données selon qu'elles sont externe ou interne au dictionnaire:

### 2.1. Données externes au dictionnaire

#### 2.1.1 Thésaurus / terminologies externes

Toute ressource externe au dictionnaire peut être importée dans la base Lingway, avec les étapes suivantes :

- Passage de la ressource au format d'import dans la base Lingway

---

<sup>1</sup> Certaines informations linguistiques telles que le mode de flexion ne sont pas liées à une brique.

- Choix (automatique ou manuel) des concepts auxquels se raccorder
- Création d'une nouvelle brique dans la base

Cette ressource peut être constituée en interne par les terminologues de Lingway, ou fournie par un client ou éditeur. Les étapes de création de la brique sont effectuées par Lingway.

### 2.1.2 Etude de corpus

En l'absence de ressource externe, ou pour en faciliter la constitution, Lingway dispose d'outils d'analyse de corpus. Ces outils permettent d'extraire :

- Les mots les plus fréquents sur lesquels le réseau est à compléter ou à nettoyer
- Les mots inconnus à ajouter au dictionnaire
- Les mots sans réseau sur lesquels le réseau est à compléter

Dans ce cas on constituera une brique initiale d'adaptation du dictionnaire aux besoins du corpus visé.

## 2.2. Données internes au dictionnaire

De nouvelles briques peuvent être créées sur la base d'informations présentes dans l'ensemble de la base. Essentiellement des informations de domaine (par exemple, travailler sur tous les mots du domaine « Construction » pour une application BTP), classe sémantique (par exemple, le travail sur les noms de fonction a été initié à partir de l'ensemble des sens de mots portant la classe fonction), mais aussi comportement morpho-syntaxique (par exemple, recherche de tous les noms finissant en -iste).

## 3 Ressources disponibles

### 3.1. Dictionnaire de la langue générale

Une brique « dictionnaire général » de la langue est utilisée comme souche commune à l'ensemble des dictionnaires livrés.

Elle contient environ 70 000 concepts lexicalisés dans les 5 langues actuellement disponibles :

- DE (Allemand)
- EN (Anglais)
- ES (Espagnol)
- FR (Français)
- NL (Néerlandais)

Cette brique contient le nombre de mots suivants :

Général	nombre de mots par langue	dont connectés au réseau
	de	102882
	en	62156
	es	79391
	fr	72233
	nl	84970
nbr de concepts		71628
nbr de relations		34624

Pour un total de 71628 concepts et 34624 relations entre concepts.

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	CONCEPTS NL	TOTAL CONCEPTS
Droit	638	848	477	847	405	1113
Militaire	614	843	407	786	302	1122
Politique et Gouvernement	773	1127	634	1053	402	1488
Agriculture, Pêche et Forêts	639	930	437	915	292	1246
Art	1161	1576	905	1528	595	2077
Gestion de l'entreprise	480	697	303	710	260	962
Industrie - Construction	448	615	280	668	214	887
Industrie - Artisanat	104	161	89	174	52	232
Sciences de la Terre	872	1632	797	1567	353	1922
Enseignement	277	359	184	307	123	514
Alimentation	1021	1415	607	1323	376	2047
Médecine et Santé	1166	1944	1124	1993	521	2534
Maison et jardin	575	771	352	757	273	1024
Sciences humaines	981	1716	955	1772	480	2141
Sciences de la vie	2382	3404	1915	3226	1058	4211
Industries manufacturières	2141	3078	1371	2989	1177	4338
Sciences Physiques et Mathématiques	1177	1679	947	1532	448	2036
Religion	513	800	473	773	276	981
Industrie - Services	1218	1573	753	1492	576	2224
Sports et Loisirs	1167	1550	772	1462	492	2189
Transport	737	979	452	898	364	1329
Sans domaine	24435	23196	12841	20939	714	82193

A la brique « dictionnaire général » s'ajoutent des briques de spécialité.

## 3.2. Briques de spécialité

### 3.2.1 Géographie

Noms de pays, villes, fleuves, relations entre adjectifs et noms de nationalité (France, français), ville (Reims, rémois), etc.

Géographie		dont connectés au réseau	
nbr de mots par langue	de	1820	1818
	en	7974	6587
	es	N/A	N/A
	fr	4970	4407
	nl	N/A	N/A
nbr de concepts		6147	
nbr de relations		5362	

### 3.2.2 Médical

Brique bilingue Anglais-Français du vocabulaire médical.

Brique MEDICAL		dont connectés au réseau	
nbr de mots par langue	de	N/A	N/A
	en	55647	55514
	es	N/A	N/A
	fr	29437	29208
	nl	N/A	N/A
nbr de concepts		20991	
nbr de relations		10382	

### 3.2.3 Juridique et fiscal

Brique français anglais

Brique JURIDIQUE et FISCAL			dont connectés au réseau
nbr de mots par langue	de	N/A	N/A
	en	4169	2960
	es	N/A	N/A
	fr	7883	5766
	nl	N/A	N/A
nbr de concepts		6102	
nbr de relations		5401	

La majeure partie des concepts de cette brique appartient aux domaines « Droit » et « Finance ».

### 3.2.4 Informatique

Cette nouvelle brique est en cours de constitution. Elle recense les principaux noms de programmes, logiciels, systèmes actuellement fréquemment cités dans les textes informatiques.

Brique INFORMATIQUE			dont connectés au réseau
nbr de mots par langue	de	1113	891
	en	1294	1005
	es	1088	878
	fr	3875	2463
		1102	N/A
nbr de concepts		1687	
nbr de relations		819	

Ces concepts appartiennent pour la plupart au sous-domaine « informatique » du domaine « Industries manufacturières ».

### 3.2.5 Ressources Humaines

Un ensemble de briques dans le domaine du recrutement :

- Intitulés d'offres d'emploi
- Intitulés de CV
- Corpus de CV (expérience, formation, etc)
- Principales grandes écoles
- Noms de fonction :
  - Recensement de l'ordre de 2000 noms de fonction
  - Affectation de pertinences
  - Gestion des formes masculin/féminin

Ressources Humaines	nombre de mots par langue		dont connectés au réseau
	de	1630	1227
	en	2529	1767
	es	1453	952
	fr	24842	16089
	nl	8769	7723
nbr de concepts		14423	
nbr de relations		11207	

### 3.2.6 Technique

Un ensemble de briques constituées dans le cadre de projets menés dans le domaine des brevets ou autres domaines techniques.

Briques Techniques	nombre de mots par langue		dont connectés au réseau
	de	1985	1672
	en	17896	12465
	es	8645	6469
	fr	54267	26031
	nl	N/A	N/A
nbr de concepts		14283	
nbr de relations		11195	

### 3.2.7 Briques d'ontologies

Différentes briques ontologiques ont été créées pour répondre notamment à des besoins de classification dans des taxinomies. Par exemple création de liens de type générique-spécifique entre fruit et fraise, pomme, banane, etc. afin de pouvoir identifier le libellé « jus de fruit » d'une classification donnée à partir de la question « jus de pomme ».

Des briques d'ontologie ont été créées dans les domaines suivants :

- Habillement (hauts, bas, pardessus, vêtements de nuit, lingerie, chaussures, accessoires, etc.)
- Plantes
  - Fleurs
  - Arbres
  - Arbustes
  - etc.
- Monde animal
  - Mammifères
  - Oiseaux
  - Insectes
  - Poissons
- Alimentation
  - Boucherie
  - Charcuterie
  - Boulangerie
  - Poissonnerie
  - fruits et légumes
  - etc.
- Véhicules
  - Aériens
  - Fluviaux / Maritimes

- Terrestres
  - Ferroviaires
- Sports et loisirs
  - Jeux
  - Sports
  - Musique

### **3.2.8 Briques terminologiques**

Pour suivre l'évolution et l'apparition de nouveaux termes, quelques briques terminologiques ont été créées dans les domaines suivants :

- Nanotechnologies
- Développement durable

# > Dictionnaires disponibles

## 4 Dictionnaire « Presse généraliste »

### 4.1. Description

Dictionnaire de la langue générale, adapté au traitement de textes de presse généraliste.

### 4.2. Contenu

- Brique Générale (voir 3.1)
- Brique géographique présentée en 3.2.1
- Briques d'adaptation des données au corpus des journaux "le monde" et "New York times" (site de démonstration "news" de Lingway)

Presse Généraliste	nombre de mots par langue		dont connectés au réseau
	de	102933	46390
	en	73531	58140
	es	84456	24199
	fr	85678	41336
	nl	88598	11785
nbr de concepts		93947	
nbr de relations		64251	

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	CONCEPTS NL	TOTAL CONCEPTS
Droit	625	1067	499	1143	545	1319
Militaire	541	840	426	804	342	1056
Politique et Gouvernement	691	1287	650	1287	599	1588
Agriculture, Pêche et Forêts	841	1428	652	2434	619	2683
Art	1077	1581	927	1607	791	1989
Gestion de l'entreprise	459	802	344	889	476	1051
Industrie - Construction	425	669	334	802	320	922
Industrie - Artisanat	109	225	101	249	93	331
Sciences de la Terre	1244	5839	834	4787	933	6082
Enseignement	250	360	188	337	175	458
Alimentation	949	1600	671	1775	668	2141
Médecine et Santé	1105	2021	1187	2180	827	2525
Maison et jardin	532	847	399	881	411	1035
Sciences humaines	1035	1785	964	1884	789	2151
Sciences de la vie	2149	3465	1936	3353	1351	3989
Industries manufacturières	1978	3740	1646	3761	1616	4794
Sciences Physiques et Mathématiques	1121	1743	1023	1701	628	2046
Religion	481	791	475	774	328	942
Industrie - Services	1119	1668	821	1799	904	2224
Sports et Loisirs	1116	1649	849	1701	736	2175
Transport	682	1111	508	1106	556	1366
Sans domaine	27814	24243	17443	33439	3579	106531

## 5 Dictionnaire « Documentation technique »

### 5.1. Description

Dictionnaire technique, adapté au traitement de documentations techniques, type brevets d'invention.

### 5.2. Contenu

- Brique Générale (voir 3.1)
- Brique technique en **Erreur ! Source du renvoi introuvable.**

Documentation Technique	nombre de mots par langue	dont connectés au réseau
	de	105405
	en	73450
	es	84462
	fr	87301
	nl	N/A
nbr de concepts		92589
nbr de relations		73338

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	TOTAL CONCEPTS
Droit	602	864	488	929	1098
Militaire	542	836	422	794	1047
Politique et Gouvernement	672	1110	637	1072	1381
Agriculture, Pêche et Forêts	720	1115	525	1169	1403
Art	1067	1563	921	1551	1937
Gestion de l'entreprise	427	710	315	750	910
Industrie - Construction	412	638	314	743	865
Industrie - Artisanat	106	217	96	225	309
Sciences de la Terre	910	1680	827	1644	1870
Enseignement	248	357	188	324	448
Alimentation	930	1517	648	1594	1965
Médecine et Santé	1100	2008	1152	2114	2468
Maison et jardin	512	795	373	807	956
Sciences humaines	987	1706	960	1780	2044
Sciences de la vie	2179	3474	1935	3342	3991
Industries manufacturières	1909	3236	1535	3424	4120
Sciences Physiques et Mathématiques	1170	1760	1002	1687	2040
Religion	478	783	473	759	931
Industrie - Services	1090	1589	790	1623	2051
Sports et Loisirs	1078	1594	818	1593	2071
Transport	681	1051	499	1020	1275
Sans domaine	29346	31256	17524	43010	121147

## 6 Dictionnaire « eRecrutement »

### 6.1. Description

Dictionnaire destiné aux applications de ressources humaines, qu'il s'agisse de recherche d'offres ou de CV.

### 6.2. Contenu

- Brique Générale
- Briques Ressources Humaines
- Brique Informatique

eRECRUTEMENT	nombre de mots par langue	dont connectés au réseau
	de	104457
	en	64155
	es	80757
	fr	82550
	nl	88707
nbr de concepts		87001
nbr de relations		50279

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	CONCEPTS NL	TOTAL CONCEPTS
Droit	797	1071	668	1312	699	1481
Militaire	547	839	408	802	337	1057
Politique et Gouvernement	753	1189	720	1186	581	1483
Agriculture, Pêche et Forêts	609	972	458	1071	438	1299
Art	1085	1585	919	1632	770	2025
Gestion de l'entreprise	648	927	525	1120	611	1284
Industrie - Construction	531	768	385	1062	413	1180
Industrie - Artisanat	118	183	101	225	92	269
Sciences de la Terre	831	1684	824	1717	512	1933
Enseignement	269	378	198	527	186	645
Alimentation	919	1461	635	1490	591	1854
Médecine et Santé	1119	1995	1166	2204	783	2543
Maison et jardin	546	838	383	901	404	1049
Sciences humaines	962	1748	993	1848	676	2110
Sciences de la vie	2153	3415	1935	3332	1353	3966
Industries manufacturières	2187	3591	1606	5002	1751	6173
Sciences Physiques et Mathématiques	1130	1724	976	1729	637	2083
Religion	479	784	472	763	315	934
Industrie - Services	1211	1705	873	2055	935	2557
Sports et Loisirs	1086	1582	795	1575	679	2050
Transport	680	1023	480	1103	508	1357
Sans domaine	26680	21910	12008	32018	3451	96109

## 7 Dictionnaire « Juridique et fiscal »

### 7.1. Description

### 7.2. Contenu

- Brique Générale

- Briques Juridique et Fiscal
- Briques d'adaptation liées au domaine

JURIDIQUE et FISCAL	nombre de mots par langue		dont connectés au réseau
	de	N/A	N/A
	en	64363	60310
	es	N/A	N/A
	fr	83039	57952
	nl	N/A	N/A
nbr de concepts		71580	
nbr de relations		50988	

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS EN	CONCEPTS FR	TOTAL CONCEPTS
Droit	3612	9513	10952
Militaire	834	783	971
Politique et Gouvernement	1204	1442	1651
Agriculture, Pêche et Forêts	928	979	1167
Art	1575	1541	1821
Gestion de l'entreprise	767	1108	1248
Industrie - Construction	628	744	843
Industrie - Artisanat	163	185	224
Sciences de la Terre	1676	1649	1811
Enseignement	357	343	434
Alimentation	1405	1369	1656
Médecine et Santé	1938	2061	2323
Maison et jardin	768	766	909
Sciences humaines	1741	1802	1997
Sciences de la vie	3384	3245	3714
Industries manufacturières	3123	3423	4050
Sciences Physiques et Mathématiques	1680	1585	1849
Religion	787	763	883
Industrie - Services	1599	2033	2344
Sports et Loisirs	1540	1470	1812
Transport	971	943	1153
Sans domaine	22024	28824	50940

## 8 Dictionnaire « eServices »

### 8.1. Description

Ce dictionnaire est destiné aux applications recensant des descriptions d'activités professionnelles, de type « Pages Jaunes ».

### 8.2. Contenu

- Brique Générale
- Brique géographique présentée en 3.2.1
- Briques d'ontologie présentées en 3.2.7

- Briques d'adaptation liées aux projets menés dans le domaine

eSERVICES	nombre de mots par langue		dont connectés au réseau
	de	102921	48020
	en	71024	65822
	es	79423	24241
	fr	81774	56966
	nl	85714	12356
nbr de concepts		89733	
nbr de relations		51773	

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	CONCEPTS NL	TOTAL CONCEPTS
Droit	624	1064	493	1135	451	1311
Militaire	542	836	407	795	302	1050
Politique et Gouvernement	689	1277	643	1242	463	1556
Agriculture, Pêche et Forêts	834	1409	631	2383	442	2640
Art	1077	1572	911	1569	602	1962
Gestion de l'entreprise	456	778	328	833	323	998
Industrie - Construction	422	649	295	775	226	899
Industrie - Artisanat	107	217	94	233	57	317
Sciences de la Terre	1244	5834	820	4769	614	6072
Enseignement	249	357	185	327	123	452
Alimentation	939	1546	642	1699	412	2075
Médecine et Santé	1089	1961	1137	2083	533	2436
Maison et jardin	529	827	373	849	299	1005
Sciences humaines	1038	1788	957	1875	508	2150
Sciences de la vie	2148	3450	1921	3314	1074	3965
Industries manufacturières	1942	3629	1459	3592	1071	4640
Sciences Physiques et Mathématiques	1121	1717	960	1664	461	2020
Religion	480	790	472	766	276	940
Industrie - Services	1099	1604	766	1670	615	2103
Sports et Loisirs	1104	1626	809	1638	528	2124
Transport	678	1097	482	1074	437	1336
Sans domaine	27852	21931	11982	28944	1035	91749

## 9 Dictionnaire « eCommerce »

### 9.1. Description

Le dictionnaire eCommerce est destiné aux portails de vente par internet.

### 9.2. Contenu

Ce dictionnaire a une base proche du dictionnaire eServices, plus quelques briques spécifiques recensant plus particulièrement des noms de produits. Il ne contient pas la brique géographique.

- Brique Générale
- Briques d'ontologie présentées en 3.2.7
- Brique recensant du matériel HiFi
- Brique spécifique de traitement des numériques pour prendre en compte des variations dans des noms d'articles (ex : Shrek deux / Shrek II / Shrek 2)

eCOMMERCE	nombre de mots par langue		dont connectés au réseau
	de	104312	47750
	en	67412	63232
	es	79417	24240
	fr	77668	53648
	nl	84976	11791
nbr de concepts		84059	
nbr de relations		51922	

La répartition du nombre de concepts par domaine et par langue est la suivante :

Libellé Domaine	CONCEPTS DE	CONCEPTS EN	CONCEPTS ES	CONCEPTS FR	TOTAL CONCEPTS
Droit	605	867	484	925	1099
Militaire	543	837	405	800	1051
Politique et Gouvernement	671	1111	634	1075	1384
Agriculture, Pêche et Forêts	764	1113	502	1177	1388
Art	1070	1570	906	1560	1952
Gestion de l'entreprise	428	711	306	759	917
Industrie - Construction	420	651	285	766	884
Industrie - Artisanat	133	240	101	267	347
Sciences de la Terre	826	1686	821	1654	1871
Enseignement	248	356	185	324	448
Alimentation	1039	1598	631	1670	1991
Médecine et Santé	1129	1982	1132	2142	2479
Maison et jardin	578	862	381	960	1103
Sciences humaines	927	1706	954	1781	2046
Sciences de la vie	2302	3497	1933	3374	3992
Industries manufacturières	1980	3683	1426	3606	4629
Sciences Physiques et Mathématiques	1144	1705	951	1634	1986
Religion	485	788	474	765	936
Industrie - Services	1119	1607	768	1667	2102
Sports et Loisirs	1107	1630	796	1656	2132
Transport	694	1064	481	1048	1306
Sans domaine	27171	25266	11990	29945	95380

## 10 Dictionnaire « Médical »

### 10.1. Description

Dictionnaire médical adapté à la classification MEDDRA.

### 10.2. Contenu

- Brique Générale
  - Dont sont supprimées les entrées ambiguës avec brique médicale (45 000 sens de mots supprimés)
- Brique MEDDRA

MEDICAL	nombre de mots par langue		dont connectés au réseau
	en	104988	101121
	fr	99245	70500
nbr de concepts		80336	
nbr de relations		45678	

## 11 Dictionnaire « Custom »

L'organisation des données dictionnaires par "briques" permet de compiler de nouveaux dictionnaires pour de nouveaux besoins. Dans ce cas, une première phase de spécifications permettra à Lingway de sélectionner les briques adaptées et de proposer un dictionnaire de base au client.

Ensuite, une adaptation plus poussée pourra être menée, et donner lieu à de nouvelle « briques ». Cette adaptation pourra s'appuyer sur :

- Des ressources existantes du client (Terminologie, thésaurus, etc.)
- Une analyse du corpus client (Ajout de mots inconnus, ajout et vérification du réseau sur les mots les plus fréquents du corpus).