Using corpora for the selection and treatment of secondary guiding elements

Rufus H. Gouws
Department of Afrikaans and Dutch
University of Stellenbosch
rhg@sun.ac.za
What to expect

• Collocations as items giving the context
• Collocations in existing dictionaries
• Guiding elements
• Corpora and collocations
• Single and complex collocations
• The lexicographic presentation.
Corpora in lexicography

- The use of corpora constitutes an integral component of lexicographic work.
- Corpora are significant and non-negotiable sources of support to obtain material for inclusion in general language dictionaries.
- This applies to both printed dictionaries and e-dictionaries.
Corpora are utilised for a variety of lexicographic purposes:

• to compile frequency lists from which the lemma candidates can be selected;
• to find orthographic variants of the lemma candidates;
• to find the necessary morphological guidance like plural and diminutive forms;
• to supply illustrative material that can be included as items giving cotextual guidance.
Cotextual items

- Illustrative material
  - Example sentences
    - the macro-syntactic environment
  - Collocations
    - the micro-syntactic environment.
- Both types of text segments usually included in the same search zone.
- No distinction.
• For text production purposes, for both mother-tongue and foreign speakers of a given language, collocations have to be seen as a necessary items and they should be presented in a clearly identifiable way.

• Their lexicographic significance may never be underestimated.

• Entered as items addressed at the lemma sign

• Should also receive their own treatment.
Collocations in existing dictionaries

• The presentation and treatment of collocations in many existing dictionaries eschew the need of users with a text production need to identify the collocations and retrieve information they require.

• This can be seen in the treatment of the lemma sign emotion in the Oxford Online dictionaries where users do not know that the cotext item she was attempting to control her emotions contains the collocation to control emotions.
emotion

Line breaks: emo|tion

Pronunciation: /ɪˈmenʃən/ /

Definition of emotion in English:

NOUN

1. A strong feeling deriving from one's circumstances, mood, or relationships with others:

'she was attempting to control her emotions'

[MASS NOUN] 'his voice was shaky with emotion'

MORE EXAMPLE SENTENCES

'She loves the fact that there is an intensity about holidays that can spark strong emotions.'

'It allows you to have strong emotions and opinions without any real risk to yourself.'

'There has been an attempt to defuse aggressive emotions and any desire for revenge.'

GET MORE EXAMPLES
• Collocations often are addressed at the lemma but they have no items explicitly addressed at them to elevate them to treatment units in their own right.

• This might be due to the fact that the use of corpora is governed by a lemma-bias and not by attempts to find data to ensure a more comprehensive information retrieval by the dictionary user.

• A much better lexicographic presentation and treatment is realised in the article of the lemma sign *money* in the *MacMillan Dictionary*:
money - definitions and thesaurus

NOUN [UNCOUNTABLE]  /ˈmʌni/

what you earn, save, invest and use to pay for things. Money can be kept in a bank, where it can earn interest. If you have a bank account, you can pay for things with a cheque.

No, I can’t come – I haven’t got any money.

make/earn money: The business has made more money this year.

spend money (on something): We’ve spent a lot of money on this house.

cost (someone) money: It would have cost us a lot of money to cancel the event.

borrow money: I have had to borrow money from my family.

save money (=avoid spending money): You can save money by taking your own lunch.

save money (=put money somewhere so that you can spend it later): They’re trying to save money so that they can have a holiday.

money coming in (=money being earned and available to spend): He had no job so there was no money coming in.

have money on you (=have money in your pocket etc): Have you got any money on you?

lose money (=earn less money than you spend): The industry is losing money and the government wants to sell it.

birthday/Christmas money (=money received as a birthday/Christmas present): I’m going to spend my birthday money on some new clothes.

raise money (=collect money for a particular purpose): Her bike ride
• Following the paraphrase of meaning a single example sentence is given.

• This is followed by a text block containing a non-grouped listing of collocations complemented by example sentences addressed at the respective collocations in procedures of non-lemmatic addressing.

• In this dictionary the collocations are elevated to secondary treatment units and secondary guiding elements.
Guiding elements in dictionary articles

- Within the article stretches of a dictionary each individual article is introduced by means of a lemma sign functioning as guiding element of that specific article.

- Data given as part of the treatment of a given word can be reached via that lemma sign as guiding element of the article.

- In printed dictionaries the consultation process typically proceeds via the lemma sign.
• In monolingual general language dictionaries items giving the context, both example sentences and collocations, typically have a lemmatic address.

• Where collocations are included for text production purposes, they help to ensure a better use of the word represented by the lemma sign.

• The need for collocations in dictionary articles also goes further than merely assisting the user in the proper use of the single word.
• For text production purposes users also need guidance regarding the proper use of collocations.
• This demands the inclusion of more items addressed at collocations.
• To enable an optimal lexicographic reflection of collocations the use of corpora becomes increasingly important.
Corpora and collocations

• How can corpora be used to improve the quality of the lexicographic reflection of collocations?
• A corpus has to be employed in a search for collocations in which the word represented by the lemma sign occurs.
• These collocations should then be retrieved as candidate items for inclusion in the specific search zone of the relevant dictionary article.

• It could lead to the inclusion of search zones that will exclusively accommodate collocations and their treatment.

• These collocations will be elevated to treatment units and to secondary guiding elements because the user can access data regarding a collocation via that collocation as an entry in a dictionary article.
• In an online dictionary the access process does not even have to proceed via the lemma but the user can identify a specific collocation as search string and enjoy rapid and direct access to this secondary guiding element.

• Via such a guiding element users can be lead to data regarding the collocation, not regarding the lemma of the specific article.
• Different types of items can be included as part of the treatment of collocations.

• The search zone structure for the treatment of collocations should include at least one or more items giving example sentences that have the collocation and not only the word represented by the lemma as significant item form.
• Dictionaries have to mark items when stylistic, geographical, chronolectic or other deviations from the default standard language variety of the dictionary occur. This also applies to collocations as items that are targets of addressing procedures.

• The corpus used for a given dictionary should also be employed to determine these different restrictions so that the required lexicographic labels, used elsewhere in the dictionary article, could be employed.
• Differences between the collocation and the single word.

• The word *invitation* is stylistically neutral; the collocation *extend an invitation* belongs to formal language use.

• Texts where collocations are found in a given corpus will assist lexicographers to employ the necessary lexicographic labels.

• Corpus consultation has to go beyond a look at the occurrence of the single word.
• The meaning of a collocation is usually a combination of the meanings of the collocating lexical items, e.g. *win a match/lose a match*.

• The data on offer from a corpus may suggest to the lexicographer that some semantic guidance should be included as part of the treatment of a collocation.

• In the article of the lemma sign *abandon* in the *Oxford Advanced Learner’s Dictionary* the following example sentence is complemented by a brief paraphrase of meaning:

  “He gave the order to *abandon ship* (= to leave the ship because it was sinking).”
In the example taken from the article of the lemma sign *money* in the *MacMillan dictionary* some collocations are also complemented by items giving semantic guidance, e.g.:

- save money (= avoid spending money)
- money coming in (= money being earned and available to spend)
• When collocations are entered as secondary guiding elements the need for brief paraphrase of meaning may become apparent.
• Working with collocations as they occur in the actual language as retrieved from a corpus will assist the lexicographer in deciding whether this data type should be included for a specific collocation.
• Finding example sentences that illustrate the typical use of collocations will also help the lexicographer to identify those instances where additional syntactic guidance needs to be given for a specific collocation.

• Compare in this regard the restricted guidance found in the article of money in the *MacMillan Dictionary*
  
  spend money (on something)
  cost (someone) money
• From the different occurrences of a given collocation in a corpus the lexicographer will quite often also find the occurrence of the collocation of two collocations with a mutual element.

• Such a combination of two collocations with one mutual element can be regarded as a complex collocation, compared to a single collocation which is the default type.
• A corpus should be used to find the relevant occurrences of both single and complex collocations.

• In the article of the lemma sign *preference* the slot for collocations could include the items *personal preference* and *express a preference*.

• The treatment of both these collocations could include an example sentence that contains the complex collocation, e.g.

  *She tried to be impartial and did not want to express a personal preference.*
• Dictionary users need to be made aware of single and complex collocations as two separate item types.
• The presentation of the items giving the context of collocations in the dictionary article should enable the user to distinguish between them.
• Complex collocations could be presented as part of the treatment of single collocations but the system employed should clearly indicate that a given item is an illustration of the specific collocation in collocation with another collocation.

• To ensure an unambiguous comprehension of such an item the lexicographer should embark on a system of presentation where different types of information can be retrieved from a single form.
Such an entry has to display a variety of items:
- an item giving the collocation express a preference
- an item giving the collocation a personal preference
- an item giving the complex collocation express a personal preference
- an item giving an example of the use of the word preference
- an item giving an example of the use of the complex collocation express a personal preference:
She tried to be impartial and did not want to express a personal PREFERENCE.

- The word represented by the lemma sign, i.e. preference, functions as base of the collocations and is given in small caps.
- The collocators are distinguished from each other by means of enrichment (the use of bold) and the use of downward expansion – with the underlining functioning as an upwardly addressed item addition.
• The entry as a whole is given in italics to identify it as an illustrative example.

• This entry includes different items with different genuine purposes, combined into a single item form.
In conclusion

- Collocations are part and parcel of the daily language use of all speakers.

- One of the reasons why dictionaries often fail to give a satisfactory presentation and treatment of collocations lies in an insufficient use of speech corpora.

- By employing corpora for searches that go beyond the mere retrieval of information directed at a single word, the lexicographer can obtain information that can enhance the nature, extent and quality of the selection, presentation and treatment of collocations.
• This demands allocating an elevated status to collocations, i.e. as guiding elements of their own search zones and as fully-fledged treatment units.
Baie dankie/Thank you very much