Click to edit Master subtitle style

Adam Kilgarriff
Lexical Computing Ltd

Sketch Engine

- Corpus tool
- Sixty languages
- Lexicography
  - OUP, CUP, Collins, Cornelsen, Le Robert
- Research and teaching
  - 200 universities worldwide
- New directions
  - Translation
  - Terminology

Parallel data

- EUROPARL
- OPUS
- 231 language pairs

Concordance
Word List

Save
View options
KWIC
Sentence
Alignment
Sort
Left
Right
Node
Shuffle
Sample
Filter
Frequency
Node forms
Doc IDs
Collocations
ConcDesc

Query love, amour 267 (4.4 per million)

Page 1 of 14 Go Next | Last

EUROPARL7, en	EUROPARL7, fr
I am speaking for the first time in this plenary part-session , so this is quite exciting for me , a little like first love , although that did last longer than two minutes .	C ' est la première fois que je prends la parole en plénière , il y a donc de quoi être un peu nerveux , un peu comme avec le premier amour , mais le premier amour a quand même duré heureusement plus de deux minutes .
And since this is St. Valentine ' s day , as a former Mayor of a regional city , I propose that we should all declare our love for all the European regions which need that love .	Et , puisque aujourd ' hui , c' est la Saint-Valentin , en tant qu' ancien maire d ' une ville régionale , je propose que nous déclarions notre amour envers les régions européennes qui en ont besoin .
And since this is St. Valentine ' s day , as a former Mayor of a regional city , I propose that we should all declare our love for all the European regions which need that love .	Et , puisque aujourd ' hui , c' est la Saint-Valentin , en tant qu' ancien maire d ' une ville régionale , je propose que nous déclarions notre amour envers les régions européennes qui en ont besoin .
Nevertheless , it is this very love that is a requirement for the healthy development of the individual .	Or c ' est précisément cet amour qui est la condition de l ' épanouissement de l ' individu , et l ' Europe n ' a que faire de droits fondamentaux progressistes si les membres de la société ne veulent pas les respecter .
Indeed , the word of God repeatedly and emphatically speaks of hospitality and mercifulness to strangers , as well as true charity as a consequence of our love for God , the Creator of all mankind .	La parole divine insiste en effet à maintes reprises et avec insistance sur la nécessité d' adopter une attitude accueillante et de témoigner des marques de charité à l' égard des étrangers , l ' amour du prochain étant le corollaire de l ' amour porté à Dieu , notre Créateur à tous

Concordance
Word List

Save
View options
KWIC
Sentence
Alignment
Sort
Left
Right
Node
Shuffle
Sample
Filter
Frequency
Node forms
Doc IDs
Collocations
ConcDesc

Query αγάπη, Liebe 77 (1.7 per million)

Page 1 of 4 Go Next | Last

EUROPARL7, el	EUROPARL7, de
Παρ ' όλα αυτά , αυτή ακριβώς η αγάπη αποτελεί προϋπόθεση για την ανάπτυξη ενός υγιούς ατόμου .	Dennoch ist gerade diese Liebe Voraussetzung für eine gesunde Persönlichkeitsentwicklung .
Ο λόγος του Θεού αναφέρεται επανειλημμένα και με έμφαση στη φιλοξενία και την ευσπλαχνία προς τον ξένο και στην πραγματική αγάπη προς τον πλησίον ως απόρροια της αγάπης του ίδιου του Δημιουργού προς τον άνθρωπο .	Gottes Wort fordert ja wiederholt und nachdrücklich Freundlichkeit und Barmherzigkeit gegenüber Fremden und gebietet tätige Nächstenliebe als Folge der Liebe Gottes , unser aller Schöpfer .
Θέλω επίσης να υπενθυμίσω σε εκείνους που ασπάζονται μια εθνοτική αντίληψη του έθνους τη ρήση του Clemenceau : " Πατριωτισμός είναι η αγάπη για την χώρα σου , εθνικισμός είναι το μίσος για τους άλλους " .	Und die Verfechter einer ethnischen Auffassung von der Nation möchte ich an die Worte von Clémenceau erinnern : " Patriotismus ist die Liebe zum Vaterland , Nationalismus ist der Haß auf die anderen Länder . "
Χρειάζεται χρόνος , υπομονή , αγάπη , για να τους ξαναδώσουμε την ελπίδα .	Es braucht Zeit , Geduld , Liebe , um ihnen wieder Hoffnung zu geben .
Όπως είπαν ορισμένοι εμπειρογνώμονες , τα θρησκευτικά θέματα δεν επιδέχονται σταθμίσεις , αυτό όμως που είναι βέβαιο είναι ότι αν μιλάμε για τις θρησκείες της Βίβλου , που κηρύττουν την αγάπη προς το Θεό και το συνάνθρωπο , τότε πρέπει να είναι δυνατόν να βρούμε μια λύση ανθρώπινη .	Wie Experten sagen , kann man bei religiösen Themen nicht vermitteln , wenn wir jedoch über die Religionen der Heiligen Schrift sprechen , die die Liebe zu Gott und zum Nächsten predigen , muss es möglich sein , dass Menschen untereinander eine Lösung finden .

- Like Linguee
  - Less data per pair
  - *More pairs*

Word sketch

- One page, corpus-based summary of a word's grammatical and collocation behaviour

- Lexicography
  - *Big impact*

  - Until now
    - monolingual

The Bilingual Word Sketch

declaration *(noun)* EUROPARL5, English-French freq = 4409

déclaration *(noun)* EUROPARL5, French-English freq = 9341

use another candidate translation: déclarations écrites écrite Déclarations

modifier		
written	149	I am surprised that on 6 May , after consultation with the World Health Organisation , the Council decided not to do this and to rely instead on checks in the country of departure and written **declarations** by interested parties .
écrite	49	Mr President , Members have had circulated to them this evening notice of my written **declaration** on alcopops which lapses at 6.30 pm . Monsieur le Président , les membres n' ont pris connaissance de ma **déclaration** écrite sur les « alcopops » que cet après-midi alors qu' elle expire précisément aujourd 'hui à 18 h 30 .
solemn	51	Solemn **declarations** and moral indignation are not enough , though ; they also , as is specified in our joint resolution , have to be backed up by a whole host of things .
solennel		EU-Africa Summit in Cairo - (FR) The solemn **declaration** of the first EU-Africa summit in Cairo opens by stating , and I quote : " Over the centuries , ties have existed between Africa and Europe ... developed on the basis of shared values of strengthening representative and participatory democracy " . Given that this secular past was a story of slavery , massacres , forced labour , plundering , colonial conquests and oppression , during which the rich European countries bled that continent dry , we can only wonder what is the most shameful aspect : the pride of the representatives of the imperialist countries or the baseness La **déclaration** solennelle du premier sommet Afrique-Europe , au Caire , commence par faire référence , je cite : aux " liens qui existent entre l' Afrique et l' Europe " ..... " depuis des siècles " qui se seraient " développés sur la base de valeurs communes telles que le renforcement de la démocratie " .
unilateral	61	He must not add fuel to the flames by threatening a unilateral **declaration** of independence for the Palestinian State .
unilatérale	9	The problem is that , after nine years of refusing to sign a border agreement with Estonia , Russia finally did so last month , but the Estonian Parliament , following typical parliamentary procedure , added a unilateral non-binding **declaration** saying that the legal continuity of the State is enforced even when territory is given up . Ce problème est le suivant : après avoir refusé pendant neuf ans de conclure un accord frontalier avec l' Estonie , la Russie a enfin accepté le mois dernier , mais le parlement estonien , au terme d' une procédure parlementaire typique , a ajouté une **déclaration** unilatérale non contraignante affirmant que la continuité juridique de l' État est assurée même quand un territoire

- Method 1
  - Parallel corpus, fully automatic
- Method 2
  - User chooses words to compare

house *(noun)* British National Corpus freq = 57976 (516.8 per million)

maison French web corpus freq = 36739 (289.6 per million)

modifier	24107	1.3	modifier	3467	0.8	object_of	9534	1.5	objet_de	5965	2.3
White	701	9.65	paternel	112	47.29	build	726	9.06	habiter	220	42.58
opera	334	8.6	hanté	47	44.74	buy	533	8.7	bâtir	136	40.33
manor	236	8.19	familial	162	41.68	sell	308	8.02	quitter	320	39.26
guest	263	8.04	universel	133	38.5	own	138	7.77	construire	220	37.76
terraced	197	8.04	voisin	100	33.12	enter	171	7.59	acheter	139	31.84
discount	212	7.96	natal	41	32.03	rent	56	7.44	clore	76	30.02
big	365	7.9	neuf	56	31.58	occupy	87	7.29	fouiller	48	29.65
clearing	167	7.77	blanc	126	29.28	search	64	7.2	louer	59	29.28
public	358	7.72	royal	55	29.25	leave	420	7.17	incendier	32	28.21

Term finding

- How does a terminologist find their terms?
  - Ask an expert
  - Copy
  - Look in a corpus

Term finding

- How does a terminologist find their terms?
  - Ask an expert
  - Copy
  - Look in a corpus

Two questions

- Is it the right shape?

- Is it distinctive of the domain?

Two questions, two methods

- Is it the right shape?
  - Is it a noun phrase?
  - Shallow parsing to find noun phrases
- Is it distinctive of the domain?
  - For each noun phrase in domain corpus
    - Compare frequency with reference corpus

Lead customer

- WIPO (World Intellectual Property Organisation)

Term	Frequency	Freq/mill	Score
station de base	28612	3292.2	3293.2
station mobile	12514	1439.9	1440.9
communication sans fil	8189	942.3	943.3
liaison montante	6561	754.9	737.5
terminal mobile	7406	852.2	709.8
liaison descendante	5434	625.3	626.3
stations de base	5010	576.5	577.5
réseau de communication	4255	489.6	490.6
communication mobile	4722	543.3	462.5
point d' accès	3907	449.6	450.6
modes de réalisation	3486	401.1	402.1
réseau d' accès	3241	372.9	373.9
réseau sans fil	2903	334.0	335.0
accès radio	2412	277.5	278.5
transfert intercellulaire	2408	277.1	278.1

Term	Frequency	Freq/mill	Score
移動局	1374	2512.5	2442.6
基地局	2324	4249.6	2048.5
無線基地局	1025	1874.3	1787.7
移動端末	702	1283.7	1284.7
無線端末	477	872.2	865.4
無線リソース	430	786.3	780.3
通信端末	435	795.4	716.2
制御部	379	693.0	656.0
送信部	337	616.2	602.8
送信電力	326	596.1	574.7
無線通信	439	802.7	569.2
無線通信端末	304	555.9	556.9
識別情報	309	565.0	539.6
制御情報	298	544.9	528.0
ハンドオーバ	270	493.7	492.7

Reference corpus

- 60 languages
  - Already available
  - Many cases
    - Already lemmatised, POS-tagged

Domain corpus

- Customer already has it
  - WIPO case
- Instant corpora from the web
  - BootCaT procedure
  - Piggyback on search engine
    - Start from ‘seed terms’
    - Send queries to search engine
    - Gather pages that search engine finds
      - Iterate, if needed

Keywords

- dioxide (415.2, 427)
- trophic (264.9, 33)
- greenhouse (238.4, 282)
- ecology (237.7, 196)
- methane (233.5, 108)
- arrhenius (232.2, 25)
- photosynthesis (230.6, 46)
- callendar (215.4, 22)
- ecosystems (211.4, 114)
- warming (193.8, 504)
- keeling (192.5, 23)
- carbon (186.8, 558)
- n't (177.1, 17)
- gases (173.9, 159)
- -oct- (169.3, 28)
- vapor (151.3, 72)
- deforestation (144.7, 38)
- ecosystem (138.6, 88)
- mutualism (75.6, 8)
- radiative (75.0, 12)
- gasses (75.0, 12)
- lca (74.4, 10)
- biotic (74.2, 10)
- acidification (74.1, 9)
- above-ground (73.6, 9)
- holism (73.5, 9)
- felzer (73.5, 7)
- carbonic (72.4, 9)
- loa (71.5, 10)
- biogeography (71.2, 9)
- organisms (70.4, 86)
- mauna (69.7, 10)
- flowering (68.4, 23)
- emitted (68.2, 27)
- suess (67.4, 7)
- infrared (65.1, 44)

Terms

- carbon dioxide (567.1)
- greenhouse effect (515.0)
- water vapor (486.8)
- global warming (298.8)
- industrial ecology (261.6)
- infrared radiation (170.9)
- carbon cycle (169.0)
- surface temperature (161.0)
- elevated carbon (156.4)
- elevated carbon dioxide (156.4)
- greenhouse gas (135.8)
- climate system (134.1)
- food web (124.3)
- amount of carbon dioxide (116.8)
- other greenhouse (114.2)
- global temperature (109.1)
- atmospheric carbon (107.1)
- human activity (106.7)

In sum

- Sketch Engine
  - Leading corpus tool
  - Til now: lexicography, research, teaching
- Now
  - Translation
    - Parallel concordancing, many language pairs
    - Bilingual word sketches
  - Term-finding
    - Ready to integrate with terminology management

Thank you

http://www.sketchengine.co.uk