MultilingualWeb Workshop Riga, 29 April 2015

Semi-automatic generation of multilingual glossaries

Ilan Kernerman
K Dictionaries Ltd, Tel Aviv

SUMMARY

K Dictionaries' semi-automated multilingual glossaries stem from our unique English multilingual dictionary:

(1) reverse engineer parts of the initial data
(2) edit the word lists and links and re-process the results
ready for 15 languages: *with 43 languages each*
(3) expand with Linked Data & Semantic Web technologies
kicking off: *lemon-based*

The glossaries serve to deal with multilingual contents on the Web and to interconnect dozens of languages.

K DICTIONARIES *TechnologyDrivenContent*

- Multi-language/multi-layer content for 50 languages
  - *monolingual, bilingual & multilingual datasets*
  - *resources for language learning & translation*
  - *morphology & pronunciation, tools & applications*
- Established in 1993, based in Tel Aviv
- Cooperation with technology, publishing & academic partners worldwide

LINGUISTIC

- macro & microstructure
- editorial & translation styleguides
- metalanguage conversion tables
- headword & word form lists
- content & format revisions
- L1 lexicographer teams & L2 translators
- technical infrastructure synchronization

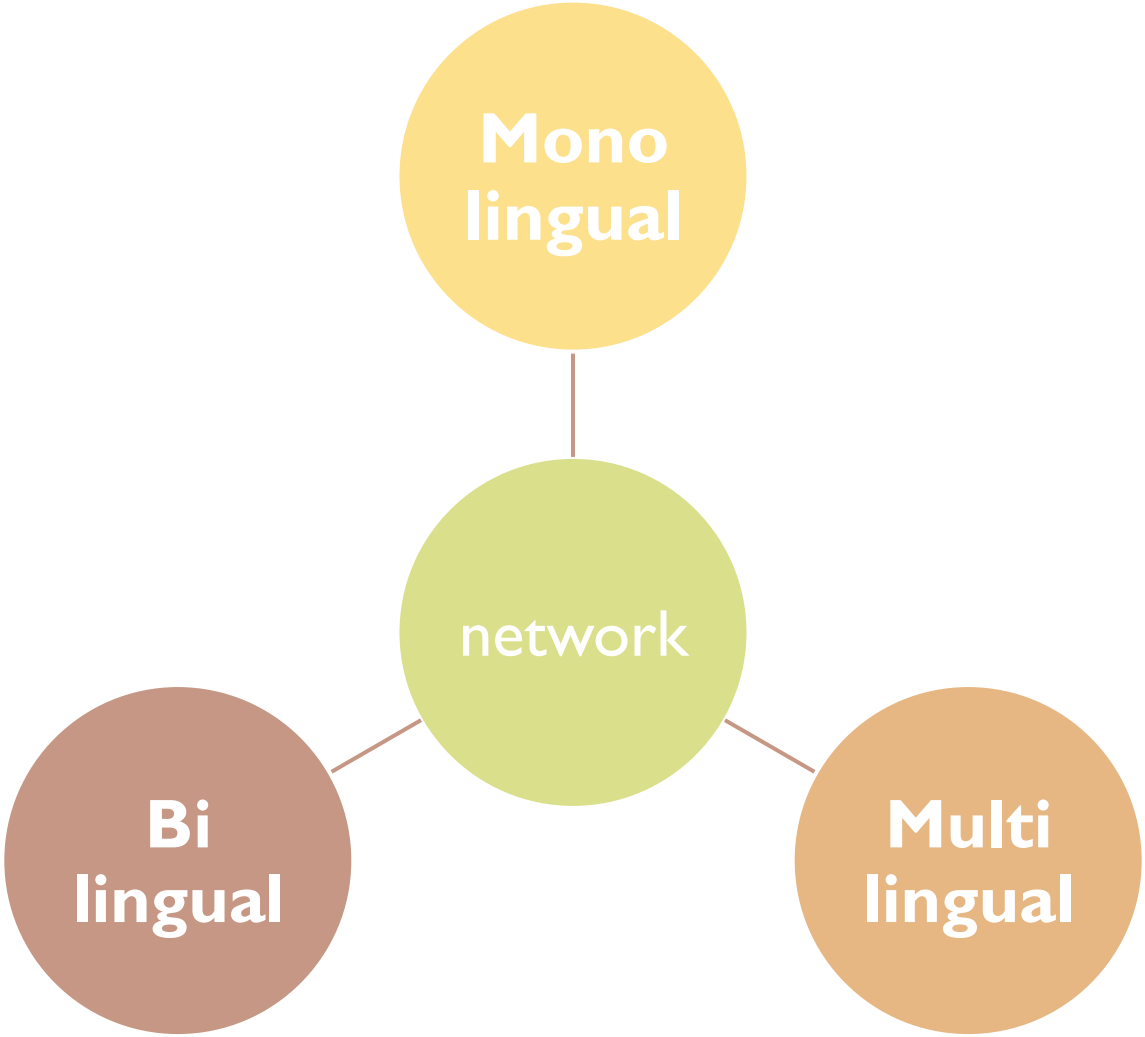
TECHNOLOGIC

- editorial, processing & publication tools
- XML-RDF configuration
- QA & statistics
- data maintenance, update & upgrade
- technical support
- digital applications
- R&D

EVOLUTION

1. monolingual English learner's dictionary
2. semi-bilingual English learner's dictionary
3. (semi-)multilingual English dictionary
4. L2-English reversed indexes
5. L2, L3 etc. multilingual dictionaries
6. L2-L3 bilingual glossaries
7. multi-language networks

MULTI-LAYER



VISION


Português
ไทย
Dansk
Deutsch
Українська
Nederlands
Čeština
ελληνικά
Türkçe
한국어
Zulu
فارسی
Hrvatski
Bahasa Indonesia
eesti
Norsk
Slovenčina
suomi
Български
中國傳統
Slovenščina
Српски
lingua Latina
Polski
עברית
Íslenska
Español
Afrikaans
Italiano
Bahasa Malaysia
Tiếng Việt
简体中文
العربية
Lietuvių
Русский
हिन्दी
پښتو
Română
Latviešu
Sesotho
Svenska
Xhosa
日本語
Magyar
English
ภาษาไทย
français

ENGLISH MULTILINGUAL

- PASSWORD semi-bilingual dictionary
- KEMD (44 languages)

  Afrikaans | Arabic | Bulgarian | Catalan | Chinese (Simplified | Traditional) | Croatian | Czech | Danish | Dutch | English | Estonian | Farsi | Finnish | French | German | Greek | Hebrew | Hindi | Hungarian | Icelandic | Indonesian | Italian | Japanese | Korean | Latvian | Lithuanian | Malay | Norwegian | Polish | Portuguese (Brazil | Portugal) | Romanian | Russian | Serbian | Slovak | Slovene | Spanish | Swedish |
  Thai | Turkish | Ukrainian | Urdu | Vietnamese

L2 MULTILINGUALS

- Extract list of Translations of any language (L2) with their corresponding English (EN) Entries & POS
- Edit the L2 Translations into L2 Headwords, keeping the default EN links
- Revise the links from the new Headword & POS to the relevant sense of the EN Entry
- Each sense of the L2 Headword now addresses its counterpart sense(s) in the EN Entries, and through it translation equivalents in all other languages
- [Expand the lexical data of the L2 Headword and turn it into a full Entry]

DATA STRUCTURE

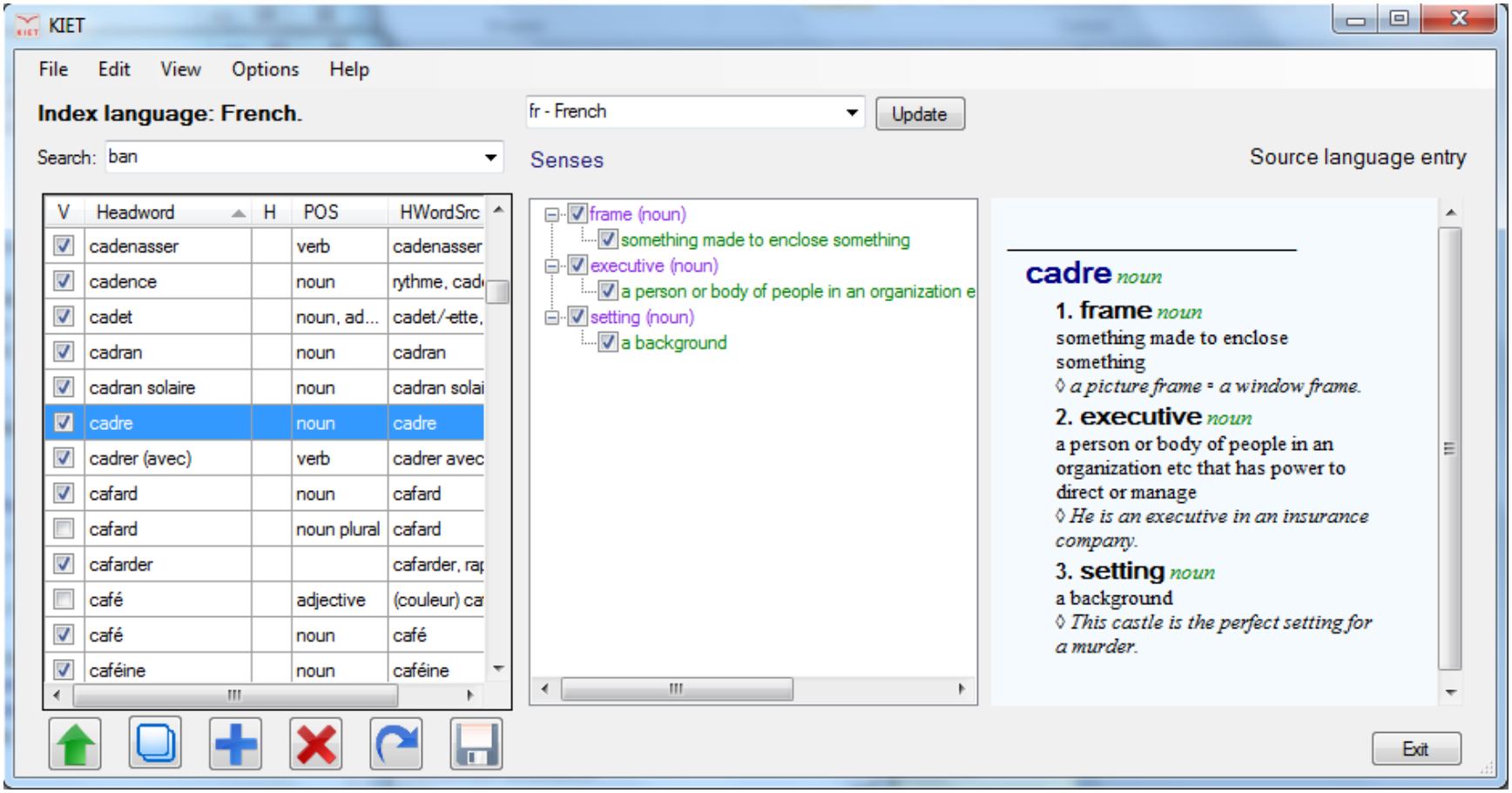
Main tables used for L2 Index generation

1. English HW table
2. Senses table
3. Translation table
4. L2 HW table
   (used in L2 Index table, generated from the English HW, Senses and Translation tables)
5. L2 Senses table
   (used for Tree and HTML preview, with English Words, Definitions and Examples tables)

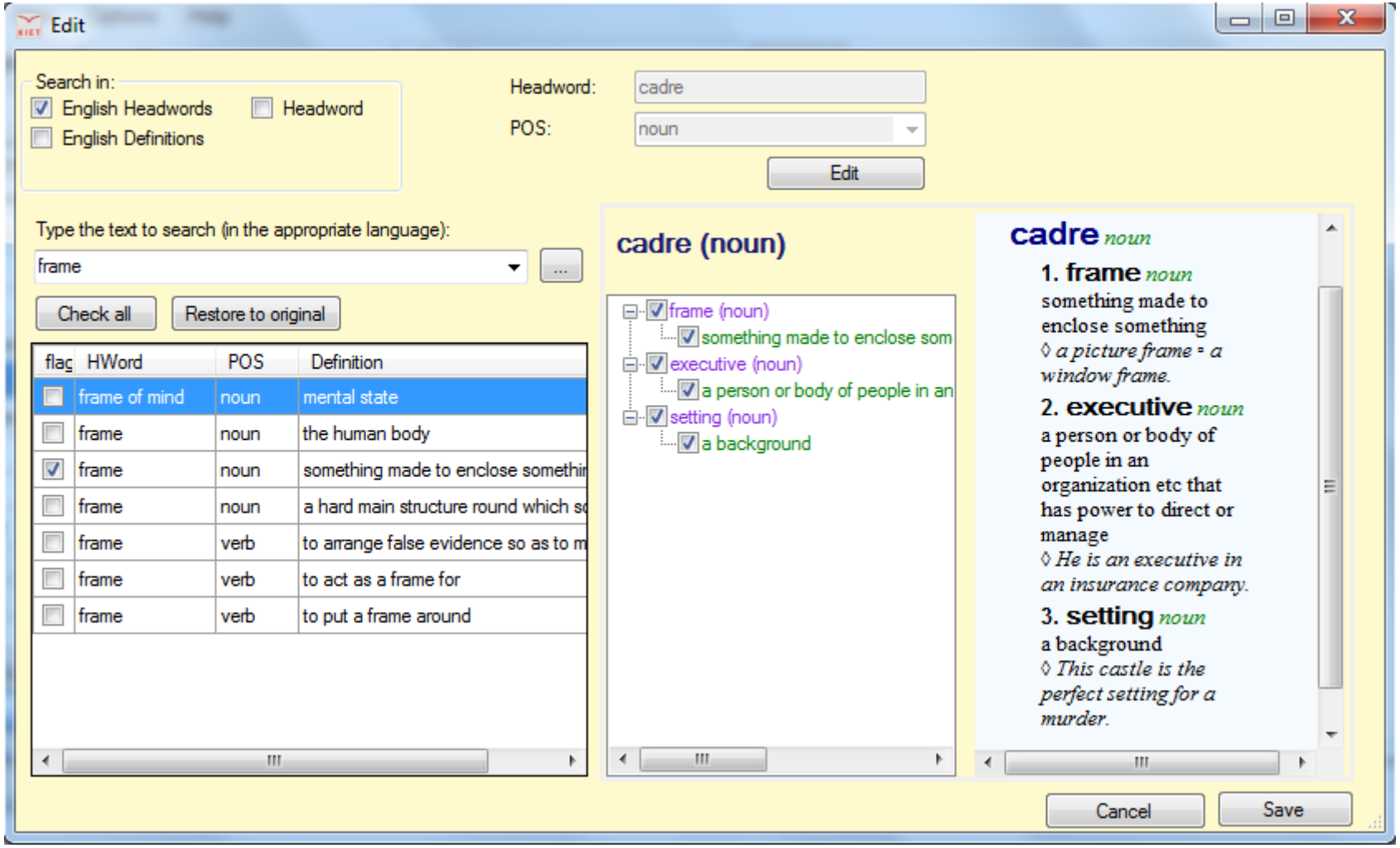
PROCESS

- Generating an L2-English Index automatically
  — produce L2 Index table
  — produce EN Senses table
- Editing the L2 Index
  — include/exclude HW in L2 Index
  — revise the L2 HW and POS
  — add new L2 HW
  — revise the Senses – add, remove, re-order
- Translating multilingually
  — link L2 HW via EN Sense to all the translations

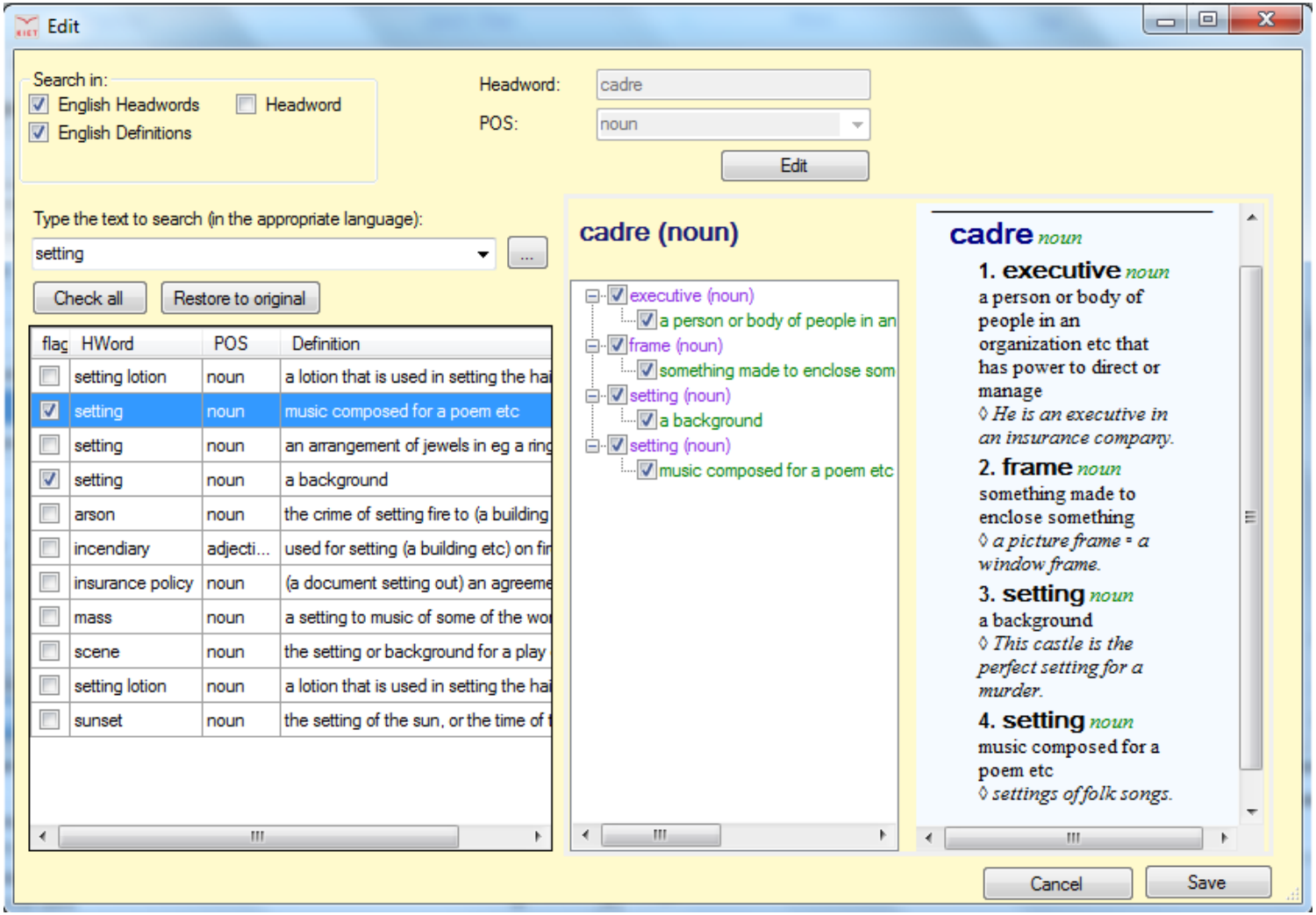
KIET. MAIN SCREEN



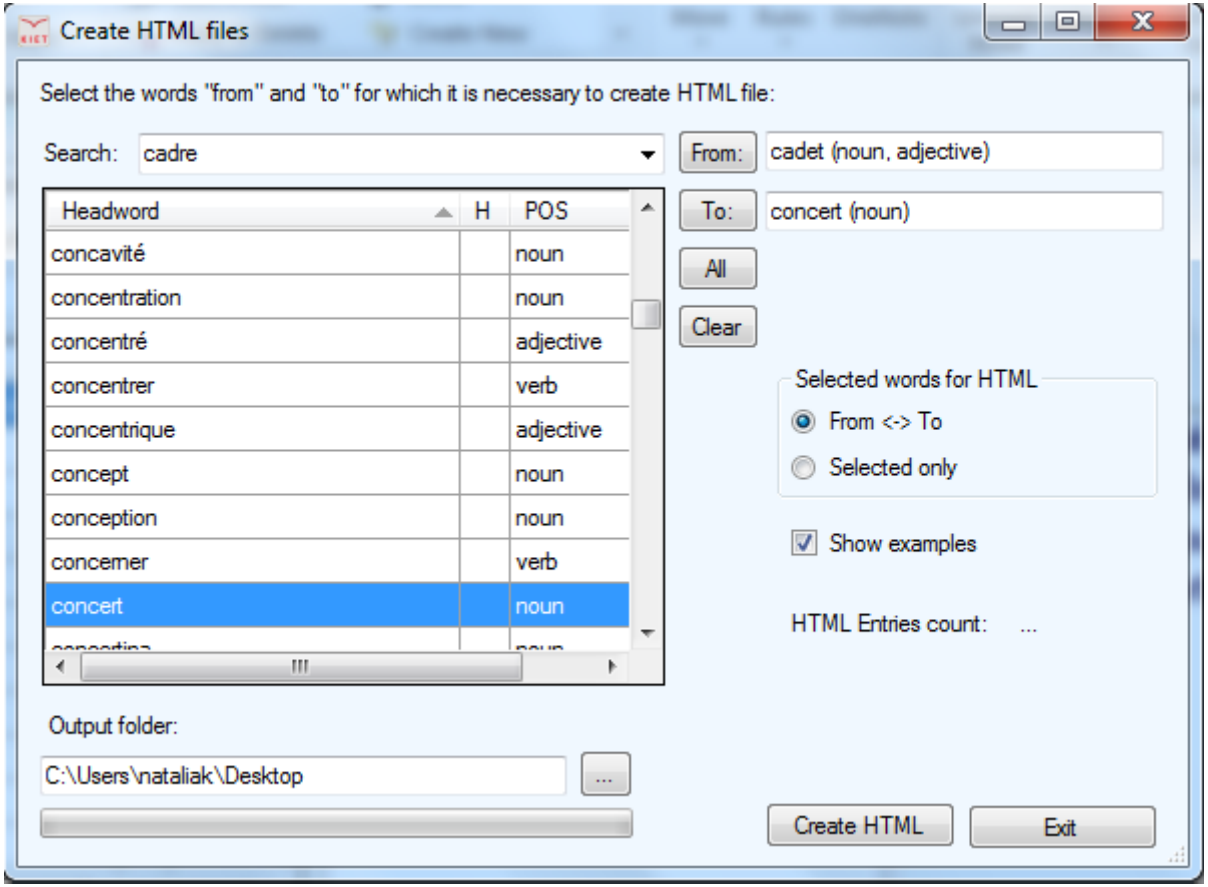
KIET. EDIT L2-ENGLISH INDEX (FRENCH)



KIET. EDIT BY DEFINITION



KIET. EXPORT TO HTML



SAMPLE. GERMAN-ENGLISH INDEX

messen *verb*

1. gauge to measure (something) very accurately
2. measure to find the size, amount etc of (sth)
3. measure to show the size, amount etc of
4. measure (with **against**, **besides** etc) to judge in comparison with
5. measure to be a certain size
6. meter to measure (*especially* electricity etc) by using a meter
7. take to make a note, record etc

SAMPLE. GERMAN MULTILINGUAL (1)

messen *verb*

1. to measure (something) very accurately

af meet | ar ي | bg измервам точно | br medir | ca mesurar, calibrar | cs (z)měřit | dk måle | el (κατα)μετρώ με ακρίβεια | en gauge | es medir, calibrar | et mõõtma | fa با دقت اندازه گیری کردن | fi mitata | fr mesurer, jauger | he לִמְדוֹד | hi प्रमाप, आयाम | hr mjeriti | hu megmér | id mengukur | is mæla | it calcolare | ja 測る | ko 정확히 측정하다 | lt matuoti | lv mērīt | ml mengukur | nl meten | no måle (opp) | pl wymierzyć | pt medir | ro a măsura | ru измерять | sk odmerať | sl izmeriti | sr izmeriti | sv mäta | th วัดด้วยมาตรวัด; เครื่องวัด | tr ölçmek | tw 精確測量 | uk виміряти | ur ناپنا کو زچ ی کس | vi đo | zh 精确测量

SAMPLE. GERMAN MULTILINGUAL (2)

messen *verb*

2. to find the size, amount etc of (something)

af meet | ar ي | bg измервам | br medir | ca mesurar | cs (z)měřit | dk måle | el μετρώ | en measure | es medir | et mõõtma | fa اندازه گیری کردن | fi mitata | fr mesurer | he לִמדוֹד | hi नापना | hr mjeriti | hu (meg)mér | id mengukur | is mæla | it misurare | ja 測る | ko 치수를 재다 | lt (iš)matuoti | lv no | ml mengukur | nl meten | no måle, ta mål av | pl (wy)mierzyć | pt medir | ro a măsura | ru измерять | sk odmerať | sl izmeriti | sr izmeriti | sv mäta | thวัดขนาด (ความยาว, ความสูง, ความเร็ว ฯลฯ) | tr ölçmek | tw 測量 | uk міряти, вимірювати | ur ناپنا، مقدار، حجم وغیرہ معلوم کرنا | vi đo lường | zh 测量

GLOBAL SERIES

- Arabic
- Chinese Simp.
- Chinese Trad.
- Czech
- Danish
- Dutch (2)
- English
- French (2)
- German (2)
- Greek
- Hebrew
- Italian (2)
- Japanese
- Korean
- Latin
- Norwegian
- Polish
- Portuguese Br.
- Portuguese Pt.
- Russian
- Spanish (3)
- Swedish (2)
- Thai
- Turkish

THANK YOU

[θæŋk juː] *interj.* I thank you: *Thank you for your attention!*

Afrikaans **dankie**
Arabic **ش**
Bulgarian **благодаря**
Chinese Simplified **谢谢（你）**
Chinese Traditional **謝謝(你)**
Croatian **hvala**
Czech **děkuji**
Danish **tak**
Dutch **dank je**
Estonian **aitäh, tänan teid**
Farsi **ممنون**
Finnish **kiitos**
French **merci**
German **danke**
Greek **(σε, σας) ευχαριστώ**
Hebrew **תודה**
Hindi **धन्यवाद देने या मना करने का एक**
Hungarian **köszönöm!**
Icelandic **þakka þér**
Indonesian **terima kasih**
Italian **grazie**
Japanese **ありがとう**
Korean **감사합니다**
Latvian **paldies; pateicos**
Lithuanian **ačiū**
Malay **terima kasih**
Norwegian **tusen takk (for)**
Polish **dziękuję**
Portuguese Brazil **obrigado/-da**
Portuguese Portugal **obrigado/-da**
Romanian **mulţumesc**
Russian **благодарю**
Serbian **hvala**
Slovak **ďakujem**
Slovene **hvala**
Spanish **gracias**
Swedish **tack [ska du/ni ha]!, tackar!**
Thai **การแสดงความขอบคุณ**
Turkish **teşekkür ederim**
Ukrainian **дякую; спасибі**
Urdu **آپ کا شکریہ**
Vietnamese **cảm ơn**