

Data structures in lexicography: from trees to graphs

Michal Měchura @ RASLAN 2016

LINGUISTICS

LEXICOLOGY

classifying and
categorizing
words

discovering the
properties
of words

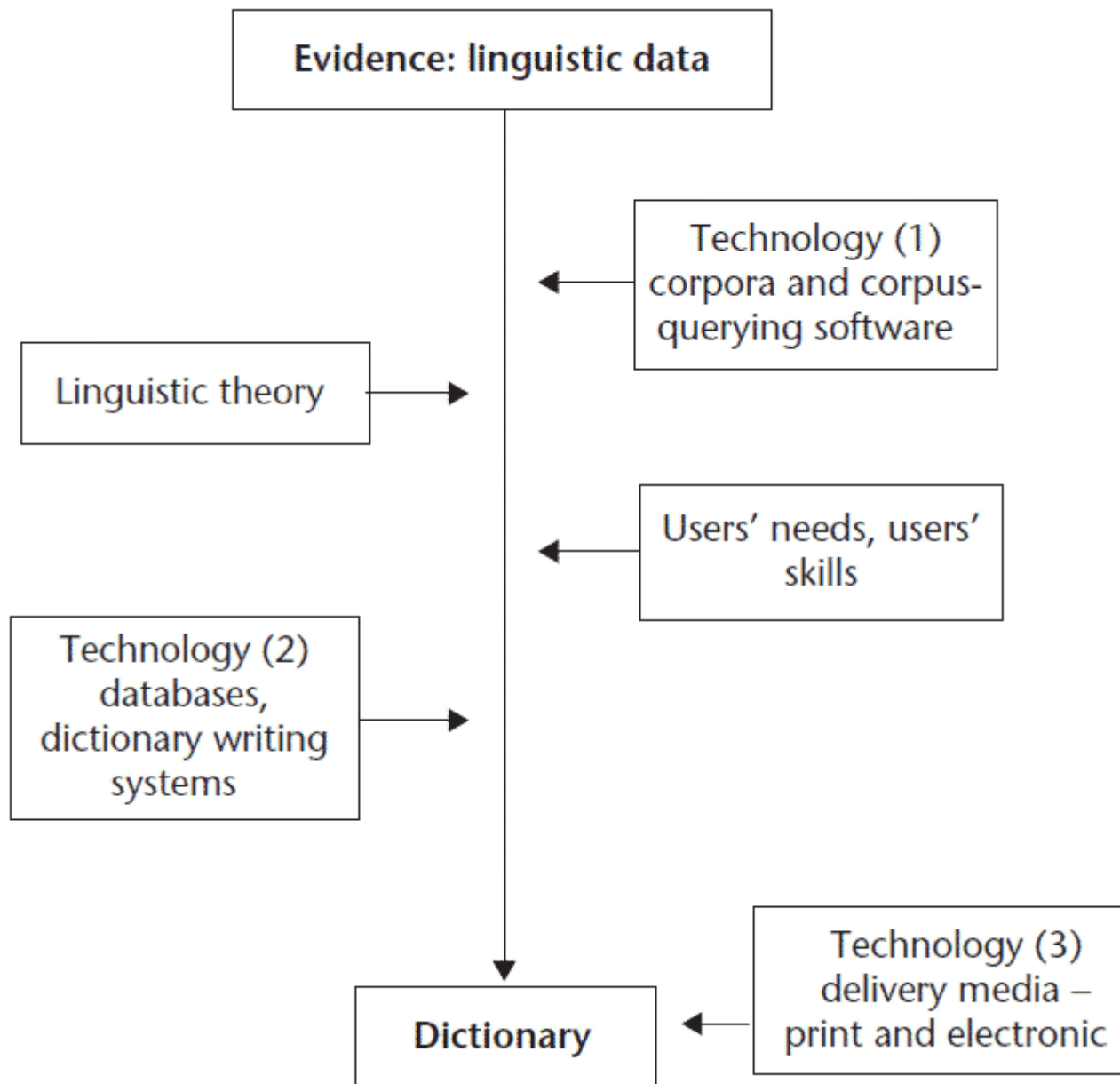
discovering the
structure of the
mental lexicon

communicating
the discoveries
to humans

communicating
the discoveries
to computers

LEXICOGRAPHY
(human-oriented)

LEXICOGRAPHY
(machine-understandable)



My Test Dictionary

www.lexonomy.eu/zo6bowdab/en/edit/entry/?id=2

My Test Dictionary Edit Configure Download Upload modhuine@dcu.ie

cancel save delete 2

starts like this

new entries: 3

able

arm


ask

```
<entry>
  <headword>ask</headword>
  <partsOfSpeech>v</partsOfSpeech>
  <sense>
    <definition>To look for an answer to a question.</definition>
    <example>I need to <h>ask</h> you a question.</example>
    <example>I don't know, <h>ask</h> your mother.</example>
  </sense>
  <sense>
    <definition>To talk to someone to see if they will do something
      for you.</definition>
    <example>She <h>asked</h> me to help her.</example>
    <example>We will have to <h>ask</h> for more money.</example>
  </sense>
</entry>
```

www.lexonomy.eu

Problem 1:
Placement of multi-word items

```
□ <entry>
  □ :: <headword>time</headword>
  □ :: <partOfSpeech>n</partOfSpeech>
  □ :: <sense>
    □ :: <translation>čas</translation>
    □ :: <translation>doba</translation>
    □ :: <phraseme>
      □ :: <source>third time lucky</source>
      □ :: <target>do třetice všeho dobrého</target>
    </phraseme>
  </sense>
</entry>
```

 This subentry also appears under third and lucky.

Problem 2:
Bilingual dictionary reversal

```
[-] <entry>
  [-] :: <headword>walk</headword>
  [-] :: <partOfSpeech>n</partOfSpeech>
  [-] :: <sense>
    [-] :: <translation>procházka</translation>
  </sense>
</entry>
```

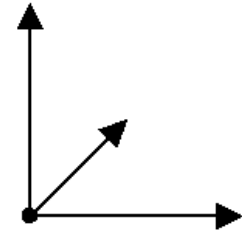
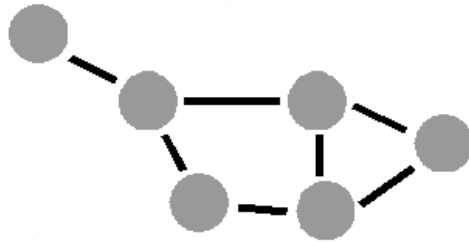
 *Walk has recently been added as a translation to vycházka in a paired dictionary.*

Do you want to add vycházka as a translation here?

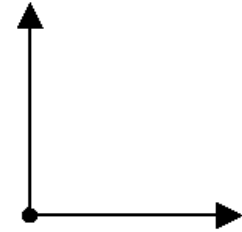
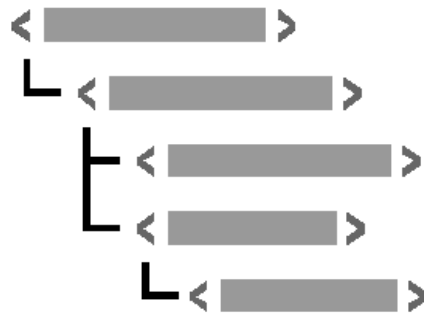
Yes

No

3 graphs



2 trees



1 formatted text

