
UCL

**Université
catholique
de Louvain**



**Has lexicography reaped the full benefit of the
(learner) corpus revolution?**

Sylviane Granger

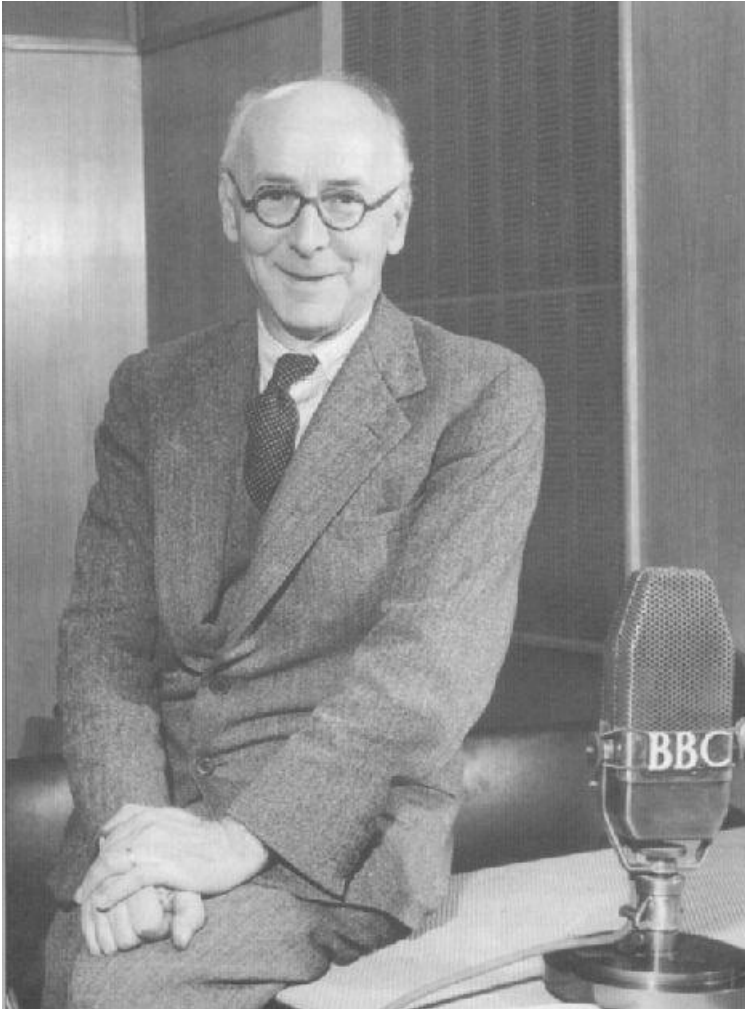
The corpus revolution

- ▶ Rundell & Stock (1992): “The corpus revolution”
 - ▶ *“The past decade has witnessed a revolution in the nature of the linguistic data available to writers of dictionaries, and this has led to the questioning – and in some cases the abandonment – of working practices established over 200 years ago”*
- ▶ Rundell (2008): “The corpus revolution revisited”
 - ▶ A second corpus revolution brought about by the arrival of the web and powerful lexical profiling software tools such as the Sketch Engine which allow lexicographers to cope with much larger corpora

Monolingual native-language corpora

- ▶ Large collections of texts produced by native speakers
- ▶ Played a key role in the development of monolingual learners' dictionaries (MLDs)
- ▶ Considerable enrichment of the information on syntactic, lexical and idiomatic usage that learners need in order to produce language
- ▶ Phraseology moved to centre stage: “*Words tend to flock together*” (Williams & Millon 2011).

Phraseology: Palmer and Hornby



- The pioneering work of Palmer and Hornby paved the way for a more phrasal approach to lexis.
- *“A knowledge of how to put words together is as important as, perhaps more important than, a knowledge of their meanings.”* (Hornby 1954)

Phraseology: Sinclair



- It is the advent of the computer corpus that has allowed the phrasal approach to lexis truly to materialise.
- Sinclair's (1991) seminal book "Corpus, Concordance, Collocation"
 - "Idiom principle"
 - Contextual approach to meaning: "*phraseology is often discriminatory in terms of sense*"

The learner corpus revolution

- ▶ “The learner corpus: A revolution in applied linguistics” (Granger 1994)
 - ▶ Learner corpora contain samples of written or spoken language produced by foreign- or second-language learners
 - ▶ Ideal resource to study learners’ use of language, identify their difficulties (errors but also over- and underuse) and, on that basis, inform pedagogical tools, in particular learners’ dictionaries
 - ▶ Most learner corpus research focuses on the more advanced stages of acquisition
- Particularly relevant for my lecture in Hornby’s honour as he was the originator of dictionaries for advanced learners

Learner corpora: targeted languages

- ▶ At first: only L2 English; now many more L2 varieties
- ▶ “Learner corpora around the world”: 141 learner corpora
 - ▶ c. 60 %: learner English
 - ▶ c. 40 %: other learner varieties: Arabic, Chinese, Czech, Dutch, Finnish, French, German, Hungarian, Italian, Korean, Norwegian, Slovene, Spanish, Swedish

<https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>

Learner corpora and lexicography

- ▶ The very first concrete application of learner corpus data is lexicography and, more particularly, MLDs
- ▶ Maingay & Rundell (1987): “Anticipating learners’ errors – Implications for dictionary writers”
 - ▶ Highlighted the benefit gained from analysing learners’ errors so as to tailor dictionary entries to learners’ attested needs
- ▶ Learner-corpus-based error notes in MLDs
 - ▶ LDOCE: Longman Learners’ Corpus
 - ▶ CALD: Cambridge Learner Corpus
 - ▶ MED: International Corpus of Learner English (ICLE) (Granger et al. 2009)

Get it right box in MED

Get It Right!: attend

When **attend** means 'to be present at an event or activity', it is not used with the prepositions **to** or **at**. It is simply followed by a direct object:

X You go to university, ~~attend to classes~~, but learn nothing about the real world.

✓ You go to university, attend classes, but learn nothing about the real world.

X They are able to ~~attend at important meetings~~.

✓ They are able to attend important meetings.

In the same way, when **attend** means 'to go regularly to school, church etc', use a direct object (not attend to or attend at):

X In some parts of our country, people still don't let girls ~~attend to school~~.

✓ In some parts of our country, people still don't let girls attend school!

X ...a peak in the number of students ~~attending at universities~~ in Sweden.

✓ ...a peak in the number of students attending universities in Sweden.

Don't confuse these meanings with the phrasal verb **attend to something**, which means 'to deal with something':

My assistant will attend to all your travel arrangements.

What about bilingual dictionaries?

- ▶ Have corpora been used? If so, what type of corpus and for what part of the dictionary-making process?
- ▶ First observation: very little information on corpus use in bilingual lexicography
- ▶ Theoretically, the most obvious type of corpus to use in bilingual lexicography is the translation corpus
 - ▶ *“such corpora have not been exploited significantly in [bilingual] dictionary compilation – unlike monolingual lexicography, where it would be unthinkable today not to use single-language corpora”* (Salkie 2008).

Why are translation corpora underused in bilingual lexicography?

- ▶ *“Lexicographers show little interest in the field of translation, at least as a source for dictionary building” (Calvo Rigual & Calvi 2014)*
- ▶ Different types of equivalence: context-dependent equivalents for translation; less context-dependent/more general equivalents for lexicography
- ▶ Lack of large, representative, readily available translation corpora
- ▶ Weaknesses of many current translation corpora:
 - ▶ Not always available
 - ▶ Limited range of registers (e.g. EuroParl: only parliamentary debates)
 - ▶ Translation direction often unknown: impossible to separate original texts from translated texts – major weakness in view of the presence of ‘translationese’, even in texts translated by professional translators
- ▶ Lack of user-friendly software tools to facilitate the processing of translation corpora by lexicographers

The key role played by monolingual corpora

- ▶ The corpora that are mainly used in bilingual lexicography are **monolingual corpora**, of the source and target languages
 - ▶ English-Irish Dictionary (Ó Mianáin & Convery 2014)
 - ▶ Monolingual corpus of English (1.7 bn words) → DANTE lexical database
 - ▶ Monolingual corpus of Irish (New Corpus of Ireland)
- ▶ Translation equivalents are usually provided by **bilingual experts** (lexicographers and/or translators).

Translation corpora are not viable

- ▶ “*the price that would have to be paid if editorial teams were to use bilingual corpora is too high*” (Atkins & Rundell 2008)
- ▶ Reasons:
 - ▶ Too many equivalence candidates
 - ▶ Every one of them seems essential to lexicographers
 - ▶ The production line grinds to a halt

Objective and focus

- ▶ **Objective: suggest two ways in which corpus use can be strengthened**
 - ▶ **General bilingual lexicography** (not domain-specific/technical)
 - ▶ **Pedagogical lexicography:** designed to meet the needs of non-native users, especially advanced learners
- ▶ **Focus**
 - ▶ **Productive use of phraseology**
 - ▶ **Two languages: English and French**
 - ▶ **Electronic dictionaries (web-based)**
 - ▶ Key advantages: customization; hybridization; efficiency of access; corpus integration; more and better data (Granger 2012)

Extended view of phraseology

- ▶ View advocated by Dobrovol'skij in his introduction to IJL's special issue on Phraseology and Dictionaries (2015): *“as a linguistic discipline concerned with fixed expressions, phraseology itself has recently significantly expanded its purview”*.
- ▶ This extended view includes
 - ▶ Semantically non-compositional units (*spill the beans, red tape, a case in point*)
 - AND**
 - ▶ Semantically compositional units (*give a talk, staff regulations, in favour of*)

Lexicography: tendency to focus on semantically non-compositional units

- ▶ Figurative idioms, sayings, proverbs, semantically non-compositional compounds
- ▶ Symptomatic: in IJL's special issue on Phraseology and Dictionaries: 4 out of 5 articles are on idioms

General bilingual lexicography: monolingual and bilingual corpora

Why focus on general bilingual dictionaries?

- ▶ The poor relation of the corpus revolution
- ▶ Bilingual dictionaries are the most popular dictionaries among users: *“Although it is commonly believed that monolingual dictionaries are superior to bilingual dictionaries in terms of their usefulness as language learning tools, attitude and ownership surveys have found that learners generally prefer to use bilingual dictionaries”* (Nesi 2014)

Phraseology in bilingual dictionaries

- ▶ Relatively good coverage of semantically non-compositional multiword units (MWUs); enhanced treatment of collocations
- ▶ Poor coverage of a wide range of (usually fixed) MWUs with grammatical or discursive/pragmatic functions
 - ▶ Multiword prepositions: *in contrast to*
 - ▶ Multiword conjunctions: *on condition that*
 - ▶ Multiword adverbs: *as a result*
 - ▶ Longer sequences: *to the best of my knowledge, the extent to which, and things like that*
- ▶ These chunks are very numerous (Granger & Lefer 2013)
 - ▶ 43 % of the occurrences of the adverb *yet* are part of MWUs: *and yet, yet another, as yet, yet again, yet more, yet to (+ INF), etc.*
 - ▶ 51% for FR *encore*: *ou (bien) encore, mais encore, là encore, (et) plus encore, encore et encore, etc.*

Word-like grammatical/discursive MWUs

- ▶ Many grammatical/discursive MWUs are **word-like**: they are “accidents of graphic continuity” (Palmer 1917) and would therefore deserve to be fully-fledged entries on a par with single words
 - ▶ *in spite of, for a long time, in view of the fact that, as a matter of fact, hardly ever* (Palmer: “polylogs”, “intimate word-groups”)
 - ▶ Often semantically compositional ⇒ less conspicuous
 - ▶ Often polysemous/polyfunctional ⇒ difficult to translate
- Worthy of lexicographical attention

Automatic extraction

- ▶ These units can easily be identified through the n-gram method, used by Biber et al. (1999) to extract “lexical bundles”, i.e. the most frequent recurring sequences of contiguous words
- ▶ This technique has not been used much in general bilingual dictionaries and yet it could considerably enrich their phraseological coverage
- ▶ Recent work: Granger & Lefer (2012, 2013, 2016), Granger (in press)

Data & methodology (Granger & Lefer 2012)

- ▶ Small one-million-word corpus of written French (Label France corpus: articles on politics, economics, culture, technology and tourism)
- ▶ Automatic extraction of 2- to 5-grams with a minimum frequency of 20 ⇒ **6,005** n-gram types
- ▶ Manual selection of dictionary-worthy phrases by the two authors ⇒ **425** dictionary-worthy MWUs
- ▶ Examples: *en matière de, il est vrai que, de ce point de vue*

Corpus/dictionary comparison


- ▶ **Coverage:** are the corpus-extracted MWUs present in general bilingual dictionaries and if so, how are they presented?
- ▶ **Translation:** are there differences between the corpus translations and the dictionary translations and if so, can significant trends be identified?
 - ▶ Translation corpora used: Europarl, Label France, PLECI (Poitiers-Louvain corpus)
 - ▶ Bilingual electronic dictionaries: : Robert & Collins (RC) & Hachette-Oxford (HO)

Coverage in bilingual dictionaries


- ▶ 85-88% are included (50-60 MWUs are totally absent)
- ▶ c. 50% are included as **sub-entries**
 - ▶ Difficult to find (even in electronic dictionaries!)
 - ▶ Shallow lexicographic treatment (e.g. polysemy, style)
- ▶ c. 25% are only included in **example sentences**
 - ▶ Less visible
 - ▶ Highly context-bound translations
- ▶ A very small percentage are granted **headword status** (5% in RC, 14% in HO)
 - ▶ Long way to go before fulfilling Sinclair's (2010) wish to see MWUs given "the same status as the usual headwords".
 - ▶ Not realistic for all MWUs but certainly desirable for word-like units


En effet: Hachette-Oxford

also: en effet
adverb phrase

soyez prudent, les routes sont en effet très glissantes 
do be careful because the roads are very slippery indeed

les résultats sont en effet excellents 
the results are indeed excellent

'tu n'étais pas chez toi hier soir?'—'en effet' 
'you weren't home yesterday evening?'—'no, I wasn't'

en effet, tu avais raison 
actually, you were right

En effet: Robert & Collins

3) **en effet** locution adverbiale

a. (introduit une explication) **because**

- cette voiture me plaît beaucoup, en effet, elle est rapide et confortable : I like this car very much because it's fast and comfortable

b. (= effectivement)

- cela me plaît beaucoup, en effet : yes (indeed), I like it very much
- c'est en effet plus rapide : it's true that it's faster

c. (dans une réponse)

- étiez-vous absent mardi dernier ? — en effet, j'avais la grippe : were you absent last Tuesday? — yes (I was) **ou** that's right, I had flu
- tu ne travaillais pas ? — en effet : you weren't working? — no, I wasn't as it happens


Translation equivalents: dictionary

- ▶ Usually one or two undifferentiated translations
- ▶ Example: translation of FR ***au bout du compte***
 - ▶ HO: *ultimately*
 - ▶ Reverso: *at the end of the day*
- ▶ Often literal translations
 - ▶ FR ***De plus en plus***: *more and more* (absent/demoted: *increasingly*)
 - ▶ FR ***De la même manière***: *in the same way* (absent/demoted: *similarly, likewise*)
- ▶ Promotion of 'general'/context-free translations
 - ▶ Not realistic for grammatical/discursive units that are inherently context-dependent


Translation equivalents: corpus






- ▶ Many translation equivalents
 - ▶ Some are infrequent and highly context-bound ⇒ excluded
 - ▶ Some are quite frequent and less context-bound ⇒ included
 - ▶ FR ***Au bout du compte***: *at the end of the day, ultimately, in the final analysis, in the end, when all is said and done*
- ▶ Advantage: translation corpus data draw lexicographers' attention to some excellent translations that they may not think of spontaneously. Even a small one-million-word corpus helps!


⇒  The use of bilingual corpus data is **useful**

 It is also **viable** if lexicographers have tools that allow them to identify the range of translation equivalents quickly and effortlessly

Reverso Context: translation candidates

 TRADUCTION | DICTIONNAIRE | CORRECTEUR | CONJUGAISON | GRAMMAIRE

au bout du compte  X Français   Anglais  

Traduction de "au bout du compte" en anglais 

Adverbe Nom

ultimately in the end eventually in the final analysis in the long run after all down the road in the last analysis

bottom line the bottom line is after all is said and done Bottom line is at the end of the day when all is said and done




the ultimate

C'est le consommateur final qui paie la taxe **au bout du compte**. It is the final consumer who ultimately pays the tax.

La vie dans les camps est démoralisante et **au bout du compte** déshumanisante. Life in camps is demoralizing and ultimately dehumanizing.

Pas très sympa, **au bout du compte**. Not a nice guy, actually, in the end.

Reverso Context: lexical bundles

en matière de    Français   Anglais  

Chercher en matière de dans: WEB IMAGES DÉFINITION DICTIONNAIRE CONJUGAISON 

Adverbe

in the area of

with regard to

in terms of

in the field of

with respect to

when it comes to

in the matter of

on the issue of

regarding

as regards

concerning

relating to

about

human

over

Autres traductions

Suggestions

- [en matière de sécurité](#) +10k - [en matière de droits](#) +10k - [en matière de santé](#) +10k - [en matière de développement](#) +10k

- [en matière de protection](#) 9322 - [en matière de gestion](#) 8079 [en matière de politique](#) 5986 [en matière de lutte contre](#) 5040

[en matière de prévention](#) 4314 [en matière de recherche](#) 3491 

Sketch Engine for bilingual lexicography

- ▶ *“Lexicographers working on bilingual dictionaries need some bilingual form of the Word Sketch tool to help them use parallel corpus data within the time constraints of commercial dictionary production.”* (Atkins 2002)
- ▶ Kovář et al. (2016): enhancement of Sketch Engine with features for bilingual and multilingual corpus data
 - ▶ Parallel concordancer
 - ▶ Automatic highlighting of translation candidates
 - ▶ Bilingual word sketches

Pedagogical lexicography: learner corpora

Learner-corpus-based error notes in MLDs

- ▶ No active work in this area in the last ten years
- ▶ Reasons?
 - ▶ General economic gloom in the dictionary-making sector
 - ▶ Not really possible to add more error notes within the framework of MLDs because a large proportion of lexical errors are L1-specific and MLDs are generic:
 - ▶ All advanced learners, whatever their mother tongue background
 - ▶ All uses (write a letter to a friend, a narrative, an academic essay, engage in an informal conversation, etc.)

Need for customization

- ▶ **L1 background:** lexical and lexico-grammatical errors vary according to the learner's mother tongue; high rate of transfer
- ▶ **Use:** One of the main tasks that advanced learners have to perform is to write academic texts, and these call for a specific style, with its own set of collocations and lexical bundles
- ▶ These two types of customization are difficult to implement in paper dictionaries but can easily be implemented in **web-based dictionaries**

Louvain **E**nglish for **A**cademic Purposes **D**ictionary (**LEAD**) (Granger & Paquot 2015)

- ▶ Integrated **web-based** tool to help non-native speakers write academic texts in English
- ▶ Focus on the **phraseology** of academic words (collocations and lexical bundles)
- ▶ **Beta version** freely accessible (comments welcome!)

<https://leaddico.uclouvain.be>

Customisation in the LEAD

▶ **Discipline customisation**

- ▶ Linguistics, literature, sociology, anthropology, education, psychology, economics, business, law, political science, medicine, general EAP

▶ **L1 customisation**

- ▶ Chinese, Dutch, French, German, Spanish learners of English

Discipline customization

Welcome

Word search

Search by function

Search by translation

Corpus search

Exercises

Help

Contact us

About

Admin zone

Logout



The Louvain EAP dictionary

The LEAD content partially adapts to user's mother tongue and field of interest. Before using the dictionary, you should therefore select a discipline and specify your language background.

Discipline:

Mother tongue:

Once you have made your

Linguistics

Literature

Sociology

Anthropology

Education

Psychology

Economics

Business

Law

Political science

Medicine

General EAP

' or use the navigation bar above to search the dictionary.

laquet & Sylviane Granger

L1 customization

Welcome

Word search

Search by function

Search by translation

Highlighter

Corpus search

Exercises

Help

Contact us

About

Admin zone

Logout



The Louvain EAP dictionary

The LEAD content partially adapts to user's mother tongue and field of interest. Before using the dictionary, you should therefore select a discipline and specify your language background.

Discipline:

Mother tongue:

Once you have made your selection, click on 'continue' or use the navigation buttons to search the dictionary.

- French
- Dutch
- Spanish
- Chinese
- German
- Other

Learners' attested difficulties with academic words

- ▶ **Meaning** (in particular, false friends – *on the contrary/au contraire*)
- ▶ **Frequency** (underuse of typical academic words and phrases; overuse of non-academic words and phrases)
- ▶ **Positioning** (initial, medial, final)
- ▶ **Punctuation** (use of comma)
- ▶ **Appropriate contextualised use** : collocation and lexical bundles
 - ▶ research: *to make research*
 to conduct, carry out research
 - ▶ to conclude: *I would like to conclude by saying that*
 it can/may be concluded that, we may

conclude that

Corpora at all stages of the dictionary-building process

Corpus data	Stages
Academic vs. fiction section of the BNC	Extract academic words & phrases (Academic Keyword List – Paquot 2010)
Several corpora of native academic writing	Identify their typical phraseology (collocations and lexical bundles)
Discipline-specific corpora of ‘model’ academic texts (top-rated journals)	Display discipline-specific examples of collocations and lexical bundles
Discipline-specific corpora & BNC-ACAD	Give direct access to concordances of words (dictionary entries or not)
International Corpus of Learner English (Granger et al. 2009)	Identify recurrent learner errors and design error notes & exercises (generic and L1 specific)

L1 customisation

1) Search via **translations** of academic words into the learner's L1 (bilingualised dictionary)

Search by translation: Chinese 类似

Welcome

Word search

Search by function

Search by translation

Highlighter

Corpus search

Exercises

Help

Contact us

About

Logout



UCL
Université
catholique
de Louvain

The Louvain EAP dictionary

Search translation:

Go!

Translation(s) of 类似

analogy (*n.*)

resemble (*vb.*)

similarity (*n.*)

Partial matches

类似



L1 customisation

1) Search via translations of academic words into the learner's L1 (bilingualised dictionary)

2) L1-specific **warnings/error notes**

Currently, French only; ongoing work on Chinese in collaboration with Fudan University

L1 customisation

1) Search via **translations** of academic words into the learner's L1 (bilingualised dictionary)

2) L1-specific **warnings/error notes**

Currently, French only; ongoing work on Chinese in collaboration with Fudan University

3) L1-specific **exercises**

Currently, French only

One major weakness

- ▶ Learners will only look up words whose use they are unsure about
 - ▶ If a learner does not know that *on the contrary* is not always the equivalent of *au contraire*, they will not look it up
- ▶ **New functionality: “Highlighter”**
 - ▶ Allows users to cut and paste their text on the interface
 - ▶ Draws their attention to potential (highlighted in orange) and real (highlighted in red) errors in their text and displays explanations in the form of pop-ups
 - ▶ Hyperlinks all the academic words used in the text to their corresponding dictionary entries

Highlighter

Welcome Word search Search by function Search by translation **Highlighter** Corpus search Exercises Help Contact us About Admin zone Logout



The Louvain EAP dictionary

Although the technical revolution has provided us with means of facilitating our work and, as a whole, making our lives much easier, it has failed at one aspect which, according to me, is the most significant one - the human soul. When different people read a same book they have probably various imaginations while reading. This process is much more obvious when we read poetry but nowadays not many people are able to read it. We were taught physics and mathematics but we were not taught how to read poems. As I said some adults are losing their ability to dream and to imagine under pressure of modern civilization. Because the imagination of one time can become the possibility or reality of another time (f. e. flight to Moon) and also reality of one time is becoming the imagination of other time. The imagination depends on what a certain society knows in certain time and what are its demands. The example of deriving the imagination from science or based on science is a science - fiction. One of a famous stories in science - fiction, because it is full of strong feelings for adult and a child, because it carries our desires and worries, is the story about a robot (golem) - a perfect technic creature which is getting away from his masters hands. After all, the supporters of this third opinion add, Hitler planned the liquidation of the bigger part of the Czech inhabitants and the war conflict would have brought that plan nearer. Thus the Government and the President decided realistically, when they accepted capitulation, because even the existence of the Czech nation was in bet. So, in this way is religion very useful but on the contrary, as I said above, it can be also misused. It happened many times that religion was misused for the purpose of one person. Especially in the terms of getting some possession or the power. It is not so difficult to manage the masses of people because people are human beings who can be easily influenced. As a conclusion it can be said that television largely replaces religion. This also seems to be the case with animals, although, it might be a bit more difficult to say whether they fight for their survival or for the fun of it. Hence, we can precede on the assumption that, generally speaking, we have to differentiate two types worth to be scrutinized more closely. Consequently, parents should have the possibility to attend teaching courses providing them with the most essential information. Besides identifying with a certain class or group of people, clothes also express our states of mind. People who feel good about themselves, are more likely to wear colourful and unusual clothes than people who feel bad. In spite of the existence of four national languages in Switzerland, namely German, French, Italian and Romansh, bilingualism is generally not encouraged in Switzerland - even though there are bilingual cities (Fribourg, Biel-Bienne) and bilingual daughters and sons (secondas and secondos, e.g. German/Turkish or French/Spanish) of immigrants. The hostile attitude towards bilingualism is mirrored in Swiss German policy towards Deaf speakers of sign languages. As a consequence of these considerations I would express the initial statement <> in different words: <>. Although there are a great deal of different relationships like companionship, friendship, love affairs and last but not least marriage, they all share certain characteristics which definitely can not be acquired by means of money. Just think about our last-year textual analysis exam and about the great number of different interpretations which arose from the same short text, and what I mean saying that a novel or a poem is open to a wide range of interpretations, will be immediately clearer. Therefore teachers should try to develop their students' skills at thinking about texts and interpreting them, supporting their points of view with relevant details to be found in the text.

PROCESS

RESET

Conclusion

Has lexicography reaped the full benefit of the (learner) corpus revolution?

- ▶ Undeniably a great deal of benefit, but not yet the full benefit

Bilingual lexicography

- ▶ Benefit of using simple techniques like n-gram extraction to improve the coverage of phraseological units
- ▶ Benefit of using translation corpora, even small ones, to improve the quality of the translation equivalents
- ▶ However, the lack of large, representative high-quality translation corpora is a major hurdle
- ▶ There is scope for a large-scale collaborative corpus collection initiative involving publishers and academic teams under the aegis of prestigious associations such as EURALEX, ASIALEX, AFRICALEX and the DSNA

Pedagogical lexicography

- ▶ Benefit of using learner corpora to tailor the entries to the needs of specific L1 populations
- ▶ Advantage of the web-based format, which allows a high degree of flexibility and brings dictionaries closer to users' needs, for example by providing automatic feedback on their texts
- ▶ Need for more learner corpora, especially error-tagged ones

Reasons to be optimistic

▶ **Bilingual lexicography**

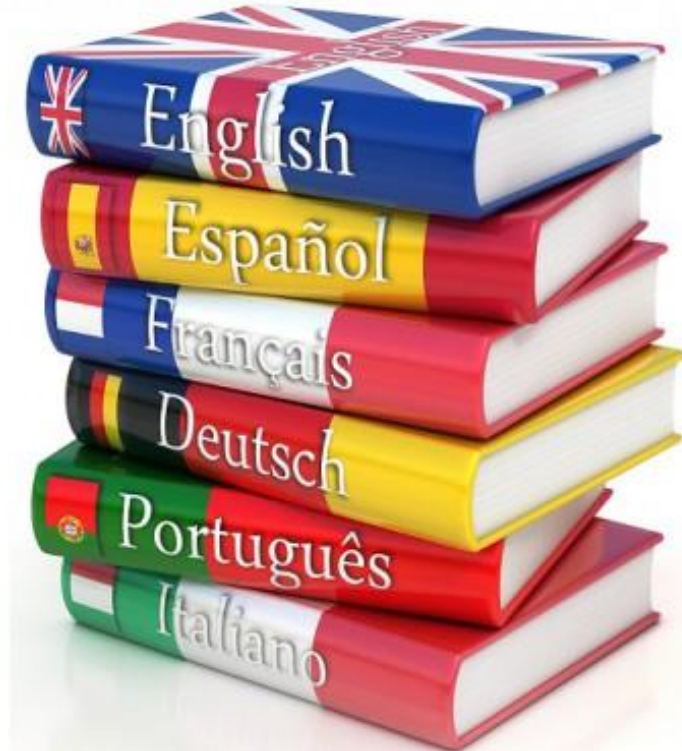
- ▶ New corpora
 - ▶ International Comparable Corpus (Kirk & Čermáková 2017): fully comparable corpora in nine languages
 - ▶ Spoken BNC2014 (McEnery et al. 2017): 11.5 million words of up-to-date British English conversation data
- ▶ New tools
 - ▶ Sketch Engine for bilingual lexicography

▶ **Pedagogical lexicography**

- ▶ New learner corpora
 - ▶ Extended web-based version of the ICLE (26 L1s – c. 5 million words)
- ▶ New tools
 - ▶ Updated and extended XML-conformant version of the Louvain Error Editor

Corpora: not the be-all and end-all of lexicography

- ▶ Strong opposition to Matthews's (2014) definition of corpus linguistics in *The Concise Oxford Dictionary of Linguistics*: “*corpus linguistics*: of investigations based on the analysis of corpora; often by linguists who discount, or seek to disparage, other sources of evidence”
- ▶ Corpora are just collections of language data – undoubtedly data with high potential value, but data nonetheless. Even pre-analysed by powerful software tools, their use in lexicography will always require human expertise.
- ▶ Moon (2008): “lexicographers still have to use intuition and judgement in selecting, interpreting, and setting out the evidence, rather than simply relaying it to the user as *quasi-scientific truth*.”



*Thanks for your attention!
Questions? Comments?*