Stellenbosch Papers in Linguistics Plus, Vol. 61, 2021, 1-16
doi: 10.5842/61-0-913

’n Verbeterde leksikografiese dataverspreiding- en inligtings-onttrekkingstruktuur

Rufus H. Gouws

Department Afrikaans en Nederlands, Universiteit Stellenbosch, Suid-Afrika
E-pos: rhg@sun.ac.za

Abstrak
Die dinamika in die leksikografie noop leksikograwe om voortdurend daarop bedag te wees dat veranderinge in hulle woordeboeke aangebring moet word. Hierdie veranderinge lê op die vlak van die woordeboekinhoud, die leksikografiese funksies asook die woordeboekstrukture. Die teikengebruikersgroep van elke woordeboek en hulle leksikografiese behoeftes en naslaanvaardighede moet deurgaans in gedagte gehou word. In hierdie artikel word verskillende aspekte van leksikografiese prosesse bespreek, met spesifieke verwysing na die aanpassing in dataverspreidingstrukture, die uitbreiding van soekplekke en die oorgang van woordeboeke vanaf die stootmedium na die trekmedium. Vir gedrukte woordeboeke word die benutting van ’n semi-geïntegreerde mikrostruktuur voorgehou om ’n beter onderskeid in die aanbieding van voorbeeldsinne en kollokasies moontlik te maak. Vir aanlyn woordeboeke is die pleidooi om die daarstelling van ’n uitgebreide soekheelal asook die inwerkingstelling van ’n datatrekkingstruktuur. So ’n struktuur bied aan woordeboekgebruikers vanuit die woordeboek toegang tot onder meer die internet asook die korpus van die betrokke woordeboek. Woordeboeke verander van eindpuntbronne na tussengangerbronne. Dit bied veel meer moontlikhede aan gebruikers, onder andere taalpraktisyns, om naas die data, onder meer koteksaanduiders, in woordeboeke ook ander data en veral voorbeelde van natuurlike taalgebruik in die woordeboek-eksterne bronne te vind.

Sleutelwoorde: datatrekkingstruktuur, inligtingsonttrekking, kollokasies, koteksaanduiders, mikrostruktuur, taalpraktisyns, woordeboekfunksies

Abstract
The dynamics in lexicography compel lexicographers to constantly be aware of the changes that need to be made in their dictionaries. These changes are on the level of the contents of dictionaries, lexicographic functions as well as dictionary structures. The target user group of every dictionary, their lexicographic needs and their reference skills should always be taken into account. In this article various aspects of lexicographic processes are discussed, with specific reference to adaptations in data distribution structures, the expansion of search venues and the transfer from dictionaries as push mediums to dictionaries as pull mediums. For printed dictionaries the use of a semi-integrated microstructure is proposed in order to ensure a better distinction in the presentation of example sentences and collocations. For online dictionaries the proposal is made to introduce a comprehensive search universe as well as a data-pulling

© 2021 The authors. This work is licensed under the Creative Commons Attribution 4.0 International License.

structure. Such a structure offers dictionary users the opportunity to access the internet as well as the dictionary corpus from within a given dictionary. Dictionaries are transformed from destination sources to transfer sources. This offers users, including language practitioners, many more possibilities to find in the dictionary-external sources, besides the data in dictionaries, including cotextual data, also other data and especially occurrences of natural language use.

Keywords: data-pulling structure; collocations; information retrieval; items giving the co-text; language practitioner; microstructure; dictionary functions

1. Inleiding

Woordeboeke is praktiese gebruiksinstrumente wat saamgestel is om in bepaalde gebruiksituasies antwoorde te verskaf op vrae van die bepaalde teikengebruikersgroep. Elke woordeboek moet 'n werklike doel hê. Die werklike doel is om die geïdentifiseerde teikengebruiker daartoe in staat te stel om aan die hand van data in die woordeboek inligting te kan onttrek ter bevrediging van daardie leksikografiese behoefte wat die betrokke woordeboekraadpleging geaktiveer het – vergelyk Wiegand (1998:299) in hierdie verband. Omdat verskillende gebruikers in verskillende gebruiksituasies verskillende inligtingsbehoeftes het, is die mite van **dié** woordeboek wat 'n enkele bron is wat alle leksikografiese behoeftes van alle lede van 'n taalgemeenskap kan oplos reeds lank in die ban gedoen. Selfs met die beplanning van 'n woordeboek wat op 'n duidelik omskrewe teikengebruikersgroep gerig is, is die vassteling van die tersaaklike behoeftes nie probleemloos nie. In hierdie verband sê Hanks (1979:35):

> The lexicographer is in the impossible position of a man who undertakes to answer peoples' questions, but since he does not know at the time of compilation what questions exactly his public will ask, he has to word his entries so as to answer all possible questions about them.

Die sterk gerigtheid in die moderne leksikografie, in sowel die teorie as die praktyk, op 'n geïdentifiseerde teikengebruiker bring mee dat daar veel meer moeite gedoen word om in die tipologiese verskeidenheid wat die leksikografiepraktyk kenmerk ook neerslagplek te bied aan 'n verskeidenheid datatipes en bewerkingsprosedures.

Dit is nie net belangrik dat gepaste woordeboeke beskikbaar moet wees nie, maar hulle moet ook toeganklik wees sodat dit moontlik is om 'n kitsonttrekking van inligting aan die data in die betrokke woordeboek uit te voer. Eweneens belangrik is dit dat die aangebode data gerig is op die bevrediging van die teikengebruikers se leksikografiese behoeftes. Met die beplanning en samestelling van enige woordeboek is dit noodsaaklik dat die leksikograaf moet weet wie die beplande teikengebruikers van daardie woordeboek sal wees en vertroud moet wees met die leksikografiese behoeftes en naslaanvaardighede van dié spesifieke gebruikersgroep. Dat leksikograwe dikwels 'n bepaalde gebruikersgroep as teikengebruikers van hulle woordeboek voorhou sonder dat die behoeftes van daardie gebruikers hoegenaamd of geensins voldoende nagevors is nie, het daartoe gelei dat Wiegand (1977:59) na die gebruiker verwys het as die "bekende vreemdeling". Almal weet wie dit is, maar niemand ken hom/haar nie.

In die samestelling van woordeboeke word daar aanvaar dat die betrokke woordeboek, afhangend van die woordeboeksoort, ook hulp moet bied met beslissings oor taalgebruik. Taalpraktisyns, waarby vertalers ingesluit word, is 'n gebruikersgroep wat woordeboeke op 'n daaglikse basis gebruik en dikwels afhanklik is van die leiding wat woordeboeke bied om hulle opdragte suksesvol te kan uitvoer. Taalpraktisyns is kundige woordeboekgebruikers wat weet wanneer om watter woordeboek te raadpleeg en hoe om die korrekte inligting aan die data in 'n woordeboekartikel te onttrek. Hulle het bepaalde verwagtinge wanneer hulle woordeboeke raadpleeg en kom ongelukkig te dikwels bedroë van 'n woordeboekraadplegingsessie af. Woordeboeke bied te dikwels nie die leksikografiese hulp en leiding wat hierdie gebruikersgroep – en ander gebruikersgroepe – nodig het nie.

Taalpraktisyns benut woordeboeke om hulp met betrekking tot 'n verskeidenheid tipes inligtingsbehoeftes, en maak daarop staat dat woordeboeke dit ten doel het om verskillende tipes leksikografiese funksies te bevredig. Hier gaan dit veral om die kommunikatiewe funksies van teksbegrip, teksproduksie, vertaling en redigering, en in 'n mindere mate die kognitiewe funksie. Vergelyk in hierdie verband onder meer Tarp (2000; 2008). Omdat taalpraktisyns dikwels onder tydsdruk werk en 'n enkele bron wil gebruik om oplossings vir verskillende probleemtipes te kry, is 'n polifunksionele woordeboek aan te beveel bo 'n monofunksionele woordeboek. Dit gaan egter nie net om die verskeidenheid funksies wat 'n woordeboek vir taalpraktisyns moet hê nie, maar ook om die datatipes en veral die doelmatige aanbieding van die betrokke aanduiders. Die tipiese woordeboekartikel moet aanduiders en struktuurmerkers bied, vergelyk Wiegand (1989:427), en daardie struktuurmerkers, sowel tipografiese as nietipografiese struktuurmerkers, moet die interne toegangsroete na die gewenste aanduider help bevestig. 'n Kitstoegang tot die gewenste data is noodsaaklik.

In hierdie artikel word daar vanuit die teoretiese leksikografie enkele voorstelle gedoen ter verbetering van die dataverspreiding- en inligtingsonttrekkingstruktuur van woordeboeke; ook met die oog op 'n verbeterde leksikografiese bewerking ter wille van taalpraktisyns as 'n woordeboekgebruikersgroep. Van die aspekte wat hier bespreek word, geld vir sowel gedrukte as aanlyn woordeboeke, maar die uiteindelike klem is sterker op aanlyn woordeboeke. 'n Datatipe waarna herhaaldelik verwys sal word, is koteksdata, dit wil sê voorbeeldmateriaal en kollokasies. Na 'n kort bespreking van koteksaanduiders verskuif die fokus na verskillende woordeboekstrukture. Daarna word voorstelle gedoen vir 'n vernuwende benutting van bepaalde leksikografiese strukture om beter leiding ten opsigte van die verkryging van onder andere koteksmateriaal te bewerkstellig.

2. Leksikografiese funksies en kotekshulp

In sowel een- as twee- en meertalige algemene woordeboeke is dit die taak van die leksikograaf om data te verskaf waaraan die teikengebruiker daardie inligting kan onttrek wat nodig is om sy/haar leksikografiese behoeftes te bevredig. Taalpraktisyns is 'n gebruikersgroep in wie se dagtaak woordeboeke dikwels 'n onontbeerlike rol speel. Vir hierdie gebruikersgroep is dit belangrik dat die data in die woordeboeke wat hulle gebruik én relevant én korrek moet wees en daar moet min twyfel wees oor die gesag van die woordeboeke om 'n getroue weerspieëling van die werklike taalgebruik te bied. Hierdie gesag is nie gesetel in 'n preskriptiewe benadering waarvolgens die leksikograaf slegs voorskryf wat hy/sy as korrekte of puristiese taalgebruik beskou nie. Veel eerder moet woordeboeke 'n deskriptiewe of, verkieslik, 'n proskriptiewe benadering hê. Teenoor die voorskriftelikheid van 'n preskriptiewe benadering waar slegs een

vorm as korrek en aanvaarbaar voorgehou word, lei ’n deskriptiewe benadering daartoe dat die woordeboek vir taalvariasie voorsiening maak en daar selfs in die geval van individuele woorde en uitdrukkings variante vorme verstrek word. In ’n deskriptiewe benadering word voorkeurvorme nie gemerk nie. ’n Proskriptiewe benadering bied die soort variasie wat ’n deskriptiewe benadering kenmerk, maar bied as bykomende leiding ’n aanduiding van watter inskrywing as voorkeurvorm beskou kan word – hetsy volgens die betrokke leksikograaf hetsy volgens die geldende taalnorme. Vergelyk Bergenholtz & Gouws (2010) vir ’n uitvoeriger bespreking hiervan.

Die weerspieëling wat algemene woordeboeke van die leksikon bied, moet verteenwoordigend wees van die werklike taalgebruik. Die spesifieke leksikografiese funksies wat vir ’n bepaalde woordeboek vasgestel word, sal bepalend wees vir die aard en omvang van datatipes wat in daardie woordeboek aangebied word. Alhoewel daar soms beweer word dat monofunksionele woordeboeke as voorkeur- leksikografiese produkte beskou moet word, dui die eise van die leksikografiepraktyk veel eerder op die wye aanvraag na, beskikbaarstelling en benutting van bi- en multifunksionele woordeboeke.

In die literatuur oor leksikografiese funksies word daar verskillende tipes funksies onderskei, onder meer die kommunikatiewe funksies van teksbegrip, teksproduksie, vertaling en redigering, die kognitiewe funksie asook die interpretatiewe en operasionele funksies. Vergelyk in hierdie verband onder meer Tarp (2000; 2008), Bothma, Bergenholtz & Gouws (2011) en Agerbo (2017).

In hierdie artikel is die fokus onder meer op koteksinskrywings in algemene verklarende en tweetalige woordeboeke en die bydrae van hierdie aanduidertipe tot die bereiking van veral die kommunikatiewe funksies van teksbegrip, teksproduksie en vertaling. Aandag word gegee aan die tipes data wat as koteksaanduiders aangebied word en veral aan die onderliggende leksikografiese strukture wat nodig is om koteksaanduiders se daarstelling in sowel gedrukte as aanlyn woordeboeke so doeltreffend en toeganklik as moontlik te maak.

Taalpraktisyns is dikwels op soek na data wat leiding kan gee ten opsigte van die tipiese gebruik van ’n bepaalde woord. Hierdie verlangde leiding kan onder meer ten opsigte van gebruiksvoorbeelde, register en kollokasies wees. Die funksie van ’n woordeboek bepaal die aard en omvang van datatipes wat opgeneem word. Koteksaanduiders is tipiese inskrywings in die artikels van woordeboeke met die kommunikatiewe funksies van teksproduksie en vertaling, alhoewel dit ook met teksbegrip kan help. Die onvoldoende aanbieding van koteksaanduiders in talle woordeboeke kan reggestel word deur skakels na woordeboek-eksterne bronne in te bou. Dit word later in hierdie artikel bespreek.

3. Koteksaanduiders en verskillende tipes mikrostrukture

Taalpraktisyns raadpleeg woordeboeke dikwels om koteksaanduiders te vind wat aan hulle leiding kan bied ten opsigte van woord- en termgebruik. Tans maak die tipiese leksikografiese bewerking in vakwoordeboeke nie naastenby, indien enigsins, voorsiening vir koteksinskrywings nie. Dit is veral kollokasies wat in die bewerking van vakterme van groot hulp vir woordeboekgebruikers kan wees. Aan hierdie leksikografiese probleem word daar nie in hierdie artikel aandag gegee nie, aangesien hierdie artikel se primêre fokus nie vakwoordeboeke is nie. Algemene verklarende en tweetalige woordeboeke bied wel hulp in

hierdie verband, maar slegs 'n beperkte aantal koteksaanduiders word gewoonlik verstrek. Gegee die tipiese behoeftes van gebruikers van algemene woordeboeke asook die ruimteprobleme in gedrukte woordeboeke, is die beperkte weergawe van koteksaanduiders te verstane. Vir woordeboekgebruikers, onder meer taalpraktisyns, wat woordeboeke raadpleeg om teksproduksiehulp te kry, is die huidige aanbieding nie voldoende nie. Dit is 'n aspek waar 'n beter vertroudheid van leksikograwe met voorstelle vanuit die teoretiese leksikografie die gebruiker tot voordeel kan wees. Veral 'n goed gemotiveerde keuse van die tersaaklike mikrostruktuur van 'n bepaalde woordeboek kan die gebruiker baie help. Vergelyk in hierdie verband onder meer Hausmann & Wiegand (1989), Wiegand (1996) en Gouws (2003).

Vir teksproduksie is dit veral geïntegreerde mikrostrukture wat die gebruik van primitiewe mikrostrukture kan vervang, vergelyk Gouws & Prinsloo (2005:141). Hierdie tipe mikrostruktuur kom dikwels in veral tweetalige woordeboeke voor. Wiegand (1996) se voorstelle vir 'n semi-geïntegreerde mikrostruktuur bied beslis moontlikhede vir 'n verbeterde aanbieding van en onderskeiding tussen voorbeeldsinne en kollokasies. In aansluiting by Wiegand (1996) stel Gouws (2015:184) voor dat in 'n woordeboek met 'n semi-geïntegreerde mikrostruktuur die voorbeeldsinne as aanduiders in die geïntegreerde komponent moet optree. In die geval van lemmas wat polisemiese leksikale items verteenwoordig, sal die betekenisparafrase of vertaalekwivalent, na gelang van die woordeboeksoort, telkens in 'n eie semantiese subkommentaar optree en in dieselfde integraat onmiddellik deur een of meer voorbeeldmateriaalaanduiders gevolg word. In die niegeïntegreerde teksblok wat na afloop van die semantiese subkommentare verstrek word, kan daar 'n geordende aanbieding van kollokasies wees. Daar kan vir sowel enkel- as komplekse kollokasies voorsiening gemaak word en die kollokasies kan tot sekondêre bewerkingseenhede verhef word waarvoor daar telkens onder meer voorbeeldmateriaal verstrek word. So 'n ruim voorsiening van koteksmateriaal met die gepaardgaande subverdeling tussen voorbeeldsinne en kollokasies sal vir taalpraktisyns van groot hulp wees. Hierdie tipe mikrostruktuur verplig die leksikograaf om ook nuut te dink oor die dataverspreidingstruktuur van die woordeboek.

4. Oor die dataverspreidingstruktuur in woordeboeke

Die dataverspreidingstruktuur van 'n woordeboek, vergelyk Bergenholtz, Tarp & Wiegand (1999), bepaal die sistematiese plasing van die data wat gekies is vir 'n bepaalde woordeboek. Die verskillende tekstuele posisies wat 'n woordeboek as 'n draer van tekssoorte vir hierdie plasing bied, is die soeksones in die artikels van die sentrale lys, asook die buitetekste in die voor- en agtertekste-afdelings van die woordeboek. Deur 'n semi-geïntegreerde mikrostruktuur met sy bykomende teksblok te gebruik, brei die leksikograaf die toepassingsgebied van die dataverspreidingstruktuur uit. Woordeboekgebruikers baat by die benutting van woordeboeke waar die dataverspreidingstruktuur 'n groter toepassingsterrein het. So 'n uitbreiding van die toepassingsterrein vind 'n mens onder meer in die benutting van 'n raamstruktuur, vergelyk Kammerer & Wiegand (1998), met buitetekste wat akkommodasie aan verskillende datatipes kan bied.

In 'n gedrukte woordeboek maak 'n benutting van buitetekste dit vir die leksikograaf moontlik om leksikografiese data in 'n woordeboek in te sluit waarvoor daar nie 'n heenkome sou wees in die verstekweergawe van die betrokke woordeboekartikels nie. Daarbenewens kan tipes tekste in die buitetekste-afdeling ingesluit word wat nie die tradisionele woordeboekinhoud weerspieël nie, maar wat volgens die leksikograaf se oordeel 'n toegevoegde waarde aan die teikengebruiker sal bied. Waar 'n woordeboek 'n uitgebreide dataverspreidingstruktuur vertoon,

dit wil sê waar die sentrale woordelys aangevul word deur buitetekste, kan die leksikograaf gelei word deur die eiesoortige behoeftes van die voornemende woordeboekgebruikers en dienooreenkomstig buitetekste insluit wat op hierdie behoeftes gerig is. ’n Algemene tweetalige gesinswoordeboek soos Grobbelaar et al. (1987) se *Reader's Digest Afrikaans-Engelse Woordeboek/English-Afrikaans Dictionary* bied ’n verskeidenheid buitetekste wat gerig is op die kommunikatiewe behoeftes van verskillende lede van ’n gesin. Dit sluit tekste in met riglyne oor die skryf van verskillende tipes briewe, aanspreekvorme, plekname en persoonsname. Daarnaas is daar verskillende geïllustreerde tekste waarin temas oor die liggaam, die huis, die werkplek, vleissnitte, die motor en sport behandel word met prente en tweetalige woordelyste. ’n Transtekstuele benadering tot leksikografiese funksies, vergelyk Gouws & Steyn (2005), bring mee dat die buitetekste daartoe bydra om die teikengebruikers van teksbegrip- en teksproduksiehulp te voorsien en ook help om die woordeboek se kognitiewe funksie te bevredig. In die *SASOL: Science & Technology Resource* (Hartmann-Petersen et al. s.j.), ’n vakwoordeboek wat daarop gerig is om hoërskoolleerders met wetenskapsverwante vakke te help, is daar buitetekste met die periodieke tabel, lyste smelt- en kookpunte van sommige elemente, asook SI-eenhede. Hierdie buitetekste weerspieël die mate waarin die leksikograaf die spesifieke behoeftes van die geïdentifiseerde teikengebruiker in ag neem.

’n Uitgebreide dataverspreidingstruktuur kan ook tot die skep van ’n geïntegreerde leksikografiese produk lei wat, byvoorbeeld in die geval van vakwoordeboeke, die sentrale woordelys aanvul met tekste wat bykomende, toeligtende of kontekstualiserende vakdata verstrek. Die woordeboek-interne mediostruktuur lê dan ook kruisverwysingsverbande tussen inskrywings in die soeksone van ’n artikel, die artikel as soekgebied, die woordelys as soekveld, vergelyk Wiegand, Beer & Gouws (2013:63), en die res van die soekstreek, dit is die woordeboek as geheel, dit wil sê die woordelys saam met alle buitetekste – vergelyk ook Gouws (2018a). In die vakleksikografie word daar tussen verskillende gebruikersgroepe ondersklei, te wete leke, semideskundiges en deskundiges – vergelyk Bergenholtz & Tarp (1995). Die leksikografievakwoordeboek *Wörterbuch zur Lexikographie und Wörterbuchforschung/ Dictionary of Lexicography and Dictionary research* (Wiegand et al. 2010) is saamgestel met semideskundiges en deskundiges as gebruikers. Dié produk is eerstens ’n vakwoordeboek, maar wil ook aan die spesifieke teikengebruikers aanvullende vakinligting bied. Twee van die tekste in die voortekste-afdeling van die eerste deel van hierdie woordeboek heet “Systematische Einführung” en “Systematic Introduction”. Albei hierdie tekste strek oor meer as 100 bladsye en bied ’n kort oorsig oor ’n hele aantal relevante onderwerpe uit die teoretiese leksikografie. Vir ’n student in die metaleksikografie is hierdie boek dus én ’n vakwoordeboek én ’n kompakte vakhandboek. In die woordeboekartikels van hierdie woordeboek is daar verskillende soeksones met aanduiders van bronne vir verdere naslaanwerk oor ’n bepaalde term. Een soeksone bevat spesifiek inskrywings wat die gebruiker verwys na bepaalde afdelings in die sistematiese inleiding waar ’n uitvoeriger bespreking van die betrokke term ter sprake kom. In hierdie woordeboek laat blyk die wisselwerking tussen woordeboekartikels en buitetekste dat die leksikograwe in hulle beplanning spesifiek aandag gegee het aan die behoeftes van die teikengebruiker van die betrokke woordeboek.

Op ’n vergelykbare manier kan taalpraktisyns gehelp word met woordeboeke wat in die sentrale woordelys leksikografiese bewerkings bied wat die soort data daarstel waaraan taalpraktisyns ’n behoefte het. Daarbenewens kan buitetekste ingesluit word, ook met kruisverwysings tussen sentrale teks en buiteteks, om aanvullende hulp te bied.

As deel van die leksikografiese proses van enige woordeboek, moet die leksikograaf in die woordeboekkonseptualiseringsfase aandag gee aan die beplande inhoud van die woordeboek, asook die dataverspreidingstruktuur. Eweneens moet die leksikograaf aandag gee aan die beplande toegangstruktuur van die woordeboek en ook die ondersteuning wat aan voornemende gebruikers gebied word om nie net die data te bereik nie, maar ook die inligting daaraan te kan onttrek. Alhoewel twee gebruikers toegang tot dieselfde data in 'n woordeboek het, onttrek hulle nie noodwendig dieselfde inligting daaraan nie omdat 'n gebruiker se vertroudheid met die stelsel van 'n bepaalde woordeboek daarvoor bepalend kan wees of die gebruiker 'n optimale inligtingsonttrekking kan behaal.

Van kundige woordeboekgebruikers, soos taalpraktisyns, kan dit aanvaar word dat hulle die stelsel van daardie woordeboeke wat hulle daaglikse gebruiksinstrumente is, genoegsaam ken om suksesvolle raadplegingsprosedures te verseker. Desondanks bied woordeboeke nie altyd al daardie data wat nodig is om die onderskeie behoeftes te bevredig nie. Kotekshulp is 'n datatipe waaraan leksikograwe met groter sorg kan aandag gee. In die hieropvolgende afdelings word aandag gegee aan 'n nuwe manier om gebruikers beter te help met betrekking tot die verkryging van koteksleiding in veral aanlyn woordeboeke. Daar word eers na enkele tersaaklike woordeboekstrukture verwys voordat kotekshulp ter sprake gebring word.

5. Dataverspreiding en bron-identifisering

Die aanlyn omgewing bied vernuwende moontlikhede vir sowel dataverspreiding as inligtingonttrekking in leksikografiese produkte. Die toepassingsterrein van die dataverspreidingstruktuur van enige gegewe woordeboek, ongeag die medium, word bepaal deur die verskillende soekplekke waar die data geplaas kan word. In 'n gedrukte woordeboek is hierdie soekplekke die soekstreek en daarbinne die soekveld, die soekgebied en die soeksone. Aanlyn woordeboeke word dikwels nie slegs as enkelprodukte ter beskikking van gebruikers gestel nie, maar tree dikwels saam met ander woordeboeke in 'n woordeboekportaal op. Waar die samestelling van die verskillende woordeboeke in een portaal gelyklopend geskied of opeenvolgend geskied maar onderhewig aan 'n oorkoepelende woordeboekkonseptualiseringsplan as deel van 'n omvattende leksikografiese proses, kan 'n komplekse dataverspreidingstruktuur ten opsigte van die hele woordeboekportaal geformuleer word. Ondergekik daaraan kan enkeldataverspreidingstrukture vir elke afsonderlike woordeboek geformuleer word. Met die uitvoering van die dataverspreiding van elke individuele woordeboek het die leksikograaf bykomende plekke, die ander woordeboeke in die woordeboekportaal, waar die data geplaas kan word. Die woordeboekportaal bied dan 'n soekheelal, vergelyk Gouws (ter perse), waarbinne elke datatipe 'n neerslagplek in een of meer woordeboeke kan kry. 'n Deel van die leksikografiese proses binne een woordeboek is dan om data wat nie noodwendig binne die verstekbewerking van die betrokke woordeboek val nie, maar waaraan die teikengebruiker dalk 'n behoefte mag hê, in 'n ander woordeboek van die portaal te plaas. Met behulp van die portaal-interne mediostruktuur moet daar dan wel 'n merker wees wat die gebruiker na die betrokke soekstreek lei waar die tersaaklike data-inskrywing as aanduider verstrek word. Woordeboekportale bevat dikwels woordeboeke wat na afloop van hulle voltooiing in die woordeboekportaal geplaas is en wat nie in afstemming op mekaar as deel van 'n oorkoepelende leksikografiese proses saamgestel is nie. Op die openingskermskoot van so 'n portaal behoort 'n spesifieke tipe identifiseringsinskrywing, te wete soekstreekmerkers, in 'n lysting van die verskillende woordeboeke in die portaal te verskyn sodat die gebruiker van een woordeboek na 'n ander, dit wil sê van een soekstreek na 'n ander,

kan beweeg. Figuur 1 en 2 bied die openingskermskote van onderskeidelik *HAT en Longman Aanlyn* en *OWID* (die *Online Wortschatz-Informationssystem Deutsch*). In albei gevalle is daar ’n aanduiding van watter ander woordeboeke ook in die portaal beskikbaar is.

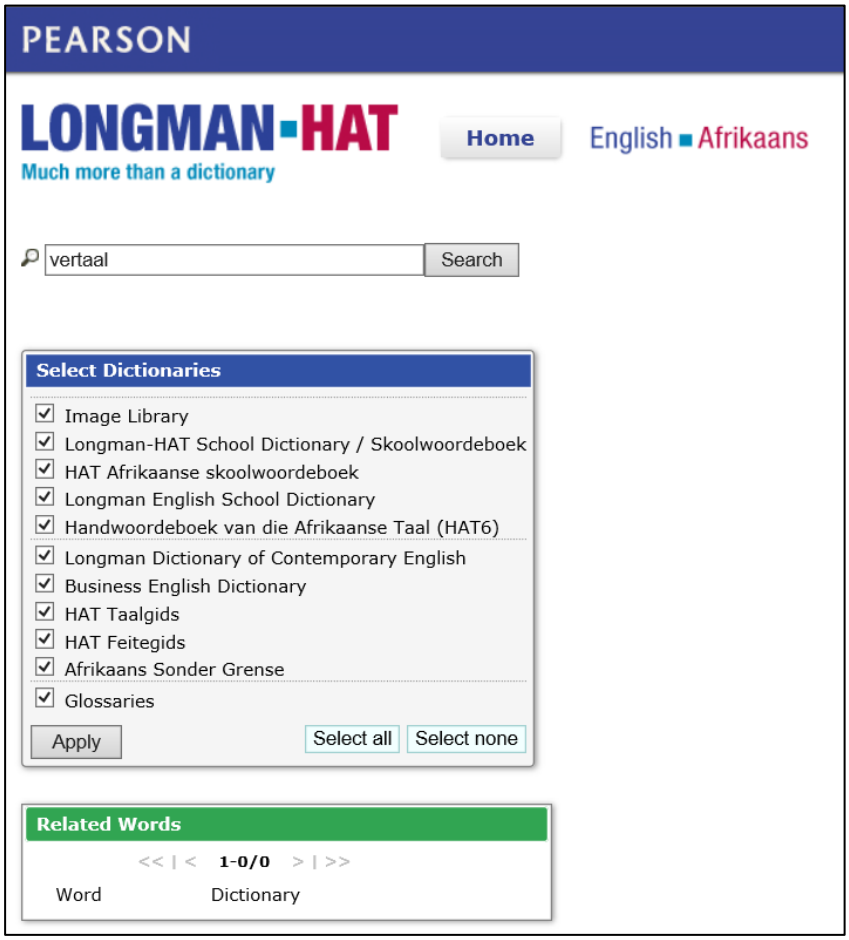


Figuur 1: *HAT en Longman Aanlyn*



Figuur 2: *OWID*

Uit figuur 3 is dit duidelik dat wanneer 'n gebruiker 'n bepaalde woord in die soekkassie van *HAT en Longman Aanlyn* intik, hy/sy ook kan aandui of die woord net in een woordeboek, in meerdere woordeboeke of in al die woordeboeke in die portaal gesoek moet word. Al is die woordeboeke in die portaal nie as deel van 'n oorkoepelende leksikografiese proses saamgestel nie, is daar wel 'n integrasie van bronne wat kitstoegang tot verskillende leksikografiese bewerkinge van dieselfde lemma lei.



Figuur 3: *HAT en Longman Aanlyn*

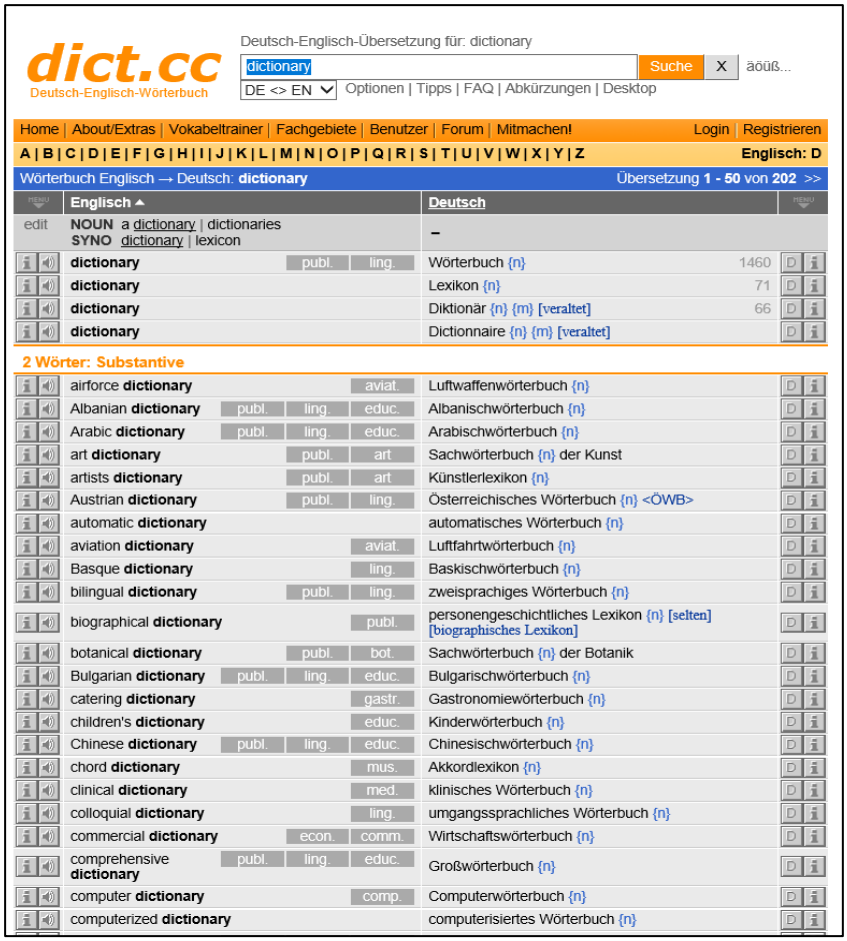
Met behulp van 'n woordeboekportaal kan die gebruiker 'n uitgebreider en meer gedifferensieerde bewerking van dieselfde woord kry en sodoende kan 'n uitvoeriger inligtingsoordrag bewerkstellig word.

6. Inligtingonttrekking

Suksesvolle inligtingonttrekking aan woordeboekdata voorveronderstel die aanwending van 'n sistematiese dataverspreidingstruktuur, asook die gebruik van data-identifiseringsinskrywings.

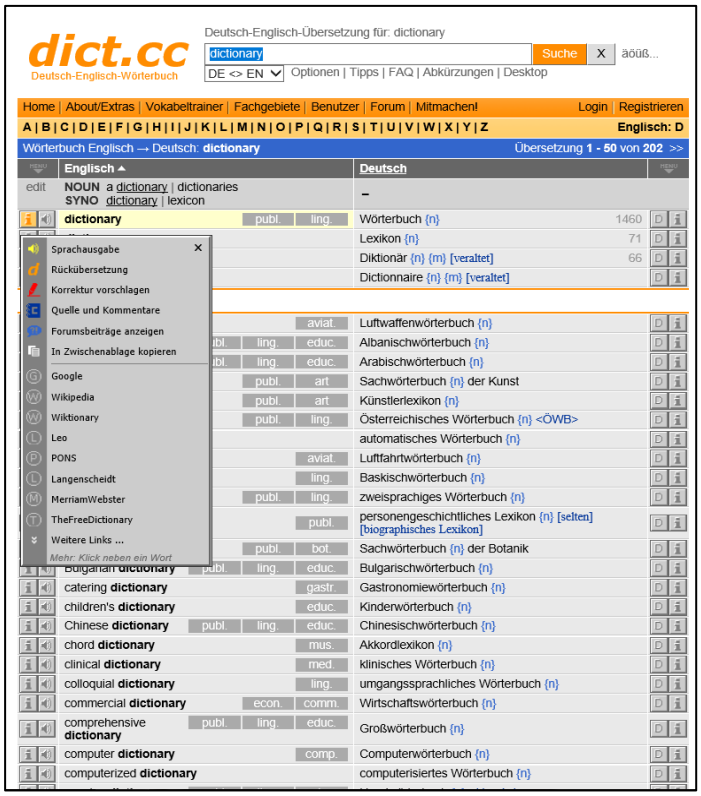
Naas die identifisering van woordeboek-eksterne bronne in 'n woordeboekportaal gaan sommige aanlyn woordeboeke reeds 'n stap verder deur gebruikers nie net na ander woordeboeke te verwys nie, maar selfs na die internet as 'n bykomende deel van die soekheelal. Sulke woordeboek- en woordeboekportaal-eksterne verwysings bring nuwe moontlikhede vir die vind van meer data – en ook beter kotekshulp. Die woordeboek *dict cc* bied aan gebruikers

nie net toegang tot dié woordeboek se eie bewerking nie, maar bevat ook data-identifiseringsinskrywings wat die gebruiker na woordeboek-eksterne bronne lei. Figuur 4 bied ’n weergawe van die skermskoot wanneer die woord *dictionary* in *dict cc* nageslaan word.



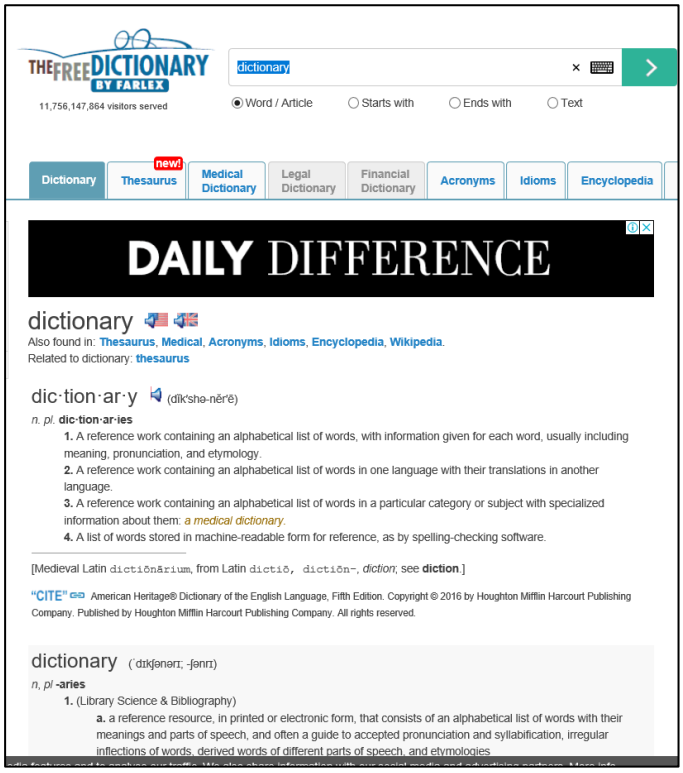
Figuur 4 uit *dict cc*

Indien die gebruiker behoefte aan ’n bykomende bewerking het, kan hy/sy op die inligtingsikoon net langs die woord klik. Figuur 5 wys die lysie woordeboek-eksterne bronne wat dan tot die gebruiker se beskikking is.



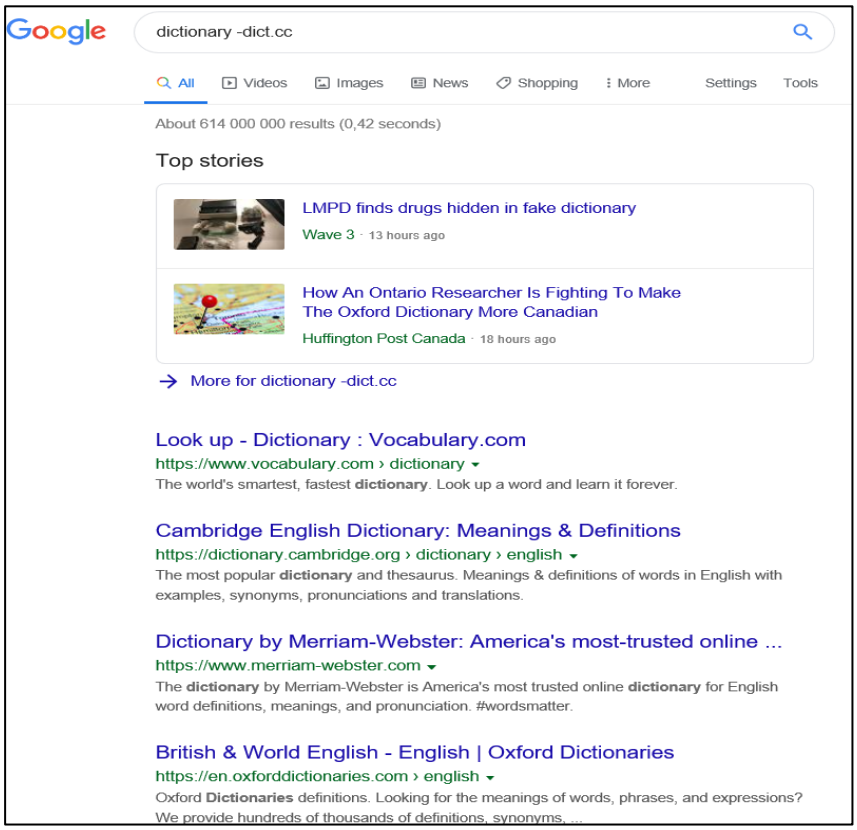
Figuur 5 uit *dict cc*

'n Klik op, byvoorbeeld, die bronidentifiseringsmerker *TheFreeDictionary* lei die gebruiker na die lemma *dictionary* in hierdie woordeboek, soos voorgestel in figuur 6:



Figuur 6 uit *TheFreeDictionary* (via *dict cc)*)

'n Klik op die bronidentifiseringsmerker *Google* lei die gebruiker regstreeks na die internet, soos gesien kan word in figuur 7 waar toegang tot verdere relevante data beskikbaar is:



Figuur 7 uit Google (via *dict cc*)

Die benutting van die internet as 'n woordeboek-eksterne bron waarna die gebruiker vanuit die woordeboek verwys word, verander die status van woordeboeke vanaf eindpuntbronne na tussengangerbronne. Soms sal 'n gebruiker die verlangde inligting aan die data in die woordeboek kan onttrek, maar gebeur dit nie word die gebruiker vanuit die woordeboek as tussengangerbron na 'n woordeboek-eksterne bron verwys. Saam met ander woordeboeke in die woordeboekportaal vorm die internet 'n bykomende komponent van die soekheelal.

7. Datatrekking

Waar 'n woordeboekgebruiker op soek is na bykomende hulp, lei die soort leksikografiese hulp wat *dict cc* bied tot vernuwende moontlikhede. Die internettoegang vanuit die woordeboek verander die tradisionele leksikografiese naslaanproses en stel leksikograwe en gebruikers voor verskeie uitdagings. In tradisionele woordeboeke besluit die leksikograaf wat aan die gebruiker gebied word en die gebruiker het geen keuse in dié verband nie. Die leksikograaf gee en die gebruiker ontvang. Woordeboeke bied gekeurde, geverifieerde en geredigeerde data. Die internet se data is nie leksikografies gekeur, geverifieer en geredigeer nie. Gebruikers wat vanuit 'n woordeboek toegang tot die internet kry, moet kundig genoeg wees om te weet dat dit dikwels rou en onverwerkte inligting is wat onttrek word en wat nie die gesag dra van die betrokke woordeboek waaruit die internettoegang gebied is nie. Elke woordeboeksoort behoort dus nie die soekheelal sodanig te vergroot nie. Taalpraktisyns se woordeboekkultuur behoort

bo verdenking te wees en hulle sal bewus wees van die verskil in status tussen woordeboek- en internetdata.

In 'n tradisionele woordeboek is daar 'n balans tussen die aangebode data en die inligting wat 'n gebruiker daaraan kan onttrek. Die dataverspreidingstruktuur en die inligtingsonttrekkingstruktuur loop parallel. 'n Gevolg van die dinamiese ontwikkeling van die leksikografie is dat die balans tussen data en inligting in 'n woordeboek versteur kan word. Deur die dataverspreidingstruktuur beslis die leksikograaf presies watter data in die woordeboek aangebied word. Aan daardie leksikografiese data kan die gebruiker dan inligting onttrek. Toegang wat woordeboekgebruikers vanuit 'n woordeboek na die internet kan kry, lei hulle na data wat nie deur 'n toepassing van die woordeboek se dataverspreidingstruktuur beskikbaar gestel is nie, maar wat wel gevind kan word as deel van 'n oorkoepelende datastruktuur wat naas die dataverspreiding ook die data-opsporing insluit. Die woordeboek se inligtingsonttrekkingstruktuur stel die gebruiker steeds daartoe in staat om alle data binne die datastruktuur te onttrek. Waar woordeboek-eksterne bronne die soekheelal van 'n woordeboek vorm, is die bestek van die inligtingsonttrekkingstruktuur omvattender as dié van die dataverspreidingstruktuur. Die leksikografiese proses maak steeds voorsiening vir die verspreiding van leksikografiese data én maak voorsiening vir die onttrekking van sowel leksikografiese as nieleksikografiese inligting.

Deur aan gebruikers toegang tot die internet te gee, verander 'n woordeboek van 'n stootmedium na 'n trekmedium – vergelyk Duan, Kartnik & Yingfei (s.j.), Müller-Spitzer (2013) en Gouws (2018b). Die leksikograaf het geen beheer oor die data wat in die internet beskikbaar is nie – en stoot dus nie data na die gebruiker toe nie. Die gebruiker kry wel via die woordeboek toegang tot hierdie data – maar sou dit ook sonder die woordeboek as tussengangerbron kon bereik het. Die gebruiker word voor die keuse gestel om vanuit die internet self data te kies – die data word deur die gebruiker getrek en nie deur die leksikograaf gestoot nie. Die leksikograaf moet wel binne die raamwerk van 'n oorkoepelende datastruktuur voorsiening maak vir die formulering van 'n datatrekkingstruktuur.

8. Bykomende koteksdata

Die uitbreiding van 'n woordeboek se soekheelal hou waarde in vir kundige woordeboekgebruikers. Die toegang tot en benutting van ander woordeboeke in dieselfde soekportaal stel gebruikers aan bykomende leksikografiese data bekend, maar 'n nog belangriker aspek is die toegang tot nieleksikografiese bronne soos die internet. Daarbenewens kan 'n omvattender soekheelal ook aan woordeboekgebruikers toegang gee tot die leksikografiese korpus wat deur die leksikograwe van 'n betrokke woordeboek gebruik is. Die beperkinge van woordeboekartikels se aanbieding van koteksmateriaal kan so oorkom word deurdat die gebruiker vanuit die woordeboek na die korpus kan gaan om daar na voorbeeldmateriaal en kollokasies te soek. Dit sal veral ook komplekse kollokasies en die optrede van kollokasies in sinsverband wees waarmee die gebruiker geholpe sal raak. Vir taalpraktisyns wat teksproduksiehulp nodig het, bied toegang tot 'n leksikografiese korpus verruimende soekmoontlikhede. Soos met internetdata sal die korpusdata waarskynlik ook nie geredigeer wees nie, maar taalpraktisyns sal wel die verlangde hulp kan vind en baat by nog meer data en voorbeelde van opgetekende taalgebruik.

9. Ten slotte

Leksikografiese korpora moet nie net ter beskikking van leksikograwe wees nie, maar woordeboekgebruikers behoort, uiteraard onderhewig aan bepaalde gebruiks- en kopieregreëls, ook toegang tot ’n korpus te kry om die sukses van hulle woordeboekraadplegingsprosedure te help verseker. Die internet moet ook ’n soekstreek binne ’n soekheelal word sodat woordeboeke met die nodige datatrekkingstrukture aan hulle gebruikers toegang tot data kan bied wat nie alles in die enkele woordeboek aangebied kan word nie.

Vir alle woordeboekgebruikers, maar veral ook vir taalpraktisyns, bied vernuwende leksikografiese benaderings opwindende naslaanmoontlikhede.

Bibliografie

Woordeboeke

Grobbelaar, P. en andere (reds.). 1987. *Reader's Digest Afrikaans-Engelse Woordeboek/English-Afrikaans Dictionary*. Kaapstad: The Reader’s Digest Association, South Africa (Pty) Ltd.

Hartmann-Petersen, P. et al. (reds.). s.j. *SASOL: Science & Technology Resource.* Kaapstad: Infosource.

Wiegand, H.E. et al. (reds.). 2010. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research.* Deel 1. Berlyn: De Gruyter. https://doi.org/10.1515/9783110474251-003

https://longman-hat-co-za.ez.sun.ac.za/?l=en (Geraadpleeg op 31 Augustus 2019)

https://www.dict.cc/ (Geraadpleeg op 31 Augustus 2019)

https://www.thefreedictionary.com/dictionary (Geraadpleeg op 31 Augustus 2019)

https://www.owid.de/ (Geraadpleeg op 31 Augustus 2019)

Ander bronne

Agerbo, H. 2017. The Interpretive Function: To Be or Not to Be, That is the Question. *Lexikos,* 27: 1-15.

Bergenholtz, H. & R.H Gouws. 2010. A Functional Approach to the Choice between Descriptive, Prescriptive and Proscriptive Lexicography. *Lexikos,* 20: 26-51. https://doi.org/10.4314/lex.v20i1.62683

Bergenholtz, H. & S. Tarp (reds.). 1995. *Manual of Specialised Lexicography.* Amsterdam: John Benjamins. https://doi.org/10.1075/btl.12

Bergenholtz, H., S. Tarp & H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern, in Hoffmann, Lothar et al. (reds.). *Fachsprachen. Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research*. Berlyn: De Gruyter : 1762-1832. https://doi.org/10.1515/9783110158847.2.21.1762

Bothma, T.J.D., H. Bergenholtz & R.H. Gouws. 2011. A model for integrated dictionaries of fixed expressions, in I. Kosen en K. Kosen (reds.). *Electronic Lexicography in the 21st Century: New applications for new users*. Ljubljana: Trojina, Institute for Applied Slovene Studies: 34-42.

Duan, Z., G. Kartik & D. Yingfei (s.j.): *Push vs. Pull: Implications of Protocol Design on Controlling Unwanted Traffic*. [Intyds] Beskikbaar: https://pdfs.semanticscholar.org/6f63/d37b4f8dd655e3594185e74daf4689f55aa1.pdf. [2019, Augustus 26].

Gouws, R.H. 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid. *Lexikos,* 13: 92-110. https://doi.org/10.5788/13-0-724

Gouws, R.H. 2015. The presentation and treatment of collocations as secondary guiding elements in dictionaries. *Lexikos,* 25: 170-190. https://doi.org/10.5788/25-1-1294

Gouws, R. H. 2018a. Expanding the data distribution structure. *Lexicographica,* 34: 225-237.

Gouws, R. H. 2018b. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos,* 28: 177-195. https://doi.org/10.5788/28-1-1461

Gouws, R. H. (ter perse). Expanding the use of corpora in the lexicographic process of online dictionaries, in J. Taborek et al. (reds.). *10. Kolloquium zur Lexikographie und Wörerbuchforschung.* Berlyn: De Gruyter. https://doi.org/10.1515/9783110716955-001

Gouws, R. H. & D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: SunMedia.

Gouws, R.H. & M. Steyn. 2005. Integrated outer texts: a transtextual approach to lexicographic functions, in I. Barz, H. Bergenholtz en J. Korhonen (reds.). *Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch.* Frankfurt a.M./Bern/New York/Paris: Peter Lang: 127-136. https://doi.org/10.7146/hjlcb.v19i37.25875

Gouws, R. H. et al. (reds.). 2013. *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlyn: De Gruyter. https://doi.org/10.1515/9783110238136.24

Hanks, P. 1979. To what extent does a dictionary definition define? *ITL,* 45/46: 32-38. https://doi.org/10.1075/itl.45-46.05han

Hausmann, F. J. et al. (reds.). 1989-1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie/Dictionaries. An International Encyclopedia of Lexicography/Dictionnaires. Encyclopédie internationale de lexicographie.* Berlyn: Walter de Gruyter. https://doi.org/10.5788/7-1-989

Hausmann, F.J. & H.E. Wiegand. 1989. Component Parts and Structures of Monolingual Dictionaries, in F.J. Hausmann et al. (reds.). 1989–1991: 328–360. https://doi.org/10.1515/9783110095852.1.4.328

Kammerer, M. & H.E. Wiegand. 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica,* 14: 224-238. https://doi.org/10.1515/9783484605930.2.187

Müller-Spitzer, C. 2013. Textual structures in electronic dictionaries. In Gouws et al. (reds.). 2013: 367-381.

Tarp, S. 2000. Theoretical challenges to LSP lexicography. *Lexikos,* 10: 189-208.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Tübingen: Max Niemeyer. https://doi.org/10.1093/ijl/ecp030

Wiegand, H.E. 1977. Nachdenken über Wörterbücher. Aktuelle Probleme, in G. Drosdowski, H. Henne en H.E. Wiegand. (reds.). *Nachdenken über Wörterbücher*. Mannheim, Vienna, Zürich: Bibliographisches Institut: 51-102. https://doi.org/10.5771/0943-7444-1978-2-123

Wiegand, H. E. 1989. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven, in F.J. Hausmann et al. (reds.). 1989-1991: 409-462.

Wiegand, H.E. 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher, in H.E. Wiegand (red.). *Wörterbücher in der Diskussion II. Vorträge aus dem Heidelberger Lexikographischen Kolloquium*. Tübingen: Max Niemeyer: 1-82. https://doi.org/10.1515/9783110937978.1

Wiegand, H.E. 1998. *Wörterbuchforschung.* Berlyn: De Gruyter.

Wiegand, H.E., S. Beer & R.H. Gouws. 2013. Textual structures in printed dictionaries. An overview, in R.H. Gouws et al. (reds.). 2013: 31-73.

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference number (UID) 85434). 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.