

MLODE

Leipzig, 2 September 2014

Generating multilingual lexicographic resources



KDICTIONARIES

Ilan Kernerman
K Dictionaries Ltd, Tel Aviv

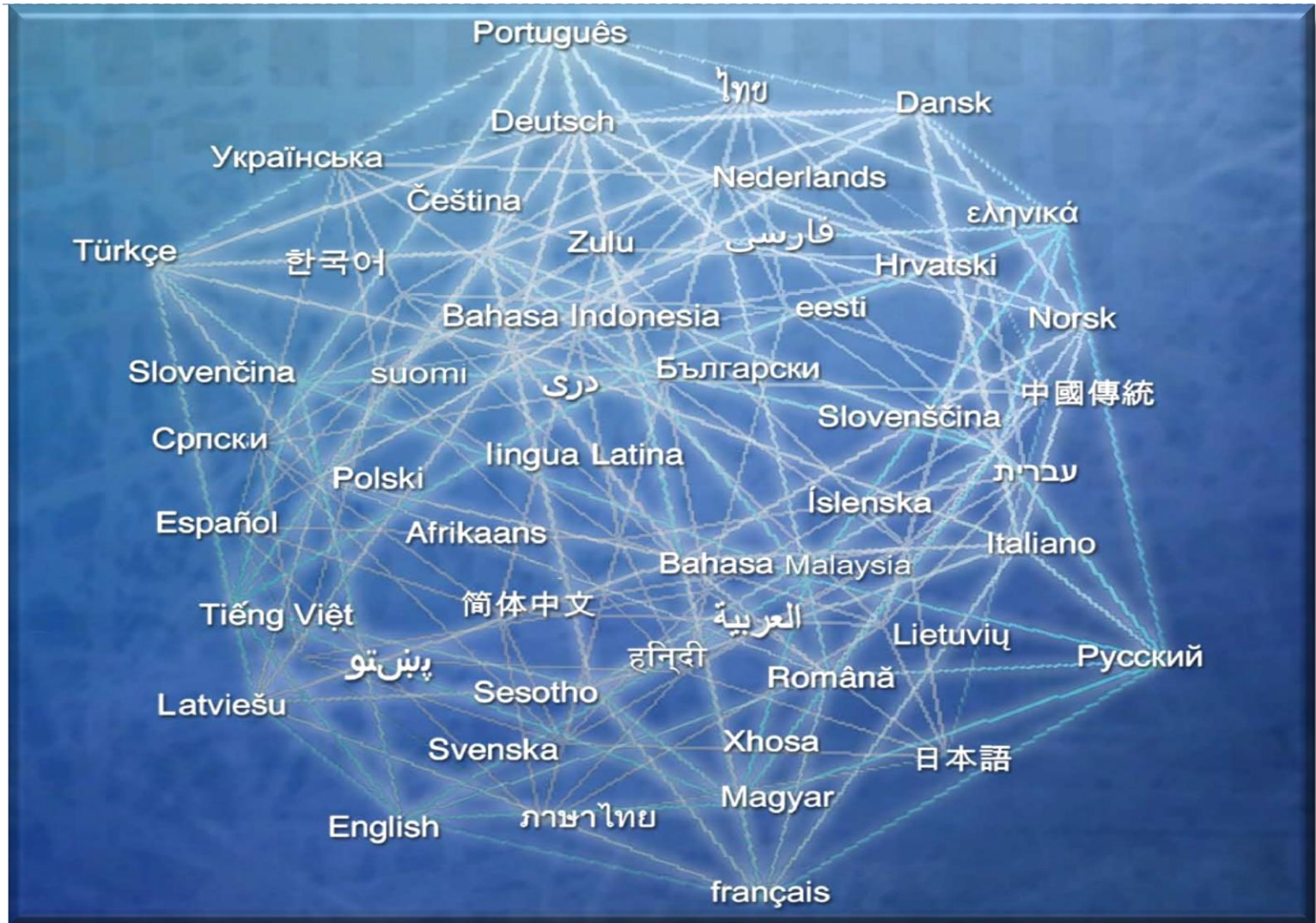
K DICTIONARIES

- ▶ Established in 1993, based in Tel Aviv
- ▶ Focus on *Technology-Driven Content*
- ▶ Create lexicographic data for 40+ languages
- ▶ Cooperate worldwide with
 - language editors, translators and technicians
 - software engineers, architects and designers
 - digital & print publishing partners and LT firms
 - the academe and professional associations

RESOURCES

- ▶ Dictionaries for English learners & native speakers
- ▶ Dictionaries for native & foreign language learning
- ▶ Dictionaries for bilingual & multilingual translation
- ▶ Multi-language/Multi-layer datasets
- ▶ Lexicographic editorial tools & applications
- ▶ Morphology & pronunciation
- ▶ Language supplements, audio & pictures

VISION



MAPPING

AlternativeScripting

AlternativeSpelling

Antonym

CompositionalPhrase

CrossReference

Definition

Example

GeographicalUsage

GrammaticalGender

GrammaticalNumber

HomographNumber

Lemma

Morphology

PartOfSpeech

Pronunciation

RangeOfApplication

Register

SenseIndicator

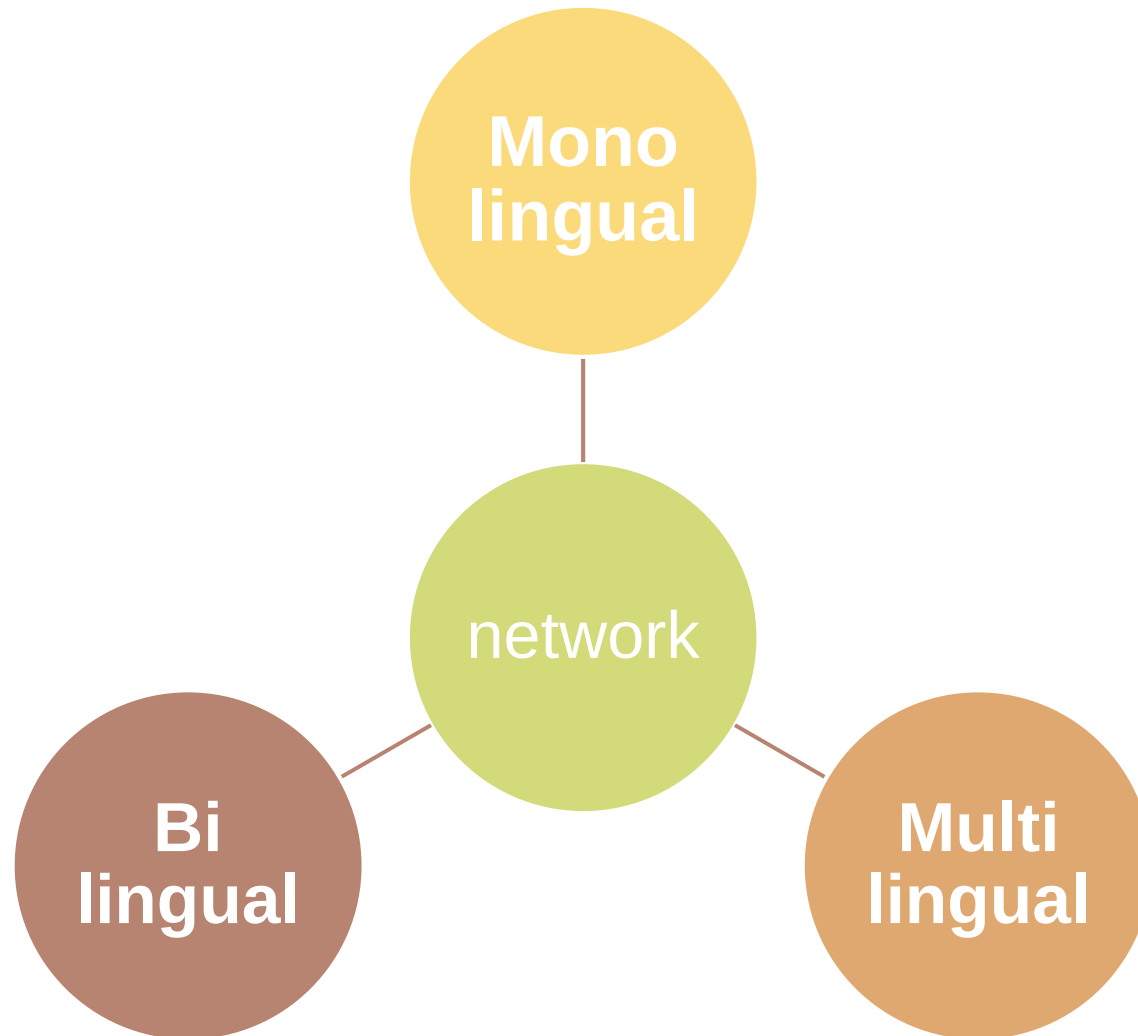
SenseQualifier

SubCategorization

SubjectField

Synonym

MULTI-LAYER



EVOLUTION

1. Monolingual English learner's dictionary
2. Bilingual English learner's dictionary
3. Multilingual English dictionary
4. L2-English reversed indices
5. L2 bilingual glossaries
6. L2, L3 etc. multilingual dictionaries

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

- ▶ **Generating L2-English Index**
 - Produce L2 Index table
 - Produce L2 Senses table
- ▶ **Editing L2 Index**
 - Include/Exclude HW in L2 Index
 - Include/Exclude Sense (checkbox in Tree preview)
 - Edit L2 HW and POS
 - Edit the Entry (modify, add, remove, re-order Senses)
 - Search Sense in English HW or Definition and add it
- ▶ **Translating Multilingually**
 - Link L2 HW via each English Sense to translations in all other languages (of the English multilingual)

SAMPLE. INDEX PREVIEW (FRENCH)

Index preview EX

L2 Index languages: French.

fr - French

Add New Edit Remove Restore Log view Settings Help About

View Checked

distraction Source language entry

flag	HWord	POS
☒	distinguer	verb
☒	distinguer de	verb
☒	distinguer entre	verb
☒	distinguer l'un de l'autre	
☒	distique	noun
☒	distraction	noun
☒	distraire	verb
☒	distrait	adjective
☒	distraitement	adverb
☒	distribuer	verb
☒	distribuer	
☒	distribuer parcimonieu...	verb
☒	distributeur automatique	noun
☒	distributif	adjective
☒	distribution	noun
☒	distributrice	

☒ absent-mindedness(noun)

☒ amusement(noun)

☒ an entertainment or interest

☒ surfing and other holiday amuseme

☒ distraction(noun)

☒ something that takes the mind off othe

☒ There are too many distractions he

☒ diversion(noun)

☒ (an) amusement.

☒ entertainment(noun)

☒ the act of entertaining.

distraction *noun*

-
- {en} - **absent-mindedness** *noun*
- an entertainment or interest
◇ surfing and other holiday amusements.
{en} - **amusement** *noun*
- something that takes the mind off other especially more serious affairs
◇ There are too many distractions here to allow one to work properly.
{en} - **distraction** *noun*
- (an) amusement.
- {en} - **diversion** *noun*

Save L2 Index Update Entry

SAMPLE. INDEX PREVIEW (RUSSIAN)

The screenshot shows the 'Index preview EX' application window. At the top, it says 'L2 Index languages: Russian.' and 'ru - Russian'. Below this is a 'View Checked' button. The main area is titled 'банк' and 'Source language entry'. On the left, there is a table with columns 'flag', 'HWord', and 'POS'. The table lists various Russian words, with 'банк' highlighted. In the center, there is a tree view showing the structure of the index entry for 'банк', including sub-entries like 'a place where money is lent' and 'a place for storing other valuable material'. On the right, there is a detailed view of the 'банк' entry, showing its part of speech as 'noun' and providing English definitions and examples. At the bottom right, there are buttons for 'Save L2 Index' and 'Update Entry'.

flag	HWord	POS
☒	банальный	adjective
☒	банан	noun
☒	банда	noun
☒	банджо	noun
☒	бандит	noun
☒	банк	noun
☒	банка	noun
☒	банкет	noun
☒	банкир	noun
☒	банкнота	noun
☒	банковская карточка	
☒	банковская книжка	

банк

банк noun

a place where money is lent or exchanged, or put for safety and/or to acquire interest
◇ He has plenty of money in the bank - I must go to the bank today.
{en} - bank noun

a place for storing other valuable material
◇ A blood bank.
{en} - bank noun

Save L2 Index Update Entry

SAMPLE. EDIT BY DEFINITION

The screenshot shows a software window titled "Edit". It has a search interface on the left and a list of words on the right. The search interface includes a "Search into English:" section with checkboxes for "Headwords" and "Definitions". Below this is a text input field containing "pastime" and a dropdown arrow. There are also "Check all" and "Restore to original" buttons. The list of words is a table with columns: flag, HWord, POS, and Definition. The word "pastime" is highlighted in blue. To the right of the table is a large text area displaying the definition of "distraction(noun)" and a list of related words with their definitions.

Search into English:

☒ Headwords
☒ Definitions

L2 Headword: distraction
L2 POS: noun

Preview

pastime

Check all Restore to original

flag	HWord	POS	Definition
☐	pastime	noun	an occupation which one enjoys and
☐	pony-trekking	noun	the sport or pastime of riding in the c
☐	professional	adjecti...	earning money by performing, or givin
☐	yachting	noun	the pastime of sailing in a yacht.
☐			

distraction(noun)

- ☒ absent-mindedness(noun)
- ☒ amusement(noun)
 - ☒ an entertainment or interest
- ☒ distraction(noun)
 - ☒ something that takes the mind of
- ☒ diversion(noun)
 - ☒ (an) amusement.
- ☒ entertainment(noun)
 - ☒ the act of entertaining.
- ☒ pastime(noun)
 - ☒ an occupation which one enjoys
 - ☒ Playing chess is his favourite

something that takes the mind off other especially more serious affairs
◇ There are too many distractions here to allow one to work properly.
{en} **distraction** noun

.
(an) amusement.

{en} - **diversion** noun

.
the act of entertaining.

{en} - **entertainment** noun

SAMPLE. SWEDISH MULTILINGUAL (RAW)

bortsprungen 1

runaway *noun*

a person, animal *etc* that runs away

◇ *The police caught the two runaways.*

■ *(also adjective)* ◇ *a runaway horse.*

af wegloper | ar هارب، شارد، جامح | bg беглец | br fugitivo | ca fugitiu |
cs uprchlík/-ice, uprchlý | de der/die Ausreißer(in); durchgebrannt |
dk bortløben | el φυγάδας | es fugitivo | et põgenik | fa فراری |
fi karkuri | fr fugitif/-ive | he בורח | hi अनियंत्रित, उच्छृंखल, बहुत सहज |
hr odbjegao | hu szökevény | id pelarian | it fuggiasco, fuggitivo |
ja 逃亡者 | ko 도망자 | lt pabėgėlis; pabėgęs | lv bēglis; izbēdzis |
ml cabut lari | nl vluchteling | pl zbiec | ps فراری | pt fugitivo |
ro evadat, fugar | ru беглец | sk utečenec/ka; na úteku, ktorý ušiel |
sl ubežnik; pobeg | sr odbegao | th ผู้หลบหนี | tr kaçak, firari |
tw 逃跑的人或動物 | uk утікач; дезертир | ur فرار ہو جانا |
vi kẻ chạy trốn | zh 潜逃者, 逃跑者

SAMPLE. SWEDISH MULTILINGUAL (RAW)

bortsprungen 2

stray adjective

wandering or lost

◇ *stray cats and dogs.*

af weglopend | ar شارد، ضال، تائه | bg изгубен | br perdido |
ca perdut, extraviat, llista de carrers | cs zatoulaný | de streunend |
dk omstrejvende; herreløs | el αδέσποτος | es perdido, extraviado,
callejero | et hulkuv | fa گمشده | fi kuljeksiva | fr errant | he נָטוּר בֵּית |
hi भटका, भूलाभटका | hr zalutao, zabludio | hu elkóborolt | id sesat |
it randagio | ja はぐれた | ko 길잃은 | lt benamis, valkataujantis |
lv noklīdis; klaiņojošs | ml terbiar | nl zwerv- | pl bezdomny |
ps ورک شوی | pt perdido | ro răătăcit | ru бездомный | sk zatúlaný |
sl klateški | sr izgubljen | th ซึ่งพลัดหลง | tr başıboş dolaşan |
tw 漫遊的 | uk бездомний | ur آوارہ یا لاوارث | vi lạc, mất | zh 漫游的

MULTILINGUAL – ANOTHER TYPE

- ▶ <http://kdictionaries-online.com/nlMLDSplus.aspx>
- ▶ <http://kdictionaries-online.com/frMLDS.aspx?Languages=ar,zh,nl,de>

LT APPLICATION

- ▶ Machine translation
- ▶ E-learning
- ▶ Word processing
- ▶ Text-mining, search engines, etc.
- ▶ User & travel guides, menus, etc.
- ▶ Text-To-Speech, STT, etc.
- ▶ Globalization & localization

THANK YOU

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

Afrikaans **dankie**

Arabic أَشْكُرُ، شُكْرًا

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 **mulțumesc**

Russian **благодарю**

Serbian **hvala**

Slovak **d'akujem**

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**