

The Presentation and Treatment of Collocations as Secondary Guiding Elements in Dictionaries

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)*

Abstract: Collocations represent an item type that is frequently presented and treated in an insufficient way in dictionaries. It is cumbersome for the user to distinguish between collocations and example sentences, and dictionaries often include no additional treatment directed exclusively at collocations. In this contribution the need is shown for the recognition of collocations as an item type. It is suggested that they should be presented in such a way in dictionary articles that they can function as secondary guiding elements. Besides their inclusion in dictionaries collocations also need to receive a treatment that makes provision for, in many instances, at least an example sentence but that could also include, when necessary, stylistic and other labels and even a brief paraphrase of meaning. It is shown how collocations are presented and treated in a few existing dictionaries. These presentations are criticised and proposals are made for an improved presentation. In this regard Wiegand's model for a semi-integrated microstructure is adapted to provide for the presentation of collocations in their own non-integrated search zone. Besides the explicit presentation of single collocations proposals are made for an implicit presentation of complex collocations. From a single item form users should then be able to retrieve different items. Although the discussion primarily focuses on printed dictionaries proposals are also made for the presentation of collocations in online dictionaries.

Keywords: ARTICLE STRUCTURE, COLLOCATION, COMPLEX COLLOCATION, COTEXT, EXAMPLE SENTENCES, INTEGRATED MICROSTRUCTURE, NON-GROUPED ORDERING, SEARCH ZONE, SEMI-INTEGRATED MICROSTRUCTURE, SINGLE COLLOCATION, SUB-ZONE

Opsomming: Die aanbieding en bewerking van kollokasies as sekondêre gidselemente in woordeboeke. Kollokasies is 'n aanduidertipe in woordeboeke waarvan die aanbieding en bewerking dikwels onvoldoende is. Dit is vir die gebruiker moeilik om kollokasies van voorbeeldsinne te onderskei en woordeboeke bied dikwels geen bykomende bewerking wat uitsluitlik op kollokasies gerig is nie. In hierdie bydrae word die behoefte aan 'n volwaardige erkenning van kollokasies as aanduidertipe bepleit. Daar word voorgestel dat die aanbieding in woordeboekartikels sodanig moet wees dat hulle as sekondêre gidselemente kan optree. Naas die opname van kollokasies moet daar ook 'n bewerking gebied word wat vir minstens 'n voorbeeldsin maar waar nodig ook vir styl- of ander etikette en selfs vir 'n bondige betekenisparafrese voorsiening maak. Daar word gewys op die manier waarop kollokasies in enkele bestaande woordeboeke

opgeneem en bewerk word. Kritiek word hierop uitgespreek en voorstelle word aan die hand gedoen vir 'n verbeterde aanbieding. In hierdie verband word Wiegand se model van 'n semi-geïntegreerde mikrostruktuur aangepas om voorsiening te maak vir die aanbieding van kollokasies in 'n eie niegeïntegreerde soeksone. Naas die eksplisiete aanbieding van enkelkollokasies word voorstelle gemaak vir 'n implisiete aanbieding van komplekse kollokasies. Vanuit een aanduidervorm moet gebruikers daartoe in staat wees om verskillende aanduiders te kan onttrek. Die bespreking is primêr op gedrukte woordeboeke gerig, maar voorstelle word ook gemaak vir die aanbieding van kollokasies in aanlynwoordeboeke.

Slutelwoorde: ARTIKELSTRUKTUUR, ENKELKOLLOKASIE, GEÏNTEGREERDE MIKROSTRUKTUUR, KOLLOKASIE, KOMPLEKSE KOLLOKASIE, KOTEKS, NIEGEGROEPEERDE ORDENING, SEMIGEÏNTEGREERDE MIKROSTRUKTUUR, SOEKSONE, SUBSONE, VOORBEELDSINNE

1. Introduction

In general language dictionaries with a text production and text reception function lexicographers often include items to illustrate the typical use of the word represented by the lemma sign of a given article. These items from which cotextual information can be retrieved, fall into different categories. Example sentences, the typical citations from corpus material or even some made up by the lexicographer, are one type and collocations are the second type. In many dictionaries both these types of text segments are included in the same search zone, i.e. a general article slot for illustrative material. In this search zone dictionaries too often do not make a distinction between the different types of cotextual items. As a result the user has no clear guidance to assist him/her in identifying collocations and to ensure a successful retrieval of information from entries representing this item type. In printed dictionaries this inferior way of presentation is partially due to a lack of space but also to an insufficient focus on collocations as fully-fledged items in dictionary articles. This insufficient focus also prevails in many online dictionaries. Collocations are significant units in the syntactic patterns of a language, cf. Hausmann (2008). Lexicographers have to negotiate the best possible ways of presenting and treating them in their dictionaries.

For text production purposes, for both mother-tongue and foreign speakers of a given language, collocations have to be regarded as necessary items, and they should be presented in a clearly identifiable way. Their lexicographic significance may never be underestimated. Collocations need to be presented as part of the cotextual guidance of a given lemma. However, their importance exceeds their occurrence as illustrative material of the word represented by the lemma sign. Consequently they should not only be entered as items addressed at the lemma sign but should also be presented as addresses, by functioning as items with their own treatment, albeit of a restricted nature.

In this paper reference is made to aspects of some of the prevailing ways in which collocations are presented and treated. Proposals are then made for a more salient lexicographic presentation and a more comprehensive treatment of collocations, especially by employing evidence from speech corpora.

2. Collocations in existing dictionaries

Unfortunately the presentation and treatment of collocations in many existing dictionaries, both printed and online dictionaries, eschew the extent of the need of users consulting the dictionary to solve a text production problem. These dictionaries often do not enable their users to achieve an unambiguous identification of collocations included in the articles and to retrieve the information they require to use these collocations in a proper way. This can be seen in the following article of the lemma sign *hart* (=heart) in the 5th edition of the Afrikaans monolingual dictionary HAT:

hart s.nw. [-e, -jie]

¹ Organ in die borskas van 'n mens of 'n gewerweld dier, wat deur sy voortdurende, ritmiese sametrekkinge die bloedsomloop aan die gang hou: *Voel hoe klop my hart.* 'n Sterk, gesonde, swak hart hê. *As 'n mens se hart gaan staan, gaan jy dood.* Lek-, sporthart.

² Deel van die liggaam waar die hart sit: *Jou hand op jou hart lê.* Iemand aan jou hart druk.

³ Setel van 'n mens se gevoelens en sedelike eienskappe; innerlike aard; wese, binneste: *Iemand met 'n sagte, verstandige, opregte hart.* 'n Harde, teer hart hê. *Die angsslaan my om die hart.* 'n Gebroke hart. *Die taal van die hart.* My hart voel lekker. *Iemand 'n goeie, kvade hart toedra.* ♣ En jy, Salomo, ... jy moet die God van jou vader ken en hom dien met jou hele hart. ... want die Here ondersoek alle harte en ken die bedoeling van elke gedagte (1 Kron. 28:9 NV). ♣ My hart ... sing soos 'n bestie (Ingrid Jonker). ♣ Huil jou hart uit, my kind, troos haar ma. Dis nie jou skuld nie (A.H.M. Scholtz).

⁴ Moed, ywer, geesdrif: 'n Mens moet hart hê om dit te kan doen. Sy hart het hom begeve.

⁵ Iets wat die vorm van 'n hart het: 'n Bruin perd met 'n wit hart voor die kop. 'n Goue hartjie aan 'n ketting om die nek dra.

^{6(a)} Middelste, binneste gedeelte van iets: *In die hartjie van die somer.* *Die ongeluk het in die hartjie van die stad gebeur.* *Die hart van 'n boomstam.* *Tot die hart van die saak deurdring.* ^(b) Gedeelte van die distillaat wat by brandewynstokery voor die naloop en na die voorloop oorstoek en brandewyn uitmaak. *vgl. naloop, voorloop.*

Example 1 from HAT 5

Although each subcomment on semantics contains a search zone populated by cotextual items, presented in italics, no distinction is made between example sentences and collocations. The user does not know whether a given cotextual item is merely an illustration of the use of the word, e.g. *Jou hand op jou hart lê* (= to put your hand on your heart), in the second subcomment on semantics, or the use of the word as component of a collocation, e.g. *As 'n mens se hart gaan staan, gaan jy dood* (= when your heart stops one dies), in the first subcomment on semantics, where *hart* collocates with *gaan staan* (= to stop). This presentation does not contribute at all to a user's need with regard to collocational guidance because the user does not know whether *gaan staan* only incidentally combines with *heart* or whether it is a frequent combination.

Similar criticism applies to the treatment of the lemma sign *emotion* in the *Oxford Online Dictionaries* where users are not informed that the cotext item *she was attempting to control her emotions* contains the collocation *to control emotions* or that *spark strong emotions* is actually a complex collocation, i.e. a combination of the two collocations *spark emotions* and *strong emotions*.

emotion



Line breaks: emo|tion
Pronunciation: /ɪ'məʊj(ə)n /

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](#)

Example 2 from *Oxford Online Dictionaries*

In both these dictionary articles the collocations are addressed at the lemma but they have no items explicitly addressed at them to elevate them to treatment units in their own right. Users have no idea that these example sentences illustrate the use of both the word represented by the lemma sign and the specific collocation or collocations.

The *lexiko* dictionary, as seen in the following two screenshots from the article of the lemma sign *Kritik* (= criticism), gives users the opportunity to click on the button "Kollokationen" (= collocations). They are then directed to a list of cotext partners:

lexiko

Kritik

Lesart: 'Beurteilung'

zur Übersichtsseite Lesarten im Überblick

Bedeutungs-erläuterung Kollokationen Konstruktionen Sinnverwandte Wörter Gebrauchsbesonderheiten Grammatik

Example 3 from *lexiko*

The screenshot shows a dictionary entry for the word 'Kritik'. At the top left is an orange square. To its right, the word 'Kritik' is displayed in a bold, black font, followed by a speaker icon. Below this, the text 'Lesart: 'Beurteilung'' is shown. A navigation bar contains several tabs: 'zur Übersichtsseite', 'Bedeutungs-erläuterung', 'Kollo-kationen', 'Konstruk-tionen', 'Sinnverwandte Wörter', and 'Gebrauchs-besonderheiten'. The 'Kollo-kationen' tab is selected. Below the navigation bar, the section 'Kollokationen: Wörter im Kontext' is displayed. It contains three sub-sections: 'Wer spricht Kritik aus?' with a list of contexts (die Grünen, Gewerkschaft, Landesrechnungshof, Opposition, Rechnungshof); 'Für wen gibt es Kritik?' with a list (Bundesregierung, Person, Regierung, Trainer); and 'Wofür gibt es Kritik?' with a list (Amtsführung, Arbeit, Äußerung, Entwurf, Führungsstil, Haltung, Kapitalismus, Kurs, Menschenrechtsverletzungen, Plan).

Example 4 from *ellexiko*

Here different collocational contexts are given albeit that the entries given as collocations are not the addresses of items presenting any form of further treatment. Questions could be raised regarding the real collocational status of these entries. A click on the button "Konstruktionen" (= constructions) guides the user to the screenshot, seen in Example 5.

These entries include some real collocations, e.g. *harsche Kritik einstecken*, *auf heftige Kritik reagieren*, etc. For the typical user it is not that easy to find these collocations.

It is a distressing fact that collocations typically occur in dictionaries exclusively as part of the treatment of the word represented by the lemma sign. Their significance as well as their potential as items that can enhance the dictionary consultation experience and can help to solve the real lexicographic needs of the target users of many dictionaries are eschewed. This might be due to a prevailing lemma-bias and insufficient attempts to present data that can ensure a more comprehensive information retrieval by the dictionary user. It may also among others be due to an insufficient theoretical approach to the selection and treatment of items in dictionaries.

The screenshot shows the OVID lexicon interface. At the top, there is a search bar with 'Suchen' and 'Erweiterte Suchen' buttons. Below the search bar is a navigation menu with letters A-Z and a 'gehe zu:' field. The main content area is titled 'Kritik' with a speaker icon and the sub-entry 'Lesart: 'Beurteilung''. There are tabs for 'Bedeutungs-erläuterung', 'Kollo-kationen', 'Konstruk-tionen', 'Sinnverwandte Wörter', 'Gebrauchs-besonderheiten', and 'Grammatik'. The 'Konstruk-tionen' tab is active, showing 'Konstruktionen: Typische Verwendungen'. This section is divided into 'Verwendungen mit Attribut' and 'Verwendungen in Verbalphrasen und Sätzen'. Under 'Verwendungen mit Attribut', there are examples like 'die Kritik an seiner Person' and 'Internationale Kritik an der Regierung'. Under 'Verwendungen in Verbalphrasen und Sätzen', there are examples like 'Kritik als Nörgelei abweisen' and 'sich heftiger Kritik aussetzen'. A left sidebar contains a list of related terms, with 'Kritik' highlighted.

Example 5 from *ellexiko*

A much better lexicographic presentation and treatment can be found in the *MacMillan Dictionary*, as seen in the reduced version of the article of the lemma sign *money* in example 6.

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, cf. Bergenholtz, Tarp and Wiegand (1999) for a discussion of the notion of grouped and non-grouped entries, 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 consequently also become secondary guiding elements within the respective dictionary articles. Unfortunately the ordering of the collocations in the text block seems to have been done in a haphazard way.

money - definitions and thesaurus ★★★



NOUN [UNCOUNTABLE]  Pronunciation /'mʌni/

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*

Example 6 from the MacMillan Dictionary

An improved approach, compared to the 5th edition, is found in the recently published 6th edition of HAT. Collocations are allocated a separate search zone where they are presented in a grouped text block. Each collocation, given in bold, is preceded by a "|" and is the address of a brief paraphrase of meaning as well as an example sentence.

ma-te *s.nw.* ♦ **in dié mate** so erg: *Die getuie het in dié mate gesnik dat die hofverrigtinge tydelik gestaak moes word.* | **'n groot mate van** baie: *'n groot mate van onsekerheid.* | **in ('n) hoë mate** baie; grotendeels: *Veiligheid op die pad word in hoë mate bepaal deur die toestand van 'n voertuig se bande.* | **in/tot 'n mate** 'n bietjie; effens: *Peter het die beoordelaars tot 'n mate beïndruk, maar hulle het op die ou end meer van die ander deelnemer gehou.* | **in meerdere of mindere mate** Ook **in mindere of meerdere mate** tot 'n sekere mate; in wisselende graad: *Enige medisyne is in meerdere of mindere mate giftig as 'n mens te veel daarvan gebruik.* | **na die mate van** in verhouding tot: *belas word na die mate van jou inkomste.* | **in ruim mate** ruimskoots. | **'n sekere mate van** 'n bietjie: *Ek bespeur 'n sekere mate van jaloesie by jou.* Vgl. MAAT²

Example 7 from HAT 6

3. Guiding elements in dictionary articles

Within the article stretches of a dictionary each individual article is introduced by a lemma sign functioning as guiding element of that specific article, cf. Wiegand and Gouws (2013: 275). Data included as part of the treatment of a given word can be reached via the lemma sign as guiding element of the article. The function of a lemma sign as guiding element is not only restricted to main lemmata but it also applies to sublemmata where these sublemmata guide the users to those entries of which they are the addresses.

In printed dictionaries a consultation process typically proceeds via the lemma sign. However, the lemma is not necessarily the only address, treatment unit or even guiding element in such a dictionary article. Procedures of non-lemmatic addressing often occur, e.g. where a cotext entry is given for a translation equivalent whilst a label directed at a microstructural item also participates in a procedure of non-lemmatic addressing. Where a synonym is given for a translation equivalent in a bilingual dictionary there is a dual addressing procedure with the second equivalent directed at the lemma in a procedure of lemmatic addressing but also at the preceding equivalent in a procedure of non-lemmatic addressing.

In monolingual general language dictionaries items giving the cotext, both example sentences and collocations, typically have a lemmatic address. Where example sentences are included for text production purposes, they help to ensure a better use of the word represented by the lemma sign. Collocations can serve the same purpose but 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 do not only need guidance regarding the choice of collocations but also about the proper use of these collocations. This demands the inclusion of more items addressed at collocations — a procedure that establishes the status of collocations as fully-fledged treatment units.

In printed dictionaries, but even more so in online dictionaries with less space restrictions adaptations to the article structure could lead to the inclusion of search zones that will exclusively accommodate collocations and their treatment, cf. Example 9 in section 5.2.2. Within such a search zone 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 to a collocation needed for text reception purposes does not even have to proceed via the lemma sign but the user can identify a specific collocation as search string and enjoy rapid and direct access to this secondary guiding element — where e.g. syntactic information needed for text production purposes can then be retrieved.

4. Data presented in the search zone for collocations

Different types of items can be included as part of the treatment of collocations.

As in the case of the microstructure of any dictionary lexicographers need to make provision for both an obligatory and an extended search zone structure. This needs to be done in accordance with the needs of the envisaged target user and the functions of the dictionary. In a general language dictionary with text production as a function the obligatory search zone structure for collocations and their treatment could include at least one or more items giving example sentences. These examples may also help to support the text reception function of a dictionary. Relying on a corpus the lexicographer needs to distinguish in the selection of example sentences between those illustrating the single word and those that have the collocation and not only the word represented by the lemma sign as significant component. Example sentences in the default slot for cotext entries as well as in a text block allocated to collocations could support the users in their text production endeavours where collocations have to be employed in actual language use. Currently too few dictionaries assist the user with regard to the proper use of those collocations included as part of the treatment of a given word.

4.1 Labels

Dictionaries have to mark items for stylistic, geographical, chronolectic or other deviations from the default standard language variety of the dictionary. This also applies to collocations as items that are targets of addressing procedures. The corpus used for a given dictionary should also be consulted to determine these different restrictions, including morphosyntactic restrictions, that apply to collocations so that the required lexicographic labels, used elsewhere in the dictionary article, could also be employed. This establishes the use of labels not only as entries directed at lemmata but also at other types of microstructural items. Looking at the possible labelling of collocations it is important to realise that there may be e.g. stylistic differences between the collocation and the single word that is the guiding element of the article and that functions as either the base or the collocate of the specific collocation, e.g. whereas the word *invitation* is stylistically neutral the collocation *extend an invitation* belongs to formal language use. Yet again, as is the case with the retrieval of information with regard to single words, the nature of the texts where collocations are found in a given corpus will assist lexicographers to select and employ the necessary lexicographic labels.

4.2 Semantic guidance

Collocations are not independent lexical items with an own meaning, but 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 search zone in which collocations are presented in the article of the lemma sign *match* does not have to

include an item giving the paraphrase of meaning of the collocation as item in the obligatory search zone structure because this collocation has a transparent meaning, constituted by the meaning of the collocating words. The retrieval of the meaning of a collocation is not always effortless. Tutin and Grossmann (2002: 8) regard collocations as semi-compositional and distinguish between regular, transparent and opaque collocations, cf. also Heid (2011). Because of these varying degrees of transparency the data on offer from a corpus may suggest to a lexicographer the need for an extended obligatory search zone structure to accommodate some semantic guidance as part of the treatment of a given collocation. In the article of the lemma sign *abandon* the *Oxford Advanced Learner's Dictionary* has the following item giving a paraphrase of meaning of this word:

2 .. to leave a thing or place, especially because it is impossible or dangerous to stay.

This dictionary does not have a search zone allocated to collocations but includes collocations embedded into example sentences where they are marked as items by a procedure of typographical enrichment according to which these sentence-internal collocations are given in bold. Within the second subcomment on semantics that accommodates the above-mentioned paraphrase of meaning, the following example sentence is included as an item giving contextual data. This 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).*

This example sentence does not only illustrate the use of the single word but also that of the collocation. The lexicographer furthermore realised that the typical user of this dictionary needs additional semantic guidance to ensure a successful comprehension of the collocation *to abandon ship*. This brief paraphrase of meaning is addressed at the embedded collocation, resulting in that collocation becoming an addressed item and therefore also a treatment unit. A procedure of typographical enrichment has been employed to identify a part of an item, i.e. the item giving the example sentence, as an item in its own right, a collocation, and to identify this embedded item as the address of a subsequent item giving a paraphrase of meaning. This presentation and treatment help with the comprehension of the collocation, in a text reception situation, and with the proper word choice in text production situations.

In the above-given example 5 from the *MacMillan Dictionary* collocations are presented in a single text block where they are elevated to secondary treatment units and consequently also become secondary guiding elements within the respective dictionary article. Enriching a single part of a more comprehensive item is therefore not needed in order to display the collocations as fully-fledged items. In the article of the lemma sign *money* 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, as is the case in this dictionary, the need for a brief paraphrase of meaning addressed at a given collocation may become apparent for some items. 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.

4.3 Syntactic guidance

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 the lemma sign *money* in the *MacMillan Dictionary*:

- spend money (on something)
- cost (someone) money

Although collocations typically consist of two components, cf. Heid (1994: 228), Tutin (2008: 1445) shows that some collocations like *to pay close attention to* include more than two elements. She regards this as a merging of two or three binary collocations, identified as *collocational chains* by Hausmann (2004) or *collocational clusters* by Spohr (2005). From the different occurrences of a given collocation in a corpus the lexicographer will quite often find the occurrence of the combination of two collocations with a mutual element, cf. Zinsmeister and Heid (2003), Spohr (2005), Heid and Gouws (2006). 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. In the search zone with a collocation as guiding element these different types of cotext items should be included where applicable, i.e. an item giving the cotext where a single collocation occurs as item form and an item giving the cotext where a complex collocation is included as item form.

The noun *habit* combines with the adjective *bad* to form the collocation *a bad habit*, and also with the verb *break* to form the collocation *to break a habit*. These two single collocations can combine, resulting in the complex collocation *to break a bad habit*. A corpus should be used to find the relevant occurrences of both single and complex collocations; yet again, looking beyond features relevant to a single word as treatment unit.

Dictionary users need to be made aware of single and complex collocations as two separate item types and the presentation of the items giving the

cotext of collocations in the dictionary article should enable the user to distinguish between them. This will be discussed in a subsequent section.

5. The presentation of collocations

5.1 Single collocations

The following discussion only focuses on the presentation of collocations in general language dictionaries. A different approach would be employed in a dictionary of collocations. Lexicographers of general language dictionaries should pay attention to the way in which dictionaries of collocations present and treat their collocations because some of these features could also be introduced in general language dictionaries. The following article for the lemma sign *Krimi* (= crime story/film) from *Feste Wortverbindungen des Deutschen* (Häcki Buhofer et al. 2014), a German collocation dictionary, illustrates the occurrence of the word *Krimi* in different collocations, arranged in the article firstly according to the different senses of the word, given in subsequent subcomments on semantics, and within each one of these subcomments on semantics according to the part of speech of the partner word:

Krimi m

Kriminalgeschichte

ADJEKTIVE/ADVERBIEN

- **unterhaltsamer** | **packender** | **spannender** K.°
- **schlechter** K.° *Die Details des Diebstahls lesen sich wie ein schlechter Krimi, in dem schon zu Beginn klar ist, wer die Tat verübt hat.*

VERBEN

- einen K. **lesen** *Wenn ich vor dem Einschlafen noch Krimi lese, habe ich manchmal Alpträume.* | einen K. **sehen**

NOMEN

- ein K. kommt/läuft im **Fernsehen** *Leider ist schon seit Ewigkeiten kein guter Krimi mehr im Fernsehen gekommen.*

ZUSAMMENSETZUNGEN

- **Fernsehkrimi** ■ **Krimiserie** *Die Polizistin ärgert sich, dass die Krimiserie die Polizeiarbeit so unrealistisch darstellt.* | **Krimiautor_in**

aufregendes Ereignis

ADJEKTIVE/ADVERBIEN

- **echter** | **richtiger** K.° *Die Verhandlung entwickelte sich zu einem richtigen Krimi.* | **wahrer** K.° *Seine Reise durch Afrika war ein wahrer Krimi.* ■ **spannend** wie ein K. | **spannender** als ein K.

Example 8 from *Feste Wortverbindungen des Deutschen*

This arrangement according to part of speech within each subcomment on semantics resembles the arrangement proposed in Wiegand (1996) for the non-integrated component of a dictionary article where a semi-integrated microstructure is used, and could be very helpful to satisfy text production needs. Compare in this regard the arrangement suggested by Wiegand (1996: 36) for the article of the lemma sign *Kranz* where the search zones in the non-integrated component, following the 6th subcomment on semantics, are ordered according to the part of speech of the words combining with the word represented by the lemma sign. The structure of this article is relevant to this discussion; not the contents.

Kranz, der; -es; Krän-ze

1. *in der Form eines Rings geflochtene od. gebundene Blumen, Zweige o.ä.* | In der Gärtnerei konnte man Kränze aller Art, z.B. Blumen-, Efeu-, Lorbeer- und Adventskränze bestellen.

2. *etw., das in einer kreisförmigen Form erscheint* | ein ~ von Haaren wuchs rings um seinen kahlen Schädel.

3. *Anzahl von etw. od. jmdm., die um eine Art Mittelpunkt gruppiert sind* | Der Kanzler präsentierte sich mit einem ~ von Ministern vor der Presse.

4. (schweiz.) *einer der ersten drei Plätze bei einem Wettbewerb* | Sie setzten sich nach den Kränzen auf die Ehrentribüne.

5. (Jägerspr.) *Abdruck der äußeren Kante der Hufe von Rotwild auf trockenem Boden* | Er konnte die Fährten deutlich erkennen, war sich aber nicht sicher, ob es sich um Kränze handelt.

6. (Kegeln) *Wurf, bei dem alle acht um den König stehenden Kegel fallen* | Ein Kranz war ihm heute abend nicht gelungen

Sub. ein ¹ **Feigen** | ein bunter ² tanzender **Kinder** | ein ² junger **Mädchen** (geh. *Gruppe, Kreis*) | ein schmackhaftes **Stück** ² (kurz für ↑Kranzkuchen)

V. einem Mädchen einen ¹ **aufsetzen** | einen ² **backen** (*Kranzkuchen*) | einen ¹ **binden** | einen ¹ **flechten** | einen ¹ **heraushängen** (bes. als Zeichen, daß Wein ausgeschenkt wird) | einen ¹ **niederlegen** (am Grab [Ehrenmal] als Zeichen der Trauer) | den Hinterbliebenen einen ¹ **schicken** (als Zeichen der Anteilnahme) | dem Sieger einen ¹ **überreichen** [**aushändigen**] | einen ⁶ **werfen** | einen ¹ **winden**

Adj. bunter [grüner, verwelkter] ¹

Präp. einen ¹ **auf** das Grab legen | einen ¹ **auf** dem Kopf tragen | Kerzen **auf** einen ¹ stecken (bes. einen Advents~) | **in** die Kränze kommen (schweiz.; umg. *unter den Siegern sein*) | das kommt nicht **in** die Kränze (schweiz.; umg. *hat keine Aussicht auf Erfolg [auf Verwirklichung]*) | den Sarg **mit** Kränzen (¹) schmücken | ein ¹ **mit** Schleifen [Kerzen] | einen ² **von** Blüten um den Teller des Ehrengastes legen | ein ¹ **von** Rosen | ein ² **von** Seen umringt die Stadt | die Blumen **zu** einem ¹ **winden** | das Haar flechten und **zum** ² aufstecken

Example 9 from Wiegand (1996)

The model used in *Feste Wortverbindungen des Deutschen* and the proposals made in Wiegand (1996) with regard to the article structure could also to be considered by lexicographers of general language dictionaries with regard to the presentation and ordering of collocations. For lexicographers who wish to give more prominence to collocations, especially for text reception purposes, their presentation as clearly identifiable items is non-negotiable. The mere inclusion of a collocation as item form within a more comprehensive item, albeit that the collocation has been identified by a procedure of typographical enriching, will no longer suffice. In a general language dictionary, especially a printed dictionary, a separate subzone for each combination of the given word with partners from different part of speech categories may perhaps seem to be less advisable due to space restrictions, but the text production function of a dictionary may eventually compel lexicographers to distinguish between different categories of collocations and to order the collocations according to these categories. The typical criterion for a distinction between categories of collocations could be the part of speech of the collocates. The typological classification of the collocation, i.e. as single or complex collocation could also play a role. The inclusion and presentation of complex collocations will be discussed in a subsequent section.

As a first step to ensure a successful system where collocations are included in a specific search zone as secondary guiding elements, a thorough planning of a data distribution structure is needed with an article structure that allows the inclusion of a search zone dedicated to collocations. This could be a fully independent search zone dedicated exclusively to the presentation and treatment of collocations or it could be a subsection of a larger search zone reserved for items giving cotextual guidance, but within such a broader search zone the subzones for collocations and for example sentences should be clearly distinguishable. As is the case with example sentences, collocations need to be considered for each subcomment on semantics of an article in which the lemma sign represents a polysemous lexical item. Seeing that the inclusion of collocations will not be compulsory in each dictionary article because all words do not participate in collocational combinations this subzone can be regarded as part of an extended obligatory article structure.

The positioning of the search zone for items giving cotextual guidance, and therefore also of the zone or subzone giving collocations, needs to be determined by the nature of the microstructure opted for in a given dictionary. A semi-integrated microstructure, cf. Wiegand (1996), could be an ideal way of presenting and treating cotextual data, with the example sentences given as items in the integrated part and collocations occupying the text block presented in the non-integrated component. In the example article from Wiegand (1996) the cotextual items in the non-integrated section are arranged according to the part of speech of the words combining with the word represented by the lemma sign. This grammatical feature of the combining words is regarded as more significant than semantic features of the core word, because it determines the presentation of a new category whereas the different senses of the word represented by the lemma sign are not allocated to different semantically deter-

mined categories. Guidance regarding the applicable sense is conveyed by a superscript number given above the tilde functioning as place-keeping symbol to indicate where the word represented by the lemma sign occurs in a given cotextual item. The superscript numbers coincide with the numbers of the sub-comments on semantics in the integrated component of the article where paraphrases of meaning for the corresponding senses are accommodated. If a lexicographer wants to adapt the Wiegand example of a semi-integrated micro-structure for the inclusion and treatment of collocations, the arrangement in the non-integrated section needs to be done in such a way that the collocations are clearly identifiable as secondary guiding elements.

An adapted version of the approach followed by Wiegand could lead to the inclusion of a search zone for items giving example sentences in each one of the sub-comments on semantics in the integrated section and to phase all collocations out to the non-integrated section. This section could display a text block containing a subzone for each polysemous sense of the word represented by the lemma sign, with each subzone populated by the relevant non-grouped ordering of collocations, typographically marked as treatment units. Such a non-grouped ordering will enhance the visibility of collocations as fully-fledged secondary guiding elements that function as targets of procedures of non-lemmatic addressing. Whereas the lemma sign functions as primary guiding element of the article, each collocation in the text block is a secondary guiding element because it guides the user to items addressed at the specific collocation. The treatment directed at the collocations will at least demand the inclusion of a typical example sentence, given as cotextual guidance, or some specific morphosyntactic features. If needed a stylistic or other label may be included and even a brief paraphrase of meaning — if the meaning of the collocation deviates from the combined meaning of the constituent words.

5.2 Complex collocations

5.2.1 Guiding elements

The introduction of complex collocations as a subcategory of secondary guiding elements in the search zone for collocations could also be considered. Although such a presentation will certainly add value to the dictionary and could be considered as the most desired way of presenting complex collocations because of the explicit acknowledgement of these combinations as fully-fledged collocations, it might perhaps, in a general language dictionary, be regarded as belabouring this data type.

The inclusion of complex collocations remains important and lexicographers should negotiate the best possible way of presenting them and of making users aware of their existence. This could be done either as guiding elements in a subcategory of the search zone for collocations or in a more implicit way as part of the treatment of single collocations, typically within an example sentence illustrating the use of the single collocation but also its occurrence as component

of a complex collocation. If the latter route is followed the lexicographer has to ensure that the complex collocations can be identified as such by the target users of the specific dictionary. To achieve a presentation that will result in an unambiguous comprehension of such a cotextual item the lexicographer should embark on a system where different types of information can be retrieved from a single item form. This will demand presentation procedures not typically employed in dictionaries. This confronts both the metalexicographer who has to devise such a system and the practical lexicographer who has to implement it with definite challenges. The following section will deal with this aspect.

5.2.2 Embedded in cotextual items

The word *preference* combines with *personal* to form the collocation *personal preference*, and it combines with *express* to form the collocation *express a preference*. These two collocations can combine to form the complex collocation *to express a personal preference*. In the article of the lemma sign *preference* the search zone or subzone for collocations should include both these single collocations as secondary guiding elements and treatment units. The treatment of these collocations can include different items but needs to include at least one or more items illustrating their use. The following item giving an example sentence in which the complex collocation *to express a personal preference* is embedded could also be presented as part of the treatment of each one of these single collocations:

She tried to be impartial and did not want to express a personal preference.

Had this sentence been given in the search zone for example sentences, and not for collocations, in the article of the lemma sign *preference*, it would have been an illustration of the use of the single word. If this sentence is given in the search zone for collocations it can be an item illustrating the use of the collocation but also that of the word represented by the lemma sign by means of a procedure of remote addressing, cf. Louw and Gouws (1996). In the search zone for collocations this item is immediately addressed at the preceding secondary guiding element and it is therefore in the first instance an item illustrating the use of that specific collocation, i.e. either *express a preference* or *personal preference*. When given as a cotextual item addressed at either of the two collocations this item also illustrates the use of the complex collocation *to express a personal preference*. This item participates in a procedure of dual addressing, i.e. both lemmatic and non-lemmatic addressing, because it is addressed at the lemma sign by means of a procedure of remote addressing and it is addressed at the preceding collocation — a procedure of non-lemmatic addressing. In its non-lemmatic addressing it has both the single and the complex collocations as address.

The user of a dictionary that includes collocations and their treatment needs to be made aware of the fact that this example sentence contains two single collocations as well as a complex collocation. The presentation of an exam-

ple sentence from which the user can retrieve all this information demands a new type of microstructural entry. Such a single entry has to display a variety of items, i.e. 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 each one of the single collocations, i.e. *express a preference*, and *a personal preference* and an item giving an example of the use of the complex collocation *express a personal preference*. To include all these items in a single entry demands a high degree of textual condensation, perhaps too high for the average dictionary user. It is an entry devised to respect the space-saving endeavours of a given dictionary. The knowledgeable user who is familiar with the system of the dictionary will be able to retrieve the necessary information from the following data presentation:

*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 collocates are distinguished from each other by means of enrichment (the use of bold) and downward expansion (the use of underlining) — 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. The small caps, bold and underlining function as structural indicators that identify collocation components. This entry includes different items with different genuine purposes, combined into a single item form. The presentation could be indicated as follows — where single collocations in the example sentence do not have the word represented by the lemma sign in small caps and the collocate in bold, whereas this is done in the example sentences illustrating complex collocations. These sentences also have the second collocate underlined:

- preference
1. The selection of something or somebody over something or somebody else
He has a preference for designer clothes.
 2. The state of being selected over somebody or something else
Her preference for luxury hotels is no secret.
 3.
 4.
- COL:
1. **decided preference:** *They have a decided preference for classical music.*
 - **express a preference** (formal): *She tried to be impartial and did not want to **express a personal PREFERENCE***
 - **personal preference:** *She tried to be impartial and did not want to **express a personal PREFERENCE***
 2. ▪ ...
 3. ▪

Example 10

In the non-integrated component, introduced by the marker "COL." single collocations are given as guiding elements in different subcomments. Each single collocation is the address of at least an item giving an example sentence. Complex collocations are not entered as secondary guiding elements but are embedded in the example sentence given for a single collocation. In this example sentence the use of small caps, bold and underlining will identify the complex collocation.

This arrangement of collocations in a non-grouped text block does not differ much from the presentation in example 6 from the *MacMillan Dictionary*. However, the *MacMillan Dictionary* includes collocations in a text block within an integrated microstructure and not in a separate non-integrated text block.

This is more clearly noticeable in the following reduced version of the partial article of the lemma sign *defence*:

defence
NOUN ...
1
[UNCOUNTABLE] actions that you take to protect someone or something that is being attacked
come/rush to someone's defence: *Two of his friends came to his defence.*
.
.
a.
[UNCOUNTABLE] the system of weapons, equipment, and people that is used to protect a country
Huge amounts of government money are spent on defence.
an attempt to reduce the defence budget
The President has promised to increase defence spending.
defence cuts (=spending less money on defence): *More defence cuts are expected next year.*
b. ...
2
[COUNTABLE/UNCOUNTABLE] something that you say to support someone or something that is being criticized
defence of: *The Minister responded with a strong defence of the government's policy.*
speak/say something in someone's defence: *Several people spoke in my defence.*
in defence of: *Monroe has been fierce in defence of the present system.*
come/spring/leap to someone's defence: *Martin immediately sprang to her defence, saying she was innocent.*
.
:
b.
the defence
LEGAL the people in a court case who try to prove that someone is not guilty. The people who try to prove that someone is guilty are called the prosecution. The defence can be followed by a singular or plural verb
The defence will begin giving evidence tomorrow.

Example 11 from the *MacMillan Dictionary*

It should however be stated that in some articles the *MacMillan Dictionary* does give an additional search zone for collocations, as can be seen in the following

reduced partial article of the lemma sign *attack*:

attack ...

1

[INTRANSITIVE/TRANSITIVE] to use violence to harm a person, animal, or place

It was shortly before midnight when the terrorists attacked.

attack someone with something: *Two prison officers were brutally attacked with a knife.*

a.

to use weapons to try to defeat an enemy

The enemy attacked at dawn.

US jets attacked five air defence sites in the north.

.

.

2

[TRANSITIVE] to strongly criticize someone or something for their ideas or actions

Opponents attacked the government's plan to increase road tax.

attack someone/something for (doing) something: *Parliament has been attacked for failing to take action.*

attack something as something: *The deputy Prime Minister last night attacked the decision as foolish.*

.

▼ Collocates: attack

attack

▪ bitterly, fiercely, openly, sharply, strongly, vigorously

Example 12 from the *MacMillan Dictionary*

Here the search zone is clearly marked but the collocations are presented without any further treatment.

The inclusion of collocations in a separate non-integrated text block, separated from the example sentences, ensures a more salient position for this item type and clearly identifies the collocations as secondary guiding elements within a dictionary article. The treatment also conveys data regarding complex collocations — a data type lacking in the most general language dictionaries. In online dictionaries where space restrictions are not as stringent the search zone for collocations could make provision for a separate subzone for complex collocations. This will diminish the high degree of textual condensation prevailing in the proposed presentation for printed dictionaries.

6. In conclusion

Collocations are part and parcel of the daily language use of all speakers. Dictionaries often still fail to give a satisfactory presentation and treatment of collocations. The nature, extent and quality of the selection, presentation and treatment of collocations need to be improved. This demands allocating an elevated status to collocations, i.e. as guiding elements within their own search zones and as fully-fledged treatment units.

Acknowledgement

1. This research is (a) conducted within the SeLA project (Scientific e-Lexicography for Africa), supported by a grant from the German Ministry for Education and Research, administered by the DAAD and (b) supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85434).
2. The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.
3. I appreciate Ulrich Heid's valuable comments on a first draft of this paper.

7. Bibliography

7.1 Dictionaries

elexiko = <http://www.owid.de/wb/elexiko/start.html>

Häcki Buhofer, Annelies et al. (Eds.). 2014. *Feste Wortverbindungen des Deutschen*. Tübingen: Narr Francke.

HAT 5 = Odendal, F.F. and R.H. Gouws (Eds.). 2005. *Handwoordeboek van die Afrikaanse taal (e-HAT)*. Fifth edition. Cape Town: Pearson Education South Africa.

HAT 6 = Luther, J., F. Pfeiffer and R.H. Gouws (Eds.). 2015. *Handwoordeboek van die Afrikaanse Taal*. Cape Town: Pearson Education South Africa.

MacMillan = <http://www.macmillandictionary.com/open-dictionary/>

Oxford Online Dictionaries = <http://www.oxforddictionaries.com/>

7.2 Other literature

Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2: 1762-1832*. Berlin/New York: Walter de Gruyter.

Hausmann, F.J. 2004. Was sind eigentlich Kollokationen? Steyer, K. (Ed.). 2004. *Wortverbindungen — mehr oder weniger fest*. Institut für Deutsche Sprache Jahrbuch 2003, 2004: 309-334. Berlin/New York: De Gruyter.

Hausmann, F.J. 2008. Kollokationen und darüber hinaus. Einleitung in den thematischen Teil "Kollokationen in der europäischen Lexikographie und Wörterbuchforschung". *Lexicographica* 24: 1-8.

Heid, U. 1994. On Ways Words Work Together — Topics in Lexical Combinatorics. Martin, W. et al. (Eds.). 1994. *Euralex '94 Proceedings. Papers Submitted to the 6th EURALEX International Congress on Lexicography in Amsterdam, The Netherlands*: 226-257. Amsterdam: Vrije Universiteit.

Heid, U. 2011. German Noun+Verb Collocations in the Sentence Context: Morphosyntactic Proper-

- ties Contributing to Idiomaticity. Herbst, T., S. Faulhaber and P. Uhrig (Eds.). 2011. *The Phraseological View of Language*: 283-307. Berlin: Mouton de Gruyter.
- Heid, U. and R.H. Gouws.** 2006. A Model for a Multifunctional Dictionary of Collocations. Corino, E., C. Marelllo and C. Onesti (Eds.). 2006. *Proceedings XII EURALEX International Congress, Turin, Italy, September 6th–9th, 2006*: 979-988. Alessandria: Edizioni dell'Orso.
- Louw, P.A. and R.H. Gouws.** 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.
- Spohr, D.** 2005. *A Description Logic Approach to Modelling Collocations*. Diploma Thesis. Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS).
- Tutin, A.** 2008. For an Extended Definition of Lexical Collocations. Bernal, E. and J. DeCesaris (Eds.). 2008. *Proceedings of the XIII EURALEX International Congress, Barcelona, 15–19 July 2008*: 1453-1460. Sèrie Activitats 20. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada.
- Tutin, A. and F. Grossmann.** 2002. Collocations régulières et irrégulières: esquisse de typologie du phénomène collocatif. *Revue française de Linguistique appliquée*. Lexique: recherches actuelles VII(1): 7-25.
- Wiegand, H.E.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger lexikographischen Kolloquium*: 1-82. Lexicographica. Series Maior 70. Tübingen: Max Niemeyer.
- Wiegand, H.E. and R.H. Gouws.** 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography: 273-314. Berlin/New York: De Gruyter.
- Zinsmeister, H. and U. Heid.** 2003. Significant Triples: Adjective+Noun+Verb Combinations. Kiefer, F. and J. Pajzs. (Eds.). 2003. *Proceedings of COMPLEX 2003: 7th Conference on Computational Lexicography and Text Research*: 92-101. Budapest: Linguistics Institute, Hungarian Academy of Sciences.