Electronic Dictionaries
Points to discuss

• History
• Typology
• Description
Computer versions of conventional dictionaries

- various systems of menu
- suitable structure of material arrangement
- various searching devices, accelerating information retrieval
- storage of information
Advantage of computerized dictionaries

One may:

• transfer from one type of information/knowledge to another one quite easily

• come up with a quick and detailed answer to one’s questions
Advantage of computerized dictionaries

One may:

• obtain mostly adequate data without forcing oneself and without browsing a lot of reference books over a long period of time

• gain new possibilities granted by information storage
1960

Webster’s Seventh New Collegiate Dictionary

was keypunched onto paper tape from paper text,
for the purpose of computational exploration
|     | NULL | SOH  | STX  | ETX  | EOT  | ENQ  | ACK  | BEL  | BS   | HT   | LF   | VT   | FF   | CR   | SI   | DLE  | DC1  | DC2  | DC3  | DC4  | NAK  | SYN  | ETB  | CAN  | ESC  | GS   | RS   | US   |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 2   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 3   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 4   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 5   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 6   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 7   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |
| 8   |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

AMERICAN NATIONAL STANDARD CODE FOR INFORMATION INTERCHANGE (ASCII) X3.4-1968
1960

• a by-product of publishing paper dictionaries
• the overall task: the modification and conversion of a flat character stream into a structured format, typically a database accessible by machine
1970-1980

Longman Dictionary of Contemporary English

the most widely used MRD (machine-readable dictionary) in NLP (natural language processing)
late-1980s

The first human readable electronic dictionaries available for the public at large:

• online (whether or not by subscription)
• on CD-ROM and other disks
• in handheld devices
Nesi’s definition for ‘electronic dictionary’

“The term electronic dictionary (or ED) can be used to refer to any reference material stored in electronic form that gives information about spelling, meaning, or use of words."
Nesi’s definition for ‘electronic dictionary’

- a spell-checker in a word-processing program
- a device that scans and translates printed words
- a glossary for on-line teaching materials
- an electronic version of a respected hard-copy dictionary are all EDs of a sort’
Nesi’s definition for ‘electronic dictionary’

All these objects are collections of structured electronic data that can be

• accessed with multiple tools
• enhanced with a wide range of functionalities
• and used in various environments
The most innovative aspect of EDs

The retrieval system
Without

- doing something about the contents
- implementing fully integrated hypermedia access structures
The mid-1990s

- The boom in scholarly discussions and the large-scale commercial production of EDs
- The topic of EDs, in addition to being discussed in journals dealing with computer science, language learning and language teaching, made its appearance in (meta)lexicographic publications
Electronic-Dictionary Typologies

- Martin’s typology
- Ide’s typology
- Sharpe’s typology
- Lehr’s typology
- Schryver’s typology
Martin’s typology

- dictionaries for human users
- computer-based dictionaries
- machine-readable dictionaries
- lexical/term banks
Martin’s typology

- machine dictionaries
- lexical databases
- artificial intelligence lexicons
Martin’s typology

No attempt is made to differentiate between computational and non-computational ‘objects’
Ide’s typology

• specific EBDs
• electronic notebooks
• CD-ROM EBDs
• ED software
Ide’s typology

A typology solely based on hard- and software does not seem to be very efficient, since it requires constant readjustments to cater for the never-ending innovations.
Sharpe’s typology

• floppy-disk based portable EBDs
• EBDs with hand-held optical character recognition (OCR) scanners
Lehr’s two-step technical-(meta)lexicographic ED typology

- In a first step, EDs are classified on technical grounds
- In a second step, each of these EDs can then be evaluated on (meta)lexicographic grounds
Nesi’s typology

Four types of Eds for language learning
<table>
<thead>
<tr>
<th>#</th>
<th>Type?</th>
<th>Source?</th>
<th>Who?</th>
<th>Profit?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Internet dictionary</td>
<td>(outdated)</td>
<td>[Netizens]</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td>copyright-free</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>material &amp; users’ contributions</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>glossary for online</td>
<td>[new material]</td>
<td>language-department staff</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>courseware</td>
<td></td>
<td>members at universities</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>learners’ dictionary</td>
<td>reputable hardcopy reference books</td>
<td>major dictionary</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>on CD-ROM</td>
<td></td>
<td>publishing houses</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>PED</td>
<td>[no-named (hardcopy) source]</td>
<td>makers of electrical goods</td>
<td>✓</td>
</tr>
</tbody>
</table>
Schryver’s three-step typology

• 1 step: “WHO accesses the dictionary?” – machines/humans

• 2 step: “WHAT is accessed?” – dictionary medium (physical-object/electronic medium). Both media can be subdivided into handheld devices/robust machines

• 3 step: “WHERE does one access the dictionary data?” – the type of storage
Paper Dictionaries versus Electronic Dictionaries

‘It is my view that the advantages of the electronic dictionary and the familiarity of today’s young people with electronic devices will eventually relegate the printed notion of “dictionary” to a secondary sense.’ (Sharpe)
# Today’s average paper and electronic-dictionary reality

<table>
<thead>
<tr>
<th>Traditional paper dictionary [PAPER]</th>
<th>Handheld dictionary (e.g. PED) [PED]</th>
<th>Robust-machine dictionary (e.g. CD-ROM) [CD]</th>
<th>Intranet dictionary [INTRA]</th>
<th>Internet dictionary [INTER]</th>
</tr>
</thead>
<tbody>
<tr>
<td>one user uses a handheld book to access a D stored on paper</td>
<td>one user uses a palmtop to access a D stored on a small disk</td>
<td>one user uses a laptop/desktop to access a D stored on a large disk</td>
<td>a group of users use laptops/desktops to access a D stored on a local mainframe</td>
<td>users worldwide use laptops/desktops to access a D stored on an online server</td>
</tr>
</tbody>
</table>
Paper Dictionary

1. familiar, reassuring
2. symbolic value as a physical object that can be owned and admired
3. easy to browse, can most readily be read recreationally, easy global reading, manhandling is part of the reading pleasure
4. easy to read, best for the eyes
Paper Dictionary

5. easy to annotate, one can physically write on them
6. durable, can be carried around the world without fear of serious damage or loss of functionality
7. has a solid independent existence (i.e. is not plugged into anything and nothing is plugged into it)
8. does not require a computer to be switched on
Handheld

1. can be used straight after acquisition
2. can be used anywhere, not site-dependent
Handheld

3. portable, small size, low weight
4. offline access
5. gadgets/status symbols
Local Disk

1. availability of audible pronunciation (no need to learn IPA) and sound files
2. take very little space and have an extremely low weight
3. record-yourself facility (i.e. comparison of one’s own pronunciation with the stored one(s))
Local Disk

4. speedy access to D data
5. alternative ways of presenting information (computer graphics)
6. fuzzy search facility, multi-access, range of look-up routes, hyper reference
Local Disk

7. (large-scale)user customisation; interactivity
8. highly user-friendly interface
9. copy-and-paste facility
Electronic

1. charm of novelty
2. paper-less office, ecologically sound
3. users are liberated from alphabetical order
Electronic

4. elimination of linear text restrictions; not everything needs to be written/visual and constant
5. optimisation of cross-referencing
Online

1. no space constraints other than the need to avoid swamping the user, huge quantities of data
2. rapid access to large amounts of lexicographical evidence in corpora
3. new types of information
Online

4. video sequences, animation
5. links with (other) software (e.g. built-in teaching/learning, games, etc.)
6. up-to-date; dynamic repository of knowledge
Online

7. one-stop consultation (i.e. simultaneous searches of hundreds of native speaker and bilingual Ds)
8. consultation cost based solely on the actual use
9. cheap if not free