

Aspekte van inligtingsonttrekkingstrukture in aanlyn woordeboeke

Rufus H. Gouws, *Departement Afrikaans en Nederlands,
Universiteit Stellenbosch, Suid-Afrika*
(rhg@sun.ac.za) (<https://orcid.org/0000-0002-3423-058X>)

en

Theo J.D. Bothma, *Departement Inligtingkunde,
Universiteit van Pretoria, Suid-Afrika*
(theo.bothma@up.ac.za) (<https://orcid.org/0000-0001-7850-3263>)

Opsomming: In gedrukte woordeboeke met 'n statiese artikel- en boekstruktuur is die verhouding tussen die verspreiding van data en die onttrekking van inligting eenvoudig en die meeste woordeboeke volg min of meer dieselfde ordeningstelsels. Aanlyn woordeboeke vertoon dinamiese strukture met data wat dikwels op verskillende vlakke en in verskillende vakke aangebied word. Dit dwing die gebruiker om afwaarts (of vertikaal) en sywaarts (of horisontaal) in 'n artikel of bepaalde artikelkommentare te beweeg en om soekprosedures te volg wat nuwe strukture soos uitgebreide kommentare en selfs nuwe soekposisies soos soektonnels in ag moet neem. Dit is uiters belangrik dat gebruikers bewus moet wees van die volle spektrum data in 'n leksikografiese produk en dat hulle daartoe in staat moet wees om die datatipe wat vir hulle ter sake is, te kan kies. Gebruikers van leksikografiese produkte het ondersteuning nodig om 'n optimale onttrekking van inligting te verseker. Gevolglik moet leksikograwe moeite doen om 'n inligtingsonttrekkingstruktuur te skep wat gebruikers kan help om suksesvolle woordeboekraadpleging te verseker. Dit kan verfyning van die databasisstruktuur van die woordeboek vereis, en die ontwikkeling van 'n gevorderde koppelvlakontwerp wat nuwe kenmerke en navigasie-opsies aan die gebruiker beskikbaar kan stel. Dit beklemtoon weer eens die noodsaaklikheid van 'n multidisiplinêre span in die ontwikkeling van 'n woordeboek om 'n geskikte finale produk aan die gebruiker te kan bied. Hierdie artikel bied voorstelle vir gestruktureerde en kitstoegang tot data en die gebruik van datamerkers om gebruikers te lei na aanduiders in komplekse woordeboekartikels. Die klem is ook op die gebruikersvriendelikheid van leksikografiese instrumente wat die gehalte van intuitiewe woordeboekgebruik kan verbeter.

Sleutelwoorde: AANLYN WOORDEBOEKE, ARTIKELSTRUKTUUR, BEPERKTE ARTIKEL, INLIGTINGSONTTREKKINGSTRUKTUUR, KONTEKSTUALISERING, OMVATTENDE ARTIKEL, SEMANTIESE SUBKOMMENTAAR, SKAKELING, SOEKSONE, SOEKTONNEL, VAKKE, VLAKKE

Abstract: Aspects of information retrieval structures in online dictionaries.

In printed dictionaries with a static article and book structure the relation between inserting data and retrieving information is not complex, and most dictionaries follow more or less the same ordering systems. Online dictionaries display dynamic structures with data often presented in

various levels and compartments that compel the user to drill down (or vertically) into articles or laterally (or horizontally) within specific comments of articles, and to embark on search procedures that must negotiate new structures such as extended comments and even new search positions such as search tunnels. It is of paramount importance that users should be aware of the full spectrum of data included in a lexicographic product and should be able to select the data types of interest to them. Users of lexicographic tools need assistance to ensure an optimal retrieval of information from the data. Consequently, lexicographers need to embark on dedicated ways to establish an information retrieval structure that can enable users to achieve an adequate dictionary consultation. This might demand a refinement of the database of the dictionary as well as a sophisticated interface that can provide new features and navigation options to the user. It emphasizes yet again the need for a multidisciplinary team in the development of a dictionary to provide the user with an appropriate final product. This paper offers proposals for structured and rapid access to data and the use of data indicators to guide users to items occurring in complex dictionary articles. The emphasis is also on user-friendliness of lexicographic instruments that can enhance the quality of intuitive dictionary use.

Keywords: ARTICLE STRUCTURE, COMPARTMENTS, COMPREHENSIVE ARTICLE, CONTEXTUALISATION, INFORMATION RETRIEVAL STRUCTURE, LEVELS, LINKING, ONLINE DICTIONARIES, RESTRICTED ARTICLE, SEARCH TUNNEL, SEARCH ZONE, SUBCOMMENT ON SEMANTICS

1. Inleiding en agtergrond

Hartmann en James (1998) bied die volgende definisie vir die lemma *lexicography*:

The professional activity and academic field concerned with DICTIONARIES and other REFERENCE WORKS. It has two basic divisions: lexicographic practice, or DICTIONARY-MAKING, and lexicographic theory, or DICTIONARY RESEARCH. ...

In hierdie bydrae gaan dit primêr om inligtingsonttrekkingstrukture in aanlyn woordeboeke; tog sal ander leksikografiese naslaanbronne wat nie woordeboeke is nie, ook ter sprake kom. Daarom is Hartmann en James se verwysing na "DICTIONARIES and other REFERENCE WORKS" ter sake.

'n Gereelde verwysing na Samuel Johnson se woordeboek van 1755 is na sy definisie van 'n leksikograaf as "A writer of dictionaries; a harmless drudge, ..." Daar word selde ook aandag gegee aan die daaropvolgende deel van die betekenisparafrase, naamlik "... that busies himself in tracing the original, and detailing the signification of words." Die aktiwiteit van besig wees met betekenisverklaring vorm 'n kerndeel van 'n leksikograaf se opdrag. In hierdie artikel sal dit juis die onttrekking van semantiese data en meer spesifiek die betekenisparafrase wees waaraan daar aandag gegee word. Dit word gedoen met die wete dat die leksikografie oor 'n wye tipologiese verskeidenheid beskik en dat hierdie woordeboeke 'n wye verskeidenheid datatipes bevat wat ook herwinbaar moet wees.

Met die aanvang van die beplanning en samestelling van 'n woordeboek moet die leksikograaf duidelikheid hê oor wie die veronderstelde teikengebruikers van die woordeboek gaan wees, asook wat hulle leksikografiese behoeftes en naslaanvaardighede is. Vergelyk hier onder andere Hartmann (1989), Tarp (2008) en Wiegand (1998). Op grond van die teikengebruiker en hulle behoeftes besluit die leksikograaf wat die leksikografiese funksies van die woordeboek gaan wees. Die leksikografiese funksies is dan van bepalende belang vir die keuse van die leksikografiese data en die woordeboekstrukture wat nodig is vir die aanbieding en huisvesting van hierdie data. 'n Wesenlike deel van die leksikograaf se beplanning is weliswaar gerig op die keuse maar ook op die aanbieding van data ter bevrediging van die gebruiker se leksikografiese behoeftes. Die uiteindelijke sukses van 'n leksikograaf se werk word bepaal deur die sukses van die teikengebruikers wanneer hulle die woordeboek raadpleeg. Om hierdie sukses te verseker, is dit belangrik dat woordeboekstrukture ontwerp en aangewend moet word om toegang tot die leksikografiese data te verhoog sodat die gepaste leksikografiese funksie bevredig kan word.

Die leksikograaf mag die suksesvolle onttrekking van inligting deur die gebruiker nie as 'n vanselfsprekendheid beskou nie. Ook hier moet 'n struktuur beplan en aangewend word om sukses te verseker. 'n Gepaste inligtingsonttrekkingstruktuur is noodsaaklik om woordeboekgebruik te kan optimaliseer. Aspekte van hierdie struktuur tipe wat nog slegs beperkte aandag in die metaleksikografiese literatuur gekry het, word in hierdie artikel bespreek. Woordeboekstrukture tree nie geïsoleerd van mekaar op nie en in 'n bespreking van die inligtingsonttrekkingstruktuur moet ander strukture, onder meer die dataverspreidingstruktuur en die artikelstruktuur, ook aan die orde gestel word.

2. Die dataverspreidingstruktuur

Die dataverspreidingstruktuur, vergelyk Bergenholtz et al. (1999), bepaal die toewysing van leksikografiese data aan spesifieke soekposisies in woordeboeke. Dié struktuur word aangewend in sowel gedrukte as aanlyn woordeboeke, asook binne 'n breër leksikografiese soekomgewing. Verskillende soekposisies kan onderskei word. Wiegand et al. (2013: 63) onderskei tussen 'n soeksone, 'n soekgebied en 'n soekveld. Die soeksone is die gleuf in 'n woordeboekartikel waarin 'n aanduiding geplaas word. Elke woordeboekartikel is 'n soekgebied en die sentrale teks van 'n woordeboek wat al die artikeltrajekte bevat, is 'n soekveld. Daarbenewens maak Gouws (2023: 383-384) ook voorsiening vir 'n soekstreek, dit is die volle woordeboek met sy sentrale teks plus die verskillende buitetekste; 'n soekdomein, dit is die portaal waarbinne aanlyn woordeboeke geplaas kan word; en 'n soekuniversum, dit is die omgewing buite die soekdomein, byvoorbeeld die internet of ander woordeboekportaal-eksterne bronne waartoe gebruikers via die woordeboek toegang het.

Dit is belangrik om daarop te let dat 'n inligtingsonttrekkingstruktuur so geformuleer moet word dat dit die gebruiker kan help om leksikografiese inlig-

ting aan die data in al die verskillende soekposisies te kan onttrek. In hierdie artikel is die fokus beperk tot die woordeboekartikel as soekgebied en die onttrekking van inligting aan data in bepaalde soeksones.

3. Die soekgebied

Die woordeboekartikel is die primêre posisie vir die toesegging van leksikografiese data in 'n woordeboek. In teenstelling met die statiese aard van woordeboekartikels in gedrukte woordeboeke het aanlyn woordeboeke 'n dinamiese artikelstruktuur (Gouws 2014: 164). Aanlyn woordeboeke mag dalk wel haas onbeperkte bergingsruimte hê, maar omdat daar beperkte vertoonruimte is, kan al die data wat aan 'n artikel toegewys word, nie noodwendig tegelykertyd gesien word nie. Aanlyn woordeboeke het dikwels omvattende artikels wat verskeie beperkte artikels insluit (Gouws 2014). Beperkte artikels is onderafdelings van 'n omvattende artikel en elkeen bied slegs 'n deel van die leksikografiese bewerking, byvoorbeeld een of meer soeksones. Die omvattende artikel kan nie as 'n geheel gesien word nie en toegang tot die data geskied via die verskillende beperkte artikels.

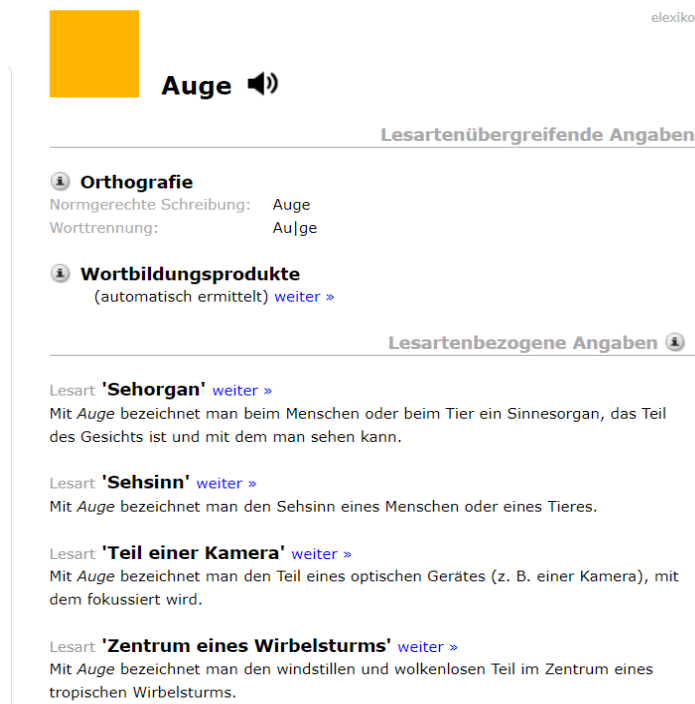
Die verspreiding van data in 'n woordeboekartikel verseker nie gebruiker-sukses nie. Wat die leksikograaf in 'n omvattende artikel plaas, moet die gebruiker kan vind en onttrek. Daarvoor is die bykomende strukture nodig.

4. Artikelstruktuur: kommentare, vlakke en vakke

Volgens Wiegand (1989: 470) bestaan 'n woordeboekartikel met 'n basiese struktuur uit twee kommentare, naamlik 'n vormkommentaar en 'n semantiese kommentaar. Alhoewel die artikelstruktuur van aanlyn woordeboeke nie noodwendig dieselfde byeenplasing van aanduiders in die onderskeie komponente vertoon nie, kan 'n vergelykbare indeling dikwels steeds herken word (Gouws 2014). Soos in gedrukte woordeboeke bevat die semantiese kommentaar van artikels met 'n lemma wat 'n polisemiese leksikale item verteenwoordig, dikwels ook verskillende semantiese subkommentare wat elk huisvesting aan die bewerking van een polisemiese onderskeiding bied. Die fokus van hierdie bespreking is vervolgens op die onttrekking van inligting aan sulke semantiese subkommentare. Dit word gedoen aan die hand van artikels uit een aanlyn woordeboek, naamlik die Leibniz-Institut für Deutsche Sprache (IDS) se *lexiko* wat op OWID se woordeboekportaal verskyn.

In 'n omvattende woordeboekartikel wat verskillende semantiese subkommentare vertoon, is daar 'n vertikale hiërargiese ordening van hierdie semantiese subkommentare en wel so dat elke subkommentaar as 'n afsonderlike vlak van die semantiese kommentaar optree en 'n eie beperkte artikel daarstel. Vlakke kan hier beskou word as soekposisies binne 'n kommentaar van 'n omvattende artikel wat elk beset word deur 'n beperkte artikel wat een semantiese subkom-

mentaar bevat. Opeenvolgende vlakke kom voor in struktureel laer artikelposisies en die verskillende vlakke se data is nie gelyktydig sigbaar nie. Toegang tot die data in 'n bepaalde vlak vereis dat daar vanuit die betrokke kommentaar se openingsaanbod afwaarts beweeg moet word. Vergelyk in hierdie verband figuur 1, 'n deelartikel van *ellexiko* se lemma *Auge* (= "oog"). In hierdie beperkte artikel, die openingskermskoot wanneer die lemma *Auge* gesoek word, is die artikelkomponent "Lesartenübergreifende Angaben" (= betekenisonderskeiding-oorkoepelende aanduiders") vergelykbaar met die semantiese kommentaar van 'n artikel. Dit bevat 'n lysing van die verskillende betekenisonderskeidinge van die woord *Auge* met 'n eenwoordbetekenissamevatting en 'n bondige betekenisparafrase:



ellexiko

Auge

Lesartenübergreifende Angaben

Orthografie
Normgerechte Schreibung: Auge
Worttrennung: Au|ge

Wortbildungsprodukte
(automatisch ermittelt) [weiter >](#)

Lesartenbezogene Angaben

Lesart '**Sehorgan**' [weiter >](#)
Mit *Auge* bezeichnet man beim Menschen oder beim Tier ein Sinnesorgan, das Teil des Gesichts ist und mit dem man sehen kann.

Lesart '**Sehsinn**' [weiter >](#)
Mit *Auge* bezeichnet man den Sehsinn eines Menschen oder eines Tieres.

Lesart '**Teil einer Kamera**' [weiter >](#)
Mit *Auge* bezeichnet man den Teil eines optischen Gerätes (z. B. einer Kamera), mit dem fokussiert wird.

Lesart '**Zentrum eines Wirbelsturms**' [weiter >](#)
Mit *Auge* bezeichnet man den windstillen und wolkenlosen Teil im Zentrum eines tropischen Wirbelsturms.

Figuur 1: Uit *ellexiko*

Elke eenwoordbetekenissamevatting word gevolg deur 'n struktuurmerker *weiter* (= "verder"). Dit is 'n skakel na 'n volgende vlak van die semantiese kommentaar en lei die gebruiker na die betrokke beperkte artikel waarin meer data oor die tersaaklike betekenisonderskeiding verstrek word. 'n Klik op die skakel *weiter* ten opsigte van die eerste betekenisonderskeiding *Sehorgan* (= "sigorgaan") van *Auge* lei tot 'n afwaartse beweging na 'n vlak wat die beperkte artikel bevat wat in figuur 2 gesien kan word:

The screenshot shows the 'Auge' entry in the 'elexiko' dictionary. The word 'Auge' is highlighted in orange, with a speaker icon indicating audio pronunciation. Below the word, it says 'Lesart: 'Sehorgan''. There are navigation links: 'zur Übersichtsseite' and 'Lesarten im Überblick'. A horizontal menu contains tabs for 'Bedeutungserläuterung', 'Kollokationen', 'Konstruktionen', 'Sinnverwandte Wörter', 'Gebrauchsbesonderheiten', and 'Grammatik'. The 'Bedeutungserläuterung' tab is active, showing the title 'Erläuterung der Bedeutung / Funktion'. The text explains that 'Auge' is the eye in humans and animals. There are links for 'Belege anzeigen' and 'Illustrationen anzeigen'. Below this is the 'Sachinformationen' section, followed by 'Weitere Informationen' which provides a detailed anatomical definition and a citation: 'Brockhaus in Text und Bild Edition (2002). Mannheim (CD-ROM)'. At the bottom, it lists 'Wortklasse: Partitivum'.

Figuur 2: Uit *elexiko*

Hierdie beperkte artikel het ook *Auge* as lemma maar hier gaan dit nie om die polisemiese leksikale item *Auge* nie, maar slegs om die lemma as gidselement vir 'n bewerking van een van die leksikale item se betekenisonderskeidings. Al die data wat in hierdie beperkte artikel voorkom, is slegs op *Auge* se optrede in een betekenisonderskeiding gerig. Die verskillende datatipes wat in een vlak voorkom en betrekking het op die enkele betekenisonderskeiding, word ook nie gelyktydig vertoon nie, maar kan wel bereik word deur binne die betrokke vlak rond te beweeg. Hierdie betekenisonderskeiding-spesifieke data word in verskillende vakke in die vlak verpak. Vakke kan hier beskou word as soekposisies binne 'n subkommentaar, 'n vlak, van 'n omvattende artikel wat elk beset word deur 'n beperkte artikel wat een datatipe bevat wat spesifiek gerig is op die betrokke subkommentaar.

Naas ander inskrywings bevat die beperkte artikel in figuur 2 ook verskeie data-merkers (*Kollokation*, *Konstruktion*, ensovoorts) wat op die boonste dwarslyn verstrek word. 'n Klik op enigeen van hierdie merkers lei die gebruiker na 'n vak binne dieselfde vlak waarin bykomende data verstrek word wat nie die lemma van die omvattende artikel as adres het nie, maar wel die lemma van die betrokke beperkte artikel. Dit is bykomende data ter sake vir die optrede van die leksikale item in daardie spesifieke betekenisonderskeiding. Om die data in

'n vak te bereik, vereis nie 'n afwaartse soekbeweging na 'n volgende vlak nie, maar slegs 'n horisontale (of sywaartse) beweging binne die betrokke vlak.

Die gebruik van verskillende vakke in een vlak help die leksikograaf om data-oormoedigheid in die aanbieding van 'n bepaalde semantiese subkommentaar te voorkom. Dit verhoog ook die gebruikersvriendelikheid van die woordeboek omdat gebruikers slegs hoef te klik op 'n merker wat relevant is vir hulle soektog en hulle ontsluit sodoende slegs die vak met data wat hulle benodig. Om hierdie funksionaliteite moontlik te maak, vereis bepaalde datastrukture in die leksikografiese databasis en komplekse koppelvlakontwerp, beide items wat deur 'n multidisiplinêre span in die ontwerp van die woordeboek aangespreek moet word. Dit vorm dus deel van 'n struktuur wat gerig is op die maklike en suksesvolle onttrekking van inligting aan die data wat aan die beperkte artikels toegewys is. Vergelyk in hierdie verband figuur 3, die beperkte artikel wat vertoon word wanneer 'n gebruiker op die datamerker *Sinnverwandte Wörter* (= "betekenisverwante woorde") in figuur 2 klik:

elexiko

Auge

Lesart: 'Sehorgan'

zur Übersichtseite Lesarten im Überblick

Bedeutungserläuterung Kollokationen Konstruktionen **Sinnverwandte Wörter** Gebrauchsbesonderheiten Grammatik

Sinnverwandte Wörter

Beziehung(en) der Bedeutungsgleichheit/-äquivalenz

Synonym(e):

Sehorgan

Beziehung(en) des Bedeutungsgegensatzes

inkompatible(r) Partner:

Gesicht

Haar

Mund

Nase

Ohr

Die Auswahl der hier angegebenen Mitspieler richtet sich nach der statistischen Signifikanz im elexiko-Korpus und deckt diejenigen Körperteile ab, die in räumlicher Nähe zum *Auge* liegen.

Figuur 3: Uit *elexiko*

Die data wat hier verstrekkend word, geld nie die ander betekenisonderskeidings van die woord *Auge* nie. Die leksikograaf gebruik die datamerkers as subkommentaaruitbreidingsaanwysers (Wiegand en Gouws 2013: 296) wat die gebruiker help om by uitbreidings van die subkommentaar uit te kom. Ook hier kan die gebruiker geordende stappe volg ter onttrekking van die inligting — dit is deel van die inligtingsonttrekkingstruktuur van die betrokke woordeboekartikel.

Uit figuur 1–3 blyk dat 'n kommentaar — hier 'n semantiese kommentaar — 'n afwaarts uitgebreide artikelkonstituent is. In hierdie semantiese kommentaar kan daar afwaarts beweeg word na beperkte artikels wat verskillende vlakke beset. Binne 'n vlak, dit wil sê 'n enkele semantiese subkommentaar, is 'n horisontale beweging moontlik om by die data in die verskillende vakke uit te kom. Die horisontale beweging binne 'n vlak skep 'n soektonnel. 'n Soektonnel (Gouws 2022) is 'n struktuur met 'n geordende reeks stappe wat in 'n artikelvlak gevolg word om 'n aanduiding in 'n bepaalde vak te bereik met spesifieke data gerig op die betrokke subkommentaar. Dit is byvoorbeeld die struktuur wat dit vir die gebruiker moontlik maak om vanaf die aanbieding in figuur 2 na die aanduiding van 'n sinoniem in die vak vir betekenisverwante woorde in figuur 3 te beweeg.

Verskillende vlakke en vakke dwing die gebruiker (indien addisionele inligting benodig word) om verby die aanduidings wat met 'n eerste oogopslag sigbaar is in 'n artikel verder te beweeg na spesifieke kommentare en subkommentare. So 'n koppelvlakstruktuur kan ook inligtingsoorlading voorkom, deur die gebruiker slegs bloot te stel aan die data wat hulle benodig om hulle inligtingsbehoefte te bevredig, en nie aan die volledige stel komplekse data in die artikel nie. In hierdie proses begeef gebruikers hulle op soekprosedures met nuwe strukture soos uitgebreide kommentare en nuwe soekposisies soos soektonnels. Dit is middele wat die leksikograaf in die aanwending van die dataverspreidingsstruktuur in werking stel. So 'n stelsel vereis wel dat die leksikograaf die gebruiker in aanmerking moet neem en moet verseker dat die nodige struktuur beskikbaar is om die soek, vind en onttrekking van inligting te verseker. Gebruikers moet bewus wees van die volle omvang van die data wat in 'n leksikografiese produk aangebied word en hulle moet daartoe in staat wees om die tersaaklike inligting te kan onttrek. Daarom moet leksikograwe nie net oor die dataverspreidingsstruktuur besin nie, maar ook oor maniere om gebruikers daartoe in staat te stel om 'n gepaste woordeboekraadpleging en inligtingsonttrekking te kan deurvoer, in samewerking met die multidissiplinêre ontwikkelingspan van die woordeboek.

5. 'n Inligtingsonttrekkingstruktuur

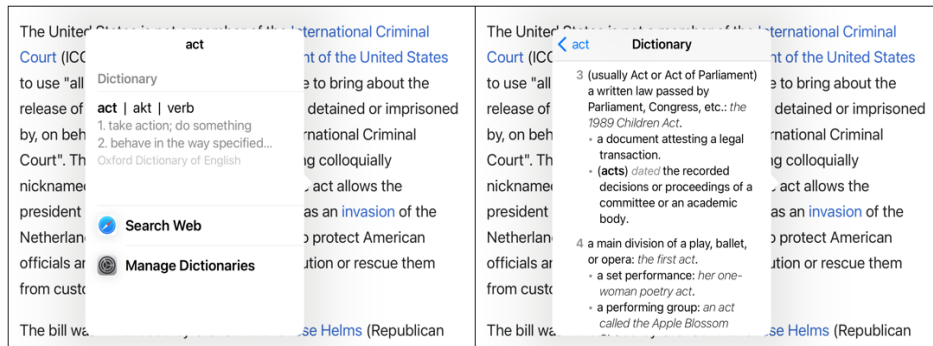
Om suksesvolle woordeboekgebruik te verseker, is die aanbieding van die verlangde data nie genoeg nie. Strukture moet geskep word wat kan verseker dat die betrokke leksikografiese funksie bevredig word en wat die gebruiker kan

help om die relevante data te vind, te interpreteer en op die regte manier te benut.

Om dit moontlik te maak, is die implementering van 'n inligtingsonttrekkingstruktuur noodsaaklik. So 'n struktuur bied 'n geordende reeks stappe wat dit moontlik maak om die gepaste inligting te onttrek aan die data wat in 'n woordeboekartikel geplaas is. 'n Woordeboek se inligtingsonttrekkingstruktuur blyk dikwels die spieëlbeeld van daardie woordeboek se dataverspreidingstruktuur te wees.

'n Inligtingsonttrekkingstruktuur kan verskillende inligtingsonttrekkingsroetes bepaal. In 'n woordeboek soos *lexiko* kan die gebruiker 'n uitvoerige toegangsroete benut om by die verlangde data uit te kom en die inligting te onttrek. 'n Gebruiker wat goed vertrouwd is met *lexiko* kan 'n korter toegangsroete volg deur bepaalde stappe in die toegangstruktuur uit te laat om gouer by die data te kom en die tersaaklike inligting te onttrek. Daar moet wel duidelike data-merkers wees om die seekroete te kan identifiseer en 'n optimale inligtingsonttrekking moontlik te maak.

E-toestelle bevat dikwels geïntegreerde woordeboeke wat uitgebreide of minder uitgebreide prosedures benut om inligting te onttrek. Gouws en Tarp (ter perse) gee aandag aan twee resultate waar 'n gebruiker van 'n e-toestel inligting aan die geïntegreerde woordeboek probeer onttrek. In albei gevalle is die gebruiker op soek na die betekenis van die naamwoord *act* soos dit in 'n spesifieke teks gebruik word. In die eerste geval klik 'n gebruiker wat 'n teks op 'n e-leser lees op die woord *act*. Die opskietvenster (figuur 4) vertoon 'n bewerking van dié woord vanuit die databasis wat aan die geïntegreerde woordeboek gekoppel is. Die eerste klik bied die eerste twee betekenisonderskeidings van die woord *act* as werkwoord. Om by die bewerking van *act* as 'n naamwoord uit te kom, is 'n verdere klik nodig. Dan word die hele artikel vertoon:



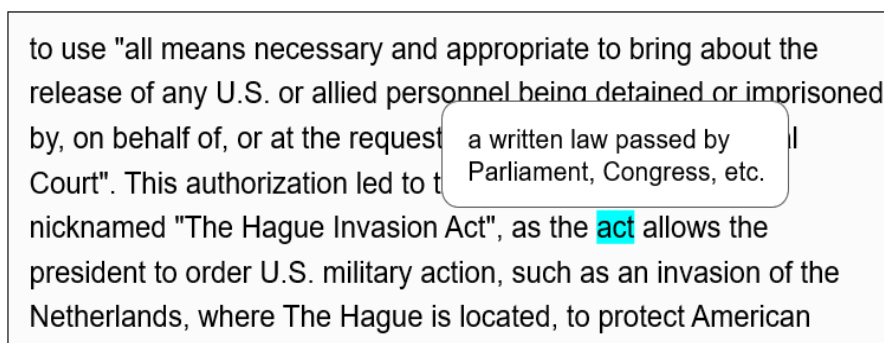
Figuur 4: Uit die teks op die e-leser en die geïntegreerde woordeboek

Die gebruiker kry steeds nie dadelik die regte betekenisparafrase nie en suksesvolle inligtingsonttrekking vereis dat die gebruiker na die tweede klik self 'n

besluit moet maak oor watter betekenisparafrese die gepaste een is. Die inligtingsonttrekkingstruktuur is hier ondoeltreffend en lei nie tot 'n vinnige en ondubfelsinnige oplossing nie.

Vir verdere voorbeelde van problematiese soektogte op die Kindle-toepassing op 'n iPad, asook gewone lopende teks in 'n Google webblaaier, sien byvoorbeeld Bothma en Fourie (2024a en 2024b), Bothma en Gouws (2020 en 2022); vir voorbeelde van tipiese foute wat algemeen in die skakeling in die Kindle-toepassing voorkom, sien Bothma en Prinsloo (2013). Skakeling na die eerste lemma wat grafies ooreenkom met die woord waarop geklik word, het wye implikasies vir die gebruiker. Dit vereis dat die gebruiker noukeurig die volledige aanbod (wat moontlik oor meerdere artikels en verskeie skerms in die aanlyn woordeboek mag strek) moet deurlees om vas te stel wat die regte betekenis van die woord in konteks is. Dit kan heelwat tyd verspil. Dit impliseer verder dat die gebruiker 'n bepaalde minimum kennis van grammatika moet besit, om minstens die korrekte woordsoort in konteks te kan bepaal. In alle gevalle moet gebruikers die verskillende betekenisopsies noukeurig evalueer om te besluit wat die korrekte betekenis in konteks is. Die ooraanbod van data — waarvan die meeste nie in die spesifieke geval relevant of korrek is nie — lei tot inligtingoerlading en frustrasie by die gebruiker (Gouws en Tarp 2017).

'n Regstreekse, verkorte en suksesvolle inligtingsonttrekkingsroete blyk uit 'n voorstel van Huang en Tarp (2021). 'n Relatief eenvoudige program laat byvoorbeeld skrywers, uitgewers, onderwysers, ensovoorts toe om op daardie woorde in 'n digitale teks te klik wat hulle as problematies vir die aangewese teikenlesers beskou. Die lemma word dan in die databasis en die geïntegreerde woordeboek nagegaan en slegs die gepaste betekenisonderskeiding en mee-gaande betekenisparafrese word aangedui. Wanneer die teikenleser dan op die betrokke woord in die teks klik, word hulle regstreeks na die verlangde betekenisparafrese gelei en 'n kitsonttrekking van die korrekte inligting is moontlik. Ter illustrasie: 'n klik op die naamwoord *act* in die bepaalde teks lei regstreeks tot die verklaring wat in figuur 5 se opskietvenster vertoon word:



to use "all means necessary and appropriate to bring about the release of any U.S. or allied personnel being detained or imprisoned by, on behalf of, or at the request of a written law passed by Parliament, Congress, etc." Court". This authorization led to the creation of a law nicknamed "The Hague Invasion Act", as the **act** allows the president to order U.S. military action, such as an invasion of the Netherlands, where The Hague is located, to protect American

Figuur 5: Uit Huang en Tarp (2021)

Dit is vernuwend en gebruikersgerigte leksikografie. 'n Bykomende merker, bv. "SEE MORE" kan ook in die opskietvenster geplaas word wat dit vir lesers met 'n behoefte aan bykomende inligting moontlik maak om nog inligting aan die databasis te onttrek. 'n Opskietspyskaart wat die beskikbare datatipes vertoon, gee die gebruiker die geleentheid om óf 'n enkele aanduiding te kies óf te vra dat die volle artikel vertoon moet word. Hierdie benadering bied verskillende inligtingsonttrekkingsmoontlikhede omdat die inligtingsonttrekkingstruktuur verskillende onttrekkingsroetes kan bepaal. Soos aangetoon word in die hieropvolgende afdeling kan hierdie stelsel egter nie generies vir alle tekste beskikbaar wees nie, aangesien moontlike problematiese woorde in 'n bepaalde teks deur die skrywer/uitgewer/leksikograaf geïdentifiseer en geannoteer moet word, en leiding is net vir hierdie woorde beskikbaar. 'n Stelsel waarin die beginsel wat hier ter sprake is, naamlik direkte, outomatiese skakeling na die korrekte betekenis vir alle woorde in die teks en in die woordeboek, is wat vereis word.

Waar 'n klik op 'n woord in 'n teks die gebruiker na die volle artikel lei, volg die inligtingsonttrekkingsroete al die stappe van die volle inligtingsonttrekkingstruktuur. Waar 'n klik na 'n beperkte artikel lei, byvoorbeeld 'n semantiese subkommentaar of dalk slegs 'n glos as betekenisparafrase soos in figuur 5, is daar 'n verkorte inligtingsonttrekkingsroete wat net 'n gedeelte van die inligtingsonttrekkingstruktuur se stappe benut. Die behoeftes en vaardighede van die teikengebruiker is hier van bepalende belang en die leksikograaf moet vir meerdere moontlikhede voorsiening maak.

In figuur 1–5 was die fokus slegs op semantiese inligting en spesifiek die betekenisparafrase. Die onderliggende leksikografiese funksie wat hier bevredig moes word, is dié van teksbegrip. 'n Inligtingsonttrekkingstruktuur moet ook vir ander datatipes en ander funksies voorsiening kan maak. Dit is belangrik dat die leksikograaf die regte data aanbied en dat die gebruiker weet wat beskikbaar is en hoe en waar om dit te vind. 'n Gebruiker wat weet watter inligting onttrek kan word, sal weet hoe om regstreeks by slegs die betrokke aanduiding uit te kom.

6. Bepaalde probleme

Die laasgenoemde opsie om 'n kitsonttrekking moontlik te maak, is nie sonder probleme nie en die leksikograaf moet hier met die nodige omsigtigheid te werk gaan. Suksesvolle begrip voorvereis dat die uitgewer, skrywer of onderwyser die teks moet annoteer. As nuwe tekste op dieselfde platform geplaas en toegang tot dieselfde databasis en geïntegreerde woordeboek bied, moet sulke tekste ook geannoteer word. Dié annotasie word gedoen op grond van dit wat die uitgewer, skrywer, onderwyser as problematies vir die teikenleser beskou. Dit is nie in die eerste plek die gebruiker se eie perspektief wat 'n rol speel nie en die beperkte bestek van die annotasies en die daargestelde inligtingsonttrekkingstruktuur gaan nie noodwendig die gebruiker help in al hulle probleemgevalle nie.

'n Alternatiewe werkswyse wat tot 'n inligtingsonttrekkingstruktuur kan lei wat alle woorde in 'n teks ter wille is, is waar die skakeling tussen die teks en die databasis konteksbewus moet wees. Die skakeling moet byvoorbeeld daartoe in staat wees om op grond van die woord waarop geklik word die regte woordsoortwaarde te kan identifiseer — in die geval van die bogenoemde optrede van *act* dat dit hier 'n naamwoord is en nie 'n werkwoord nie. Ter wille van 'n ondubbelsinnige inligtingsonttrekking moet die gepaste betekenisonderskeiding binne die konteks en koteks van die woord in die bepaalde teks geïdentifiseer kan word. In hierdie verband bied Inligtingkunde hulp aan die aanlyn leksikografie om met gesofistikeerde soek- en filtreertegnologie die regte afsparring tussen 'n geklikte woord en die tersaaklike betekenisonderskeiding te verseker. Vergelyk in hierdie verband Bothma (2011), asook Bothma en Gouws (2022), Bothma en Fourie (2024a, b) en Tarp en Gouws (2020).

7. 'n Mensgerigte inligtingsonttrekkingstruktuur

Suksesvolle woordeboekgebruik is gewoonlik afhanklik van die naslaanvaardighede van die woordeboek se aangewese teikengebruikers. 'n Wesenlike deel van 'n tradisionele woordeboekkultuur, vergelyk Hausmann (1989), is dat die gemeenskap by die leksikografie moet aanpas en onder meer woordeboekvaardighede bemeester. Soos Bothma (2011) toon Tarp en Gouws (2020) ook aan dat die aanlyn omgewing nuwe moontlikhede geskep het om gebruikers op beter maniere by te staan. 'n Nuwe woordeboekkultuur wat onder andere op mensgerigte ontwerpbeginsels steun, kan die gebruik van naslaanbronne, insluitende woordeboeke en ander leksikografiese produkte, makliker maak vir gebruikers. Volgens Tarp en Gouws (2020: 2) plaas dit die verantwoordelikheid op die skouers van die leksikograaf om produkte te ontwerp wat intuïtiewe gebruik moontlik maak.

Die strewe na 'n verwerking van intuïtiewe gebruik geld ook ten opsigte van die ontwerp en implementering van 'n inligtingsonttrekkingstruktuur. Waar 'n mensgerigte inligtingsonttrekkingstruktuur gebruik word, lei 'n klik op 'n bepaalde woord in 'n teks die gebruiker onmiddellik na die aanduiding wat die gebruiker as verstektipe geïdentifiseer het. Data vanuit die geïntegreerde woordeboek wat van die leksikografiese databasis afgelaai is, laat gebruikers toe om oplossings vir hulle probleme te vind sodat hulle feitlik sonder onderbreking verder kan lees. Die soeksone waarheen gebruikers gelei word, bevat geen aanduiders wat hulle nie nodig het nie en fokus, waar dit die verstekwaarde is, slegs op die betekenisparafraze ter bevrediging van 'n teksbegripfunksie. So 'n woordeboek beskik oor 'n kitsinligtingsonttrekkingstruktuur.

8. Kontekstualisering en skakeling

Intuïtiewe gebruik van 'n mensgerigte produk wat kitstoegang tot die aangebode data en 'n regstreekse inligtingsonttrekking bied, vereis slim e-toestelle en

slim sagteware. Bothma en Gouws (2022) het reeds daarop gewys dat meer aandag gegee moet word aan skakelingprosedures tussen byvoorbeeld e-lesers, die onderliggende databasisse en die geïntegreerde woordeboeke. Die rol van die databasis raak in hierdie verband toenemend belangrik. Indien die gebruiker slegs 'n teksbegripbehoefte het, moet 'n klik op 'n woord in 'n teks op 'n e-leser nie 'n skakeling met die gepaste woordeboekartikel bewerkstellig nie, maar met die gepaste aanduiders in die onderliggende databasis. Om dit moontlik te maak, is gesofistikeerde prosedures van kontekstualisering nodig.

Die skakelingsprosedure tussen 'n woord wat in 'n bepaalde konteks optree en die tersaaklike aanduiders in die databasis voorveronderstel dinamiese artikels en leksikografiese strukture. Bothma en Gouws (2022) dui ten regte aan dat die geïntegreerde woordeboek moet aanpas by die buite-leksikografiese behoefte wat tot die betrokke naslaanhandeling gelei het; sien ook Bothma en Fourie (2024a, b). Suksesvolle skakeling vereis hoër kontekstualiseringsvereistes. Gekontekstualiseerde skakeling skep dan 'n verhouding tussen woorde in 'n e-tekst en gebruiker-gespesifiseerde leksikografiese bronne. Die gebruiker word gelei na die presiese en relevante aanduiders en 'n gepersonifieerde inligtingsonttrekkingstruktuur word gevestig.

Suksesvolle skakeling verg die benutting van hoëvlaktegnologie. Slim sagteware, byvoorbeeld woordsoortmerkers, en die gebruik van menslike taaltegnologie, kunsmatige intelligensie en masjienleer moet in werking gestel word. Sodoende kan 'n gegewe teks op die e-leser geïnterpreteer word en na wisselwerking met die databasis kan die gebruiker van 'n gekontekstualiseerde resultaat voorsien word. Die toepassing van mensgerigte ontwerpbeginsels bring mee dat die sagteware geen, of hoogstens beperkte, manipuleringsdeur deur die gebruiker nodig het. Die inligtingsonttrekkingstruktuur vertoon homself as 't ware aan die gebruiker.

Kunsmatige intelligensie — beide "tradisionele" kunsmatige intelligensie, wat tipies van natuurliketaalprosessering (NLP, "natural language processing") gebruik sou kon maak, en generatiewe kunsmatige intelligensie, gebaseer op groottaalmodelle (LLMs, "large language models") — kan heel moontlik 'n belangrike bydraende rol speel in die kontekstualiseringsproses. Dit word egter nie in hierdie artikel bespreek nie, omdat die fokus nie hier op die implementering is van die voorstelle wat hier gemaak word nie. Verdere navorsing sou met vrug hieraan kon aandag gee.

9. Ten slotte

Leksikograwe moet vernuwend dink en beplan, en poog om die gebruiker die geleentheid te bied om moeiteloos 'n optimale onttrekking van inligting te kan bewerkstellig. In die beplanning van 'n inligtingsonttrekkingstruktuur moet leksikograwe nie die artikel as mikpunt van 'n soektog sien nie, maar eerder 'n spesifieke aanduiders wat daardie data bied waarna 'n spesifieke gebruiker in 'n spesifieke gebruikssituasie en met inagneming van 'n spesifieke konteks op soek is.

Data en maklike toegang tot data is die goudstandaard van die moderne leksikografie en daarom mag die rol van die databasis waaraan 'n geïntegreerde woordeboek gekoppel is, nooit onderskat word nie. Trouens, dit plaas die beplanning van die leksikografiese databasis op die voorgrond, sien in hierdie verband ook Fuertes-Olivera et al. (2018) en Tarp (2022). Die leksikograaf moet besin oor beide die dataverspreidingstruktuur en die inligtingsonttrekkingstruktuur, en dan bepaal hoe die leksikografiese databasis gekonstrueer moet word om optimaal aan gebruikers se inligtingsbehoefes te kan voldoen. So 'n leksikografiese databasis vereis 'n hoë partikelaard om te verseker dat data maklik in vlakke en vakke gegroep kan word. Die ontwerp en implementering van so 'n komplekse databasis vereis uiteraard die insette van die leksikograaf (soos reeds aangedui), maar aangesien die leksikograaf in die meeste gevalle nie ook 'n programmeerder of stelselontwikkelaar is nie, word die insette van rekenaarwetenskaplikes vereis. Die belangrikheid van 'n mensgerigte koppelvlak is ook reeds aangetoon. Die leksikograaf is uiteraard ook nie 'n koppelvlakontwerper nie, en moet dus met spesialiste op die gebied van mens-rekenaarinteraksie (HCI = "human computer interaction") en gebruikerservaring (UX = "user experience") saamwerk om die koppelvlak te ontwerp. Dit beklemtoon die noodsaaklikheid van 'n multidissiplinêre span om die leksikografiese databasis en die beplande dataverspreidingstrukture en inligtingsonttrekkingstrukture te ontwerp. Indien hierdie beplanning nie vooraf in diepte gedoen word nie, sal dit onmoontlik wees om in 'n latere stadium gesofistikeerde inligtingsonttrekkingstrukture te implementeer wat aan 'n verskeidenheid gebruikersbehoefes voldoen.

In die leksikografie was daar 'n verskuiwing vanaf diskrete teks na taaldata en 'n beweging van 'n gerigtheid op die publikasie van boeke na 'n voorsiening van taaldienste (Ogilvie 2021: 91). Hierdie taaldienste sluit die erkenning van verskillende leksikografiese funksies asook die daarstelling van 'n verskeidenheid datatipes in. Dit is gebaseer op 'n mensgerigte benadering wat intuïtiewe woordeboekgebruik moontlik maak. Daarom moet die taaldienste op so 'n manier voorsien word dat gebruikers op 'n ondubbelsinnige en regstreekse manier daardie data kan bereik waaraan hulle die benodigde inligting kan onttrek. Dit kan slegs gebeur as 'n woordeboek se dataverspreidingstruktuur aangevul word met 'n inligtingsonttrekkingstruktuur.

Bronnelys

Aanlyn woordeboeke

lexiko: <https://www.owid.de/docs/elex/start.jsp>

Gedrukte woordeboeke

Hartmann, R.R.K. en G. James. 1998. *Dictionary of Lexicography*. Londen/New York: Routledge.
Johnson, S. 1755. *Dictionary of the English Language*. Londen: J. & P. Knapton.

Ander bronne

- Bergenholtz, H., S. Tarp en H.E. Wiegand.** 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Reds.). 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.* Berlyn: De Gruyter.
- Bothma, T.J.D.** 2011. Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. Fuertes-Olivera, P.A. en H. Bergenholtz (Reds.). 2011. *e-Lexicography: The Internet, Digital Initiatives and Lexicography.* Londen/New York: Continuum: 71-102.
- Bothma, T.J.D. en I. Fourie.** 2024a. Contextualised Dictionary Literacy, Information Literacy, and Information Behaviour in the e-Environment. *Library Management* (ter perse).
- Bothma, T.J.D. en I. Fourie.** 2024b. Enhancing Conceptualisations of Information Behaviour Contexts through Insights from Research on e-Dictionaries and e-Lexicography. *Information Research: An International Electronic Journal* 29(2): 179-197.
- Bothma, T.J.D. en R.H. Gouws.** 2020. e-Dictionaries in a Network of Information Tools in the e-Environment. *Lexikos* 30: 29-56.
- Bothma, T.J.D. en R.H. Gouws.** 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that Are Linked to e-Texts. *Lexikos* 32(2): 53-81.
- Bothma, T.J.D. en D.J. Prinsloo.** 2013. Automated Dictionary Consultation for Text Reception: A Critical Evaluation of Lexicographic Guidance in Linked Kindle e-Dictionaries. *Lexicographica* 29: 165-198.
- Fuertes-Olivera, P.A., S. Tarp en P. Sepstrup.** 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Diccionarios Valladolid-UVa. *Lexikos* 28: 152-176.
- Gouws, R.H.** 2014. Article Structures: Moving from Printed to e-Dictionaries. *Lexikos* 24: 155-177.
- Gouws, R.H.** 2022. Extended Article Comments in Online Dictionaries. *Lexicography. Journal of ASIALEX* 9(2): 111-138.
- Gouws, R.H.** 2023. Nuwe raamstruktuurtypes in aanlyn woordeboeke. *Lexikos* 33(1): 382-403.
- 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.
- Gouws, R.H. en S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Gouws, R.H. en S. Tarp.** Ter perse. Despite Current Challenges: Lexicography Has a Bright Future. *Lexicographica* 40.
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Reds.). 1989-1991:103-111.
- Hausmann, F.J.** 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. Hausmann, F.J. et al. (Reds.). 1989-1991: 1-19.
- Hausmann, F.J. et al. (Reds.).** 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlyn: De Gruyter.
- Huang, F. en S. Tarp.** 2021. Dictionaries Integrated into English Learning Apps: Critical Comments and Suggestions for Improvements. *Lexikos* 31(1): 68-92.

- Ogilvie, S.** 2021. The Future of Dictionaries and Lexicography. *Dictionaries: Journal of the Dictionary Society of North America* 42(2): 91-94.
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Berlyn/New York: Max Niemeyer.
- Tarp, S.** 2022. Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. *Lexikos* 32(1): 66-87.
- Tarp, S. en R.H. Gouws.** 2020. Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 470-498.
- Wiegand, H.E.** 1989. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. Hausmann, F.J. et al. (Reds.). 1989-1991: 462-501.
- Wiegand, Herbert Ernst.** 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlyn/New York: De Gruyter.
- Wiegand, H.E., S. Beer en R.H. Gouws.** 2013. Textual Structures in Printed Dictionaries. An Overview. Gouws, R.H. et al. (Reds.). 2013: 31-73.
- Wiegand, H.E. en R.H. Gouws.** 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H. et al. (Reds.). 2013: 273-314.