Compilation Techniques for Pedagogically Effective Bilingual Learners’ Dictionaries

presented by
Jack Halpern, CEO
The CJK Dictionary Institute
日中韓辞典研究所

ASIALEX 2016 PHILIPPINES
The 10th International Conference of the Asian Association of Lexicography
June 1-3, 2016

Shortcomings of existing dictionaries

- Archaic headwords and senses
- Inappropriate examples
- Overly prescriptive approach
- Learner-unfriendly sense ordering
- Omission of important multiword expressions

Pedagogically effective dictionaries by CJKI

- The CJKI Chinese Learner’s Dictionary
- The Kodansha Kanji Learner’s Dictionary
- The CJK Arabic-English Learner’s Dictionary

Summary

- Background and aims
- Presentation of senses and logico-semantic ordering
- Semantic interrelatedness and core meanings
- Semantic transparency and morphological productivity
- Lexical categories in Chinese
- Arabic headword selection criteria
- Multiword expressions in Arabic
- Pedagogical selection criteria
- Interactive Parallel Text – a new paradigm
- Future Work
Ordering of senses

- Chronological
- Frequency-based
- Logico-semantic

Chronological ordering

A Learner’s Chinese Dictionary

The Oxford Chinese Dictionary

Frequency ordering

Entry for 本 běn in A Chinese-English Dictionary; a Basic Dictionary for Chinese Language Learning (Foreign Languages Press, 2007)

本 běn
1. <名 n.> booklet
2. (～儿) <名 n.> capital; principal
3. <名 n.> root or stem of a plant; generally referring to the origin of the matter
4. <名 n.> version; edition and book binding
5. <量 meas.> used to indicate the copy of books
6. <代 pron.> one’s own; native

Logico-semantic ordering

▲ bound morpheme  ◭ bound or free morpheme △ free morpheme

- bound morpheme
- bound or free morpheme
- free morpheme
Interrelatedness of senses

List A
1 a KEEP in place, ask to stay
b KEEP for future use, reserve, leave behind
c KEEP in custody, detain
d KEEP one’s mind on, concentrate on

List B
1 keep in place, ask to stay
2 reserve, leave behind
3 detain
4 concentrate on

Semantic transparency in Japanese

<table>
<thead>
<tr>
<th>English</th>
<th>Etymology</th>
<th>Kanji</th>
<th>Roman</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>hydrophobia</td>
<td>Greek</td>
<td>恐水病</td>
<td>Kyōsuibyō</td>
<td>fear+water+illness</td>
</tr>
<tr>
<td>aquarium</td>
<td>Latin</td>
<td>水族館</td>
<td>Suizokukan</td>
<td>water+family+building</td>
</tr>
<tr>
<td>water wheel</td>
<td>Anglo-Saxon</td>
<td>水車</td>
<td>Mizugeruma</td>
<td>water+wheel</td>
</tr>
</tbody>
</table>

Transparency and productivity

- Lack of POS codes
- Limited POS codes
- Incomplete POS codes
POS codes for 留 in CCLD

- remain in a given condition: 留 at home
- stay abroad: 留

Character Meanings

- The Equivalent
- Core Meaning
- Interrelatedness of Senses

CCLD Entry for 留

KKLD entry for 留

- stay abroad
- remain in a given state or condition
- keep some meaning, stay in position
- stay in custody, detain
- stay for future use, leave behind
- stay abroad
- remain in a given state or condition
- keep some meaning, stay in position
- stay in custody, detain
- stay for future use, leave behind
Traditional criteria for selecting Arabic headwords

- Quran and classical literature
- Copying from past dictionaries
- Analyzing authentic texts found in corpora
- Subjective judgment of lexicographer
- ‘Importance’ based on frequency and lexicographer’s discretion

Multiword expressions in Arabic

- Omission of multiword expressions
  (الشرق الأوسط 'assharqu l'awSaTu 'the Middle East')
- Omission of derived words
  (پسرغة bisur'atin ‘quickly’)
- Inclusion of infrequent words
  (خمد Hamd 'praise')
- Exclusion of common words
  (مسامحة musaamaHa ‘forgiveness’)

Pedagogical selection criteria

- Potential headwords drawn from corpora, textbooks, learners’ dictionaries, websites, and other sources
- Special efforts made to include modern words such as blog and ATM
- Rare/archaic words omitted by relying on modern corpora and sources
- Dialectal words omitted regardless of their corpus frequency
- MWEs systematically included as main entries or subentries rather than as examples
- Important derived words, especially adverbs, systematically included
- Headword/senses culled by lexicographers to maximize relevance

Future Work

- Lexical categories
- Headword selection
- Multiword expressions
- Pedagogically effective tools
Learner's Interactive Bitext
Electronic Reading Application

Libera

Interactive Parallel Text (IPT)
- Allows learners to effortlessly read vast amounts of L2 texts
- New medium for accessing comprehensible L2 texts
- Libera is cutting edge platform for reading enjoyably
- Texts precisely linked on segment level
- Interactive links highlight L2 segment and activate glosses, grammar notes, audio, and transcriptions

Four-Panel Sentence Mode
- Texts precisely linked at segment (word or phrase) level
- Tapping a segment in any one text simultaneously highlights linked segment in all other texts

Thank You
Salamat po

The CJK Dictionary Institute
Nii-za-shi, Saitama, Japan
www.cjk.org