
Human-machine interaction in the lexicographic process for digital lexical information system

Konan Kouassi (FAU)

konan.kouassi@fau.de

Euralex 2022: 12-16 july

Content

- 1. PofiLex research project**
- 2. Dictionary examples for lexical information systems: *Creation perspective***
- 3. Analysis of dictionary example items: *Reverse perspective***

I. About the ProfiLex Research project

Project name: The Professional Lexicographer
= ProfiLex Project

Project type: Metalexicographic and scientific project, funded
by the German Research Foundation (DFG)

Project place: Interdisciplinary Centre for Lexicography, Valence
and Collocation Research at FAU Erlangen-
Nürnberg, URL: <http://www.lexi.uni-erlangen.de/de>

Core of project: Functions and tasks of the professional
lexicographer as distinct from the automatic
procedures in the modern lexicographic
process for lexical information systems

General considerations:

1- Dictionaries of today and tomorrow are rather digital products than print dictionaries.

2- User's perspective: electronic dictionary applications and especially "digital word information systems" are coming to the fore alongside Google searches.

3- Today's and tomorrow's dictionaries are computer-guided or automatically generated.

4- the automatically generated items can be corpus-based well or poorly chosen or formulated (cf. Kirkness 2016, Mollica 2017, Bielińska/Schierholz 2017, Schierholz 2019).

5- the new role of the lexicographer in the modern lexicographic process is questionable.

Practical case: Systematical interaction between technology and human action: creating and presenting example items

**Reason
for
choice:**

Example items are, among other things, areas in which this interaction can be described and demonstrated very precisely

**Relevant
questions**

1. Meeting point between technology and lexicographer in the creation of example items
2. Features characterising both the technical and the intellectual working process
3. The meaning of a fully automatic output for examples

**II. How dictionary examples for lexical information systems are created:
Creation perspective**

Starting point  **Complementarity between technology and lexicographer**

The automatic part of the creation of dictionary examples items

1. Linguistic data preparation for the dictionary base
2. Technically labelled example formats in terms of corpus example
3. Computer equipment
4. Collect, arrange, sort data
5. Automatic analysis and language correction: spelling, grammar
6. Data presentation

Facts about automatic processes

- 1. The specific lexicographic activities today rely on technical and technological mechanisms**
- 2. Automatic work processes are based on word forms**
- 3. Dynamic and unstable dictionary basis**
- 4. Automatic creation of dictionary examples quite possible, but uncontrollable**
- 5. Human cannot do automatic tasks**

The intellectual effort of the lexicographer

1. Definition and application of the example concept

2. Adequate use of the technical infrastructure

3. Cognitively guided recognition of dictionary subject-related relevant evidence or examples

4. Intellectual consideration of dictionary function and potential users addressed

5. Intellect-based exploration of the semantic spectrum of the lemma

⊘ Points **1** to **5** are considered as preliminary steps for the concrete creation of example items

6. Choice of lexical units:
conversion/transformation of the preliminary steps to concrete example items

7. Intellect-based evaluation of example quality

Facts about the lexicographer input

- 1. Historical view: dictionary examples are not automatical created or provided**
- 2. For general monolingual dictionaries, examples are obligatory items**
- 3. The efforts of the lexicographer cannot be done by computer technologies in proper way**
- 4. Computer tasks cannot be done by human, although possible in some cases with 'but'**
- 5. The lexicographer is responsible for the quality of the example items, not the computer or its technologies**
- 6. The intellectual input accompanies the whole example creation process: conceptions, implementations and critical observations of the automatic example creation process**

III. Reverse perspective

Backtracking of the creation of
example entries



Goal: observe the respective
contribution of lexicographer
and computer technology



different/metalexigraphic
perspective

Example items
produced by
lexicographer
(Duden)



Automatically
produced example
items (DWDS)

Editorially produced EI (Duden)

GEDANKE 1. a) etwas, was gedacht wird, gedacht worden ist; Überlegung

BEISPIELE

- gute, vernünftige Gedanken
 - dieser Gedanke liegt mir fern, verfolgt mich, tröstet mich
 - ein Gedanke ging mir durch den Kopf
 - mir drängt sich der Gedanke auf, dass das nicht stimmt
 - einen Gedanken fassen, aufgreifen, fallen lassen, in Worte kleiden, zu Ende denken, nicht mehr loswerden
 - Gedanken an jemanden, etwas verschwenden
 - auf einen Gedanken kommen, verfallen
 - es ist mir ein schrecklicher Gedanke (*eine schreckliche Vorstellung*), dass du verärgert bist
 - seine Gedanken sammeln (*sich konzentrieren*)
 - seinen Gedanken nachhängen, sich seinen Gedanken überlassen (*[nach]sinnen*)
- (...)

Eng**THOUGHT**: something that is thought,
has been thought; reflection

 Reduced exemple group

→ GEDANKE in Duden online: <https://www.duden.de/rechtschreibung/Gedanke>

Editorially produced EI (Duden)

Collocations (i.e. *gute, vernünftige Gedanken*)

Full sentence example (i.e. *ein Gedanke ging mir durch den Kopf*)

Single example groups (i.e. *dieser Gedanke liegt mir fern, verfolgt mich, tröstet mich*)

Infinitive constructions (i.e. *auf einen Gedanken kommen*)

With pragmatic marks (i.e. „seine Gedanken sammeln (*sich konzentrieren*)“)

Fig. shows a quantitatively hybrid example complex, which can be attributed to the free decision of the lexicographer

Possible lexicographer functions:

- ❖ the decision function of the lexicographer
- ❖ the interpretation function for semantic analysis
- ❖ The redaction function
- ❖ The validation function for attesting the correctness of data

The creation of usage examples alone involves the use of several functions of the lexicographer

Automatically produced
usage example (DWDS)



Homographs

VERBAND¹ = **eng** BANDAGE

VERBAND² = **eng** CONNECTION /
ASSOCIATION / UNION

- *Im vergangenen Jahr schätzte der Verband ihre Zahl auf 284.000.*
- *Unterstützt werden die Verbände von etwa 2 Millionen ehrenamtlichen Helfern.*
- *Die verschiedenen Verbände gehen sich aber meistens aus dem Wege.*
- *Mittlerweile hatte der Allgemeine Verband eine andere Taktik für zweckmäßig erachtet.*
- *Als sie den Verband angelegt hatte, griff er nach ihrer Hand und küsste sie dankbar.*

**automatically produced
usage example (DWDS)**

**About the
usage
examples |**

1. They are automatically extracted

2. No consideration of lemma polysemy

3. EI Generated for both homographs and
are not assigned to VERBAND¹ & VERBAND²

4. The last EI belongs to VERBAND¹ and the
others to VERBAND²

5. The users must carry out the interpretation assignment
themselves

6. Users may be irritated by the automatically
generated data or may not even notice the errors
and take the wrong information from the data

automatically produced usage example (DWDS)

About the usage examples II

7. Automatic language data analysis and machine provision of lexicographic data are limited to word form
8. Automated work processing is managed by useful tools, but also show how quickly the extraction software reaches its limits
9. The analyses of further dictionary articles on homographs provide comparable results
10. The input of the professional lexicographer is missing
11. The automatic provision of example entries; quantitatively rich content, but poor in quality

In conclusion

- ❖ Example items can be used to illuminate the interaction between humans and computer technology
- ❖ Both components interact in the creation of example items or usage examples respectively
- ❖ Human and machine meet
 - ✓ where the completion of a certain lexicographic task is impossible or too difficult and tedious for humans
 - ✓ where the task cannot be reliably completed by the computer alone
- ❖ Automatic procedures are useful, especially for the lexicographical data of the form commentary
- ❖ A secure automatic data provision requires for semantic aspects both components as obligatory and in a complementary relationship

Thank you!

ProfiLex