
Aktuelle Aussprachewörterbücher des Deutschen und des Englischen im Vergleich

Krzysztof Nycz, *Institut für Germanistik, Philologische Fakultät,
Universität Rzeszów, Polen*
(knycz@ur.edu.pl)

(<https://orcid.org/0000-0003-2976-1746>)

und

Zygmunt Tęcza, *Institut für Germanistik, Philologische Fakultät,
Universität Rzeszów, Polen*
(ztęcza@ur.edu.pl)

(<https://orcid.org/0000-0002-9543-4674>)

Zusammenfassung: In dem vorliegenden Artikel werden insgesamt fünf, zwei deutsche und drei englische Aussprachewörterbücher, die den aktuellen Entwicklungsstand der Phonolexikographie des Deutschen und des Englischen repräsentieren, einer kritisch-vergleichenden Betrachtung unterzogen. Das Ziel dieser Gegenüberstellung besteht nicht nur darin, die Stärken und die Schwächen der einzelnen Werke vor dem Hintergrund der aktuellen lexikographischen Praxis zu erörtern und ggf. auch Impulse für zukünftige Entwicklungen in der Darstellung und Vermittlung von Ausspracheinformationen zu geben, sondern auch, einen interlingualen Diskurs anzuregen, der zur Reflexion über universelle und sprachspezifische Herausforderungen in der phonetischen Standardisierung und Kodifizierung beitragen könnte.

Den Ausgangspunkt bildet eine allgemeine Charakteristik der zu untersuchenden Wörterbücher, bei der Aspekte wie ihre Entstehungsgeschichte, ihre Zielgruppen sowie Konstruktion und Umfang des jeweiligen Wörterverzeichnisses und des einführenden Textteils zur Sprache kommen. Die eigentliche Analyse fokussiert sich auf vier Schwerpunkte: die Auffassung dessen, was als Standardaussprache zu verstehen und zu kodifizieren ist, den Umgang mit phonostilistischen und regionalen Aussprachevarianten, ferner die grafische Gestaltung der verglichenen Nachschlagewerke und schließlich ihre multimedialen Erweiterungen. Im letzten Abschnitt werden die in den genannten Bereichen festgestellten Gemeinsamkeiten und vor allem Unterschiede zwischen den analysierten Wörterbüchern zusammenfassend geordnet und abschließend wird noch kurz ein subjektiver Ausblick auf die Zukunftsperspektiven der Aussprachewörterbücher gewagt.

Schlüsselwörter: AUSSPRACHEWÖRTERBUCH, AUSSPRACHEKODIFIZIERUNG, DEUTSCHE AUSSPRACHE, ENGLISCHE AUSSPRACHE, STANDARDAUSSPRACHE, PHONETISCHER STANDARD, AUSSPRACHEVARIANTEN, PHONOLEXIKOGRAPHIE, DIGITALE LEXIKOGRAPHIE, ELEKTRONISCHE LEXIKOGRAPHIE

Abstract: **Current Pronunciation Dictionaries of German and English Compared.** The following paper takes a critical and comparative look at a total of five dictionaries of

Lexikos 35(2) (AFRILEX-reeks/series 35A: 2025): 232-255

pronunciation, two German and three English, which represent the current state of development in German and English phono-lexicography. The aim of this comparison is not only to discuss the advantages and disadvantages of the individual works in the context of current lexicographical practice and, possibly, to provide ideas for future developments in presenting and conveying pronunciation information, but also to stimulate an interlingual discourse that could contribute to reflection on universal and language-specific challenges in phonetic standardisation and codification.

As a starting point, a general description of the dictionaries under investigation is provided, covering aspects such as their origins, target groups, as well as the construction and extent of the respective word lists and front matters. The actual analysis focuses on four main areas: the concept of what is to be considered and codified as standard pronunciation, the treatment of phono-stylistic and regional pronunciation variants, the graphic design of the dictionaries compared, and finally their multimedia extensions. The last section summarises the similarities and, above all, the differences between the dictionaries analysed in the above areas, and concludes with a brief subjective outlook on the future prospects of pronunciation dictionaries.

Keywords: PRONUNCIATION DICTIONARY, PRONUNCIATION CODIFICATION, GERMAN PRONUNCIATION, ENGLISH PRONUNCIATION, STANDARD PRONUNCIATION, PHONETIC STANDARD, PRONUNCIATION VARIANTS, PHONO-LEXICOGRAPHY, DIGITAL LEXICOGRAPHY, ELECTRONIC LEXICOGRAPHY

1. Einleitung

Im Zeitalter globalisierter Kommunikation und zunehmender Medialisierung stellen Aussprachewörterbücher nach wie vor ein zentrales Instrument zur Normierung und Standardisierung der Aussprache dar. Der vorliegende Beitrag widmet sich einer kritisch-vergleichenden Analyse gegenwärtig in Gebrauch befindlicher orthoepischer Nachschlagewerke des Deutschen und des Englischen. Hierbei werden zwei aktuellste Werke der deutschen Phonolexikographie, nämlich *Deutsches Aussprachewörterbuch* (2009) sowie *DUDEN — Das Aussprachewörterbuch* (Bd. 6 der DUDEN-Reihe, 2023), den drei führenden englischsprachigen Referenzwerken dieser Art in der jeweils jüngsten Ausgabe gegenübergestellt, und zwar dem *CAMBRIDGE English Pronouncing Dictionary* von Daniel Jones (2022), dem *LONGMAN Pronunciation Dictionary* von John Wells (2008) sowie dem *OXFORD Dictionary of Pronunciation for Current English* (2001/2003).

Vier wesentliche Aspekte stehen im Mittelpunkt der Untersuchung: Zum einen wird die jeweilige Auffassung des phonetischen Standards dargestellt, d. h. die Auffassung darüber, was unter Standardaussprache zu verstehen ist. Zum anderen wird der Frage nachgegangen, inwieweit der Unterschied zwischen normativen Vorgaben und sprachlicher Realität im Sinne phonostilistischer bzw. phonopragmatischer Abweichungen von der Standardaussprache, respektive nationale bzw. regionale Aussprachevarianten, in den fraglichen Wörterbüchern berücksichtigt werden. Ein weiterer Schwerpunkt liegt auf der grafischen Gestaltung, die nicht nur als visuelle Repräsentation phonetisch-phonologischer Informationen fungiert, sondern auch deren Zugänglichkeit und Verständlich-

keit beeinflusst. Schließlich werden die multimedialen Komponenten der untersuchten Nachschlagewerke, etwa in Form digitaler Anwendungen und interaktiver Elemente, kritisch hinterfragt — ein Aspekt, der gerade im digitalen Zeitalter sehr stark an Bedeutung gewinnt.

Das Ziel der vergleichenden Betrachtung ist, die Stärken und die Schwächen der einzelnen Werke im Kontext der aktuellen lexikographischen Praxis zu erörtern und Impulse für zukünftige Entwicklungen in der Darstellung und Vermittlung von Ausspracheinformationen zu geben. Durch die Gegenüberstellung deutscher und englischer Ansätze wird zudem ein interlingualer Diskurs angeregt, der zur Reflexion über universelle und sprachspezifische Herausforderungen in der phonetischen Standardisierung und Kodifizierung beiträgt.

2. Gegenstand der Betrachtung. Genese, Ziel und Konstruktion der untersuchten Wörterbücher

Gegenstand des Vergleichs sind im vorliegenden Beitrag — wie bereits erwähnt — insgesamt fünf (zwei deutsche und drei englische) Aussprachewörterbücher, die zum Teil unterschiedliche Traditionen und Ansätze repräsentieren, jeweils aber den aktuellsten Stand der wissenschaftlichen und phonodidaktischen Auseinandersetzung mit der Standardlautung des Deutschen und des Englischen widerspiegeln. Die zu untersuchenden deutschen Aussprachewörterbücher sind:

- *Deutsches Aussprachewörterbuch* von Eva-Maria Krech, Eberhard Stock, Ursula Hirschfeld und Lutz Ch. Anders, erschienen im Jahr 2009 (nachstehend DAWB), und
- *DUDEN — Das Aussprachewörterbuch* (Bd. 6 der DUDEN-Reihe), bearbeitet von Stefan Kleiner und Ralf Knöbl, in seiner jüngsten, aktualisierten 8. Ausgabe von 2023 (nachstehend DU-23).

Die Vorgeschichte des einen wie des anderen Wörterbuchs ist mittlerweile recht lang, sie reicht in beiden Fällen bis in die 1960er Jahre zurück. Das DAWB knüpft an das *Große Wörterbuch der deutschen Aussprache* (GWDA) von 1982 an, das wiederum eine erweiterte Fassung des 1964 in der damaligen DDR zum ersten Mal herausgebrachten *Wörterbuchs der deutschen Aussprache* (WDA) darstellt. Wie im Einleitungsteil des DAWB ausdrücklich betont wird, verfolgt dieses Nachschlagewerk ähnliche Ziele wie das GWDA und es basiert auf denselben konzeptionellen wie methodologischen Ansätzen (vgl. DAWB: 15). Die Kontinuität des DUDEN-Aussprachewörterbuchs nicht nur zu seinem direkten Vorgänger, der ebenfalls von St. Kleiner und R. Knöbl stammenden siebten Auflage aus dem Jahr 2015, sondern auch zu allen früheren Ausgaben (in umgekehrt chronologischer Reihenfolge: von 2005, 2003, 2000, 1990, 1974 und 1962), bei denen Max Mangold die Federführung hatte, ist hingegen allein schon durch die redaktionelle Betreuung und den festen Platz des Wörterbuchs in der zwölfbändigen *DUDEN-Reihe* gewährleistet.

Bei den Aussprachewörterbüchern des Englischen handelt es sich um die folgenden Werke:

- *CAMBRIDGE English Pronouncing Dictionary* von Daniel Jones in seiner 18. Ausgabe von 2022 (nachstehend EPD),
- *LONGMAN Pronunciation Dictionary* von John Wells in seiner 3. Ausgabe von 2008 (nachstehend LPD) und
- *OXFORD Dictionary of Pronunciation for Current English* von Clive Upton, William A. Kretzschmar und Rafal Konopka, das zuerst als Hardcover-Ausgabe im Jahr 2001 und später als Paperback-Version im Jahr 2003 erschien (nachstehend ODP).

Vor allem das erstgenannte Nachschlagewerk blickt auf eine lange Tradition zurück: Bereits 1917 war nämlich die erste Ausgabe von „The Jones EPD“ erschienen, während *LONGMAN Pronunciation Dictionary* 1990 erstmals herausgebracht wurde und *OXFORD Dictionary of Pronunciation for Current English* noch gut ein Jahrzehnt jünger ist. Nach der Erstausgabe Anfang der 2000er Jahre kam die zweite, aktualisierte Ausgabe 2017 unter dem Titel *The Routledge Dictionary of Pronunciation for Current English* (nachstehend RDP) mit Clive Upton und William A. Kretzschmar als Autorenteam auf den Markt.¹

Alle fünf Wörterbücher — d. h. sowohl die deutschen als auch die englischen — definieren ihren Empfängerkreis möglichst weit und nennen dabei explizit zwei verschiedene Zielgruppen, an die sie sich richten: die Mutter- oder ErstsprachlerInnen einerseits und die NichtmuttersprachlerInnen, v. a. Sprachlernende, andererseits (vgl. DAWB: 4. Umschlagseite, DU-23: 11 f., EPD: iii, LPD: xvii, ODP: ix). Die Ersteren, darunter oft Personen, die von Berufs wegen an phonetischer Korrektheit interessiert sind,² suchen meist Auskunft zur Aussprache von Fremdwörtern oder (fremdsprachigen) Eigennamen, ggf. auch zu möglichen Aussprachevarianten einheimischer Wörter; die Letzteren, für die die Sprache des jeweiligen Wörterbuchs eine Fremdsprache ist, benötigen dagegen vor allem Informationen zur Standardaussprache des darin kodifizierten heimischen (Kern-)Wortschatzes, seltener zu dessen Aussprachevarianten (vgl. DU-23: 11 f., LPD: xvii, ODP: ix).

Das Größenverhältnis zwischen den beiden Zielgruppen dürfte sich allerdings für die deutschen und die englischen Wörterbücher unterschiedlich gestalten. Zieht man nämlich zum einen in Betracht, dass sich im Englischen die Aussprache von Wörtern in einem viel geringeren Maße aus deren Schreibung folgern lässt, als es im Deutschen der Fall ist, und zum anderen, dass die Anzahl von Menschen, die Englisch als Fremd- oder Zweitsprache verwenden, im Verhältnis zu der Anzahl der Muttersprachler sehr hoch ist und diese (anders als beim Deutschen) auch deutlich übersteigt, so liegt die Schlussfolgerung nahe, dass grundsätzlich nach einem deutschen Aussprachewörterbuch proportional weitaus weniger NichtmuttersprachlerInnen greifen als nach einem englischen.

Ein Vergleich der Aussprachewörterbücher zeigt weiterhin, dass für das Deutsche und für das Englische z. T. unterschiedliche redaktionelle Ansätze ver-

folgt wurden, die sich allein schon im Umfang der Werke widerspiegeln. Bei den deutschen Wörterbüchern bewegt sich die Gesamtzahl von Seiten, Wort-einträgen und angeführten Transkriptionen in einem recht engen Rahmen. So umfasst das DAWB 1.076 Seiten, etwa 130.000 Lemmata und insgesamt 150.000 Transkriptionen, während DU-23 mit seinen 944 Seiten rund 136.000 Stichwörter und 144.000 Transkriptionen auflistet.

Bei den Wörterbüchern des Englischen ist das Bild differenzierter. Das EPD umfasst etwa 606 Seiten und enthält 80.000 Einträge mit insgesamt über 230.000 Transkriptionen. Im Vergleich dazu zeigt das LPD bei rund 920 Seiten eine deutlich höhere Anzahl von Stichwörtern (138.000) und etwas geringere Gesamtzahl an Transkriptionen, die sich auf ca. 225.000 beläuft. Auffällig ist jedoch vor allem das ODP bzw. das RDP mit entsprechend 1208 und fast 1600 Seiten. Bei ca. 100.000 Einträgen (und zusätzlich 2.000 Neueinträgen im Fall des RDP) werden hier ca. 300.000 Transkriptionen aufgeführt. Der signifikant größere Umfang der beiden Ausgaben im Vergleich zu EPD und LPD resultiert vor allem aus der konsequenten Aufnahme von zwei separaten Transkriptionen für das *British English* und das *American English* — auch dann, wenn beide Varianten identisch sind.

Die Wortlisten der Wörterbücher, bei denen es sich nicht um die Erstausgabe handelt, also: DU-23, EPD und LPD, basieren naturgemäß auf den Wörterverzeichnissen der jeweiligen Vorgängerauflagen. In gewissem Maße betrifft dies auch das DAWB, das sich, wie oben erwähnt, in der Nachfolge des gut ein Vierteljahrhundert älteren GWDA sieht. Die Quellen des in die aktuellen Ausgaben neu aufgenommenen Wortmaterials sind in den deutschen Aussprachewörterbüchern direkter und auch übersichtlicher als in den englischen aufgelistet. DU-23 nennt in diesem Zusammenhang das Rechtschreib- und das Fremdwörterbuch aus dem Hause Duden (vgl. DU-23: 16), in DAWB ist die Quellenliste, da die Aktualisierung und vor allem Erweiterung des Wörterverzeichnisses in diesem Fall ein unvergleichlich größeres Ausmaß haben musste, wesentlich länger. Außer den beiden schon genannten Titeln umfasst sie eine Online-Datenbank der Universität Leipzig, ferner einige Universalwörterbücher und Wörterbücher für Deutsch als Fremdsprache, ein Neologismen- und ein Herkunftswörterbuch, ein rückläufiges Wörterbuch und noch einiges mehr (vgl. DAWB: 17 f.). Demgegenüber fällt die einschlägige Information des in der ersten Ausgabe erschienenen englischen ODP: „[T]he headword list is tied closely to the lists of all the major dictionaries of Oxford University Press“ (ODP: vii) eher spärlich aus.

Die Transkription der Einträge im Wörterverzeichnis reflektiert prinzipiell die Aussprachevariante(n), die das jeweilige Wörterbuch seinen BenutzerInnen als standardgemäß empfiehlt (mehr dazu in Abschn. 3). Bei der Beurteilung und Auswahl der jeweils zu empfehlenden Aussprache werden von den WörterbuchautorInnen oft empirische Quellen analysiert, zu denen diverse Datenbanken, Aufnahmen von Sendungen der elektronischen Medien, im Internet verfügbare Audio- und Videoaufnahmen und nicht zuletzt selbst durchgeführte Akzeptanzuntersuchungen zu konkreten Aussprachevarianten gehören (zu den Letzteren s. Abschn. 4), wobei die Quellen auch in diesem Fall in den deutschen

Wörterbüchern meist expliziter genannt werden als in den englischen (vgl. DAWB: 16, DU-23: 17, LPD: xviii). Wie es dem auch sei, lässt sich wohl das folgende Bekenntnis des EPD auch auf die übrigen Wörterbücher beziehen:

Ultimately, however, the decisions about which pronunciation to recommend, which pronunciations have dropped out of use, and so on, have been based on the editors' intuitions as professional phoneticians and observers of the pronunciation [...] over many years. The opinion of many colleagues and acquaintances has also been a valuable source of advice. (EPD: vi)

Allen Wörterbüchern ist gemeinsam, dass der jeweiligen Wortliste ein mehr oder weniger informativer einführender Textteil (*front matter*) vorausgeht. Im Hinblick auf dessen Beschaffenheit divergieren die untersuchten Nachschlagewerke allerdings erheblich. Bei den Aussprachewörterbüchern des Deutschen fällt vor allem die enorme Ausführlichkeit der Darstellung auf: So umfasst der einführende Teil des DAWB beeindruckende 280 Seiten, was durchaus dem Umfang eines eigenständigen Phonetik-Lehrbuchs entsprechen dürfte. Im Vergleich dazu ist die Einführung im DU-23 mit ca. 150 Seiten zwar kürzer, aber immer noch sehr umfangreich und informativ. Beide Werke setzen damit auf eine intensive theoretisch-phonodidaktische Aufbereitung, die insbesondere für Studierende und Fachleute von hohem Wert ist. In beiden Fällen werden im einführenden Teil, wenn auch in unterschiedlichem Ausmaß und unterschiedlicher Anordnung, u. a. folgende Schwerpunkte behandelt:

- Erklärungen zur Konstruktion des Wörterverzeichnisses bzw. Anleitung zu dessen Gebrauch,
- Kriterien der Aufnahme von Wörtern ins Wörterverzeichnis; Behandlung von Flexionsformen und Wortbildungskonstruktionen,
- die IPA-Transkription und deren jeweilige Handhabung,
- phonologisch-phonetisches Grundwissen: Laut und Phonem, Phonemvarianten; Kriterien der Beschreibung von Vokalen und Konsonanten (distinktive Merkmale); Silbe und Silbengrenze,
- die Standardaussprache, deren historische Entwicklung und aktuelles Verständnis; Variation in der Standardaussprache; im DAWB zusätzlich die Spezifik der (Standard-)Aussprache in Österreich und in der Deutschschweiz, jeweils in einem separaten Hauptkapitel des Einführungsteils erörtert,
- Einzelbeschreibungen der (Standard-)Artikulation sämtlicher Vokale und Konsonanten mit entsprechenden Grafiken des Ansatzrohres,
- Informationen zur Phonem-Graphem-Zuordnung und zur Distribution einzelner Vokale und Konsonanten,
- Aussprachehinweise zu ausgewählten Fremdsprachen.

Somit ermöglichen also die deutschen Aussprachewörterbücher mit ihrem jeweiligen Textteil als einem lehrbuchartigen Werk eine vertiefte Auseinandersetzung mit den Schlüsselfragen der Phonetik und Phonologie.

Im Gegensatz dazu setzen die Aussprachewörterbücher des Englischen auf ein schlankes und pragmatisches Konzept, das vor allem dem schnellen Zugriff auf phonologisch-phonetisches Grundwissen und praktische Transkriptionsregeln dienen soll. Der einführende Textteil ist hier in jedem Wörterbuch vergleichbar lang (bzw. vergleichbar kurz) und umfasst nur ca. 20 Seiten. Folgende Schwerpunkte werden darin behandelt:

- Erklärungen zum Aufbau des Wörterverzeichnisses,
- Informationen zu den Aussprachemodellen bzw. -varianten, die der Kodifizierung zugrunde liegen und im jeweiligen Wörterbuch berücksichtigt werden,
- grundlegende Informationen zur Aussprache von Vokalen, Diphthongen und Konsonanten, zur Wortakzentuierung, zu Assimilationen und Elisionen, Silben und Silbentrennung,
- Hinweise zur Aussprache von fremdsprachigen Lauten,
- die IPA-Transkription und deren jeweilige Handhabung.

Die 18. Ausgabe des EPD enthält zusätzlich eine Sammlung von sechs kurzen Aufsätzen zu unterschiedlichen phonetischen Einzelfragen (zusammengefasst unter der Überschrift *The World of Pronunciation*), die von führenden ExpertInnen im Bereich der Phonetik verfasst wurden. Es handelt sich dabei um folgende Beiträge: *Pronunciation in Spontaneous Speech* von Richard Cauldwell, *Pronunciation for English as a Lingua Franca* von Jennifer Jenkins, *The Daniel Jones Legacy* von Jack Windsor Lewis, *Teaching and Learning Pronunciation* von Jonathan Marks, *The BBC, Its Pronunciation Unit, and 'BBC English'* von Catherine Sangster sowie *Shortcuts in Casual English Pronunciation* von Linda Shockey.

3. Die Auffassung des phonetischen Standards, oder wessen Aussprache kodifiziert wird

In den Aussprachewörterbüchern des Deutschen, DAWB und DU-23, beruht die Auffassung des phonetischen Standards auf einem vielschichtigen Konzept, das einerseits die normativen und regulativen Funktionen der Standardaussprache berücksichtigt, andererseits aber auch ihrer Realitätsnähe und innerer Differenziertheit Rechnung trägt.

In diesem Punkt verfolgen die beiden Wörterbücher zum Teil unterschiedliche Ansätze. So wird in DAWB die Standardaussprache bereits zu Beginn des Einführungsteils als „die mündliche Form der Standardvarietät in der Bundesrepublik Deutschland“ definiert, die „in geografischer und sozialer Hinsicht über eine weite Geltung [verfügt] und [...] insbesondere durch die elektronischen Medien verbreitet [wird]“ (6). Die Standardaussprache sei demnach

- dialektneutral, d. h. sie verzichte bewusst auf mundartliche oder regional gefärbte umgangssprachliche Varianten, sodass sie in allen sozialen Gruppen und von allen MuttersprachlerInnen überregional verstanden werden könne;

- in offiziellen, öffentlichen Situationen bevorzugt zu verwenden;
- durch unterschiedliche Grade der Artikulationspräzision charakterisiert, was bedeute, dass ihre Anwendung auch außerhalb öffentlicher Kontexte möglich sei;
- kodifiziert und erfülle damit eine normative bzw. regulative Funktion; dabei variere der Grad ihrer Verbindlichkeit, die wohlgemerkt in erster Linie für die soziale Gruppe der BerufssprecherInnen gelte (ebd. 7).

Das DU-23 erweitert noch diese Perspektive, indem es zwar den überregionalen Charakter der Standardaussprache unterstreicht, zugleich aber Raum für Varianten zulässt, die im deutschsprachigen Gebiet nur (sub-)national oder großregional verbreitet sind. Trotz ihrer Normativität orientiert sich die so verstandene Standardaussprache eng an der tatsächlichen Sprechwirklichkeit und berücksichtigt neben der als überregional geltenden Aussprache geschulter BerufssprecherInnen auch die in der Bevölkerung üblichen Aussprachevarianten. Somit wird der Standardaussprache auch das Merkmal der inneren Differenziertheit bzw. Variabilität zugeschrieben. Sie zeigt sich darin, dass die in der Bevölkerung tatsächlich verwendete Standardaussprache, selbst in formellen Situationen, in unterschiedlichem Ausmaß von der überregionalen Norm abweicht.

In DAWB werden dagegen, was die innere Differenziertheit bzw. Variabilität angeht, drei konkrete phonostilistische Ebenen der Standardaussprache, die sog. *Grade der Artikulationspräzision* (vgl. 99 ff.), unterschieden:

- (1) hohe bis mittlere Artikulationspräzision (v. a. beim reproduzierenden Sprechen wie Vorlesen von Nachrichten im Radio und Fernsehen);
- (2) sehr hohe Artikulationspräzision (v. a. beim reproduzierenden Sprechen wie z. B. einem feierlichen Vortrag);
- (3) verminderte Artikulationspräzision (v. a. beim freien Sprechen, z. B. bei öffentlich geführten Gesprächen im Radio und Fernsehen).

Als Kodifizierungsgrundlage wird dabei die unter (1) genannte hohe bis mittlere Sprechspannung und Artikulationspräzision gesetzt. Diese Wahl ist übrigens als eine klare Prioritätenverschiebung im Verhältnis zu früheren Aussprachewörterbüchern, etwa dem *Großen Wörterbuch der deutschen Aussprache* (1982), zu bewerten, in dem die sehr hohe Artikulationspräzision die zentrale Bezugsebene war, und findet nun zwangsläufig sowohl bei der Formulierung der Ausspracheregeln im Einführungsteil des Wörterbuchs als auch bei den Ausspracheangaben im Wörterverzeichnis ihren Niederschlag.

In dieser Frage gehen die Autoren des DU-23 teilweise einen anderen Weg. Sie vertreten die Auffassung, dass es für eine möglichst realitätsnahe Dokumentation der deutschen Aussprache unerlässlich sei, den Kreis der als kodifikationsrelevant angesehenen Situationen und SprecherInnen zu erweitern. So seien Sprachformen, die von Millionen deutschen MuttersprachlerInnen in formellen Sprechsituationen — etwa im schulischen Kontext, sowohl von Lehrkräften als auch von SchülerInnen — verwendet und als situationsangemessen eingestuft

werden, als standardsprachlich im Sinne eines ‚Gebrauchsstandards‘ zu werten. Dabei wird direkt auf das im englischsprachigen Raum bereits etablierte Konzept Bezug genommen, den Sprachgebrauch der *educated speakers* als Basis für standard-sprachliche Aussprachewörterbücher und Grammatiken zu nutzen (vgl. DU-23: 30 f.). Diese Vorgehensweise betrachten die Autoren auch für das Deutsche als einen „zeitgemäßen Schritt zu einem erweiterten und damit realitäts- und gebrauchsnäheren Standardsprachkonzept“ (ebd. 31).

Offensichtlich ist ihnen daran gelegen, ein möglichst differenziertes Bild der deutschen Aussprache in formellen Kontexten zu zeichnen. Anders als in DAWB wird hierbei auf eine hierarchische phonostilistische Registeraufspaltung verzichtet. Stattdessen bietet DU-23 in einem speziellen Kapitel einen kompakten und dennoch recht informativen Überblick über relevante Varianten und Variationsphänomene innerhalb der Standardaussprache; alles in allem werden etwa vierzig solcher Erscheinungen auf dem Gebiet der Vokale, Konsonanten und der Realisation von Nebensilben diskutiert. Darüber hinaus sind im Wörterverzeichnis bei einer Reihe von Lemmata auch regional gebräuchliche Aussprachevarianten mit entsprechender Angabe (wie *österr.*, *schweiz.*, *nordd.*, *vorw. nordd.* oder *bes. nordd.*, *südwestd.* etc.) aufgeführt, eine Lösung, die als recht funktional zu bewerten ist und die man in DAWB, das bei einheimischen Wörtern prinzipiell nur eine Transkription angibt, gänzlich vermisst.³

Was nun die Wörterbücher der englischen Aussprache angeht, so fokussieren sie sich traditionell in erster Linie auf das britische Englisch. Dabei galt jahrzehntelang die Aussprache der britischen bzw. südenenglischen Elite der öffentlichen Schulen (*Public School Pronunciation*), später die sog. RP, *Received Pronunciation*, als Modell, auf das sich die orthoepischen Nachschlagewerke bei der Kodifizierung der Aussprache stützten.

In ihrer ursprünglichen Form war die RP eine Art Standard, dem die Aussprache gebildeter SprecherInnen in London und den sog. *Home Counties* zugrunde lag. Ihr Gebrauch war jedoch nicht auf diese Region beschränkt, sondern war im neunzehnten Jahrhundert für die Sprache der Oberschicht im ganzen Land charakteristisch. In England und in Wales gilt die RP weiterhin als Mustermodell für die korrekte Aussprache, insbesondere in der formalen Sprache der gebildeten SprecherInnen (vgl. EPD: v).

Allerdings sind sich die Autoren aller drei aktuellen Aussprachewörterbücher des Englischen einig, dass solch eine konservative Auffassung der RP heutzutage kaum noch haltbar ist. In EPD wird daher programmatisch vorgeschlagen: „The time has come to abandon the archaic name *Received Pronunciation*“ (ebd.). Stattdessen wird eine modernisierte RP-Variante („a modernized version of RP“), ein breiter angelegtes und leichter zugängliches Aussprachemodell für das britische Englisch vorgeschlagen. In EPD und LPD wird dieses Aussprachemodell als *BBC English* bezeichnet (vgl. EPD: v; LPD: xix).

[T]his is the pronunciation of professional speakers employed by the BBC as news-readers and announcers on BBC1 and BBC2 television, the World Service and BBC Radio 3 and 4, as well as many commercial broadcasting organisations such as ITN. (EPD: v)

Es wird als ein wichtiges Merkmal der modernen RP betont, dass die Sprache der BBC-SprecherInnen von den meisten Menschen nicht mit der hohen sozialen Gesellschaftsschicht und mit Privilegien assoziiert wird, wie dies in der Vergangenheit bei der PSP (*Public School Pronunciation*) und der RP der Fall war (vgl. ebd.). Darüber hinaus wird hervorgehoben, dass die moderne RP überregional ist, d. h. nicht an eine bestimmte Stadt oder Region gebunden.

Das ODP geht diesbezüglich zum Teil seinen eigenen Weg und unterscheidet zwischen *marked* und *mainstream/unmarked* RP. Die *marked* RP ist demnach eine Aussprachevariante, die, wenn sie von den meisten britischen MuttersprachlerInnen gehört wird, dazu führt, dass der Benutzer oder die Benutzerin als altmodisch, affektiert oder präventios eingestuft wird; sie wird mit einer kleinen Gruppe älterer SprecherInnen der Mittel- und Oberschicht assoziiert, die eine enge Verbindung zum Südosten Englands haben. Währenddessen ist die *mainstream/unmarked* RP eine Aussprachevariante, die mit Ausbildung und Kultiviertheit/Eleganz assoziiert wird, dabei keine besonderen regionalen Untertöne enthält und keine negativen Bewertungen evoziert. Sie wird von den meisten nationalen Radio- und FernsehnachrichtensprecherInnen, von Menschen mittleren Alters und jüngeren Berufstätigen verwendet („broadcast RP“). Die *unmarked* RP ist überregional und nicht an eine bestimmte soziale Schicht gebunden. Es ist die Aussprache eines breiten Spektrums von MuttersprachlerInnen aus vielen Teilen des Landes, die einen sehr unterschiedlichen beruflichen Hintergrund (aber ein eher überdurchschnittliches Bildungsniveau) aufweisen (vgl. ODP: xii). Die *unmarked* RP bildet in ODP die Kodifizierungsgrundlage und entspricht im Allgemeinen der in den zwei anderen Aussprachewörterbüchern beschriebenen modernen RP.⁴

Was nun die amerikanische Aussprache betrifft, so wird sie in allen drei Wörterbüchern mehr oder weniger explizit zu der als *General American* bekannten Sprachvarietät in Verbindung gesetzt, die von der Mehrheit der AmerikanerInnen gesprochen wird, genauer: von denjenigen, die keinen auffälligen östlichen oder südlichen Akzent haben. Das *General American* ist normalerweise bei professionellen SprecherInnen in nationalen Nachrichten- und Informationssendungen zu hören (vgl. LPD: xx; EPD: v–vi).

In EPD und in ODP wird des Weiteren darauf hingewiesen, dass keine einzige Varietät — ob regional oder sozial — als amerikanischer Standard herausgehoben wurde (vgl. EPD: vi; ODP: xiv). So gibt es selbst in den nationalen Medien mit professionell ausgebildeten SprecherInnen auch solche mit regional gefärbter Aussprache. Dennoch könne nach Auffassung des EPD dieses mediale „Network English“ in seiner farblosesten Form schon als eine relativ homogene Variante der gesprochenen Sprache (im Orig. „dialect“) beschrieben werden, welche die fortschreitende Entwicklung der amerikanischen Dialekte widerspiegele. Darauf Bezug nehmend erklären die EPD-Autoren, in ihrem Wörterverzeichnis die Vielfalt der Aussprachevarianten mit dem geringsten Maß an regionaler oder sozialer Markierung auflisten zu wollen, „while still being sensitive to the traits of the individual word“ (EPD: vi).

Regionale Varianten, und zwar sowohl für das britische als auch für das amerikanische Englisch, werden in keinem der analysierten englischen Wörterbücher aufgeführt. In EPD und in ODP wird darauf explizit hingewiesen, in dem Letzteren mit folgender Begründung:

It would of course be both impracticable and confusing to attempt to present transcriptions for even a fraction of the variety of regional and social accents which characterize British and American English. For both pragmatic and pedagogic reasons model accents have to be sought, transcriptions of which may be generally accepted as embodying the major elements of the pronunciation-types under consideration. (ODP: xix)

Mit praktischen Gründen wird auch von den Autoren des EPD argumentiert, wobei sie eher den enormen Arbeitsaufwand in den Vordergrund stellen, den die Aufnahme von regionalen Aussprachevarianten ins Wörterbuch erfordern würde:

A pronouncing dictionary that systematically presented the pronunciations of a range of regional accents would be very valuable, but it would be very much bigger than the present volume and the job of ensuring an adequate coverage which treated all accents as equally important would have taken many years. (EPD: vi)

Die (Nicht-)Berücksichtigung von regional verbreiteten Spielarten der jeweiligen Aussprache ist als einer der markantesten Unterschiede zwischen den beiden deutschen und den drei englischen Wörterbüchern anzusehen, die hier Gegenstand des Vergleichs sind. Zieht man jedoch in Betracht, dass Deutsch hauptsächlich in drei Ländern mit insgesamt etwa 100 Millionen Einwohnern gesprochen wird, Englisch dagegen außer in Großbritannien und den USA auch in Irland, Kanada, Australien, Neuseeland sowie in einigen weiteren Dutzend Ländern der Welt als Mutter- und/oder Amtssprache in Gebrauch ist, so muss der Anspruch, die gesamte Vielfalt von gegenwärtig vorkommenden Aussprachevarianten des Englischen zu erfassen, in der Tat als recht unrealistisch bewertet werden.

4. Grafische Gestaltung und multimediale Komponenten

4.1 Grafische Gestaltung

Die hier untersuchten Aussprachewörterbücher sowohl des Deutschen als auch des Englischen sind im Vergleich zu den Wörterbüchern früherer Generationen durchaus fortschrittliche Nachschlagewerke, die durch eine klare, benutzerfreundliche und didaktisch orientierte grafische Gestaltung auffallen und neuartige, multimediale Komponenten — dem digitalen Zeitalter entsprechend — enthalten. Bis auf das ODP setzen alle Werke in ihrem jeweiligen Wörterverzeichnis visuell hervorstechende Informationskästen ein, in denen das phonetisch-phonologische

In den beiden Aussprachewörterbüchern des Deutschen werden dabei ähnliche gestalterische Lösungen verwendet. Im Wörterverzeichnis des DAWB und des DU-23 finden sich nämlich ca. 190 bzw. knapp 180 farbige Infokästen (vgl. Abb. 1), die jeweils an das direkt vorausgehende Stichwort anknüpfen, auf Aussprachebesonderheiten oder -schwierigkeiten hinweisen und — wie in der Anleitung von DAWB versprochen — „Anregungen zur produktiven Erschließung weiterer Ausspracheformen durch den Nutzer“ (279) liefern.

großkalibrig			Wisch[n]f Wischoten <i>niederl.</i> ˈwɪnskɔtə Winslei <i>vinzo</i> ˈaj Wins[e]ler ˈvɪnz[a]le winseln ˈvɪnzlən Winsen ˈvɪzn Winsford <i>engl.</i> ˈwɪnzfəd Winslow <i>engl.</i> ˈwɪnzləʊ Winsor <i>engl.</i> ˈwɪnzə Winsten <i>engl.</i> ˈwɪnstən Winter ˈvɪntə Winter (Name) ˈvɪntə, <i>engl.</i> ˈwɪntə, <i>tschech.</i> ˈvɪntər Winterberg ˈvɪntərbɛrk Winterburger ˈvɪntərbʊrgə Winterburgerin ˈvɪntərbʊrgərɪn Winterfeldt[d] ˈvɪntəfɛlt Wintergerst ˈvɪntəɡɛst Winterhalter ˈvɪntəhɛltə winterlang ˈvɪntəlɔŋ winterlich ˈvɪntəlɪç Winterling ˈvɪntəlɪŋ wintern ˈvɪntɐn winters ˈvɪntəs Winters <i>engl.</i> ˈwɪntəz Wintersberger ˈvɪntəsbɛrgə Wintersbergerin ˈvɪntəsbɛrgərɪn Winterstein ˈvɪntəʃtaɪn Winterstetten ˈvɪntəʃtɛtən Wintersüber ˈvɪntəʃbɛːrə
großkalibrig ɡʁˈoːskal̩iːbrɪç	Großrussland ɡʁˈoːʁʊslant Großsachsen ɡʁoːʂˈaksən Großschlatten ɡʁoːʃlˈatn Großschönau ɡʁoːʃˈoːnaʊ Großsegel ɡʁˈoːʂzɛːɡl̩ Großsiegelbewahrer ɡʁoːʂz̩ˈiːɡlbavəːrə Großstadt ɡʁoːʂtɑt Großstadtklimasphäre ɡʁoːʂtɑtˈatmoːˌsfɛːrə	Am Wort- und Silben- auslaut werden für Le- nis-Frikative und -Plo- sive stimmlose Fortes gesprochen. Die Aus- lautverhärtung entfällt, wenn vor folgendem Suffix <D>, <N> oder <R> steht. → A. 5.5.2.3	Wirbel <i>wirb</i> wir <i>wir</i> wir Das Personalpronomen <i>wir</i> wird in unbetonter Position spontansprachlich oft artikula- torisch reduziert. Die Redukti- onformen [va], [və] sind in Deutschland allgemein verbrei- tet und treten insbesondere in der Position nach dem finiten Verb auf, mit dem sie eine enklitische Verbindung einge- hen können: <i>Dann sind wir</i> [ˈzɪndvɪr] <i>gegangen</i> . Beson- ders in enklitischer Stellung nach einem Verb mit nasalem Stammsilbenanlaut <i>wie kom- men</i> (bzw. oft auch nach der Reduktionsform [ham] von <i>ha- ben</i>) tritt in Mittel- und Süd- deutschland sowie in Öster- reich zudem [ma], [me] auf. In der Schweiz ist die Reduktion des Pronomens in standard- sprachlicher Rede dagegen weitgehend unüblich.
G Großkampftag ɡʁˈoːskampftːaːk Großkanischa ɡʁoːskˈanɪʃaː Großkopferte ɡʁˈoːskɔpfɛtə	Großstädter ɡʁoːʂtɛtɐ od. ...ʃtɛtə großstädtisch	Wirbel <i>wirb</i> wirbel[ig] <i>wirb</i> [a]llig, -e ...ɪɡə wirbeln <i>wirb</i> lən	

Abb. 1: Infokästen in DAWB (links) und in DU-23 (rechts)

In Wirklichkeit ist aber die Anzahl der Infokästen in DAWB wesentlich geringer, da viele davon, mitunter mit leichten Anpassungen, mehrfach wiederholt werden. So kommen beispielsweise die beiden Sonderregeln: (1) dass [h] im Silbenanlaut vor dem Schwa-Laut und vor den Suffixen *-en*, *-er*, *-ig*, *-ung* nicht gesprochen wird, und (2) dass das Graphem *e* der Endung *-ie* in den Fremdwörtern, in denen diese Endung als [i:] zu realisieren ist, im Singular zwar nur als Dehnungszeichen fungiert, im Plural aber schon als Schwa-[ə] gesprochen werden muss, insgesamt je ca. zwanzigmal vor. Eine Reihe weiterer Informationen, etwa zu Stimmlosigkeitsassimilation von Konsonanten, Konsonantengemination an Silben- und Wortgrenzen, Frikatisierung von [t] vor einem unsilbischen i-Laut, unsilbischer Realisation bestimmter Vokale in manchen Fremdwörtern, auch zu Auslautverhärtung sowie vokalischer Realisation des Suffixes *-er*, die bei einer Silbengrenzverschiebung (z. B. infolge weiterer Suffigierungen) zur konsonantischen Realisation wechseln kann, wird über zehnmal wiederholt.

Die Informationsmodule des DU-23 wirken zwar auf den ersten Blick ähnlich wie jene des DAWB, unterscheiden sich jedoch von ihnen grundlegend in ihrer Struktur und dem vermittelten Inhalt. Sie werden auch jeweils nur einmal eingesetzt; im Gegensatz zu DAWB gibt es also keine Wiederholungen. Rund 86 von insgesamt 177 Infokästen präsentieren prozentuale Ergebnisse einer Online-Umfrage zur Akzeptanz alternativer Aussprachevarianten ausgewählter Wortformen, bei der Angemessenheitsurteile von 573 Probanden aus verschiedenen Regionen Deutschlands, Österreichs und der deutschsprachigen Schweiz berücksichtigt wurden. Dabei beziehen sich die meisten dieser Statistiken auf die Aussprache von konkreten Fremd- und Lehnwörtern, und gut ein Drittel davon thematisiert Schwankungen des Wortakzents (vgl. Abb. 2).

Bei den wenigen einheimischen Wörtern, deren Aussprachepräferenzen statistisch erhoben wurden, lag der Fokus der Untersuchung vor allem auf der Wahl zwischen einem langen gespannten und einem kurzen ungespannten Vokal (wie bei *Nische*, *Walnuss* und einigen weiteren) bzw. auf der rein qualitativen [ɛ:]/[e:]-Schwankung (wie bei *Mädchen*), und lediglich in einem Fall handelte es sich um den variablen Wortakzent (*liebkoosen*).

Anis *nordd./westmd. vorw.:*
a'ni:s, österr./ostmd./
schweiz./südd. vorw.: 'a:nis,
auch: 'anis

Anis			
Umfrage: Wortakzent			
• Norden und Mitte des deutschen Sprachraums			
– a'ni:s			
+81%	4%	–15%	
– 'a:nis			
+29%	12%	–59%	
• Süden des deutschen Sprachraums			
– a'ni:s			
+31%	15%	–54%	
– 'a:nis			
+82%	9%	–9%	

Abb. 2: DU-23: Infokästen zum Lemma *Anis*

Die übrigen Infokästen, die keine statistischen Daten enthalten, liefern jeweils einen Kommentar zur Wortaussprache, der aber im Unterschied zu DAWB nur vereinzelt präskriptiven Charakter hat; in den meisten Fällen werden Aussprachevarianten kurz diskutiert und problematisiert, und sehr oft wird auch hier auf konkrete regionale Differenzen hingewiesen.⁵

Was nun die Aussprachewörterbücher des Englischen angeht, so bieten zwei von ihnen, EPD und LPD, ähnliche gestalterische Lösungen. In ihrem jeweiligen Wörterverzeichnis sind nämlich auch farbige Infokästen, die sog. *information panels*, verstreut.

Da die Korrespondenz zwischen Graphie und Aussprache im Englischen sehr oft inkonsistent ist und die Schreibung eines Wortes häufig keinen eindeutigen Hinweis darauf gibt, wie dieses auszusprechen ist, werden sowohl in LPD als auch in EPD zunächst die sog. *Pronouncing-the-letter-* bzw. *Spelling-to-sound*-Informationsmodule verwendet. Sie sind jeweils der Liste von Einträgen vorangestellt, die mit dem betreffenden Buchstaben des Alphabets beginnen. Der Einsatz gerade solcher Module wird von den Autoren des LPD wie folgt begründet:

English spelling is notorious for its shortcomings. Knowing the orthography of a word does not enable one to predict its pronunciation with any confidence. Nevertheless, certain general principles do govern the relationship between spelling and sound (grapheme and phoneme), even though they may be subject to exceptions and uncertainties. Although many handbooks of English pronunciation ignore them entirely, on the implicit grounds that these rules are so chaotic that it is better to learn the pronunciation of each new word separately, it nevertheless seemed helpful for LPD to offer the user something rather than nothing. (LPD: xviii)

Dementsprechend werden zu jedem Graphem bzw. einer Graphemkombination präzise Ausspracheregeln oder klare Hinweise gegeben, wie diese in unterschiedlicher lautlicher Umgebung auszusprechen sind. Dabei wird in jedem Fall auf sämtliche Unregelmäßigkeiten und etwaige Ausnahmen hingewiesen.

Obwohl die fraglichen Infokästen der beiden Wörterbücher ähnlich aussehen und ähnliche Inhalte vermitteln, sind sie teilweise unterschiedlich aufgebaut. Während in LPD die entsprechenden Grapheme und ihre möglichen Kombinationen samt Ausspracheunregelmäßigkeiten bzw. -ausnahmen in einem Modul behandelt werden, sind diese Inhalte in EPD in separaten Infokästen untergebracht. So findet man in EPD direkt unter A das Modul *Pronouncing the letter A* (s. Abb. 3), auf den folgenden Seiten kommen aber noch vier weitere Module zu verschiedenen Graphemkombinationen, nämlich *Pronouncing the letters AE*, *Pronouncing the letters AEO*, *Pronouncing the letters AI*, *AY* und *Pronouncing the letters AU*, *AW*. Dies hat zwei Folgen: Da in LPD sämtliche Ausspracheregeln zu einem Graphem bzw. dessen Kombinationen in nur einem Infokasten zusammengestellt sind, fallen manche dieser Infokästen recht umfangreich aus (siehe z. B. *O Spelling-to-sound* oder *S Spelling-to-sound*). Im Gegensatz dazu haben die Informationsmodule in EPD — da sie bestimmte Einzelfragen in separaten Modulen behandeln — zwar einen geringeren Umfang, dafür sind sie aber teil-

weise informativer: Zum einen werden in (fast) jedem Modul Angaben zur amerikanischen Aussprache gemacht und zum anderen werden in den Abschnitten *In addition* und *In weak syllables* zusätzliche Aussprachevarianten diskutiert. Die Anzahl von Infokästen dieser Art variiert dementsprechend von 26 in LPD bis insgesamt 69 in EPD.

1
abduct

A

A

Pronouncing the letter A

→ See also **AE, AEO, AI/AY, AU/AW**

The vowel letter **a** has two main strong pronunciations linked to spelling: a 'short' pronunciation /æ/ and a 'long' pronunciation /eɪ/. In the 'short' pronunciation, the **a** is usually followed by a consonant which closes the syllable, or a double consonant before another vowel, e.g.:

tap	tæp
tapping	'tæp.ɪŋ

The 'long' pronunciation usually means the **a** is followed by a single consonant and then a vowel, e.g.:

tape	teɪp
taping	'teɪ.pɪŋ

When there is an **r** in the spelling, the strong pronunciation is one of three possibilities: /ɑː/ US /ɑːr/, /eə/ US /er/ or /æ/ US /eɪ/, e.g.:

car	kɑː US kɑːr
care	keə US ker
carry	kæ.rɪ US ker.i, kær.i

In addition

There are other vowel sounds associated with the letter **a**, e.g.:

ɑː	father /'fɑː.ðər/ US /'fɑː.ðɜː/
ɑː US æ	bath /bɑːθ/ US /bæθ/
ɒ US ɑː	swan /swɒn/ US /swɑːn/
ɔː US ɑː, ɔː	walk /wɔːk/ US /wɑːk/
	warm /wɔːm/ US /wɑːrm/

And, in rare cases:

e many /'men.i/

In weak syllables

The vowel letter **a** is realized with the vowels /ə/ and /ɪ/ in weak syllables, and may also not be pronounced at all in British English, due to compression, e.g.:

above	ə'baʊ
village	'vɪl.ɪdʒ
necessary	'nes.ə.səri US -ser.i

a indefinite article: strong form: eɪ; weak form: ə

Note: Weak-form word. The strong

abnegate /'æb.nɪ.ɡeɪt/, in which it is stressed, and **abduct** /əb'dʌkt/, where it is unstressed.

Abbeville in France: 'æb.vɪl, US æb'vɪl; in US: 'æb.ɪ.vɪl

abbrev. A- 'æb.i -s -z

Abb. 3: EPD: Infokasten zu *Pronouncing the letter A*

Das phonetisch-phonologische Wissen wird in LPD in rund 40 sogenannten *Language Panels* vermittelt (vgl. Abb. 4). Die behandelten Themen reichen von der Definierung phonetisch-phonologischer Grundbegriffe wie z. B. Phonem und Allophon, Diphthong, Affrikate, Geminate, über die Beschreibung phonetischer Einzelfragen wie Aspiration, Assimilation, Elision, Glottisschlag, bis hin zu supra-segmentalen Erscheinungen wie Akzentuierung. Einige wenige Module behandeln Schwerpunkte wie australisches Englisch oder die Aussprache von Sonderzeichen, die in E-Mails oder auf Webseiten verwendet werden, wie @, #, /. Der Übersichtlichkeit halber sind alle *Language Panels* in einem Verzeichnis mit dem Titel *Index to notes on pronunciation and phonetics* mit Seitenangabe aufgelistet, das sich im einleitenden Teil des Wörterbuchs befindet.

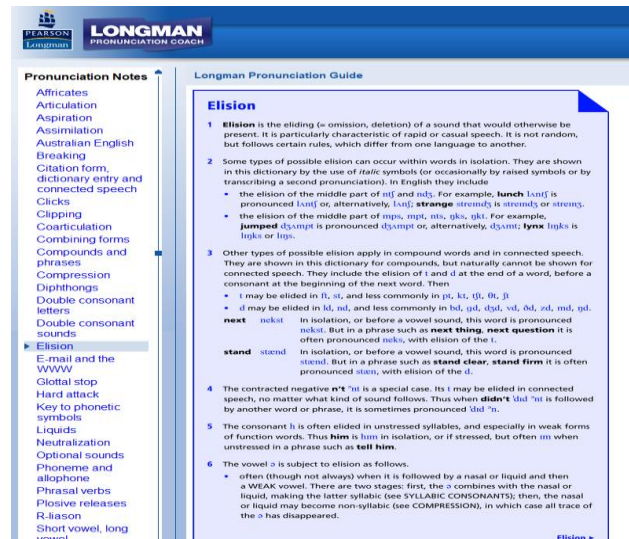


Abb. 4: LPD: *Language Panels* zur phonetisch-phonologischen Terminologie (CD-ROM)

Ähnlich wie in dem deutschen Aussprachewörterbuch DU-23 sind auch in LPD Infokästen mit Statistiken vorhanden (vgl. Abb. 5). Sie bringen prozentuale Ergebnisse von fünf Meinungsumfragen, den sogenannten *opinion polls* (aus den Jahren 1988, 1993, 1998, 1999-2002 und 2007), zu Präferenzen in Bezug auf bestimmte Aussprachevarianten ausgewählter Wörter. Für Einzelheiten zu den Umfragen sowie dazu, welche Wörter in welcher Umfrage enthalten waren, wird auf die Website des Autors verwiesen (vgl. LPD: xviii).

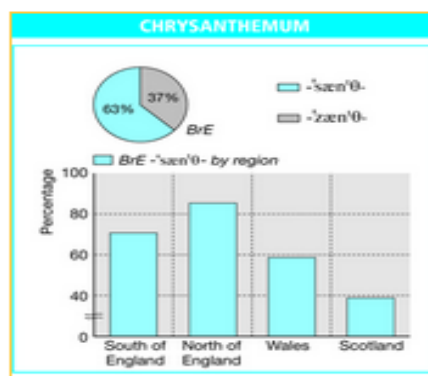


Abb. 5: LPD: *Preference Polls* zur Akzeptanz bestimmter Aussprachevarianten (CD-ROM)

EPD hingegen verzichtet in seiner aktuellen, 18. Auflage auf die Infokästen zu den phonetisch-phonologischen Grundfragen. Stattdessen wird am Ende des Wörterverzeichnis ein Glossar mit der einschlägigen Terminologie angeboten. Darüber hinaus finden sich in EPD über das gesamte Wörterverzeichnis verstreut Verwendungshinweise, sogenannte *usage notes* (vgl. Abb. 6). In den meisten Fällen werden hier kurz Aussprachevarianten diskutiert oder auch komplexere Ausspracheregeln erläutert (vgl. EPD: vii).

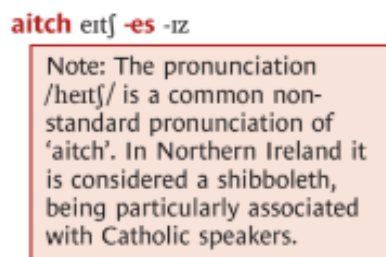


Abb. 6: EPD: Infokasten vom Typ *Usage Note*

Das dritte der behandelten englischen Aussprachewörterbücher: ODP, das, wie bereits in Abschn. 2 erwähnt, als einziges in seinem Wörterverzeichnis für jedes Lemma sowohl britische als auch amerikanische (mitunter gleich lautende) Aussprachevarianten aufführt, geht auch bei der grafischen Gestaltung seinen eigenen Weg. Das Layout des gesamten Wörterbuchs ist so klar und einfach wie möglich gehalten, es gibt darin weder Grafiken noch Informationsmodule, und die laufende Wortliste wird durch keine Sonderelemente unterbrochen.

4.2 Multimediale Komponenten

Mit der Integration der unter 4.1 beschriebenen Informationsmodule in das Wörterverzeichnis der untersuchten Aussprachewörterbücher des Deutschen und des Englischen wurde ein grafisches Mittel eingesetzt, das nicht nur zusätzliche Informationen zur Aussprache vermittelt, sondern auch die visuelle Monotonie einer herkömmlichen Wortliste auflockert. Darüber hinaus bieten aber die Wörterbücher noch einige weitere innovative medientechnische Funktionen, darunter Audiomaterial, was bei diesem Wörterbuchttyp von besonderem Interesse und hoher Relevanz für die BenutzerInnen ist.

So ist DAWB mit einer Audio-DVD ausgestattet, auf der die standardisierte (bundesdeutsche) Aussprache ausgewählter Wortbeispiele aus dem Einführungsteil präsentiert wird. Sie enthält insgesamt 98 Audiodateien im wav-Format. Im Wörterbuch selbst sind die Abschnitte, für die Audiodateien zur Verfügung stehen, deutlich gekennzeichnet und durchnummeriert. Die Entscheidung der AutorInnen, insgesamt ca. 2.000 Worteinheiten und Wortgruppenlexeme aus dem Textteil für die Vertonung auszuwählen, ist wohldurchdacht, denn

gerade diese Beispiele illustrieren in den einführenden Kapiteln verschiedene Ausspracheregeln und besitzen somit einen hohen Verallgemeinerungs- bzw. Analogiewert.

Darüber hinaus ist die gesamte Druckausgabe des DAWB auch als E-Book im PDF-Format verfügbar, sodass es ebenso gut auf einem PC oder anderen elektronischen Geräten genutzt und leicht durchsucht werden kann. In einem engen Zusammenhang mit dem Wörterbuch steht die online zugängliche *Deutsche Aussprachedatenbank* (DAD), die auf dem Wörterverzeichnis des DAWB basiert und dieses durch Ergänzungen, Erweiterungen und Korrekturen weiterführt. Nach Angaben der Projektbetreibenden soll zudem eine (synthetische) Vertonung der in der DAD enthaltenen Stichwörter geplant sein.⁶

Demgegenüber ist das DUDEN-Aussprachewörterbuch nicht nur als Hardcover-Buch und PDF-Datei, sondern auch — wie schon die Vorgängerausgabe von 2015 — als eine richtige Software (Desktop-App) erhältlich, die von der Homepage des Verlages im Rahmen der sog. *Duden-Bibliothek* gegen entsprechendes Entgelt heruntergeladen werden kann.⁷

Die aktuelle, 8. Auflage wurde nach Angaben des Verlags um über 4.000 Stichwörter erweitert und enthält somit 144.000 Aussprachen zu Wörtern und Eigennamen, einschließlich der im Deutschen gebräuchlichen Fremdwörter und fremdsprachigen Namen. Alle Stichwörter wurden vertont und die Audios stehen als Download zur Verfügung, wobei die Vertonungen nur beim Kauf der Printausgabe oder der Software verfügbar sind. Da die neue Auflage erst seit Oktober 2023 auf dem Markt ist, muss noch mit gewissen technischen Unzulänglichkeiten gerechnet werden; darauf wird auch auf der Webseite des Verlages direkt hingewiesen. Einige Audiodateien sind nämlich noch nicht verfügbar, sie sind in der Datenbank mit einem grauen Lautsprechersymbol gekennzeichnet.⁸ Auf jeden Fall ist aber die Bereitstellung von Audios zu allen Einträgen des Wörterbuchs als ein großer Fortschritt im Vergleich zu der Vorgängerauflage zu werten, da jene insgesamt nur 13.000 Vertonungen enthält, was knapp 10 % aller Stichwörter ausmacht.

Von den drei englischen Aussprachewörterbüchern bieten zwei, nämlich EPD und LPD, neben der gedruckten Buchausgabe zusätzlich Audiodateien auf einer CD-/DVD-ROM an.⁹ In beiden Fällen enthält die CD-/DVD-ROM die Vertonung jedes Wörterbucheintrags jeweils in zwei nationalen Aussprachevarianten, das heißt in britischem und amerikanischem Englisch. Beide CD-/DVD-ROMs bieten außerdem interaktive Suchfunktionen, mit denen die Einträge nach verschiedenen Kriterien durchsucht werden können. So kann man z. B. die Wortliste nach bestimmten Lauten oder Lautsequenzen filtern, um gezielt ähnliche Wörter oder Minimalpaare ausfindig zu machen. In LPD erscheinen in der Registerkarte *Longman Pronunciation Guide* wiederholt sämtliche *Spelling-to-sound*-Informationsmodule sowie die Module zur phonetisch-phonologischen Terminologie. Auch die *Preference Polls* werden erneut angeboten, diesmal in digitaler Form.

Was die beiden Werke von den Aussprachewörterbüchern des Deutschen stärker unterscheidet, sind interaktive Übungsmodule verschiedener Art (s. Abb. 7 u. 8), die es u. a. ermöglichen, die eigene Aussprache aufzuzeichnen und sie mit

der zugänglichen modellhaften britischen oder amerikanischen (Standard-)Aus-
sprache zu vergleichen. Dies ist eine technische Lösung, die es in den deutschen
Wörterbüchern nicht gibt. Und was die englischen Aussprachewörterbücher für
ihre BenutzerInnen noch interessanter macht, ist die Möglichkeit, in Hunderten
von interaktiven Übungen die phonetische Lautschrift, die Perzeption und die
Identifikation von Lauten und Lautkombinationen sowie die Laut-Graphem-
Korrespondenz zu trainieren. Bei LPD ist es darüber hinaus möglich, Handouts
mit Übungen aus dem *Teachers' Resource Centre* auszudrucken.

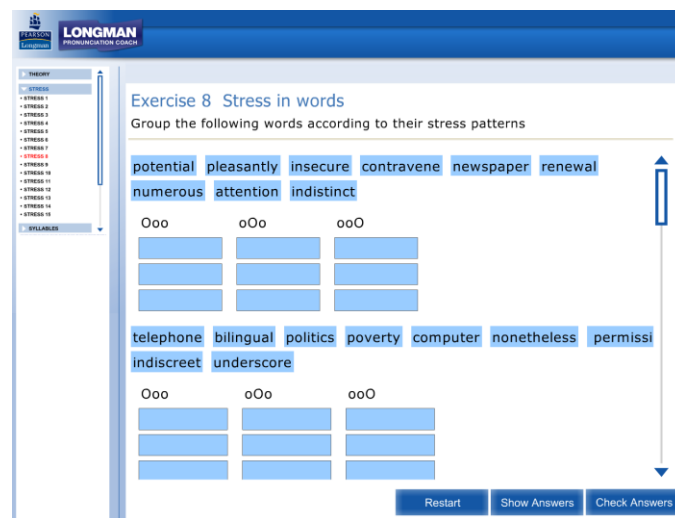


Abb. 7: LPD: Interaktive Übungen (1)

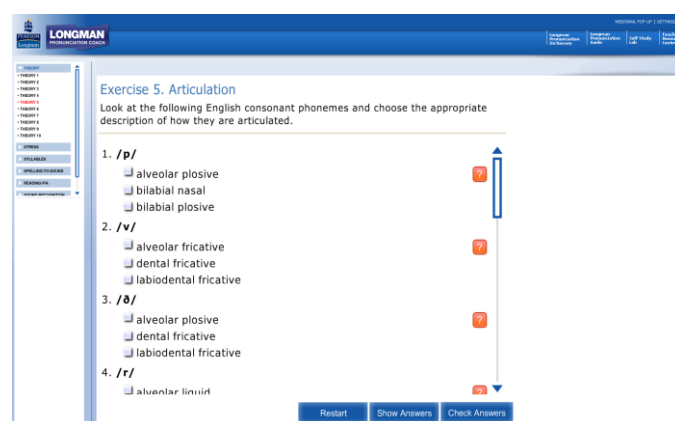


Abb. 8: LPD: Interaktive Übungen (2)

Im Jahr 2015 wurde auch eine digitale Version der 18. Ausgabe des EPD veröffentlicht. Es handelt sich dabei um eine App, die für das iPhone und das iPad von Apple verfügbar ist und über den Apple Store vertrieben wird. Eine Android-Version erschien dann zwei Jahre später.

Das einzige der drei untersuchten Aussprachewörterbücher des Englischen, das kein Begleitmedium in Form einer CD-/DVD-ROM mitliefert, ist das ODP. Andererseits ist es aber auch das einzige, das in seiner zweiten Ausgabe von 2017 (*The Routledge Dictionary of Pronunciation for Current English*, RDP) zugleich als E-Book erhältlich ist.

5. Zusammenfassung und Ausblick

Die untersuchten Aussprachewörterbücher des Deutschen und des Englischen sind ausgereifte Nachschlagewerke, die nicht nur durch ihren Umfang, sondern mehrheitlich auch durch ihre grafische Gestaltung und ihre multimedialen Komponenten überzeugen.

Die festgestellten erheblichen Unterschiede im Umfang des einführenden Textteils (*front matter*) zwischen den Aussprachewörterbüchern des Deutschen und denen des Englischen reflektieren zum Teil unterschiedliche redaktionelle Ansätze. Während die Einführungskapitel der deutschen Wörterbücher — mit insgesamt 150 bis 280 Seiten — als quasi-lehrbuchartige Werke eine vertiefte Auseinandersetzung mit den phonetisch-phonologischen Fragen ermöglichen, liegt bei den Aussprachewörterbüchern des Englischen der Fokus stärker auf der direkten Anwendung und dem schnellen Nachschlagen, während die theoretische Fundierung eine geringere Priorität hat. Auch scheint in den englischen Wörterbüchern weniger Wert darauf gelegt zu werden, den BenutzerInnen genauere erstellungstechnische Hintergrundinformationen, etwa über die Zusammenstellung des Wörterverzeichnisses oder empirische Quellen für die zu transkribierenden Aussprachevarianten, in dem jeweiligen Einführungsteil zu bieten.

Gemeinsam ist dafür allen untersuchten Werken, dass sie — abgesehen von der einzelsprachlich bedingten Spezifik — den (ortho)phonetischen Standard im Vergleich zu den jeweils früheren Ansätzen weniger elitär, also deutlich breiter und damit auch realitätsnäher definieren. Sollte hier dennoch unter diesem Gesichtspunkt hierarchisiert werden, so lässt sich wohl am ehesten das DAWB hervorheben, und zwar als etwas konservativer als die übrigen Wörterbücher, da es die Standardaussprache ausdrücklich in erster Linie mit offiziellen Kommunikationssituationen in Verbindung setzt und den präskriptiven bzw. regulativen Charakter seiner Transkriptionen stärker betont (Abschn. 3).

In unterschiedlichem Ausmaß bzw. mit unterschiedlicher Tiefgründigkeit wird in den fraglichen deutschen und englischen Wörterbüchern der Differenziertheit der jeweiligen Aussprache, darunter vor allem ihrer regionalen Variation Beachtung geschenkt. Dass die Aussprachewörterbücher des Englischen auf die Darstellung von regional oder gar großregional verwendeten Aussprache-

varianten weitgehend verzichten, lässt sich jedoch, wie im Obigen (Abschn. 3) erörtert, durch die außerordentlich große Ausbreitung und Vielfältigkeit der englischen Sprache durchaus plausibel erklären.

Bezüglich der grafischen Gestaltung ist vor allem festzuhalten, dass gleich vier der fünf untersuchten Wörterbücher in ihrem jeweiligen Wörterverzeichnis die sog. Infokästen (*information panels*) mit diversen Inhalten einsetzen, die visuell attraktiv wirken und allesamt nützliche Ergänzungsinformationen bringen. Als Einziges fällt hierbei das ODP (auch in der jüngeren Routledge-Ausgabe, RDP, von 2017) aus dem Rahmen, da es sowohl in seinem Einführungsteil als auch in der Wortliste auf jegliche grafische Mittel verzichtet und überhaupt ein sehr traditionelles Erscheinungsbild bietet, das aus der heutigen Sicht schon recht spartanisch wirkt.

Ein entscheidender Aspekt bei der Erstellung und Analyse von Aussprachewörterbüchern ist heutzutage die fortschreitende technologische Entwicklung, die die Lexikographie zunehmend in digitale Formate überführt. Mit Ausnahme des ODP/RDP, das auch in dieser Hinsicht den übrigen Wörterbüchern nachsteht, enthalten jetzt schon die untersuchten Werke multimediale Erweiterungen, sei es als eine richtige Desktop- oder Smartphone-Anwendung mit dem gesamten Inhalt des Papierwörterbuchs und Audios zu allen Stichwörtern (DU-23, EPD), sei es in Form einer beigefügten CD-/DVD-ROM, die entweder Audiodateien zu ausgewählten Stichwörtern enthält (DAWB), oder aber die Vertonung aller Stichwörter und darüber hinaus erweiterte Suchfunktionen sowie interaktive Ausspracheübungen bietet (EPD, LPD). Hinzu kommt die Fortführung des DAWB durch eine Online-Aussprachedatenbank (vgl. Abschn. 4.2).

Angesichts der rasanten Fortschritte in der Sprach- und Spracherkennungstechnologie sowie der zunehmenden Verfügbarkeit interaktiver Online-Wörterbücher ist es jedoch offensichtlich, dass die derzeit angebotenen Printausgaben mit multimedialen Komponenten (wobei Datenträger wie CD- oder DVD-ROM bereits als überholt gelten müssen) lediglich eine Übergangsphase in der Entwicklung der Phonolexikographie darstellen.¹⁰ Digitale Plattformen bieten nämlich nicht nur den Vorteil der ständigen Aktualisierung, sondern auch die Möglichkeit einer noch anschaulicheren und flexibleren Darstellung phonetisch-phonologischer Informationen, etwa durch interaktive IPA-Transkriptionen oder KI-gestützte Aussprachehilfen.

Vor diesem Hintergrund ist davon auszugehen, dass die Weiterentwicklung von KI-Technologien tiefgreifende Auswirkungen auf die Zukunft von Aussprachewörterbüchern haben wird. KI-gestützte Systeme können eine nahezu unbegrenzte Menge an Sprachdaten auswerten, fortlaufend aktualisieren und auch individualisierte Rückmeldungen zur Aussprache geben, indem sie Benutzeraufnahmen mit Standardvarianten vergleichen. Daraus erwächst die Möglichkeit, dass Aussprachewörterbücher, heute noch primär statische Referenzwerke, sich mit der Zeit zu speziellen interaktiven Informations-, Lern- und Trainingsplattformen weiterentwickeln (mehr zur aktuellen Beschäftigung mit den Auswirkungen der KI auf die Lexikographie siehe z. B. De Schryver 2023: 355 ff., McKean und Fitzgerald 2024: 7 ff., Ptasznik und Lew 2025: 140 ff.).

Denkbar ist aber auch ein anderes Szenario. Zieht man nämlich in Betracht, dass online zugängliche renommierte Universalwörterbücher wie die deutschen *DWDS* und *Duden online*, die britischen *Oxford English Dictionary (OED)* und *Cambridge English Dictionary* sowie das amerikanische *Merriam Webster* Aussprache-Audios (und vor allem die letzteren drei auch Transkriptionen) zunehmend in ihre Lemmata integrieren bzw. es bereits weitgehend getan haben, dass dies auch auf diverse Lernerwörterbücher oder zwei- und mehrsprachige Wörterbücher zutrifft und es zudem seit längerem auf die Aussprachevermittlung spezialisierte Internetplattformen wie die *Forvo (forvo.com)* gibt, so muss wohl damit gerechnet werden, dass der Benutzerkreis von orthoepischen Fachwörterbüchern mittel- und längerfristig deutlich schrumpfen wird.

Bereits jetzt fallen die deutschen und die englischen Aussprachewörterbücher im Vergleich zu modernen Universal- oder Lernerwörterbüchern etwa in Bezug auf die Zugriffsstruktur und die angebotenen Verlinkungen eher bescheiden aus. Ihre Zugriffsstruktur beschränkt sich im Wesentlichen auf die alphabetische Lemmaliste in den Printausgaben bzw. auf einfache Suchfunktionen in den digitalen Ausgaben. Erweiterte Optionen wie phonetische Suchmöglichkeiten sind nur in Einzelfällen und in begrenztem Umfang vorhanden. Auch die Verlinkungsangebote bleiben in den untersuchten Werken auf punktuelle Querverweise zwischen unterschiedlichen Aussprachevarianten oder Flexionsformen beschränkt. Eine systematische Vernetzung mit externen Ressourcen wie Thesaurus, Kollokationen oder (ggf. phonetisch relevanten) Grammatikhinweisen, wie sie in heutigen Universal- oder Lernerwörterbüchern üblich ist,¹¹ wird von den aktuellen Aussprachewörterbüchern nicht angeboten.

Sollten nun Universalwörterbücher wie die oben aufgezählten in Zukunft noch mit weiteren benutzerfreundlichen Funktionalitäten angereichert werden, etwa der Möglichkeit, die eigene Aussprache mit dem Standard-Audio des Wörterbuchs zu konfrontieren und online zu üben, dann lässt sich u. U. nicht vermeiden, dass eigenständige Aussprachewörterbücher schließlich auf ein Nischenphänomen reduziert werden, das hauptsächlich nur noch für Fachleute von Interesse bleibt.¹²

Endnoten

1. Zu einer kritischen Auseinandersetzung mit dem Wörterbuch siehe Lewis 2020: 147 ff.
2. In diesem Zusammenhang zählt DAWB auch einige konkrete Berufsgruppen auf, und zwar „Pädagogen, Mediensprecher, Schauspieler, Sprecherzieher, Rhetoriklehrer, Logopäden, klinische Linguisten“ (4. Umschlagseite).
3. Ausführlicher dazu siehe Nycz und Tęcza 2023: 168 ff., Nycz und Tęcza 2020b: 465 ff., Tęcza und Nycz 2016a: 384 ff.
4. Zur Kritik des *Oxford Dictionary of Pronunciation for Current English* siehe Picard 2004: 192.
5. Mehr zu formal-editorischer Beschaffenheit der deutschen Aussprachewörterbücher siehe Nycz und Tęcza 2020b: 76 ff., Tęcza und Nycz 2016b: 364 ff.

6. Vgl. <https://dad.sprechwiss.uni-halle.de/dokuwiki/doku.php/about/start>, Zugriff am 30.8.2025. Eingehender zu der DAB, deren Aufbau, geplanter Weiterentwicklung sowie besonders phonodidaktischem Potenzial siehe Ebel und Skoczek 2021: 39 ff., Skoczek und Ebel 2021: 73 ff.
7. Siehe <https://shop.duden.de/Duden-Das-Aussprachewoerterbuch/9783411040681>, Zugriff am 30.8.2025. Eine genauere Darstellung der Konstruktion und der Funktionsweise dieser Anwendung ist in Nycz und Tęcza 2020a: 85 ff. zu finden.
8. Vgl. <https://shop.duden.de/Duden-Das-Aussprachewoerterbuch/9783411914326>, Zugriff am 30.8.2025.
9. Zur Kritik des *Longman Pronunciation Dictionary* siehe Sobkowiak 2009: 191 ff., Lewis 2009: 238 ff.
10. Eine interessante und dabei noch in hohem Maße aktuelle Diskussion dieser Problematik wurde schon vor einem Jahrzehnt in Sangster (2015: 307 ff.) vorgenommen.
11. Vgl. z. B. das LDOCE, *Longman Dictionary of Contemporary English Online* (<https://www.ldoceonline.com>).
12. Wir möchten an dieser Stelle den beiden anonymen GutachterInnen für ihre wertvollen Kommentare und Anregungen danken, die die Qualität unseres Artikels erheblich verbessert haben.

Literaturverzeichnis

I. Deutsche Aussprachewörterbücher (mit Siglen)

- [DAWB] Krech, E.-M., E. Stock, U. Hirschfeld und L. Ch. Anders (mit Beiträgen von W. Haas, I. Hove und P. Wiesinger). 2009. *Deutsches Aussprachewörterbuch*. Berlin/New York: De Gruyter.
- [DU-23] Kleiner, St. und R. Knöbl (bearb. in Zusammenarbeit mit der Dudenredaktion). 2023. *DUDEN. Das Aussprachewörterbuch* (= DUDEN-Band 6; achte, komplett überarbeitete und erweiterte Auflage). Berlin: Dudenverlag.

II. Englische Aussprachewörterbücher (mit Siglen)

- [EPD] Jones, D. (Hrsg. von P. Roach, J. Setter and J. Esling). 2022. *Cambridge English Pronouncing Dictionary* (18. Aufl.). Cambridge: Cambridge University Press.
- [LPD] Wells, J.C. 2008. *Longman Pronunciation Dictionary* (3. Aufl.). Harlow: Pearson Education.
- [ODP] Upton, C., W.A. Kretschmar Jr and R. Konopka. 2001/2003. *Oxford Dictionary of Pronunciation for Current English*. Oxford: Oxford University Press.
- [RDP] Upton, C. and W.A. Kretschmar Jr. 2017. *The Routledge Dictionary of Pronunciation for Current English* (2. Aufl.). Abingdon: Routledge.

III. Sekundärliteratur

- De Schryver, G.-M. 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.
DOI: <https://doi.org/10.1093/ijl/ecad021>
- Ebel, A. und R. Skoczek. 2021. Die Deutsche Aussprachedatenbank (DAD) — Potenzial für Deutsch als Fremd-, Zweit- und Muttersprache. *Korpora Deutsch als Fremdsprache* 1(2): 31-53.

- Lewis, J.W.** 2009. Wells, J.C. *Longman Pronunciation Dictionary* (3. Aufl.). *Journal of the International Phonetic Association* 39(2): 238-240.
DOI: <https://doi.org/10.1017/S0025100309003934>
- Lewis, J.W.** 2020. Clive Upton and William A. Kretzschmar Jr. (Editors), *The Routledge Dictionary of Pronunciation for Current English*, 2nd edn. 2017. Pp. xxxvii + 1561. ISBN 978-1-138-12566-7 (hardback); ISBN 9-78131545969 (ebook). *Journal of the International Phonetic Association* 50(1): 147-149.
DOI: <https://doi.org/10.1017/S002510031700038X>
- McKean, E. und W. Fitzgerald.** 2024. The ROI of AI in Lexicography. *Lexicography* 11(1): 7-27.
DOI: <https://doi.org/10.1558/lexi.27569>
- Nycz, K. und Z. Tęcza.** 2020a. On the Pronunciation Dictionaries of Contemporary German: Lexicographic Construction and Multimedia Components. *International Journal of Lexicography* 33(1): 73-89.
DOI: <https://doi.org/10.1093/ijl/ecz026>
- Nycz, K. und Z. Tęcza.** 2020b. On the Pronunciation Dictionaries of Contemporary German: The Concepts of Phonetic Standard and Differences in Specific Phonetic Issues. *International Journal of Lexicography* 33(4): 463-487.
DOI: <https://doi.org/10.1093/ijl/ecz032>
- Nycz, K. und Z. Tęcza.** 2023. Phonostilistische und regionale Variation der deutschen Aussprache in den orthoepischen Nachschlagewerken der dritten Generation. *Lexikos* 33: 166-183.
DOI: <https://doi.org/10.5788/33-1-1804>
- Picard, M.** 2004. Review of *The Oxford Dictionary of Pronunciation for Current English*. *Language* 80(1): 192-192.
DOI: <https://dx.doi.org/10.1353/lan.2004.0040>
- Ptasznik, B. und R. Lew.** 2025. Dictionaries versus AI Tools through the Eyes of English Majors. *International Journal of Lexicography* 38(2): 140-158.
DOI: <https://doi.org/10.1093/ijl/ecaf005>
- Sangster, C.** 2015. Pronouncing Dictionaries. Durkin, Ph. (Hrsg.). 2015. *The Oxford Handbook of Lexicography*: 292-309. Oxford: OUP.
- Skoczek, R. und A. Ebel.** 2021. German Pronunciation Database and Its Possible Applications in the Age of Homeschooling. *Lublin Studies in Modern Languages and Literature* 45(3): 117-127.
- Sobkowiak, W.** 2009. John C. Wells. *Longman Pronunciation Dictionary*. *International Journal of Lexicography* 22(2): 191-209.
DOI: <https://doi.org/10.1093/ijl/ecp003>
- Tęcza, Z. und K. Nycz.** 2016a. Deutsche Aussprachewörterbücher im Vergleich: Die Auffassung des phonetischen Standards und Divergenzen in phonetischen Einzelfragen. *Germanica Wratislaviensia* 141: 383-398.
DOI: <https://doi.org/10.19195/0435-5865.141.25>
- Tęcza, Z. und K. Nycz.** 2016b. Deutsche Aussprachewörterbücher im Vergleich: Die lexikographische Architektur und Prinzipien der Aufnahme lexikalischen Materials. *Germanica Wratislaviensia* 141: 361-382.
DOI: <https://doi.org/10.19195/0435-5865.141.24>

Where Tools Are Few: Constructing Limited-Word Bilingual Learner's Dictionaries in a Low-Resourced Community in Malawi

Ian David Dicks, *Ciyawo–English Dictionary Project (CEDP), Malawi;*
Tabor College, Adelaide, Australia
(iandicks@icloud.com)
(<https://orcid.org/0000-0003-0110-9596>)

Abstract: Many minority language communities in sub-Saharan Africa are required to navigate national education systems dominated by languages in which they are not largely proficient, and which they lack resources and tools to learn. When taken together with a low political will for resourcing minority languages, these become formidable barriers for learning and development (Matiki 2006: 244, 246; Prinsloo 2017: 20). Many minority languages are also not understood or spoken outside of these communities, and there are few resources to assist people to gain capacity in them, which results in large communication gaps. This paper examines how a team of non-professional Yawo lexicographers in Malawi worked to overcome these barriers in order to produce two limited-word bilingual learner's dictionaries (English–Ciyawo; Ciyawo–English) in book and smartphone application forms. More specifically, it explains how the team was trained, the development of a limited organic corpus from oral interviews as a method for supplementing an unbalanced corpus, and the importance of collaboration. The paper also articulates the linguistic and pedagogical theories that are foundational to the project, including that second-language learners develop second-language competence more efficiently through engaging with a curated headword list of high-frequency and high-relevance words (Nation 2022: 15-17; Bayetto 2018: 12).

Keywords: CIYAWO, POVERTY, BILINGUAL, LEARNER, DICTIONARY, EDUCATION, MALAWI, AFRICA, OXFORD, LEXICOGRAPHY, SMARTPHONE, APPLICATION, CORPUS, LANGUAGE

Opsomming: Waar hulpbronne skaars is: Die maak van beperkte tweetalige aanleerderwoordeboeke in 'n Malawiese gemeenskap met geringe hulpbronne.

Baie minderheidstaalgemeenskappe in sub-Sahara Afrika moet in nasionale onderwysstelsels werk wat oorheers word deur tale waarmee hulle onvertroud is en waarin hulle nie hulpbronne en gereedskap het om te leer nie. Wanneer dit saamval met 'n swak politieke wil om minderheidstale te befonds, word dit gedugte hindernisse vir leer en ontwikkeling (Matiki 2006: 244, 246; Prinsloo 2017: 20). Baie minderheidstale word ook nie buite hierdie gemeenskappe verstaan of gepraat nie, en daar is min hulpbronne om mense te help om taalvaardiger te word, wat tot groot kommunikasiegapings

lei. Hierdie artikel ondersoek hoe 'n span amateur-Yawo-leksikograwe in Malawi gewerk het om hierdie hindernisse te oorkom om twee tweetalige aanleerderwoordeboeke (Engels–Ciyawo; Ciyawo–Engels) met beperkte woorde in boek- en slimfoonformaat te produseer. Meer spesifiek word daar verduidelik hoe die span opgelei is, asook hoe 'n beperkte organiese korpus uit mondelinge onderhoude ontwikkel is as 'n metode om 'n ongebalanseerde korpus aan te vul, en die belangrikheid van samewerking. Die artikel neem ook die taalkundige en pedagogiese teorieë fundamenteel vir die projek onder die loep, insluitend dat tweedetaalleerders tweedetaalbevoegdheid meer doeltreffend ontwikkel deur gebruik te maak van 'n saamgestelde trefwoordlys van hoëfrekwensie- en hoogs relevante woorde (Nation 2022: 15-17; Bayetto 2018: 12).

Slutelwoorde: CIYAWO, ARMOEDE, TWEETALIG, LEERDER, WOORDEBOEK, ONDERWYS, MALAWI, AFRIKA, OXFORD, LEKSIKOGRAFIE, SLIMFOON, TOEPASSING, KORPUS, TAAL

Introduction

The Ciyawo–English Dictionary Project was first conceived in the late 1990's by Australian Baptist intercultural workers working in Malawi who saw the need to communicate with the Yawo in Ciyawo, but struggled to locate adequate tools to help them do so. Although there are strong social, cultural and missiological reasons for wanting to learn Ciyawo, the decision to learn Ciyawo instead of Chichewa, the most dominant language in Malawi, was not an easy decision as there was a dearth of available language learning resources in Ciyawo at that time. The only Ciyawo grammar books and dictionaries in existence were between fifty and one hundred years old and out of print, with only a few collector's versions available in specialty book shops at exorbitant prices. The problem with these tools, however, was more than a lack of availability and cost. They also lacked elements and qualities that would enable second-language (L2) learners to grow their capacity to understand, speak, read and write Ciyawo efficiently.

Initially, to overcome the communication gap, Ciyawo word lists with English equivalents were created and utilised by intercultural workers. These proved to be inadequate as they did not provide definitions for multiple senses or example sentences to show how headwords are naturally used, as well as other useful information for L2 learners. Another problem with the initial Ciyawo word lists were that they were largely translated from English headword lists. These lists represented ideas that native English speakers wanted to talk about rather than the ideas the Yawo had about the world and their life situations.

In 2005, Ian Dicks, an Australian intercultural worker working with Baptist Mission Australia (formerly Global Interaction) and Steffi Nchembe, a Yawo language and culture assistant, began to conceive of developing a Ciyawo learner's dictionary. This dictionary would consist of Ciyawo headwords and other elements that L2 learners of Ciyawo need to grow their capacity to communi-

cate with the Yawo. The notion for a limited-word learner's dictionary was based on research regarding the needs of L2 learners, which shows that L2 learners grow their second language competence most efficiently by learning the high frequency and high relevance words first (Coxhead 2000: 213; Nation 2001: 17-18).

The size of language vocabularies varies greatly. For example, the Oxford English Dictionary (OED) claims to explain the meaning of 500 000 English headwords and phrases (<https://www.oed.com/information/about-the-oed>). There is, however, a vast difference between the number of words in a language and the number of words that the average native speaker knows and uses. Studies show that "native speakers of English typically learn between 15–20 000 word families"¹ (Webb and Nation 2017: 7, 14). It is also recognised that an L2 learner can communicate adequately with a much smaller curated vocabulary. Studies show that high-frequency words are of the greatest value to L2 learners (Webb and Nation 2017: 6). By learning the 2 000 highest-frequency word families in English, as designated by the British National Corpus (BNC), a person has the ability to understand 86–91 percent of discourse and text in a novel, 81–84 percent in a newspaper, 89 percent of a conversation, 90 percent in a film and 89 percent of a lecture (Webb and Nation 2017: 11-12).²

It has also been shown that it is less beneficial for L2 learners to learn additional high-frequency words from the third, fourth, fifth, etc. sets of 1 000 word families than it is to learn additional low-frequency words of high relevance (Nation 2001: 17-18). The Oxford 3 000 and Oxford 5 000 word lists have been developed on this theory. The headwords in the Oxford 3 000 list are selected using two criteria — frequency and relevance for learners (<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#>). Oxford University Press (OUP) identified the highest frequency words, drawing on the BNC and the OED corpus (Shaffer 2005: 136). To this high-frequency list they added words of 'high relevance' to English language learners by selecting high-frequency words from "a specially created corpus of Secondary and Adult courses published by Oxford University Press" (<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#>).

Dicks and Nchembe saw the value of learning high-frequency and high-relevance words for L2 learners of Ciyawo. Initially, this interest was in the concept of developing a high-frequency and high-relevance word list in Ciyawo for L2 learners of Ciyawo. However, through research it became obvious that there were not adequate resources available to do this. There was neither a Ciyawo general corpus or a high-relevance corpus from which to develop such a list, nor were there human resources to utilise the data if such lists emerged.

During exploration of the value of high-frequency and high-relevance words for L2 Ciyawo learners, Dicks and Nchembe became acutely aware of the Yawo community's language learning context. They noted the lack of resources and tools to help Yawo students learn English as a second language, which at the time was the language medium of instruction for upper-primary, secondary and tertiary education in Malawi.

The Yawo education experience

The Yawo of Malawi are disadvantaged within the Malawian education system, as they are constantly playing catch-up throughout their educational experience. This is due to Chichewa being used as the primary medium of communication in primary school. The Yawo are also disadvantaged when it comes to engaging with government social services in the areas of health, law and order, civil society and development initiatives, as well as religious activities. The majority of professionals who come to work in Yawo-dominant areas are not proficient in Ciyawo, nor are they able to grow their capacity in the language even if they wanted to, as there are few resources available to help them.

This situation is not unique to the Yawo. It is also true of many minority language communities in Africa and around the world. Trudgen (2000: 68) says that in many minority language communities there is "a crisis in being understood". What Trudgen (2000) says of the Yolŋu people, an Aboriginal community in Northern Australia, reflects in many ways the reality of the Yawo and other minority language communities throughout southern Africa. The Yawo sit in meetings, listening to Chichewa and/or English, but often do not fully understand these languages. Alternatively, they may understand but do not feel competent to engage in the dominant language being spoken and so they remain quiet. This results in a tangible isolation in their own country. Trudgen (2000: 69) similarly points out:

Poor communication stops people receiving almost all news or knowledge from outside their language and cultural domain. This includes day-to-day news and general information. It also includes what may well be life-saving information from health professionals. It stops them knowing what they are giving consent for, how to comply with medical instructions and how to intervene in their own health problems. In this way, poor communication directly impacts on high mortality rates.

In Malawi it is taken for granted that Chichewa is universally understood; however, this is not the reality. For example, it is common for the Yawo to return from a visit to the hospital not fully knowing their diagnosis or the treatment regimen. This is not because it was not explained, but because there is often a communication gap. The professional did not understand or speak even a modicum of Ciyawo and the patient was not conversant in English, or fully conversant or confident enough in Chichewa to ask questions.

Trudgen (2000: 77) states that the communication crisis operates two ways: firstly, professionals are unable to communicate with people in minority language communities about matters that impact their lives and well-being, even when discussing basic concepts using basic terms, and secondly, people in minority language communities struggle to share their needs and ideas with professionals in a dominant language.

The Yawo context in Malawi

There are 17 languages spoken in Malawi (Matiki 2001: 203). Four languages are considered official languages, namely Chichewa, Citumbuka, Chitonga and Ciyawo, together with English (<https://www.immigration.gov.mw/citizenship/list-of-languages-in-malawi/>). Chichewa and English were declared the official languages of Malawi in 1968, with 50,2 percent of the population claiming to be first-language (L1) speakers of Chichewa and 4,9 percent claiming to understand English at that time (Matiki 2001: 201). According to Matiki (2001: 215), English became the "instrumental, regulative, interpersonal and imaginative/innovative language" as it was the only language accepted in education, judiciary, parliament and the media. However, due to English not being widely understood or spoken, "Chichewa acts as the de facto national language and English is the de facto official language in the country" (Reilly 2019: 32).

Although it has been widely reported that sub-Saharan languages are not in danger of being supplanted by languages of the colonisers, they are at risk of being supplanted by other dominant languages in their countries. Brenzinger (in Kandybowicz and Torrence 2017: 2) says, "the most immediate threat to minority African languages are posed by other local languages or sub-national languages", which is the case with Chichewa in relation to the other languages in Malawi.

Ciyawo is a widely spoken language in South Eastern Africa. It is the first language of more than 3,2 million people in Malawi, Mozambique and Tanzania (Baldauf and Kaplan 2004: 85, 155). In Malawi, more than 2,3 million people identify as Yawo and they make up 13,3% of the population (Malawi Data Portal 2018).

Malawi, however, is one of the least developed countries in the world with half of the population living below the poverty line when measured against the Multidimensional Poverty Index, which assesses people's lives across four equal dimensions of Health and Population, Education, Environment, and Work.³ Significantly, under the Education dimension, low levels of literacy and issues with schooling contribute to the nationwide level of multidimensional poverty in the country. However, other indicators reveal the depth and breadth of poverty for the Yawo, as the incidence of multidimensional poverty is 65,7 percent for people in rural areas, which is where the Yawo predominantly live, compared to 20 percent in urban areas (National Statistical Office Malawi 2022: vi). Moreover, children and youth are among those most impacted by poverty in Malawi — 63,8 percent of children between 0 to 9 years and 61,8 percent of children aged 10 to 19 years record the highest incidence of poverty in the country (National Statistical Office Malawi 2022: vi). The significance of these statistics is amplified at a district level as Mangochi (78,4 percent) and Machinga (78,2 percent) are the districts in Malawi that have recorded the highest percentage of people experiencing multidimensional poverty. These are areas in which the Yawo are the majority ethnolinguistic community (National Statistical Office Malawi 2022: vi).

This data indicates that Malawi is suffering an education crisis. This is supported by other data that shows that the percentage of children unable to read at a level appropriate for their age is estimated to be 89 percent in Malawi (UNESCO 2024). Successive governments have attempted to address the education issue in various ways. For example, tuition fees were abolished in 1994 for all government primary school education. In 2018, the government also announced its intention to abolish fees for government-provided secondary education.

However, the education crisis in Malawi is more nuanced than merely poor literacy and numeracy skills, and truancy. The education system has largely benefited first language Chichewa speakers since 1968 when English and Chichewa were adopted as the national languages of Malawi (Trudell 2016: 44). In 1969, Chichewa was established as the medium of instruction for the first four years of primary school and English became the language of instruction from Standards 5–8 in primary school and throughout secondary school (Matiki 2001: 206). English is also a mandatory subject in which a pass mark is required for the Primary School Leaving Certificate examinations in Standard 8; the Junior Certificate examinations in Form 2, and the Malawi School Certificate of Education (MSCE) in Form 4 (Matiki 2001: 206).

In 1996, once multiparty democracy was established, the education policy was changed to recognise and structurally include other languages in the educational life of Malawi. It was hoped that this would lead to a new dawn in education for the speakers of minority languages. A new education policy was written that directed mother tongue languages to be used as the medium of instruction in the first four years of education (Reilly 2019: 33). English would continue to be used as the medium for instruction from Standard 5 onwards. Apart from a few pilot programmes, however, the Ministry of Education did not implement this policy in any significant way, upholding the status quo.

Although people who are ethnically Chewa only make up 40 percent of the Malawian population, Chichewa has become the *de facto* first language of several other large tribes in Malawi, including the Ngoni and the Lomwe. When all first language speakers of Chichewa are counted together, they represent approximately 70% of the population of Malawi, which is a large and influential block (Matiki 2006: 250).

In 2014, the language policy for education changed again. This time it stipulated that the language medium of instruction would be English from Standard 1 in primary school through to Form 4 in secondary school (Trudell 2016: 45). This policy was meant to bring more equity to the education system, however it has also not been widely supported or implemented.

There are several reasons for the lack of implementation of the minority language policy as well as other minority language initiatives, including the social significance of Chichewa in Malawi. Chichewa serves an interpersonal function and operates as a link between people with different cultural and linguistic backgrounds in Malawi (Matiki 2001: 212). According to Matiki (2001: 212), Malawians are more likely to use Chichewa than any other language to bridge linguistic

differences, including in the education system. Issues of low language competence among teachers in minority languages and English is also a factor, as is linguistic centrism within the Ministry of Education and District Education Departments, and the overall paucity of educational resources for the population (Matiki 2006: 246-247; Trudell 2016: 39). Malawi does not have sufficient resources for its current education system, let alone a "translanguaging" approach to education (Kamwendo 2000: 6; Reilly 2019: 34). The extent of the education crisis was made evident to the project team when visiting 60 primary and secondary schools in the Mangochi District during 2024 to conduct seminars on how to use a dictionary and dictionary smartphone application (app). One primary school on the Eastern shore of Lake Malawi had 1 363 students and 8 teachers, resulting in a student-to-teacher ratio of 170:1. Moreover, the high number of students means that in some schools, classes are conducted in the open, under trees, as there are insufficient classroom blocks. In regard to the paucity of language resources, at one large primary school in the Mangochi District the project team was shown a small closet containing approximately 20 books, which is vastly inadequate for a primary school of more than 2 000 students. The consequences of this are seen in the Malawi 2018 Census, which indicates that the percentage of Yawo completing primary, secondary and tertiary education is lower than any other ethnic group in Malawi (National Statistical Office Malawi 2019).

The impetus for a wider project

It was from this understanding of the Yawo education context that the Ciyawo–English Dictionary Project (CEDP) was created as a vehicle for developing two bilingual dictionaries for two different end-user groups. One dictionary, the Ciyawo–English Learner's Dictionary (CELD), would be developed primarily for L1 English speakers who want to learn Ciyawo for communicating with the Yawo in their language, thereby promoting dignity and equity in dialogue and conversation (Dicks 2022; 2024a). The primary end users of the CELD are non-Yawo, English-speaking adult learners. The other dictionary, the English–Ciyawo Learner's Dictionary (ECLD), would be developed for L1 Ciyawo speakers, primarily students who need to develop their competence in English for educational and employment purposes (Dicks 2018; 2024b). The target audience for the ECLD would be students in Standards 6–8 of primary school and Forms 1–2 of secondary school in Malawi.

The need for both of these dictionaries was seen as important and urgent. However, since the project team lacked adequate resources with which to develop the CELD — namely, a word list of high-frequency and high-relevance headwords in Ciyawo developed from a corpus, as well as the identification of key word senses and natural example sentences — the team decided to embark on developing the ECLD first, as resources were available to begin this part of the project.

Methods of development

It is well documented that certain resources are deemed essential for the construction of modern, bilingual learner's dictionaries. At centre stage is the corpus, consisting of machine-readable written texts for determining headword lists, sense hierarchies, definitions and example sentences that are to be included in the dictionary articles (Atkins and Rundell 2008: 53). However, even the best and largest corpus is of little use without people with educational, linguistic, lexicographic and technological expertise to design a dictionary for specific end users, and to construct it by utilising the data appropriately. Moreover, anyone who has been involved in the management of a dictionary project knows that they are multi-year, even decades-long projects that require significant funding for equipment, software, rent, salaries, consultancy, publishing, etc. A project is also dependent on oversight and government and community support to ensure that the end product is constructed to an acceptable standard and aligned with the ideals and needs of the community.

In the Yawo context of southern Malawi, virtually all of these resources were unavailable in any significant manner at the conception of the project, except for community support. Prior to 2005, there was not a recognised standard orthography for Ciyawo (Centre for Language Studies (CLS) 2005), and up until 2005, Ciyawo was written idiosyncratically in Malawi and other countries. Even after a cross-border initiative to standardise the Ciyawo orthography involving linguists from Malawi, Mozambique and Tanzania, some linguistic and education departments continued to write Ciyawo idiosyncratically, a decision that hampers the development of literature in Ciyawo to the present day (Banda et al. 2008).

A formidable barrier to creating a high-frequency Ciyawo word list was the dearth of machine-readable text available in Ciyawo with which to develop a Ciyawo corpus. The available texts in Ciyawo were more than a century old. One of these was Yohanne Abdallah's (1973) historical account of the Yawo, written in Ciyawo in 1909. There were also several collections of Yawo-sacred stories and proverbial tales documented and translated by Duff Macdonald dating from 1881 and 1882 (Macdonald 1881; 1882). The only recent texts of significance were a Yawo Bible translation produced by the Bible Society of Malawi, and portions of translated Scripture and Scripture teachings that were being produced by translation teams in Malawi and Mozambique in the late 1990's and early 2000's.⁴ The only other recent text was a collection of traditional Yawo proverbs and proverbial stories, *The Wisdom of the Yawo*, published by Ian Dicks in 2006 (Dicks 2006).

Also available, but not usable for corpus construction were Yawo word lists and older dictionaries, including Edward Steere's Collections for a *Handbook of the Yao Language* (1871), Alexander Heatherwick's *A Handbook of the Yao Language* (1902) and Meredith Sanderson's *A Dictionary of the Yao Language* (1954). Apart from not being in a machine-readable state, none of these dictionaries were constructed from a corpus and they contained archaic terms and definitions

that are no longer commonly used by the Yawo. These texts, however, would still be useful as refining tools once a corpus was developed, as would the *Mgopolela Malowe Jwa Ciyawo* (CLS 2013), a monolingual Ciyawo dictionary.⁵

Aware that the previous dictionaries were dated and largely developed without a corpus, the project team decided to do everything possible to produce a general corpus in Ciyawo from which to construct the CELD. However, there is much debate regarding what constitutes a valid corpus (De Schryver and Prinsloo 2000: 91). A corpus is commonly defined as,

a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research (Sinclair 2004).

Moreover, "a general corpus is typically designed to be balanced, by containing texts from different genres and domains including spoken and written, private and public" (De Schryver and Prinsloo 2000: 91).

Even the task of producing a general corpus was a challenge for an African language with so little available written text. Moreover, the written text that was available was largely limited to a single semantic domain — Christian Scriptures. Nevertheless, it has been suggested that "any corpus — however unbalanced — is to be a source of information and indeed inspiration. Knowing that your corpus is unbalanced is what counts" (Atkins et al. 1991). The unbalanced nature of the Yawo corpus is one fact that the project team were fully aware of.

The aim then became to create what De Schryver and Prinsloo (2000: 92) call a "structured corpus", which is the first step towards creating an organic corpus. Available machine-readable written texts in Ciyawo were gathered, mainly in the domain of Christian Scriptures and teachings. To these were added additional transcribed texts from recorded oral interviews in a variety of semantic domains. The recorded interviews were conducted in domains including but not limited to the following: (i) work (well digging, farming, food preparation, fishing, grafting, business, harvesting, hunting, piece work, making local fermented drinks, clay pot making, carpentry, preparing groundnuts, making hoes, making metal pots); (ii) travelling; (iii) family life (marriage, traditional rules for pregnancy, food, cooking); (iv) traditional medicine (role of a healer, use of charms); (v) Yawo traditional religion (practices and beliefs, sorcery, witchcraft); (vi) Islam (practices and beliefs, religious differences); (vii) historical events (war in Mozambique); (viii) health and well being; (ix) Yawo Games; (x) initiation (boys, girls, pregnancy, chieftaincy songs); (xi) stories (traditional Yawo tales, histories); and (xii) death (funerals).

The bank of text consisting of Scripture and Scripture teachings were mainly in the narrative genre and constituted 448 199 words of the corpus. In regard to the oral interviews, more than 190 interviews were conducted, adding more than 322 000 words to the corpus. The total number of words from all sources in the Ciyawo corpus came to 770 871.⁶

The corpus, however, was unbalanced, as 60 percent was made up of text from Scripture and Scripture teachings and 40 percent from transcribed oral interviews. Some people have raised questions regarding the unreliability of an unbalanced corpus, as well as the unreliability of a corpus developed from spoken data, especially unscripted conversations (cf. De Schryver and Prinsloo 2000: 94). However, it is "any corpus compiler's task to attempt to assemble a representative corpus for his/her specific needs" (De Schryver and Prinsloo 2000: 92). The specific need identified by the project team was to identify the 3 000 high-frequency and high-relevance words in Ciyawo, and to define them for an English-speaking audience in the CELD. Considering this, the oral nature of a large portion of the source material is seen as a strength, since it captures speech that is natural and frequent, and not at a higher and more rarified literary level. Ultimately, corpus design is constrained by what is available, and if there is no written text with which to create a more so-called balanced corpus, the compilers must rely on whatever means are available to them.

It has also been argued that different genres of speech events are of different value to a corpus, with unscripted conversations at one end and monologues and lectures at the other (De Schryver and Prinsloo 2000: 94). The project team used what is known as an "informal interview" approach when conducting oral interviews on selected domains, which placed the genre of speech at the monologic end of the spectrum. The people interviewed were those whom the community identified as having knowledge of a specific topic. The interviewers used various questioning methods to elicit extended responses, including the grand tour question, which allows a person to give an overview of the topic in response. This type of question was used predominantly because it encouraged a conversational tone and allowed participants to talk about what they considered important in regard to the topic, rather than focusing on what the interviewer deemed significant (Fetterman 2010: 40).

Once the interviewing and transcribing was complete, a corpus query tool, Wordsmith Tools 6.0, was used to process the text files and provide statistical information about the database. It also gave concordance lines showing how words are used in context, and enabled the tagging and assigning of word classes. Knowing the unbalanced nature of the corpus required the team to diligently sort, demote and remove lemmas that were overly represented, such as religious terminology. While Wordsmith works well as a query tool, it does not help in identifying low-frequency but high-relevance words. For this, the team consulted privately produced word lists and monolingual dictionaries, including the *Mgopolela Malowe Jwa Ciyawo* (CLS 2013), as well as older Ciyawo Dictionaries mentioned earlier, and the Yawo community.

The need for developing two dictionaries

The project was ambitious from the start. The aim was to produce two dictionaries for two different end-user groups. Some individuals inquired as to why

one bilingual dictionary could not be produced and then reversed or back-translated for the second language community. The reason this is not possible is because bilingual dictionaries are not like eggs — they do not flip easily, especially in the case of Ciyawo and English. This is because the high-frequency word lists are different for both groups of end users. However, even when actions or objects have gloss equivalents in both languages, the meanings applied to them can be different for each "languaculture" community (Agar 2002: 60). Even when the reality being defined seems to be ubiquitous to both language and culture communities — such as *bicycle/njinga*, *dog/mbwa*, *farmer/mlimi*, *to pray/kuswali* — it was found that L1 speakers of Ciyawo and English apply different meanings to these realities. This means that appropriate headword lists are required for each languaculture community, applying L1 definitions to each sense so that L2 learners can grow their understanding to include the L1 meaning of the headword rather than their own L2 cultural meaning.

The English–Ciyawo Dictionary

The purpose of the English–Ciyawo Learner's Dictionary (ECLD) is to assist L2 Ciyawo speakers of English in growing their English proficiency so they can navigate the Malawian education system, gain employment, and interact with L1 and L2 English speakers internationally.

Therefore, the headword list for the ECLD needed to consist of the 3 000 most high-frequency and high-relevance words for L2 learners of English. The project team did not have the ability to develop such a list. OUP was approached and asked if their 3 000 word list could be used, as well as definitions up to four senses, and example English sentences for these senses. While the agreement took time to arrange, OUP was helpful, generous, and accommodating throughout the entire process of interaction. Permission was granted to translate the OUP definitions into Ciyawo. For this task, a meaning-based method of translation was used. This method differs from a literal translation method in that it aims in "retelling, as exactly as possible, the meaning of the original message in a way that is natural in the language into which the translation is being made" (Barnwell 1986: 9).

The target audience for the ECLD is upper-primary and lower-secondary school students, unlike the *Oxford Advanced Learner's Dictionary* (OALD), which has been primarily designed for adult learners. This meant that a number of example sentences in the OALD 8th edition needed to be changed, as the ideas expressed in them were not easily recognisable by the Yawo end users (Hornby 2010).

To assist learners further, the translators sought to use only the 3 000 key English headwords in the example sentences. This enables Yawo learners to look up the words used in the English example sentences that they do not understand. Unfortunately, the OALD does not adhere to this principle strictly and includes words in a number of example sentences that are outside of the 3 000 key word list. To overcome this problem, permission was sought to use a consulting editor to locate replacement example sentences for all the senses

missing appropriate example sentences. This included senses in which the example sentences expressed an idea uncommon in a Yawo context, example sentences that used words outside of the Oxford 3 000 word list, and instances where an example sentence was not provided.⁷ Overall, the consulting editor provided 1 386 appropriate example sentences.

Training non-professional lexicographers

At the beginning, the project lacked people with the skills and abilities required to produce a dictionary, in particular lexicographers. Lacking available expertise, Dicks and Nchembe set to work learning the art and craft of dictionary construction. Dicks attended an intensive introductory course on lexicography at Stellenbosch University. Following this, Dicks and Nchembe researched practical lexicography using the works of Landau (2001), and Atkins and Rundell (2008). They also attended AFRILEX conferences to learn from people who were working on other dictionaries in the African context.

Drawing on this learning, research, and personal experience as learners of Ciyawo and English, Dicks and Nchembe created a style manual and included elements in the dictionary article that would be most helpful for L2 learners of both languages.

The dictionary articles in both dictionaries include the following elements:

- (1) Headword: The word that an L2 learner would be searching for.
- (2) Senses: This includes up to four key senses of meaning in the ECLD and up to three in the CELD.
- (3) Translation equivalent in L1 of learner: This applies if a lexical equivalent exists.
- (4) Definition: The definition of the headword in the learner's L1.
- (5) Example sentence in L2 of learner: This sentence or phrase is meant to show how the headword is used naturally according to the sense.
- (6) Translation of example sentence in L1 of learner: This would include the translation equivalent if a lexical equivalent exists.
- (7) Other tense forms: Verbs are shown in three tenses, namely present, continuous and past tense.
- (8) Other word forms: Other forms of the headword are shown if the word is an adjective.
- (9) Plurals: The plural form of a noun is given.
- (10) Part of speech: Ciyawo grammar terms are used in the ECLD, such as *lina* (noun), *msali* (verb), *mlondecesya* (adjective), *mjonjecesi* (adverb), *mpecesi* (preposition), etc. English grammar terms are used for parts of speech in the CELD.

After receiving feedback from end users on the ECLD that was published in 2018, several changes were made to the dictionary article for the construction of the CELD. A translation of the Ciyawo definition in English was added so that learn-

ers could read the definition in two languages. The number of senses was also reduced in the CELD from four to three key senses.

The lexicographical team

The original Yawo lexicographers were chosen primarily for their language capacity in both Ciyawo and English. They were also required to have completed the Malawi School Certificate of Education (MSCE), which is a recognised competence marker in English. Included in the first group of lexicographers and field testers employed were several primary and secondary school teachers who provided insight into the end users of the ECLD.

All of the lexicographers were provided with initial training, which included teaching on the nature of a dictionary, different types of dictionaries, the hallmarks of a learner's dictionary, the importance of understanding the end user: including their needs, abilities and their perspectives, the objectives of the Ciyawo–English Dictionary Project, and most importantly the principles and methods of defining words. This teaching covered various styles of defining words, including genus and differentia, synonym, typical, and complete sentence styles, as described in several practical guides to lexicography.

A key aspect of a learner's dictionary is the provision of an example sentence in which the headword is used appropriately according to the sense of the word. The ECLD used primarily example sentences from the OALD 8th edition and from the consulting editor, as already mentioned. However, for the construction of the CELD, the newly created Ciyawo corpus was the main source of these example sentences. It was not always possible to use the corpus, since available sentences and phrases were not always appropriate or clear. For the senses without appropriate example sentences, recorded role-plays were conducted by the team to capture the way a word is used in natural dialogue. This method was found to be helpful and also proved to be a lot of fun.

Collaboration

In a low-resource community collaboration is essential — not just for dictionary construction, but in every aspect of life. The Yawo realise this and have many proverbs that talk about the necessity of working together to achieve a purpose, such as 'Mtwe umpepe wangatwicila cipagala' ('One head does not carry a small thatched roof') (Dicks 2006: 96).

In many ways collaboration is what made it possible to undertake and complete this dictionary project. From the very beginning, collaboration was required between L1 English and L1 Ciyawo speakers, as neither community had the language abilities to construct a bilingual dictionary on their own. Collaboration extended further to virtually every aspect of the project. Within the lexicography office, people operated collaboratively: Knowing that individually people

were under-skilled, the lexicographical team always worked in pairs when creating and reviewing dictionary articles, and role-plays were conducted as a group.

Knowing that the team did not have the skills required for designing and constructing a single dictionary, let alone two, collaboration was sought with OUP, which resulted in permission to use designated parts of the OALD 8th edition. OUP is a large multinational company that could have declined these requests. Instead, to their credit, OUP helped this small project team to achieve its goals, while also protecting their intellectual property and brand.

The team collaborated with the CLS at the University of Malawi, Zomba. The CLS provided education and training on the Ciyawo Orthography. Furthermore, two Yawo-speaking academics from the CLS became consulting editors on the project. They contributed by reviewing and providing editorial insights on both dictionaries. The CLS also helped locate several Yawo graduates in linguistics who joined the CEDP lexicographical team.

Collaboration also occurred with the wider Yawo community. Yawo-speaking school teachers and school principals helped to facilitate field testing with students and teachers in primary and secondary schools. The Paramount Yawo chief, Ce Kawinga, visited the project on several occasions, as did other senior chiefs and local village headmen and women to understand the project and give encouragement. Senior Yawo chiefs also sent representatives for training as reviewers, and gave the team feedback and approval to publish the finished work.

The Malawi Ministry of Education provided support. This came after publishing the first English–Ciyawo Learner's Dictionary in 2018. The Minister of Education at the time, Hon. Bright Msaka SC, was the keynote speaker at the launch of the ECLD in Mangochi, and encouraged parents to purchase the dictionary for their children and the Ministry of Education to support the initiative. The recommendation of the Minister of Education facilitated collaboration with the Ministry of Education in Mangochi, enabling the third phase of the project to move ahead, which was to conduct seminars in primary and secondary schools in the Mangochi District on 'How to Use a Dictionary and Dictionary Smartphone Application'.

Project learnings

There were many lessons learnt from this project. The first is that dictionary construction takes longer than anticipated. The ECLD took 12 years from conception to publication, and an additional 4 years to complete and publish the CELD. Thereafter, it took an additional 2 years to release them as smartphone apps, add sound files, and conduct sensitisation workshops in primary and secondary schools in the Mangochi District on how to use them.

Another learning relates to the difference between the needs of dictionary end users and their ability to purchase one. Although Yawo students wanted and needed to learn English, very few had the financial means to do so even at a sub-

sidised price. The ECLD was published and printed in book versions in 2018 and the CELD in 2022. The initial print runs were small: Only 2 000 copies of the ECLD were printed, and of these 1 500 copies were distributed free of charge to primary and secondary schools in Yawo-speaking areas of Malawi, while only 500 copies were sold to the general public. For the CELD, 1 000 copies were printed, of which 512 copies were distributed without cost to schools and 488 copies were sold through various bookstores in Southern Malawi for between US\$5–\$10. Although the demand for the dictionary was great, the ability of individuals and the Malawian Government to purchase them was extremely low. Due to the poor economic circumstances of many Yawo and Malawians, books and dictionaries are rarely purchased because of their cost.

Dictionary smartphone apps

With the introduction of smartphones in Malawi and the high adoption of them by many households, even among the Yawo, it was decided to investigate publishing the ECLD and CELD in smartphone apps that could be operated without the need for cellular data once the dictionaries were downloaded. The 2018 census indicates that 51,7 percent of Malawian households own a mobile phone. Significant for this project, 47,1 percent of households in the Mangochi District own a cellular phone, which is the fifth highest level of cellular phone ownership of the 13 districts in Southern Malawi, apart from cities (National Statistical Office Malawi 2019). Moreover, the number of cellular connections continues to rise. In 2025 there are 13,2 million cellular connections, which represents 60,3 percent of the general population (Datareportal 2025).

Once again, the team lacked the technical expertise or financial resources to undertake such a venture independently. The idea of creating smartphone apps was initially discussed with OUP; however, their suggested pathway was technically and financially out of reach. To find a solution, the team investigated the Dictionary App Builder (DAB) programme, developed by the Summer Institute of Linguistics (SIL), an international Scripture translation organisation. The DAB enables non-professional app developers equipped with an appropriate lexicon data file to build a dictionary app at an extremely low cost. The two CEDP dictionaries were constructed using the dictionary writing system TshwaneLex, which provides multiple options for exporting a lexicon database. The DAB is a sophisticated programme that enables the construction of a dictionary smartphone app with all elements of a dictionary article that are beneficial to L2 learners.

The completed dictionary apps have several advantages over the paper form of the dictionaries. The app enables the end user to search for words in multiple ways, including without knowing how to spell the word fully. Morris (2021: 36) highlights this as a benefit, saying that "electronic dictionaries are structured in a completely different way and no longer rely on the alphabetical access structure as the only access, the outer alphabetical access structure loses its status as the default access structure". The DAB app also allows the end user to search

for words that are only partially known, and to search for and find words that are misrepresented orthographically, which according to Morris (2021: 39) is the sign of a good electronic dictionary interface.

The DAB app also displays each article on a single screen, unlike an e-reader version of a dictionary. This reduces the overpowering affect that a page of articles can have on a novice. The DAB app also allows the dictionary to have an audible dimension, which greatly adds to its pedagogical function. In the second version of the ECLD and CELD apps, sound files were added so that end users can hear the headwords spoken by L1 speakers of Ciyawo and English, which assists with their speech reproduction. Once again, the construction of the apps was made possible through collaboration with the SIL who provided the DAB programme free of charge. The only costs associated with publishing the dictionaries as smartphone apps, have been paying a technology consultant, who has directed and assisted the building of the apps for Android and iOS. Both dictionaries are now available on the Google Play Store and the Apple App Store free of charge.

Finally, the Ciyawo–English Dictionary Project has been a collaboration between the Dictionary making team and financial supporters, who were predominantly ordinary people in Baptist churches of Australia who wanted the Yawo to have more opportunities to complete primary, secondary and tertiary education. My own role, as senior editor and project manager has always been part-time and has been supported largely through people in Australian Baptist churches. Baptist Mission Australia have supported this project and overseen the finances and tax deductible giving, which has enabled this project to be completed.⁸

The cost of this project is hard to establish, as the costs associated with the senior editor and project manager role were not accounted for between 2008–2016. These roles were part-time and conducted in conjunction with other roles and responsibilities in Malawi. Overall, the senior editor and project manager roles were undertaken approximately two days a week for 14 years, and then one day a week for another 4 years.

It is estimated that USD \$385 000 was spent to develop, publish and print two dictionaries in book form, as well as to produce smartphone apps for both dictionaries so that they could become available on Android and iOS.⁹

Conclusion

Many minority language communities in sub-Saharan Africa are under pressure of becoming redundant as mediums of communication, as they are undervalued and under-resourced in language learning tools. The result is seen in low education outcomes, inadequate delivery of social services, a sense of disempowerment, and isolation. What has been shown in this article is that a team of non-professional lexicographers can produce language learning tools through collaborative

effort. There are many other minority language communities in similar situations to the Yawo. Realistically, dictionaries for these communities are unlikely to be profitable for large dictionary corporations in the short term. However, collaboration with minority communities will greatly benefit communities, making learning more equitable and assisting in closing the communication gap.

Endnotes

1. "A word family consists of a headword (for example, 'assume'), its inflects ('assumes', 'assumed', 'assuming'), and its derivations ('unassuming', 'unassumingly')" (Webb and Nation 2017: 7, 14).
2. Others have suggested that a better threshold for an L2 learner is 3 000 highest-frequency word families as this would increase their ability to understand 98 percent of most reading materials and 95 percent of vocabulary used in spoken discourse. Dang and Webb (2016) have suggested that high-frequency words lists should count lemmas and lemma headwords instead of word families as "lemma headwords are the most commonly used unit of counting both inside and outside the classroom" (Webb and Nation 2017: 11-12). Either way, there is agreement that learning high-frequency words is key to closing the communication gap for L2 learners.
3. There are 13 subcategories, referred to as indicators, under these four headings: Food Security, Drinking Water, Nutrition, Sanitation, School Attendance, Literacy and Schooling, Asset Ownership, Housing, Rubbish Disposal, Electricity, Child Labour, Job Diversity, Unemployment (National Statistical Office Malawi 2022: 7).
4. There were some other texts that were very old, and that were not accessible to be transcribed. The most significant texts in the Ciyawo language were mainly other Bible translations, including several New Testament books translated by Alexander Hetherwick in the 1890's, including *Utenga Wambone Wa Luka*, *Masengo Ga Wandumitume*, *Utenga Wambone Wa Marko*, *Utenga Wambone Wa Matayo*, *Utenga Wambone Wa Yohana*, *Achikalata Jua Paolo Jua Ndumitume Kwa Wa Korinti* and *Kalata Jua Paolo Jua Ndumitume Kwa Wa Rumi* (Houston 2022: 9). There were also some primers, produced by R.S. Hynde in 1892 and 1894 (Houston 2022: 6).
5. The CLS dictionary is a monolingual dictionary produced as a tool for the hoped-for initiative of vernacular education in Ciyawo, which was made policy by the Malawian government in 1996 but never developed beyond a few pilot projects.
6. There is conjecture as to what constitutes a word, especially in Bantu languages, as it depends on the orthographical style (De Schryver and Prinsloo 2000: 100). What takes three words in English, "I am going", is written conjunctively as a single word in Ciyawo — Ngwawula.
7. Lorna Morris (née Hiles) was a tremendous help to our project as she had experience as an editor working on the *Oxford South African Illustrated School Dictionary* (2008).
8. I want to particularly thank my colleagues at Baptist Mission Australia, formerly Global Interaction, who saw the value of this project and supported it over many years, particularly Mrs. Alison Nissley and Dr. John Davis, who oversee projects in the organisation and who championed it and encouraged me along the way, as well as the many people who supported this project financially.
9. This figure was calculated from SFI giving records between 2009–2025.

References

- Abdallah, Y.B.** 1973. *The Yaos: Chiikala cha WaYao*. Edited and translated by M. Sanderson. London: Frank Cass.
- Agar, M.** 2002. *Language Shock: Understanding The Culture of Conversation*. New York: Harper Perennial.
- Atkins, S., J. Clear and N. Ostler.** 1991. *Corpus Design Criteria*.
<http://www.natcorp.ox.ac.uk/archive/vault/tgaw02.pdf> [16 May 2025]
- Atkins, B.T.S. and M. Rundell.** 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Baldauf, R.B. and R.B. Kaplan.** 2004. *Language Planning and Policy in Africa: Botswana, Malawi, Mozambique and South Africa*. Vol. 1. Clevedon: Multilingual Matters.
- Banda, F., A. Mtenje, L. Miti, V. Chanda, G. Kamwendo, A. Ngunga, M. Liphola, C. Manuel, B. Siteo, S. Simango and M. Nkolola.** 2008. *A Unified Standard Orthography for South-Central African Languages: Malawi, Mozambique and Zambia*. Second edition. Monograph Series No. 229. Cape Town: CASAS.
- Barnwell, K.** 1986. *Bible Translation: An Introductory Course in Translation Principles*. Third edition. Dallas: Summer Institute of Linguistics.
- Bayetto, A.** 2018. *The New Oxford Wordlist Research Report*.
https://www.oup.com.au/__data/assets/pdf_file/0016/120841/SCHL_WORDLIST_Report-2018_FA_2-web.pdf
- Centre for Language Studies (CLS).** 2005. *The Orthography of Ciyawo*. Chileka: E+V Publications.
- Centre for Language Studies (CLS).** 2013. *Mgopolela Malowé jwa Ciyawo: Ciyawo Dictionary*. Blantyre: Dzuka.
- Coxhead, A.** 2000. A New Academic Word List. *TESOL Quarterly* 34(2): 213-238.
- Dang, T.N.Y. and S. Webb.** 2016. Making an Essential World List for Beginners. Nation, I.S.P. (Ed.). 2016. *Making and Using Word Lists for Language Learning and Testing*: 153-167. Amsterdam: John Benjamins.
- Datareportal.** 2025. *Digital 2025: Malawi*.
<https://datareportal.com/reports/digital-2025-malawi> [17 May 2025]
- De Schryver, G.-M. and D.J. Prinsloo.** 2000. The Compilation of Electronic Corpora: With Special Reference to the African Languages. *Southern African Linguistics and Applied Language Studies* 18(1-4): 89-106.
- Dicks, I.D.** 2006. *Wisdom of the Yawo People: Under the Elephant's Belly You Can't Pass Twice*. Zomba: Kachere.
- Dicks, I.D.** 2018. *English-Ciyawo Learner's Dictionary*. Mzuzu: Mzuni Press.
- Dicks, I.D.** 2022. *Ciyawo-English Learner's Dictionary*. Mzuzu: Mzuni Press.
- Dicks, I.D.** 2024a. *Ciyawo-English Learner's Dictionary*. Smartphone Application. Mzuni Press / Google Play Store / Apple App Store.
- Dicks, I.D.** 2024b. *English-Ciyawo Learner's Dictionary*, Smartphone Application. Mzuni Press / Google Play Store / Apple App Store.
- Fetterman, D.M.** 2010. *Ethnography: Step-by-Step*. Third edition. Los Angeles: SAGE.
- Hornby, A.S. (Ed.).** 2010. *Oxford Advanced Learner's Dictionary*. Eighth edition. Oxford: Oxford University Press.

-
- Houston, T.J.** 2022. Utenga Wambone — The "Good News": An Exploration of Historical Ciyawo Bible Translations and Linguistic Texts. *Studia Historiae Ecclesiasticae* 48(3): 1-18.
- Kamwendo, G.H.** 2000. Interfacing Language Research with Policy: The Case of Language in Education in Malawi. *Nordic Journal of African Studies* 9(2): 1-10.
- Kandybowicz, J. and H. Torrence.** 2017. Africa's Endangered Languages: An Overview. Kandybowicz, J. and H. Torrence (Eds.). 2017. *Africa's Endangered Languages: Documentary and Theoretical Approaches*: 1-10. Oxford/New York: Oxford University Press.
- Landau, S.I.** 2001. *Dictionaries: The Art and Craft of Lexicography*. Second edition. New York/Cambridge: Cambridge University Press.
- Macdonald, D.** 1881. *East African Tales*. Edinburgh: William Blackwood and Sons.
- Macdonald, D.** 1882. *Africana; or, The Heart of Heathen Africa*. Vol. 1. London: Simpkin Marshall & Co.
- Malawi Data Portal.** 2018.
<https://malawi.opendataforafrica.org/gesljee/population-2018> [17 May 2025]
- Matiki, A.J.** 2001. The Social Significance of English in Malawi. *World Englishes* 20(2): 201-218.
- Matiki, A.J.** 2006. Literacy, Ethnolinguistic Diversity and Transitional Bilingual Education in Malawi. *International Journal of Bilingual Education and Bilingualism* 9(2): 239-254.
- Morris, L.H.** 2021. *A Model for a Comprehensive Electronic School Dictionary for South African Primary School Learners*. Unpublished Ph.D. Dissertation. Stellenbosch: Stellenbosch University.
- Nation, I.S.P.** 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I.S.P.** 2022. *Learning Vocabulary in Another Language*. Third edition. Cambridge: Cambridge University Press.
- National Statistical Office Malawi.** 2019. *2018 Malawi Population and Housing Census. Main Report*.
<https://www.nsomalawi.mw/census/2018> [17 May 2025]
- National Statistical Office Malawi.** 2022. *The Second Malawi Multidimensional Poverty Index Report — Nov 2022*.
https://www.undp.org/sites/g/files/zskgke326/files/2023-08/Malawi-Multidimensional%20Poverty%20Index%20Report_0.pdf [17 May 2025]
- Oxford Learner's Dictionaries.** 2025. *The Oxford 3000 and the Oxford 5000*.
<https://www.oxfordlearnersdictionaries.com/about/wordlists/oxford3000-5000#> [17 May 2025]
- Prinsloo, D.J.** 2017. Africa's Response to the Corpus Revolution. Xu, Hai (Ed.). 2017. *Proceedings of the 11th International Conference of the Asian Association for Lexicography (ASIALEX 2017), 10-12 June 2017, Guangzhou, China: Lexicography in Asia: Challenges, Innovations and Prospects*: 20-31. Guangzhou, China: ASIALEX.
- Reilly, C.** 2019. Attitudes Towards English as a Medium of Instruction in Malawian Universities. *English Academy Review* 36(1): 32-45.
- Shaffer, D.E.** 2005. Review of *Oxford Advanced Learner's Dictionary of Current English* (7th ed.). Korea TESOL Journal 8(1): 135-141.
- Sinclair, J.** 2004. Corpus and Text: Basic Principles. Wynne M. (Ed.). 2004. *Developing Linguistic Corpora: A Guide to Good Practice*: 1-16. Oxford: Oxbow Books.
<https://users.ox.ac.uk/~martinw/dlc/> [16 May 2025]
- Trudell, B.** 2016. *The Impact of Language Policy and Practice on Children's Learning, Evidence from Eastern and Southern Africa*. Nairobi: UNICEF.
<https://www.unicef.org/esa/sites/unicef.org/esa/files/2018-09/UNICEF-2016-Language-and-Learning-FullReport.pdf> [16 May 2025]

<http://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2108> (Volle uitgawe / Full volume)
<https://lexikos.journals.ac.za>; <https://doi.org/10.5788/35-2-2089> (Project)

Where Tools Are Few: Constructing Limited-Word Bilingual Learner's Dictionaries 275

Trudgen, R. 2000. *Why Warriors Lie Down and Die*. Darwin: Aboriginal Resource and Development Services Inc.

UNESCO. 2024. *Malawi: Education Country Brief*.

<https://www.iicba.unesco.org/en/malawi> [17 May 2025]

Webb, S. and P. Nation. 2017. *How Vocabulary is Learned*. Oxford: Oxford University Press.

Generative Artificial Intelligence (GenAI) as Information Tool for Lexicographic Information Needs

Theo J.D. Bothma, *Department of Information Science,
University of Pretoria, South Africa*
(theo.bothma@up.ac.za)

(<https://orcid.org/0000-0001-7850-3263>)

and

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, South Africa*
(rhg@sun.ac.za) (<https://orcid.org/0000-0002-3423-058X>)

Abstract: The focus of this paper is on GenAI, chatbots and some implications for lexicography and dictionary use. The use of GenAI as an information tool to provide information to end-users (readers) who have a specific information need when reading a text is discussed. GenAI could previously provide content similar to that of a dictionary but could not provide contextualised answers. This has changed and most chatbots now have two modes, a "search" mode and a "thinking/reasoning mode", i.e., it is able to argue logically about its different proposed meanings in context and tends to offer a solution. This feature is discussed at the hand of examples, and it seems as if the chatbots can now provide correct contextual meaning and logically motivate the choice of meaning in context, based on their critical analysis and thinking skills although it still "hallucinates" if it has no answer. Hereafter, the role of dictionaries and lexicographers in an AI-enhanced world is discussed. The traditional assignment of dictionaries to document the status and history of a language remains a very important function and needs to be encouraged, especially in environments with limited language resources. However, exploring new commercial ventures, incorporating latest technologies, would be essential to the future of the discipline and industry.

Keywords: CHATBOT, CONTEXTUALISATION, GENERATIVE ARTIFICIAL INTELLIGENCE, INTERFACE, HALLUCINATE, GENAI, INFORMATION NEEDS, INFORMATION TOOLS, MEANING, TEXT RECEPTION

Opsomming: Generatiewe Kunsmatige Intelligensie (GenKI) as inligtings-werktuig vir leksikografiese inligtingsbehoefte. Hierdie artikel fokus op GenKI, kletsbotte en sommige implikasies daarvan vir leksikografie en woordeboekgebruik. Die gebruik van GenKI as 'n inligtingswerktuig wat inligting moet verskaf aan eindgebruikers (lesers) met spesifieke inligtingsbehoefte wanneer hulle 'n teks lees, word bespreek. GenKI kon vroeër inhoud vergelykbaar met dié van 'n woordeboek verskaf, maar kon nie gekontekstualiseerde antwoorde verskaf nie. Dit het verander en die meeste kletsbotte het tans twee modi, naamlik 'n "soekmodus"

en 'n "dink/redeneermodus". Hulle kan hulle verskillende betekenisvoorstelle logies binne die gepaste konteks beredeneer en neig om 'n oplossing aan te bied. Hierdie kenmerk word aan die hand van voorbeelde bespreek, en dit lyk asof kletsbotte nou korrekte kontekstuele betekenis kan verskaf en die keuse logies binne die konteks kan motiveer, gebaseer op hulle kritiese analise en denkvaardighede, alhoewel hulle steeds "hallusineer" as hulle nie 'n antwoord het nie. Hierna word die rol van woordeboeke en leksikograwe in 'n KI-versterkte wêreld bespreek. Die tradisionele opdrag van woordeboeke om die status en geskiedenis van 'n taal te dokumenteer, is steeds 'n belangrike funksie wat aangemoedig moet word, veral in omgewings met beperkte taalhulpbronne. Die verkenning van nuwe kommersiële ondernemings, insluitend die nuutste tegnologieë, is egter noodsaaklik vir die toekoms van die dissipline en die industrie.

Sleutelwoorde: BETEKENIS, GENERATIEWE KUNSMATIGE INTELLIGENSIE, GENKI, HALLUSINEER, INLIGTINGSBEHOETES, INLIGTINGWERKTUIG, KLETSBOT, KONTEKSTUALISERING, KOPPELVLAK, TEKSBEGRIP

1. Introduction

1.1 Background to artificial intelligence

Currently, artificial intelligence (AI) can be found in numerous and diverse environments. Many newspapers, popular tech journals, tech newsletters, academic articles, books, etc. discuss the uses of AI, with widely divergent opinions. On the one hand, there is a very positive approach, arguing that everybody should be AI literate and should actively use AI. On the other hand, there is a negative approach, arguing, *inter alia*, that AI heralds the end of humanity, the end of jobs and that it is bad for the environment. There are also various issues regarding ethics in the work environment, for students, school children and researchers. Problems of bias, plagiarism, privacy, lack of learning and "hallucinations" are contrasted with the apparent usefulness of this tool, even if it is used correctly. Current AI detection software is not able to differentiate unambiguously between AI generated content and authors' own work, as well as between mis-/disinformation, fake news and deepfakes. The potential positive impact of the AI revolution can, however, not be negated.

A significant aspect of GenAI has been the speed of its development. The first release of the first commercial chatbot was at the end of 2022. It was App- or browser-based and since then AI and GenAI have become common tools for users in all walks of life for the satisfaction of multiple types of information needs. It also enabled typical natural language search tasks from the GenAI interface. AI has also been embedded in many application software packages, social media software, system software (e.g., on smartphones such as iPhones and Samsung phones), and it has been embedded in operating systems, browsers, etc. An AI summary is often provided as the default answer to a Google search, presenting an AI Overview, followed by relevant links. Google's latest development, AI Mode, will have a further significant influence on the general AI-based

provision of information to all Google search users. AI Mode is characterized by Germain (2025) as follows:

People use Google Search five trillion times a year — it defines the shape of the internet. AI Mode is a radical departure. Unlike AI Overviews, AI Mode replaces traditional search results altogether. Instead, a chatbot effectively creates a miniature article to answer your question. As you read this, AI Mode is rolling out to users in the US, appearing as a button on the search engine and the company's app. It's optional for now, but Google's head of Search, Liz Reid, said it plainly when launching the tool: "This is the future of Google Search." [...] There is little doubt AI Mode is an impressive piece of technology. It deploys a "fan out method" where the AI breaks your question into subtopics and does multiple searches simultaneously. Google says this lets AI Mode recommend more diverse sources, produce deeper answers to more complex queries, dives deeper — and you have the ability to ask follow-up questions.

1.2 Terminology issues

When discussing AI, one has to be aware of some potential terminology problems. Industry tends not to specify the exact nature of the different AI developments and usages, but rather only employs the generic term *AI* to refer to all these developments. One should, however, strictly and clearly differentiate between:

- AI as an umbrella term for multiple technologies, such as NLP (natural language processing), neural networks, machine learning, deep learning, and symbolic reasoning.
- GenAI, based on large language models (LLMs)
- Agentive AI, which refers to autonomous agents
- General Purpose AI, Artificial General Intelligence (AGI), currently still an ill-defined industry term for the point at which human intellect is outpaced by algorithms
- Hybrid AI systems which combine different types of AI techniques to create more powerful and versatile AI solutions

The focus in this paper is on GenAI, chatbots and some implications for lexicography and dictionary use, with special reference to the end-users (readers) who have a specific information need when reading a text, i.e., a text reception information need.

Since the recent emergence of GenAI, its development has been phenomenal. This can be seen in the reliability of responses when using chatbots. Chatbots originally tended to "hallucinate", but the use of much larger training databases resulted in chatbots becoming more accurate. In this development, multiple iterations of chatbots from multiple software companies have been released. In addition, specialised chatbots for different environments have been developed, e.g., for general queries, coding, for research, text summarization and image, video and audio generation. Of major significance is the recent devel-

opment of two different modes, viz., to be the equivalent of a search engine, and the ability to "think" and to "reason," i.e., to argue about the validity of its answers.

1.3 AI and chatbots in lexicography

AI and chatbots have also been a focus in recent lexicographic research. The focus has mostly been on possibilities for dictionary compilation and the preparation of writing assistants. A prominent feature in modern-day lexicography is the user-perspective, postulating that "the dictionary user must be an important consideration in dictionary-making" (Hartmann 1989: 102). The main implication of the user-perspective is that the compilation of any dictionary should be preceded by the identification of the envisaged target user, the lexicographic needs of this user and their reference skills. However, this is often neglected. Wiegand (1977: 62) already urged that lexicography should strive towards a much closer relationship with the intended target user, and that this user should no longer remain the familiar stranger. Unfortunately, the user has yet again been neglected in recent developments regarding AI and chatbots in lexicography. Too little attention has been directed at the users and their needs and expectations.

Some recent research relevant to AI, lexicography and the linking between text, dictionaries and other information tools can be found in Bothma and Gouws (2022), De Schryver (2023), Bothma and Fourie (2024), Huete-García and Tarp (2024), Lew (2024), Fuertes-Olivera (2024), and Bothma and Fourie (2025).

1.4 Lack of context

Research directed at the use of AI and chatbots in lexicography has shown that GenAI could already provide content similar to a dictionary, but that it cannot provide contextualised answers. Consequently, the reader is still dependent on their own evaluation of the GenAI-provided content to determine the meaning of the word or phrase in context. This is to the detriment of the reader because even when sufficient context is provided in the prompt, the chatbot often provides only a single meaning/sense of a given word. The chatbot sometimes misunderstands the context provided in the prompt and could then easily provide an incorrect meaning. If then queried through a follow-up prompt why it chose a specific meaning, it could not provide any explanation. This causes doubt in the mind of the user and diminishes the user-friendliness and usefulness of the chatbot.

2. Empirical work

In this section, we provide a brief outline of our experimental methodology and the prompts that were used in the experiments.

2.1 Methodology

The experiments were done during March and April 2025. It was evidently not feasible to test and evaluate all chatbots, as the purpose of this paper is to show tendencies and development of chatbot capabilities (compared to 2024). The selection of GPT o3-mini was based on an AI Overview of the different versions of ChatGPT in Google that stated: "o3 and o3-mini: These are reasoning models designed for solving complex problems, especially in domains like research, strategy, coding, math, and science. They are good for structured, logic-driven thinking and layered reasoning" (March 2025). The choice for DeepSeek was because it was recently released, and according to popular reviews, was as powerful as or even more so than OpenAI's ChatGPT and other equivalents.

The choice of examples was based on previous experiences of the first author of this article with ChatGPT-3.5 and 4 during 2024, in which the chatbot "hallucinated" (i.e., provided incorrect answers), and when it was pointed out that the answers were incorrect, it could not provide any rationale or reasoning for those answers. In some of the current experiments, we used modified versions of two of the prompts of the 2024 experiments, supplemented with further examples to test the scope of the chatbots' ability to analyse prompts and reason.

In the examples and analyses below, GPT o3-mini provided the full analysis directly in its answer. DeepSeek, however, first provided its full thinking process (in a somewhat lighter font, but still easily accessible to the user), and then gave its detailed answer, based on its preceding thinking processes. Both chatbots also provided the number of seconds they "thought/reasoned" about a problem, in the current examples between 28 and 106 seconds.

2.2 Prompts

The full prompts we used to interact with the chatbots are given below and discussed in detail in Sections 2 and 3 of this article.

2.2.1 Prompt 1

"When asked about the meaning of the word 'Barbarossa' in the quoted text below, ChatGPT earlier suggested that it referred to the German army and their Barbarossa campaign in WWII. This was evidently wrong. The text reads: "'The Redbeard's German army approaches our northern borders and he offers me twenty thousand dinars!' He threw the scroll aside, where Imad al-Din picked it up and gazed at it earnestly, as if close scrutiny might reveal some previously concealed zeros. The German army. At the door of the war tent, Malek's heart dropped like a stone. Amid the glee of their recent successes, he had forgotten about Barbarossa and his advancing horde." If you disagree with ChatGPT's earlier explanation, why do you do so? What would be the correct meaning and why do you say so? Please provide references.

Prompt 1 tested whether the chatbot could analyse the example in its historical context, and identify a plausible historical context based on the available keywords. It also tested whether it could argue why one of the potential historical contexts is more likely than the other.

2.2.2 Prompt 2

"When asked about the meaning of the phrase 'Timothy's Book' in the quoted text below, ChatGPT earlier suggested that it referred to 'the first letter of Paul to Timothy, commonly known as 1 Timothy'. This was evidently wrong. The text reads: 'He always paused at the tomb of Prior Philip. It was the largest monument there. Philip, a twelfth-century monk, was a figure of legend, though not much was known about him. According to Timothy's Book, a history of the cathedral started in the Middle Ages and added to later, Philip had organized the rebuilding of the cathedral after it was destroyed in a fire.' If you disagree with ChatGPT's earlier explanation, why do you do so? What would be the correct meaning and why do you say so? Please provide references."

Prompt 2 tested whether the chatbot could analyse the example in a fictional historical context and argue from the formulation in the text of the novel, why the previous answer was illogical.

2.2.3 Prompt 3

"What are the different meanings of the word 'level'? Provide example sentences to illustrate the different meanings and include references."

With this prompt, we tested to what extent the chatbot could generate a valid dictionary article (or, at least, the outline of such an article) and illustrate each meaning with example sentences.

2.2.4 Prompt 4

"Can 'borrow' be used as a double transitive verb?" and "'Borrow' is often used as a double transitive verb in South African English. Is this an acceptable variant, or simply wrong?"

This prompt tested both senses and syntactic features of a word, and we tried to establish whether the chatbot could distinguish between regional variations in the use of words. For this example, we used Google AI overview (based on Gemini).

2.2.5 Prompt 5

"Provide a list of GenAI policies for research, teaching and learning in iSchools across the world (with references) and summarise the content of these policies"

and "Please provide a similar list and summaries for iSchools in the Global South, again with examples and references."

This is evidently not a text reception or purely lexicographic example. However, it does establish to what extent the chatbot is able to analyse a number of different, and typically unrelated, concepts to formulate a correct answer.

2.2.6 Prompt 6

"The sentence 'Time flies like an arrow' has multiple syntactic and semantic interpretations. List the different interpretations with, in each case, an analysis of the part of speech of each word in the sentence, together with references. Which is the most likely interpretation in your opinion?"

This is a complex problem with potentially multiple interpretations due to the ambiguity of the sentence. It requires the chatbot to use the potential meanings and morphological analyses of all the words in the sentence, based on linguistic and lexicographic criteria, to establish the most probable meaning of the ambiguous sentence.

Full details of the GPT o3-mini and Deepseek answers for this prompt are provided in the Addendum, available online (click [here](#)).

2.3 Examples of human-like communication

It is well known that chatbots communicate with users in narrative format, emulating human language characteristics. In this section, we provide examples of such characteristics under different headings.

2.3.1 Informal communication

Both chatbots in the experiments sometimes used informal language to create the impression of human communication, such as in the fairly extensive use of words such as "alright," "okay," "right," by both chatbots, as well as "Interesting!" and "Absolutely!" by DeepSeek.

2.3.2 Analysing prompts

The chatbots provided the users with an idea of how they went about analysing the problem and arriving at an answer, as in the following examples:

"Alright, let's break this down. [...] I need context from historical references [...]," "I initially thought about the WWII context, but no, that doesn't seem to fit," "This makes sense given the historical context of the narrative. I think I misunderstood earlier," "I need to provide sources," "I could refer to general works [...], but I must avoid fabricating references," "I remember several possi-

ble interpretations," "The sentence 'Time flies like an arrow' is famously ambiguous! Let's break it down." (GPT o3-mini)

"Okay, let's tackle this question [...] So I need to figure out why that answer was wrong and what the correct interpretation is," "Let me recall," "Let me read the quoted text again carefully," "The key here is [...]," "I need to explain [...]," "I need to structure all these aspects without overwhelming the user," "Let me make sure I cover every possible angle here." (DeepSeek)

2.3.3 Use of humour

DeepSeek sometimes included humorous statements in its answers, in the following examples:

"Buckle up — this will be a granular journey. 🤖," "Let me know if you'd like to venture into string theory next ... or maybe just another sentence! 😊" (DeepSeek)

No such examples were observed in the current tests with GPT o3-mini.

2.3.4 Identifying contextual keywords

Chatbots identified the relevant keywords correctly, and used this in their analyses to substantiate their solutions, as in the following examples:

2.3.4.1 Prompt 1, the Barbarossa example

GPT o3-mini identified the relevant keywords correctly: "The text uses terms like 'dinar,' and names such as 'Imad al-Din' and 'Malek,' and it depicts scenes like a war tent — elements that point to a medieval or early modern Islamic context rather than the modern era."

DeepSeek also identified all the relevant keywords: "References to dinars (a historical currency), Arabic names (Imad al-Din, Malek), and a 'war tent' suggest a medieval Middle Eastern context, likely during the Crusades."

2.3.4.2 Prompt 2, the Book of Timothy example:

GPT o3-mini stated: "The passage explains that 'Timothy's Book' is a source stating that 'a history of the cathedral started in the Middle Ages ...' and goes on to recount that Prior Philip 'had organized the rebuilding of the cathedral after it was destroyed in a fire.' This language is typical of a chronicle or local historical account, not of a letter offering pastoral guidance."

DeepSeek stated: "The context is about Prior Philip, a 12th-century monk, and his role in rebuilding a cathedral after a fire. The text mentions 'Timothy's Book' as a history of the cathedral started in the Middle Ages and added to

later. So, this book is a historical record of the cathedral's history, begun in medieval times and expanded over time."

2.3.5 Lexicographic and linguistic terminology

Chatbots could contextualise prompts in the specific subject/discipline, in the following examples from Prompt 6, "Time flies like an arrow." Both bots could provide the correct parts of speech for all five words in the sentence for all possible interpretations they listed. They also used relevant grammatical terminology, such as "plural," "imperative," "3rd person singular present" etc.

DeepSeek provided a fairly detailed answer in the first iteration (answer 1), but each time suggests to the user that it could give further analyses (answers 2–4), in which it analysed the following for each of the possible interpretations:

In answer 1, for each of the three interpretations, it provided Analysis, Meaning, Structure (Syntactic Tree) and a Reference;

In answer 2, the Syntactic Structure (Sentence Type and Phrase Structure), Semantic Analysis (Meaning and Plausibility);

In answer 3, Parts of Speech & Syntactic Roles (including Syntactic Tree (X-Bar Theory)), Semantic Analysis (Formal Semantics) and Pragmatic Factors;

In answer 4, it summarised the topics, viz., "phonetics, morphology, formal syntax/semantics, neurolinguistics, and even prosody." It then provided the following detail: Phonetic & Morphological Breakdown (Phonetic Transcription (IPA) and Morphology), Syntactic Analysis via Minimalist Program, Interpretation (Derivation and Features), Formal Semantics with Lambda Calculus (logical form), and, very briefly, Pragmatics & Gricean Maxims (Maxim of Relevance and Maxim of Manner), Psycholinguistic Processing (Eye-Tracking Predictions), Neurolinguistic Correlates, Computational Parsing Models, Historical & Cross-Linguistic Analysis, Prosodic Disambiguation, Quantum Linguistics (with a question mark and a smiley) and Final Conclusion at the Planck Scale. It concluded with the remark: "Let me know if you'd like to venture into string theory next ... or maybe just another sentence! 😊"

2.3.6 References

In each of the prompts, we requested the chatbot to provide references. We checked that the references actually existed and were relevant; in contrast with earlier bots, they did not "hallucinate," and the references existed. In some cases, the references were fairly broad and did not refer to the actual example, but provided more detail about the "setting" of their answer in a broader context, as is evident from the analyses of prompt 6, where the chatbots provided references to a number of well-known linguistics textbooks, or historical reference sources, in the case of prompts 1 and 2 (DeepSeek); GPT o3-mini tended to use more popular sources, e.g., to *Wikipedia* (in the case of prompt 1), and *Encyclopaedia Britannica* (as well as a specialist history book) (for prompt 2).

For the less successful (prompts 3 and 4) and unsuccessful (prompt 5) examples, there were some irrelevant and "hallucinated" sources, as detailed in Sections 3.2 and 3.3.

3. Discussion of examples

In this section, we discuss the successful, less successful and unsuccessful examples of chatbot consultations.

3.1 Successful

Both chatbots dealt successfully with prompts 1, 2 and 6.

Both chatbots were remarkably successful in analysing prompts 1 and 2. Both were able to identify relevant keywords in both examples and use these keywords to guide their interpretations. They argued convincingly that the keywords in prompt 1 pointed to a medieval setting, and not to a modern setting, which excluded the Barbarossa campaign of World War II. In prompt 2, both pointed out that the context of the Book of Timothy excluded a biblical interpretation, viz. a reference to the Pauline letter of 1 Timothy. For both prompts, both chatbots were therefore able to place the texts in a correct period and point out that the later interpretation (World War II) and earlier interpretation (a biblical text) were contextually not correct.

3.1.1 Prompt 1

GPT o3-mini provided the following summary answer for prompt 1:

"The earlier explanation was incorrect because it ignores several key contextual clues in the quoted text. Here's why:

"1. Historical Setting:

"The text uses terms like "dinar," and names such as "Imad al-Din" and "Malek," and it depicts scenes like a war tent — elements that point to a medieval or early modern Islamic context rather than the modern era. In contrast, Operation Barbarossa (the German invasion of the Soviet Union in 1941) is a 20th century event, making that explanation anachronistic.

"2. Meaning of "Barbarossa":

"The term "Barbarossa" comes from the Italian *barba rossa*, meaning "red beard." This nickname was historically associated with figures such as Frederick Barbarossa, the 12th century Holy Roman Emperor known for his red beard. In literary or historical texts set in medieval times, using "Barbarossa" would more likely evoke the memory or legend of such a figure rather than referring to a modern military campaign.

"3. Narrative Distinction:

Notice that the text separately mentions "The Redbeard's German army" and later "Barbarossa and his advancing horde." This implies that the invading force led by Barbarossa is a distinct entity from the other German army mentioned. In a WWII context, "Barbarossa" is synonymous with the Nazi operation, not as a separate leader or faction. The separation here reinforces that the name is being used in a medieval sense — as a proper noun evoking a legendary invader.

"In summary:

"Given the language and imagery, the correct interpretation is that "Barbarossa" in this text is a reference to a medieval invader — likely evoking the historical persona of Frederick Barbarossa (or a similarly legendary figure) — and not the WWII campaign. This interpretation fits the cultural and historical setting of the narrative much better than the modern association.

"For more details on the historical figures and events associated with the name, see:

- [Frederick Barbarossa – Wikipedia](#)
- [Operation Barbarossa – Wikipedia](#)"

3.1.2 Prompt 2

DeepSeek provided the following summary answer for prompt 2:

"The phrase "Timothy's Book" in the quoted text does **not** refer to the biblical First Epistle of Paul to Timothy (1 Timothy). Here's why:

"1. Contextual Discrepancy:

- The text explicitly describes "Timothy's Book" as "a history of the cathedral started in the Middle Ages and added to later."
- **1 Timothy**, by contrast, is a New Testament letter written in the **1st century AD** addressing pastoral and ecclesiastical issues. It has no connection to medieval cathedral histories or chronicles.

"2. Fictional or In-Universe Reference:

- The passage appears to be from **Ken Follett's *The Pillars of the Earth***, a historical novel centered on the construction of a medieval cathedral. In this context, "Timothy's Book" is a fictional chronicle created by Follett to serve as an in-universe historical record.
- The novel often references fictional documents to ground its storytelling in a sense of historical authenticity. For example, Prior Philip (a key character) is depicted as a legendary figure whose deeds are recorded in this fictional "Timothy's Book," which spans generations of cathedral history.

"3. Why ChatGPT's Initial Answer Was Incorrect:

- ChatGPT likely conflated the name "Timothy" with the biblical epistle, missing the **medieval and fictional context** of the text. The mention

of a cathedral's history being compiled over centuries (post-12th century) rules out a 1st-century biblical text.

"4. References:

- In *The Pillars of the Earth*, Follett uses fictional chronicles and manuscripts to enrich the narrative. "Timothy's Book" is one such device, akin to the *Kingsbridge Chronicles* (the series to which the novel belongs).
- The novel itself is set in the 12th century, and the fire/reconstruction of the cathedral aligns with historical events like the real-life destruction of medieval churches (e.g., the 1174 Canterbury Cathedral fire).

"Conclusion:

"'Timothy's Book' is **not a biblical text** but a fictional medieval chronicle within the novel's universe, documenting the cathedral's history. This distinction highlights the importance of **context** when interpreting references in historical or literary works."

It is evident that the reasoning and argumentation are similar to the thinking and reasoning processes of a human who is interested in the historical details of the text.

3.1.3 Prompt 6

The discussions of Prompt 6, about the different meanings of the sentence "Time flies like an arrow," were equally very successful. Both bots analysed the sentences in detail, providing part of speech, syntactic and semantic analyses. Both also provided three different possible interpretations, adapting the part of speech, syntactic and semantic analyses to reflect the specific potential interpretation. They furthermore provided a motivation for their specific choice and motivated this in detail. Both chatbots provided a number of relevant references, all of which existed (i.e., no "hallucinations").

The full answers of both chatbots are provided in the Addendum, available online (click [here](#)), and make for quite interesting reading.

GPT o3-mini introduced and discussed its reasoning process as follows:

"The sentence 'Time flies like an arrow' is famously ambiguous! Let's break it down.

"Interpretation 1: It can mean that time quickly passes, much like an arrow's flight.

"Interpretation 2: 'Time flies' might refer to insects (as a noun), with 'like an arrow' being a comparison.

"There are multiple ways to interpret the parts of speech here: 'flies' as a noun (the insects) or verb (the action), and 'like' as a comparison. It's a classic case of ambiguity.

"I remember several possible interpretations."

This is followed with a detailed discussion of more than 800 words. For a full transcript, see the Addendum, available online (click [here](#)).

At the end of each answer for this prompt, DeepSeek asked whether further, more detailed analyses were required, and provided further details that may potentially be relevant for a specialist, or, citing some of the potential motivations and contents.

DeepSeek introduced and discussed its reasoning process as follows:

Answer 1:

"Okay, so I need to figure out the different interpretations of the sentence 'Time flies like an arrow.' I remember that this sentence is often used as an example of syntactic ambiguity. Let me start by breaking down each word and considering their possible parts of speech;" "First, the straightforward interpretation is probably 'Time passes quickly, similar to how an arrow flies.';" "But the user mentioned there are multiple interpretations. Let me think about other possible ways to parse this;" "Wait, another possibility ...;" "Wait, this is getting complicated;" "But that's a stretch ...;" "I need to make sure I'm not missing any other interpretations. Let me think again."

Answer 2:

"The user might be interested in the linguistic theories behind ambiguity. Chomsky's deep vs. surface structure comes into play here. The same surface sentence has different deep structures based on parsing. Radford's analysis of syntactic ambiguity would support this. Also, mentioning lexical ambiguity with homonyms is important;" "Also, consider the user's possible deeper needs. They might be studying linguistics, preparing for an exam, or just curious about language. By elaborating, I should provide enough detail to satisfy academic curiosity but keep it understandable. Avoid jargon where possible, or define terms if used," "Let's delve deeper into the syntactic and semantic ambiguities of 'Time flies like an arrow, expanding on linguistic theories, structural analyses, and contextual plausibility."

Answer 3:

"Okay, the user has already asked for a more granular analysis after I provided a detailed breakdown. They seem really interested in the linguistic aspects of the sentence 'Time flies like an arrow. Let me make sure I cover every possible angle here," and "Also, are there other possible interpretations I haven't considered? Maybe 'like' as a conjunction, though that's less likely. Or 'arrow' as a modifier? Probably not. But I should check;" "I need to make sure each analysis is thorough, referencing linguistic theories, providing syntactic trees with proper notation, and discussing semantic plausibility. Also, check if there are any other references or studies that

discuss this specific sentence — maybe classic examples in ambiguity literature."

Answer 4:

"Let me think about how to go deeper. First, maybe I should look into phonetics and morphology. But the sentence is in English, and the words are straightforward." It then briefly mentions "homonyms and their phonetic representations," "syntactic theories beyond X-Bar. Maybe using Minimalist Program structures, discussing movement or feature checking;" "Semantic roles could be expanded. Theta roles were mentioned;" "Pragmatic aspects could include Gricean maxims in more depth;" "computational linguistics;" "psycholinguistics;" "historical linguistics;" "morphological analysis;" "lexical semantics;" "neurolinguistics;" "prosody." It concludes its reasoning as follows: "I need to structure all these aspects without overwhelming the user. Organize them into sections with clear headings, use linguistic notation where applicable, and reference theories and studies for each point. Ensure that each detail adds value and connects back to the main interpretations. Check for coherence and avoid redundancy from previous answers. Highlight how each tiny detail contributes to the overall understanding of the ambiguity."

The four answers of DeepSeek result in a detailed discussion of approximately 6,000 words. For a full transcript, see the Addendum, available online (click [here](#)).

3.2 Partially successful

The answers to prompts 3 and 4 were less successful, primarily due to poorly structured groupings and "hallucinations" in the answer to prompt 3, and inconsistencies in the answers to prompt 4.

3.2.1 Prompt 3

In prompt 3, the chatbots were required to construct the equivalent of a dictionary article for the meanings/senses of a single word, *level*, as well as provide references.

GPT o3-mini introduced its answer as follows: "The word 'level' is a versatile term in English, used in different contexts with varying meanings. Below are the main meanings of 'level', each with examples and references." In each case, an example paraphrase or sentence from a dictionary is provided, in this case, the *Oxford English Dictionary* (OED), *Merriam-Webster*, *Cambridge Dictionary*, *Collins English Dictionary* and *Longman Dictionary*. The examples are in all cases provided in quotation marks, which creates the impression that they are taken verbatim from the specified dictionary, but in none of the examples this

is the case — chatbot in each case provided a paraphrase. In addition, the proper dependencies between meanings/senses (as is evident in all the cited dictionaries) were not maintained. In this case, the chatbot is therefore a less dependent source of meanings/senses.

The chatbot also offered the option that it could provide examples from only one specific dictionary: "Let me know if you'd like definitions from a specific dictionary like OED or Merriam-Webster with direct quotes." We asked it to provide an answer from only the OED. None of the examples in this answer, however, were direct quotes from the OED — unless it referred to one of the other dictionaries of English published by Oxford, and not the "*The historical English dictionary*," typically known as "*Oxford English Dictionary*," and abbreviated as "OED" (<https://www.oed.com/>).

This is a less successful answer than could be obtained by typing "level meaning" in a Google search bar, which provides a comprehensive article with examples, synonyms etc.

3.2.2 Prompt 4

Prompt 4 (a query in the Google search bar, the answer of which is an AI Overview based on Gemini) is included to highlight confusion which sometimes occurs in GenAI answers. We asked whether the verb "borrow" could be used as a double transitive verb. In a first answer, it said: "Yes, the verb 'borrow' can be used as a double transitive verb, also known as a ditransitive verb." The rest of the explanation and examples seemed counterintuitive. We therefore repeated the question about an hour later (to check consistency in its responses), and the answer it gave was: "No, the verb 'borrow' is not typically used as a double transitive verb. While 'borrow' does take a direct object (what is being borrowed), it doesn't usually take an indirect object (to whom or for whom)." Since "borrow" is often used as a double transitive verb in (informal) South African English (i.e., as an equivalent of "lend"), we asked whether this usage is an acceptable variant in South African English, or simply a mistake. The answer read as follows: "While the use of 'borrow' as a double transitive verb (e.g., 'Can I borrow you the car?') is common in South African English, it's generally considered a nonstandard usage in formal English. It's often a loan translation from Afrikaans, where 'leen' (lend) and 'leen by' (borrow from) are used in a way that translates directly to 'borrow' in the double transitive construction. While it's widely understood and accepted in informal contexts, it's best avoided in formal writing or speech." This is the expected answer (which may confirm the authors' biases). Interestingly, this possible meaning is not listed in the *Dictionary of South African English* (DSAE, <https://www.dsae.co.za/>). The correct/incorrect reply from Google Overview AI highlights the importance of evaluating any answer GenAI provides, and of triangulating it, especially if it is counterintuitive.

3.3 Unsuccessful

Prompt 5 is not a lexicographic example, and is included to show that, even though chatbots have improved considerably since earlier versions, they still hallucinate if their training data does not contain the required data to answer the question.

3.3.1 Prompt 5

We used two prompts for this question. In the first prompt, we required the chatbot, in this case, DeepSeek, to be able to identify the different iSchools in the world (see the iSchools website at <https://www.ischools.org/>) and then find on the websites of the iSchools any policies for the use of GenAI in research, teaching and learning. The answer DeepSeek provided seemed to be correct, but very general. It provided references from 2023, however, without URLs to enable the verification of the existence of the sources, or more information. Some of these documents could be found, but in most cases, documents with the titles that were listed, were not available. It also provided a number of references under the heading "Hypothetical References," which is very strange.

We also needed information on the same topic, restricted to iSchools in countries in the Global South. This would imply that DeepSeek should be able to understand which countries the Global South consisted of, to match this list with the list of iSchools worldwide and then find the relevant policies. It was unfortunately not successful at all. We knew that, at that stage, there were four iSchools in Global South countries, one each in Brazil and Indonesia, and two in Africa. DeepSeek identified six iSchools in Global South countries, five of which were wrong and only one which was correct. It again provided incorrect or non-existent references and again added "Hypothetical References." The discussions were very general and none of the content could be verified.

These two examples again highlight the importance of evaluating and triangulating any answer GenAI provides.

4. GenAI versus dictionaries

The preceding discussion and examples show that GenAI cannot fully replace "conventional" dictionaries — at this stage. Access to actual dictionary sources provides better assistance, e.g., a search in online dictionaries that are available via open access, or even a Google search that provides "Definitions from Oxford Languages," results in retrieving better and more comprehensive definitions, as well as additional lexicographic information on demand. Currently, GenAI cannot be regarded as an authoritative lexicographic source, even when it claims to cite online dictionaries (as in prompts 3 and 4); it can, however, be an additional information tool for more complex information needs, as in the cases of

prompts 1, 2 and 6. In addition, it still "hallucinates" if it does not have an answer, as in the case of prompt 5. The potential unreliability of any GenAI-generated answer again emphasises the importance of evaluating any such answers and triangulating the results with other reputable (or curated) sources.

GenAI-enhanced tools are capable of responding to lexicographic information needs and to combine the obtained data with additional resources, constituting a network of information resources, as is evident in prompts 1, 2 and 6.

Irrespective of the advantages and disadvantages of AI, one has to accept that AI and GenAI are here to stay, at least for the foreseeable future. One cannot ignore these developments and should rather strive to understand what it can and cannot do in order to embrace it intelligently and ethically.

5. Suggestions for lexicographic research

5.1 Traditional lexicographic research

A focus in "traditional" dictionary research has always been, and continues to be, important, as dictionaries can best document and display the status and history of a language. This is witnessed in dictionaries such as the OED (*Oxford English Dictionary*, <https://www.oed.com>), WAT (*Woordeboek van die Afrikaanse Taal*, <https://www.wat.co.za>), the DSAE (*Dictionary of South African English*, <https://www.dsae.co.za>) and many others. These endeavours need to be encouraged, especially in environments with limited language resources. The lexicography of the African languages in South Africa could also benefit from such an approach. The dynamic nature of language demands that dictionary definitions should continuously be refined.

No single dictionary can be everything for everyone, but in the absence of an established typological infrastructure, existing dictionaries should increase their scope beyond the traditional typological boundaries so that they can present a wider linguistic coverage to their target users. One option the lexicographer has, is to plan and compile multifunctional online dictionaries with added functionalities, e.g., language-learning exercises, lexicotainment, etc. They could also offer access to additional lexicographic and non-lexicographic data, and/or links to such data, even to data in dictionary-external sources. This demands the appropriate interfaces and interface design that allow the filtering of data as well as drilling down and retrieving information on demand. Interface design is an exceptionally important aspect for the development of a successful online dictionary, and typically requires a number of specialists to collaborate, including lexicographers, UX (user experience) designers, programmers/computer scientists, information scientists and a number of others, as well as users to test the efficiency of the proposed interface.

When planning the interface design, it would be helpful to be reminded of Ben Shneiderman's Visual Information-Seeking Mantra (1996): "Overview first, zoom and filter, then details on demand." This should be understood as:

- "Overview: Gain an overview of the entire collection"
- "Zoom: Zoom in on items of interest"
- "Filter: Filter out uninteresting items"
- "Details-on-demand: Select an item or group and get details when needed"

To this could be added "Relate", "History" and "Extract", realised as:

"Relate: View relationships among items."

"History: Keep a history of actions to support undo, replay, and progressive refinement."

"Extract: Allow extraction of sub-collections and of the query parameters."

These aspects of interface design will not be discussed in this paper. Existing online dictionaries already offer an interface design that can guide the user within a comprehensive article to both a number of restricted articles, cf. Gouws (2022: 120), in which various aspects of the treatment of the lemma are presented or to dictionary-external data like other dictionaries in the same portal, or even to Google and other search engines. An option that modern-day lexicographers should consider is to integrate access to a chatbot, enabling dictionary users to retrieve complementary information to enhance the quality and extent of their dictionary consultation experience. This integration of GenAI and lexicography requires, among others, the appropriate interface design and data distribution structure.

5.2 Lexicographic research and technology

Lexicographic databases, their compilation and use are focal areas in modern-day lexicographic research. These databases and the lexicographic products relying on them can be improved with more attention paid to the concepts of granularity and (re)usability. Granularity is required for the fine-grain filtering and display of lexicographic data, whereas (re)usability should play a pertinent role in deriving different lexicographic products, including dictionaries for specific user groups from the data on offer. Different applications, e.g. translation tools, writing assistants and tools that help users to improve their writing can employ the same data but applied in different ways.

Lexicographic data becomes the gold standard in the planning and compilation of lexicographic products, and the data assessment, distribution and presentation need to be done in such a way that different target groups can use different tools to retrieve the information they require in specific extra-lexicographic situations. The optimal use of these data can be improved if lexicographers, AI experts and other collaborators can successfully explore and license new commercial ventures, incorporating the latest technologies, including AI and GenAI. Such ventures could also elevate the level of contextualized linking for text reception.

6. Conclusion

Based on the limited number of examples that have been examined, and contrary to earlier research, it seems as if the latest versions of GenAI chatbots can often provide correct contextual meaning. It can logically motivate the choice of meaning in context, based on its critical analysis and thinking skills, typically associated with humans.

However, the reliability of the results of GenAI answers varies, covering the full spectrum from very reliable to complete hallucinations. A mixture of reliability criteria prevails. As a result, the reader remains responsible to evaluate the answers. In case of uncertainty, they need to check other resources in order to verify and triangulate the results.

When expanding the information retrieval structure of a dictionary to allow users access from any point in a given dictionary to one or more chatbots, lexicographers need to make their users aware of the fact that GenAI provides non-curated data that might not always live up to the correctness, appropriateness and validity of the information they can retrieve from the curated data in dictionaries.

However, good lexicography demands interdisciplinary collaboration. An interactive relation between lexicographers and developers of GenAI could be beneficial for both groups. Lexicography could benefit by improved versions of GenAI becoming sophisticated lexicographic information tools and GenAI could benefit from the expertise in lexicography regarding the user-perspective, information needs, lexicographic functions, appropriate contents and relevant structures to distribute and access data.

7. Addendum

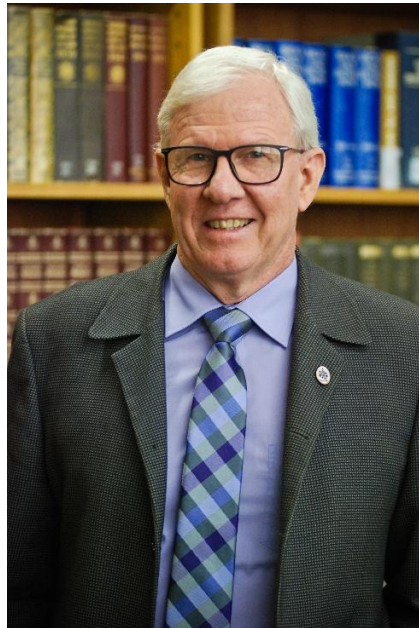
The full answers of both chatbots for prompt 6 are provided in the Addendum, available online (click [here](#)).

8. References

- Bothma, T.J.D. and I. Fourie.** 2024. 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.
<https://publicera.kb.se/ir/article/view/55127> (accessed 15 June 2025)
- Bothma, T.J.D. and I. Fourie.** 2025. Contextualised Dictionary Literacy, Information Literacy, and Information Behaviour in the e-Environment. *Library Management* 46(1-2): 14-28.
<https://doi.org/10.1108/LM-08-2023-0082>
- Bothma, T.J.D. and 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.
<https://doi.org/10.5788/32-2-1697>

- De Schryver, G.-M.** 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.
<https://doi.org/10.1093/ijl/ecad021>
- Fuertes-Olivera, P.A.** 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34: 123-140.
<https://doi.org/10.5788/34-1-1883>
- Germain, T.** 2025. Is Google about to Destroy the Web? *BBC*, 13 June 2025.
<https://www.bbc.com/future/article/20250611-ai-mode-is-google-about-to-change-the-internet-forever> (accessed 15 June 2025)
- Gouws, R.H.** 2022. Extended Article Comments in Online Dictionaries. *Lexicography. Journal of ASIALEX* 9(2): 111-138.
<https://doi.org/10.1558/lexi.24289>
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*: 102-111. Berlin: De Gruyter.
- Huete-García, Á. and S. Tarp.** 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34: 21-40.
<https://doi.org/10.5788/34-1-1862>
- Lew, R.** 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11(1): 1-8.
<https://doi.org/10.1057/s41599-024-02889-7>
- Shneiderman, B.** 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages, Boulder, CO*: 336-343.
<https://doi.org/10.1109/VL.1996.545307>
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim: Bibliographisches Institut.

Frederik Johannes Lombard 1953 – 2025



Dit was met groot leedwese dat die personeel van die Woordeboek van die Afrikaanse Taal (WAT) verneem het van die afsterwe van 'n geliefde kollega en vriend, dr. Frikkie Lombard, op 19 April 2025.

Dr. Frederik Johannes Lombard, of Frikkie soos ons hom almal geken het, het op 1 April 1982, ná drie jaar as onderwyser by Paul Roos Gimnasium, as mederedakteur by die WAT in diens getree. Hy het van vroeg af 'n passie vir taal, boeke en nuwe woorde en uitdrukkings gehad en 'n loopbaan in die leksikografie was nie 'n moeilike keuse nie.

Op 1 April 1993 is Frikkie tot senior mederedakteur bevorder. Dié pos het hy beklee tot hy op 1 Januarie 2001 in die pos van eindredakteur aangestel is. As daar 'n eienskap ten opsigte van woordeboekmaak is waaroor Frikkie by uitstek beskik het, was dit sy vlymskerp insig om betekenis akkuraat, helder en bondig te verwoord. Verder kon hy sitate moeiteloos knip en rangskik sodat hulle die definisie netjies ondersteun. Vanweë sy breë algemene kennis het die mederedakteurs hom met groot vrymoedigheid geraadpleeg — veral wanneer dit oor geskiedenis, weermagsake, fotografie, motor- en motorfietsterme, musiek of sport gehandel het — en almal is altyd met deernis en begrip behandel. Hy

het ook oor die jare as 'n mentor vir jonger kollegas opgetree en hulle altyd met die grootste geduld in die proses van manuskripmaak en -redigering begelei. Hy het boonop nie gekroom om erkentlikheid te bewys waar dit geregverdig was nie.

Vir 25 jaar, van 1995 tot 2020, het Frikkie in die Taalkommissie (TK) gedien, die eerste vyftien jaar as sekretaris en by geleentheid as waarnemende ondervoorsitter. As een van die langsdienende TK-lede het hy 'n betekenisvolle bydrae daartoe gelewer om die Afrikaanse spelreëls te help standaardiseer. Hiervoor is hy ook deur die Suid-Afrikaanse Akademie vir Wetenskap en Kuns met twee toekennings vereer.

Die onderwys het Frikkie ook altyd ná aan die hart gebly. Hy het sy doktorsgraad verwerf met die proefskrif '*n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke*. Frikkie het dan ook uiteindelik die geleentheid gekry om van hierdie teoretiese beginsels in die praktyk toe te pas deur mee te werk aan twee skoolwoordeboeke vir Oxford University Press, naamlik die eentalige *Oxford Afrikaanse Skoolwoordeboek* en die tweede uitgawe van die *Oxford Afrikaans-Engels English-Afrikaans Skoolwoordeboek School Dictionary*. Beide woordeboeke is kommersieel baie suksesvol en lewer steeds 'n belangrike bydrae om met die leesbegripkrisis in die Suid-Afrikaanse skoolstelsel te help.

Wat die werk aan die WAT betref, was Frikkie betrokke by tien van die sestien boekdele en ook die T-trajek wat slegs aanlyn bekendgestel is. Die laaste WAT-artikel wat Frikkie kort voor sy aftrede in Oktober 2023 vir redigering onder oë gehad het, was dié van die lemma **trek**, 'n polisemiese lemma met nie minder nie as 149 betekenisonderskeidings. In daardie stadium het ons nie gedink dat dit profeties sou wees nie. Vir sy aftrede het **trek** potensieel beteken "van huis, woning, kantoor of werkplek verander" en ook "'n tog wat met onderbrekings gepaard gaan, aanpak of onderneem". Dit was hom egter nie beskore nie. Uiteindelik het dit beteken dat hy finaal moes "versit of verskuif".

As mens buite die werkplek was sy liefde vir sy familie en vriende 'n inspirasie. Hy was baie mededeelsaam en sou altyd, wanneer hy iets as geskenk gekry het wat hy met ander kon deel, nie gekroom het om dit te doen nie. As grap- en storieverteller het Frikkie ook uitgeblink. Hy het 'n hele arsenaal grappies en anekdotes gehad waarmee hy vriende en kollegas vermaak het. Hy het 'n skerp humorsin en 'n ongelooflike geheue gehad en kon 'n storie op 'n baie interessante manier en in die fynste detail vertel.

Die personeel van die WAT het net die hoogste waardering vir die 42 jaar wat Frikkie Lombard die WAT met onderskeiding gedien het. Sy menslikheid, kollegialiteit en humorsin sal baie gemis word.

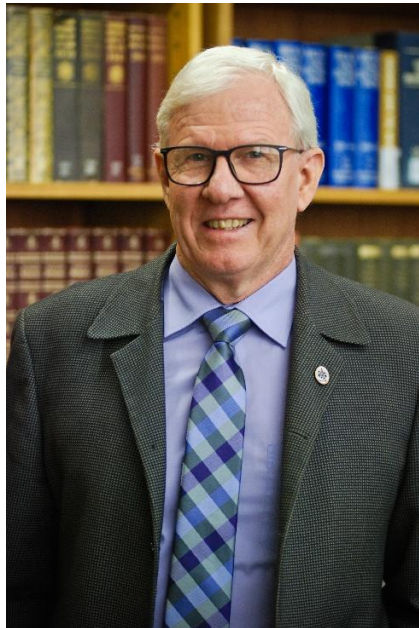
Alet Cloete
Woordeboek van die Afrikaanse Taal
Stellenbosch
Suid-Afrika
(aec@sun.ac.za)

Phillip Louw
Woordeboek van die Afrikaanse Taal
en
Departement Afrikaans en Nederlands, Universiteit Stellenbosch
Stellenbosch
Suid-Afrika
(pal@sun.ac.za)
(<https://orcid.org/0009-0007-2002-3081>)

en

Gerhard van Wyk
Woordeboek van die Afrikaanse Taal
Stellenbosch
Suid-Afrika
(gjvw@sun.ac.za)

Frederik Johannes Lombard 1953 – 2025



It was with great sadness that the staff of the *Woordeboek van die Afrikaanse Taal* (Dictionary of the Afrikaans Language, henceforth the WAT) learned of the passing of a beloved colleague and friend, Dr Frikkie Lombard, on the 19th of April 2025.

Dr Frederik Johannes Lombard, or Frikkie as we all knew him, joined the WAT as co-editor on the 1st of April 1982, after three years as a teacher at Paul Roos Gymnasium. From an early age, he had a passion for language, books and new words and expressions, and a career in lexicography was not a difficult decision.

Frikkie was promoted to senior co-editor on the 1st of April 1993. He held this position until he was appointed as final editor on 1 January 2001. If there was one quality in dictionary-making at which Frikkie excelled, it was his razor-sharp insight into expressing meaning accurately, clearly and concisely. Furthermore, he could effortlessly cut and arrange example sentences to neatly support the definition. Because of his broad general knowledge, co-editors consulted him with great confidence — especially concerning history, military affairs, photography, motoring and motorcycle terms, music or sport — and everyone

was always treated with compassion and understanding. Over the years, he also acted as a mentor to younger colleagues and guided them through the lexicographic process with the utmost patience. Moreover, he did not hesitate to show gratitude where it was warranted.

Frikkie served on the Taalkommissie (the Language Commission, hereafter TK) for 25 years, from 1995 to 2020, with the first fifteen of those years as secretary and occasionally as acting vice-chairman. As one of the longest-serving TK members, he made a significant contribution to the standardisation of Afrikaans spelling rules. For this, he was honored with two awards by the Suid-Afrikaanse Akademie vir Wetenskap en Kuns (South African Academy for Science and Arts).

Education also remained close to Frikkie's heart. He obtained his doctorate with the dissertation *'n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke* (A Metalexicographical Foundation of Afrikaans School Dictionaries). Frikkie later had the opportunity to apply some of these theoretical principles in practice by collaborating on two school dictionaries for Oxford University Press, namely the monolingual *Oxford Afrikaanse Skoolwoordeboek* (Oxford Afrikaans School Dictionary) and the second edition of the *Oxford Afrikaans–English English–Afrikaans School Dictionary*. Both dictionaries are commercial successes and continue to make an important contribution toward addressing the reading comprehension crisis in the South African school system.

As for the work on the WAT, Frikkie was involved in ten of the sixteen printed volumes as well as the stretch of the letter T, which was exclusively published online. The last WAT article edited by Frikkie shortly before his retirement in October 2023, was that of the lemma **trek**, a polysemic lemma with no less than 149 senses. At the time, we did not think that it would be prophetic. For his retirement, **trek** potentially meant "to change house, residence, office or workplace" as well as "to embark on or undertake a journey that involves interruptions". However, this was not meant to be his destiny. In the end, he had to "move or be moved" for good.

Outside of the workplace, his love for his family and friends was an inspiration. He was very generous and when he received something as a gift that he could share with others, he would always do so without hesitation. Frikkie also excelled as a joker and storyteller. He had an arsenal of jokes and anecdotes with which he entertained friends and colleagues. He had a sharp sense of humour and an incredible memory, and he could tell a story with remarkable detail and captivating flair.

The WAT's staff have only the highest appreciation for the 42 years that Frikkie Lombard served the WAT with distinction. His humanity, collegiality and sense of humour will be greatly missed.

Alet Cloete
Woordeboek van die Afrikaanse Taal
Stellenbosch
South Africa
(aec@sun.ac.za)

Phillip Louw
Woordeboek van die Afrikaanse Taal
and
Department of Afrikaans and Dutch, Stellenbosch University
Stellenbosch
South Africa
(pal@sun.ac.za)
(<https://orcid.org/0009-0007-2002-3081>)

and

Gerhard van Wyk
Woordeboek van die Afrikaanse Taal
Stellenbosch
South Africa
(gjvw@sun.ac.za)

Thematic Classification in Polish Electronic Dictionaries — From Tradition to New Tools and Opportunities*

Barbara Batko-Tokarz, *Philological Faculty,
Jagiellonian University, Krakow, Poland*
(barbara.batko-tokarz@uj.edu.pl)
(<https://orcid.org/0000-0001-9147-5180>)

Abstract: The main topic of this article is the thematic classification of vocabulary in two Polish dictionaries that were compiled at the beginning of the 21st century. The first is the *Wielki słownik języka polskiego PAN* (WSJP PAN) [The Polish Academy of Sciences Great Dictionary of Polish], which describes contemporary vocabulary and can function as a thematic dictionary. The second is the *Słownik pojęciowy języka staropolskiego* (SPJS) [Conceptual Dictionary of Old Polish], which collects vocabulary up to the end of the 15th century and arranges it in conceptual fields. The primarily electronic format of these dictionaries has created completely new possibilities for classification according to meaning, which is worth briefly describing in the context of the thematic lexicography used in Polish paper dictionaries to date. Therefore, the most important Polish thematic dictionaries over the centuries are listed first (Section 2), followed by a more detailed description of the SPJS and WSJP PAN (Section 3). In Section 4, the following is described based on these two Polish electronic dictionaries: the ease of switching between onomasiological and semasiological orders, the significantly greater possibilities for classifying lexical units, and the ability to search fields using advanced search functions. These considerations are important given the growing number of electronic dictionaries increasingly adopting onomasiological orders.

Keywords: ONOMASIOLOGICAL LEXICOGRAPHY, THEMATIC LEXICOGRAPHY, THEMATIC DICTIONARIES, CONCEPT DICTIONARIES, PRIMARILY ELECTRONIC DICTIONARIES, SEARCH TOOLS IN ELECTRONIC DICTIONARIES, ONOMASIOLOGICAL AND SEMASIOLOGICAL ORDER, THEMATIC CLASSIFICATION OF VOCABULARY

Zusammenfassung: Thematische Klassifikation (des Wortschatzes) in polnischen elektronischen Wörterbüchern — von der Tradition bis hin zu neuen Werkzeugen und Möglichkeiten. Der vorliegende Artikel befasst sich mit der thematischen Klassifizierung des Wortschatzes in zwei zu Beginn des 21. Jahrhunderts erarbeiteten Wörterbüchern

* This publication and its translation has been supported by a grant from the Priority Research Area (POB Heritage) under the Strategic Programme Excellence Initiative at Jagiellonian University .

der polnischen Sprache. Das erste ist *Wielki słownik języka polskiego PAN* (WSJP PAN) [Großes Wörterbuch der polnischen Sprache der Polnischen Akademie der Wissenschaften], das den zeitgenössischen Wortschatz ab dem Jahr 1945 erfasst und zugleich als thematisches Wörterbuch genutzt werden kann. Das zweite, *Słownik pojęciowy języka staropolskiego* (SPJS) [Begriffswörterbuch des Altpolnischen], dokumentiert den bis zum Ende des 15. Jahrhunderts verwendeten Wortschatz, geordnet nach Begriffsfeldern. Die primär elektronische Form beider Wörterbücher eröffnet völlig neue Möglichkeiten für die Klassifikation nach Bedeutungen. Diese Entwicklungen verdienen eine kurze Darstellung im Kontext der thematischen Lexikografie, wie sie bislang in gedruckten polnischen Wörterbüchern praktiziert wurde. Daher werden zunächst im zweiten Abschnitt die wichtigsten thematischen Wörterbücher des Polnischen im Verlauf der Jahrhunderte vorgestellt. Im dritten Abschnitt folgt eine detaillierte Charakterisierung des SPJS und des WSJP von PAN. Der zentrale vierte Abschnitt des Artikels widmet sich den verschiedenen Zugriffsformen auf das Wörterbuchmaterial in diesen beiden polnischen elektronischen Wörterbüchern. Besonders hervorgehoben wird dabei die Möglichkeit des einfachen Wechsels zwischen onomasiologischer und semasiologischer Ordnung. Beschrieben werden zudem deutlich erweiterte Optionen für die Klassifizierung des sprachlichen Materials sowie für die feldspezifische Suche mithilfe erweiterter Suchfunktionen. Diese Überlegungen gewinnen an Relevanz angesichts der wachsenden Zahl elektronischer Wörterbücher, die zunehmend auf einer onomasiologischen Ordnung basieren.

Schlüsselbegriffe: ONOMASIOLOGISCHE LEXIKOGRAFIE, THEMATISCHE LEXIKOGRAFIE, THEMATISCHE WÖRTERBÜCHER, BEGRIFFSWÖRTERBÜCHER, PRIMÄR ELEKTRONISCHE WÖRTERBÜCHER, SUCHFUNKTIONEN IN ELEKTRONISCHEN WÖRTERBÜCHERN, ONOMASIOLOGISCHE UND SEMASIOLOGISCHE ORDNUNG, THEMATISCHE KLASSIFIKATION DES WORTSCHATZES

1. Introduction

The aim of this article is to present the thematic classification of vocabulary used in two Polish dictionaries that were compiled at the beginning of the 21st century (and are only available online), as well as new possibilities for its use as offered by their electronic versions. The first of these is the *Wielki słownik języka polskiego PAN* (WSJP PAN) [The Polish Academy of Sciences Great Dictionary of Polish], which describes contemporary vocabulary from 1945 onward. The second is the *Słownik pojęciowy języka staropolskiego* (SPJS) [Conceptual Dictionary of Old Polish], which collects vocabulary up to the end of the 15th century. These dictionaries were created independently of each other, as they were developed by two different teams of lexicographers, and they differ in the types and amounts of linguistic material covered. However, they also have much in common, in that they are primarily electronic dictionaries that use both semasiological and onomasiological orders. Dictionary entries are also available thematically, as both dictionaries employ semantic classification of large, collected vocabulary groups. Due to the pioneering nature of these solutions, the authors responsible for this kind of classification in each dictionary have presented both the methodological framework and practical aspects of applying these classifications, situating

them within the broader context of Polish and European thematic or onomasiological lexicography (Batko-Tokarz 2019; Sieradzka-Baziur 2020).

Dictionaries that do not follow alphabetical but semantic order constitute a subtype of onomasiological lexicons (Hartmann and James 2002: 101-102). In European lexicography (including Slavic), the following terms are used to describe such works: *analogical*, *conceptual*, *ideographical*, *ideological*, *thematic*, *topical*, *systematic*, *semantic (dictionary)*, *nomenclator*, and *thesaurus* (e.g., Jedlička 1977: 222-223; Sierra 2000: 224; Nikolić-Hoyt 2004: 7-8, 21; Hüllen 2006: 13; França 2010: 1291-1292; Obstová, Tichý and Klégr 2022: 77).¹ There are also differences in the terminology used in both dictionaries, as the WSJP PAN uses the terms *thematic classification* or *thematic qualification*, *thematic dictionary*, *thematic zones*, *fields*, and *subfields*, while the SPJS uses the term *conceptual* in its title, *conceptual categories* in its classification scheme, *semantic unit* (Sieradzka-Baziur 2020: 46) and *semantics* in entries.

In this article, terms containing the adjective *thematic* — such as *thematic classification*, *thematic dictionary*, and *thematic lexicography* — are used as overarching in relation to other types of dictionaries and classifications that arrange vocabulary according to its semantic properties. This is firstly related to the tradition of such terminology in Polish and Slavic lexicography (although other terms are also used in this tradition).² Additionally, this decision is compliant with new lexicographic trends because, as analyses of the interest in onomasiological dictionaries at the turn of the 20th and 21st centuries indicate, the adjective *thematic* is becoming increasingly more popular due to its general meaning (Kawalya and De Schryver 2013: 178). These terms will be understood broadly (McArthur 1998; Hartmann and James 2002: 142; Burkhanov 2010: 242-243 — the first meaning);³ that is, they refer both to dictionaries with a simple structure (such as those arranging words according to specific topics important in teaching, or presenting a certain type of vocabulary) and to general language dictionaries with precisely built hierarchical structure (such as thesauruses or conceptual lexicons, as their subtypes).

SPJS and WSJP PAN are available only in electronic format, which firstly means that it is easy to switch from the order typical of semasiological dictionaries to that used in onomasiological dictionaries, and secondly, advanced search options are available. It is worth showing, using the example of these two Polish dictionaries, how the use of electronic tools makes the thematic classification used in dictionaries much more effective and functional. The electronic form of dictionaries enables a completely new way of functioning for Polish thematic lexicography, which until now has only existed in printed form. Furthermore, it eliminates many of its previous limitations and difficulties, which is why more attention to the connections between lexicography and new disciplines, especially information technology, is necessary (Hartmann 2005: 15-16).

However, it would be difficult to demonstrate this without a brief background overview. This overview includes a short description of the tradition of Polish thematic lexicography over the centuries, with particular emphasis on Polish printed dictionaries of the 20th and 21st centuries, arranged in whole or

in part thematically in line with the words: "Yet, it seems high time that non-alphabetical dictionaries should at least regain their history" (Hüllen 2006: 21). Although the tradition of lexicographic works arranged in such a way may differ depending on language, the cross-sectional approach to a specific type of dictionaries in a given country is justified. Such an approach makes it possible to identify what is universal, not only in terms of historical practices, but also in relation to future applications. Therefore, the most important Polish printed thematic dictionaries in history will be briefly presented below (Section 2), followed by a description of contemporary electronic (online) dictionaries using a thematic layout, namely the SPJS and WSJP PAN (Section 3). In the penultimate and particularly important part of this article using these two dictionaries as examples, new possibilities and tools for the thematic classification of vocabulary offered by electronic lexicography will be presented (Section 4).

2. Brief history of Polish thematic lexicography⁴

2.1 From the 16th century to the first half of the 20th century

The tradition of arranging vocabulary in dictionaries according to meaning is older than the European culture of writing — dictionaries arranged thematically rather than alphabetically constitute the foundations of European lexicography (cf. Nikolić-Hoyt 2004: 39-62; Hüllen 2006: 32-54). The beginnings of Polish lexicography are also associated with the thematic arrangement of vocabulary, which can be seen in the first printed dictionaries (with Polish materials modeled on foreign works). This tradition began with the Polish adaptation of the dictionary by Dutch philologist Johannes Murmellius (Hüllen 2006: 324-326), namely the German–Latin–Polish dictionary titled *Dictionarius Ioannis Murmelli variarum rerum* (Murmeliusz 1526), which was frequently reprinted back then. It is now believed that the Polish part was prepared by Hieronim of Wieluń (Spiczyński), but in the Polish tradition this work is referred to as the "Murmellius Dictionary". In 1528, Francis Mymer's *Dictionarium trium linguarum* was published with the same language combination (Mymer 1528). The author modeled it on a Czech dictionary from 1513 and, in part, on Murmellius' dictionary (Gruszczyński and Saloni 2013: 207-208). This type of vocabulary arrangement was also used in the so-called *Słownarz* (1532), an adaptation of the quadrilingual dictionary (Gruszczyński and Saloni 2013: 208). The last thematically organised dictionary of the 16th century was the Latin–German–Polish *Nomenclator selectissimas rerum appellationes* [...] by Petrus Artomius (Artomiusz 1591), which was published as a reworking of the Latin–German version. This became the most frequently reprinted Polish dictionary in the 17th century, though most of the editions were erroneously signed with the name of Murmellius (Gruszczyński and Saloni 2013: 209).

From the 17th century onward, the alphabetical arrangement began to dominate European lexicography, including that of Poland (Gruszczyński and Saloni 2013: 209 ff.). Thematically arranged dictionaries from the 17th to the mid-20th

centuries were rare and considered anachronistic by the lexicographic standards of the time — most often, these were glottodidactic dictionaries. However, it was then that a work of great importance for Polish thematic lexicography was created, namely, the *Dobór wyrazów. Słownik wyrazów bliskoznacznych i jednoznacznych. Do praktycznego użytku* [Selection of Words. Dictionary of Synonymous and Mono-semantic Words. For Practical Use] (Zawiliński 1926). The author based his work on the 1912 edition of Roget's *Thesaurus*. However, this dictionary was underestimated by its first users and researchers, who were unprepared for its reception and identified it merely as a dictionary of synonyms, despite its much higher quality.

2.2 From the second half of the 20th century to modern times (printed dictionaries)

Since the middle of the 20th century, there has been a revival of the tradition of classifying vocabulary based on meaning, alongside its increasing popularity in linguistic research and lexicography. Furthermore, the influence of foreign achievements in this field on Polish lexicography is evident, as is the development within the country itself. The paper dictionaries created at that time using thematic arrangements can be divided into three types.

First, they were local dialect dictionaries. In the 1960s, the *Porównawczy słownik trzech wsi małopolskich* [Comparative Dictionary of Three Lesser Poland Villages] was published (Kucala 1957). The researcher followed the scheme of R. Hallig and W. von Wartburg (1952) but adapted it to the collected vocabulary. This dictionary gained recognition in Polish and Slavic dialectology. Although the lexicography of dialects departed from thematic arrangements in later years, there has recently been a return to this method of ordering the dictionary macrostructure. In this respect, at least two large dictionary projects are worth mentioning, with the first being the ongoing ethnolinguistic *Słownik stereotypów i symboli ludowych* [Dictionary of Folk Stereotypes and Symbols] (Bartmiński and Niebrzegowska-Bartmińska 1996–2024). According to the plan, the work was to consist of seven overarching subject areas, starting with nature and ending with man, culture, and categories that organise the world. Another dictionary of this type is the recently completed 12-volume *Słownik gwar lubelszczyzny* [Dictionary of Lublin Dialects] (Pelcowa 2012–2024). The overarching thematic arrangement of the individual volumes was intended to reflect the image of the world and man as preserved in the consciousness of the rural population. At present, increased attention is being devoted to electronic versions of dictionaries in Polish dialect lexicography (cf. Rembiszewska 2016: 49, 80–81), as well as to the challenges and advantages associated with employing thematic organisations within these resources (cf. Rembiszewska 2016: 42, 223–225, 273). However, to date, no dictionary has yet combined thematic classification with an electronic format.

Second, thematic arrangements were used in the dictionaries of the different language varieties in Polish lexicography in the second half of the 20th century, for example the *Słownik gwary studenckiej* [Lexicon of Students' Dialect] (Kaczmarek, Skubalanka and Grabias 1994),⁵ and the *Słownik polszczyzny potocznej* [Colloquial Polish Dictionary] (Anusiewicz and Skawiński 1996). Because the entry orders in these dictionaries were based on hierarchical classification schemes created from the bottom up, they showed the image of non-linguistic reality and the value systems that were encoded in these language varieties. At the beginning of the 21st century, the use of thematic organisation in Polish lexicography in phraseological dictionaries and dictionaries of proverbs became prominent, which made it easier to show the valuations that were encoded in these language varieties, for example the *Wielki słownik frazeologiczny* [Great Phraseological Dictionary] (Lebda 2009).

Third, pedagogical dictionaries are an important branch of thematic lexicography in Poland (as is the case in world lexicography) (Burkhanov 1999; Hartmann and James 2002: 107), even though many of them are arranged alphabetically. Bilingual thematic dictionaries that include Polish have been very popular in Poland since the early 1990s. However, these types of dictionaries are not as popular in other Slavic countries, such as Serbia and Croatia, as they are in Poland, since thematic arrangements are more commonly found in the form of picture dictionaries in these countries (cf. Daković 2011: 40-45). In Polish lexicography, however, these dictionaries are rare, although an original and highly rated glottodidactic publication of this type does exist, titled *A co to takiego? Obrazkowy słownik języka polskiego* [What's This? A Picture Dictionary of Polish] (Seretny 2008), which was first published in 1993. It is also worth mentioning monolingual pedagogical dictionaries that contain basic vocabulary for learners of Polish, which are organised thematically, for example the *Ilustrowany słownik podstawowy języka polskiego* [Illustrated Elementary Dictionary of Polish] (Kurzowa 2020),⁶ first published in 1999. Monolingual pedagogical dictionaries intended primarily for native speakers interested in improving their language skills should also be mentioned, such as the *Słownik tematyczny języka polskiego* [Thematic Dictionary of Polish] (Kita and Polański 2002).

There remains a need for electronic thematic dictionaries (monolingual, bilingual, or multilingual) of a glottodidactic nature, despite the existence of online thematic dictionaries that are added to language portals or multilingual thematic portals that are created by users (Daković 2011: 44-45). However, printed formats still dominate among dictionaries published by established publishing houses, and their electronic versions are rarely made available.⁷ Linguists working in Polish scientific centres have been developing various ideas and projects related to bilingual (Daković 2011: 45-47) and multilingual corpora-assisted thematic dictionaries (Koutny 2005). Researchers are also paying more attention to both the theoretical and practical potentials of pedagogical thematic dictionaries regarding the enhancement of the linguistic or cultural skills of their users (Koutny 2012; Kononenko and Mytnik 2017; Sękowska 2020; Greń 2023).

3. Newest electronic dictionaries in which thematic classification of vocabulary is applied

In a survey on onomasiological lexicons in 20th-century Europe, Hartmann (2005: 11) indicated the potential of two particular Slavic languages in relation to onomasiological lexicography: Russian and Polish, which achieved the best results among the Slavic languages. As far as Russian is concerned, this lexicography is particularly related to its very large user base, but it is also related to the academic tradition of creating such dictionaries (Hartmann 2005: 11). In the case of Polish (which has a significantly smaller number of users), Hartmann (2005: 13) clearly emphasised the potential of Polish lexicographers. He wrote the following in the paragraph that was dedicated to Polish onomasiological lexicons: "Compilers and theoreticians have been open to ideas and innovations from both East and West, a development that is likely to continue into the future" (Hartmann 2005: 13). These predictions turned out to be correct, because work began on two Polish electronic dictionaries soon after the publication of this article, and they employed the thematic classification of vocabulary that is described in this article.

3.1 The *Słownik pojęciowy języka staropolskiego* (SPJS) [Conceptual Dictionary of Old Polish]

In 2015, a project was completed at the Institute of Polish Language of the Polish Academy of Sciences in Krakow (which began in 2011). Its outcome was the creation of the SPJS. It is an electronic dictionary that includes the lexis of the earliest period of Polish language development up until the end of the 15th century. Based on the printed dictionary of old Polish language in alphabetical order, the first comprehensive and systematic lexicographic presentation of old Polish semantics was created (Sieradzka-Baziur 2020: 9). As emphasised by the dictionary's editor-in-chief, Bożena Sieradzka-Baziur (2020: 41), the SPJS is a "digitally born dictionary" because "it retains its distinctiveness from the *Old Polish Dictionary* — it has a different structure, and the material taken from that work has undergone extensive modifications, which include both the macrostructure of this dictionary and the microstructure of the entries".

In the SPJS, more than 23 000 medieval words and several thousand expressions were classified, but the classifications (sometimes even to several categories) were made on the level of meanings, which often resulted in several distinct classification within a single dictionary entry (Sieradzka-Baziur 2020: 49, 72). For practical reasons, the lexicographers decided to present the old Polish linguistic universe within the field of contemporary concepts. Again in Polish thematic lexicography (Kucala 1957), the scheme that was applied in the dictionary by R. Hallig and W. von Wartburg (1952) became the starting point for creating a classification structure in the SPJS. However, the original scheme was modified and adjusted to old Polish vocabulary, so the differences between them were clear (Sieradzka-Baziur 2020: 69-75). The classification structure consisted of more than 200 con-

ceptual categories within the hierarchised classification scheme. Similarly to the original dictionary, this comprised of three general areas: *Wszechświat i świat* [Universe and the World]; *Człowiek* [Human], and *Człowiek a wszechświat i świat* [Man, the Universe and the World]. Although the work on the SPJS is finished, the lexicographers are open to its possible modifications (Sieradzka-Baziur 2020: 83).

In the bookmark named *Słownik* [Dictionary], users can search for words by using the conceptual classification scheme (which is the essence of the project). However, this can also be done according to an alphabetical arrangement or grammar characteristics. A general search window is also available in the dictionary. The dictionary's website has recently made the English version of the classification scheme available, including information on the number of semantic units assigned to each category (<https://spjs.ijppan.pl/spjs/strona/classificationScheme>).

3.2 The *Wielki słownik języka polskiego Polskiej Akademii Nauk* (WSJP PAN) [Polish Academy of Sciences Great Dictionary of Polish]

The WSJP PAN was created in 2007 as the first primarily electronic general dictionary of contemporary Polish. This lexicographical project is still under development. Its history and principles of preparation are also described in English (Żmigrodzki 2018).

In the description, modern solutions implemented in the project are outlined, including the thematic classification introduced into newly created dictionary entries. In July 2024, the total number of all the entries in the dictionary was more than 90 000. Meanwhile, work has also been completed on a dictionary application for mobile devices, which can be downloaded and used on a tablet or mobile phone (cf. <https://wsjp.pl/page/historia>). The vast majority of the autosemantic entries (i.e., those in which semantic classification was possible), received thematic categorisation, although sometimes they could be assigned to the *athematic* category. Thematic classification was also carried out on the level of particular meanings that were separated within the dictionary entries, and it was possible to classify them into several thematic categories (Batko-Tokarz 2008: 34-39; 2018: 34-36).

For the purpose of the thematic classification of the entries that are being created in the WSJP PAN, an original classification scheme was created that is hierarchical and consists of three degrees. This comprises seven general thematic groups: (1) Man as a Physical Being; (2) Man as a Psychological Being; (3) Everyday Life of a Man; (4) Man in Society; (5) Man and Technology; (6) Man and Nature; and (7) Physical Categories. Within these groups, 45 fields are separated, followed by 205 subfields, which are the basic place of the lexical units classification (although some units may be classified on higher levels). The thematic classification was intended to cover newly created dictionary entries, therefore the classification scheme had to be designed *a priori*. To improve the functionality of such a classification, the names of certain categories were revised, and their scope was made clearer and more understandable for users in 2013.

The rules of the classification and the classification scheme have also been described in detail (Batko-Tokarz 2008; 2019: 135-293).

The dictionary makes it possible to access the lexicographic material in different ways. One can check words according to an alphabetical arrangement, but there are also various advanced search options, with the possibility of marking various categories (including thematic ones). Thanks to this solution, it was possible to find dictionary entries that were classified into particular groups and related fields as well as thematic subfields for many years (Żmigrodzki 2018: 215-216; Żmigrodzki 2022: 157-160). At the end of 2022, the bookmark *Słownik tematyczny* [Thematic dictionary] was added to the top right corner of the home page, which shows how important such access to the vocabulary was for the lexicographers. The rules for using the dictionary are clearly presented on the website in the *Jak korzystać* [How to use] section, which also includes instructional videos demonstrating how to use the various tools available to users. They were created as part of a modern online guide to this dictionary. One of them focuses to the thematic dictionary (<https://www.youtube.com/watch?v=0tySvF1ZKBA>)

4. Polish thematic lexicography in the electronic form — New tools and opportunities

The advantages and disadvantages of primarily electronic dictionaries are increasingly more interesting to various researchers (cf. Granger and Paquot 2012; Ferrett and Dollinger 2021) — including those that deal with Slavic languages (cf. Żmigrodzki 2008: 98-123; Balalaieva 2021). The electronic format results in there being no limits to volume as well as the linear arrangement of dictionaries no longer being necessary. Instead, lexicographers are dealing with multimedia data-bases and open, flexible, and interactive systems with various search tools. User may adjust systems' structures to their needs by opening the desired modules. Moreover, electronic dictionaries (especially those that are included on the internet) are available to everyone, and their forms make it possible for lexicographers to implement the necessary modifications and enhancements. Thematic classification therefore seems to be much more effective in an electronic dictionary than in a paper one. Considering this, the advantages and new opportunities of this combination are presented below on the base of the WSJP PAN and SPJS.

4.1 Combining a semasiological dictionary with an onomasiological dictionary

In an electronic dictionary, there are no longer differences among various dictionaries that used to function according to the principle of opposition. In modern printed dictionaries, the divisions into semasiological, onomasiological, and picture dictionaries have also become increasingly more blurred (Kawalya and De Schryver 2013: 180, 182). In electronic lexicography this is much simpler,

which is why Kawalya and De Schryver (2013: 185-186) suggest a new useful model of a multilingual dictionary. This model is called the "alphaconceptual+dictionary", which combines the advantages of these three types of dictionaries and eliminates the faults of each of them separately (cf. Kawalya and De Schryver 2013: 185-186).

In both of the aforementioned Polish electronic dictionaries, the lexicographers managed to combine the arrangement that is typical of thematic dictionaries with information that can be found in a semasiological dictionary,⁸ and moving from a single entry to a thematic category (and the other way around) is very easy. In both of these dictionaries, the structure of an entry makes it possible to check into which fields of the classification scheme a given unit was classified. In the WSJP PAN the arrangement within an entry, or within its individual meaning (subentry), is structurally organised in the form of expandable tabs. The *Kwalifikacja tematyczna* [Thematic qualification] tab appears below the *Definicja* [Definition] tab (cf. Żmigrodzki 2018: 216; Żmigrodzki 2022: 157-158). When opened, it displays information about the thematic classification (Figure 1).

The screenshot shows the WSJP PAN website interface. At the top, there's a navigation bar with links: 'Historia słownika', 'Jak korzystać', and 'Podstawy naukowe'. The main content area displays the entry for 'dom'. The 'Kwalifikacja tematyczna' tab is active, showing a hierarchical classification of 'dom' into categories like 'CODZIENNE ŻYCIE CZŁOWIEKA' (Daily life of a person), 'CZŁOWIEK I TECHNIKA' (Human and technology), and 'Budownictwo' (Construction). The right sidebar contains tabs for 'Definicja', 'Kwalifikacja tematyczna', 'Połączenia', 'Cytaty', 'Odmiana', and 'Pochodzenie'.

Figure 1: The *Kwalifikacja tematyczna* [Thematic qualification] tab in entry *dom* [house] (1st subentry 'building') in the WSJP PAN

In the SPJS, all relevant information within an entry (subentry) is displayed on the right-hand side, including classification into a conceptual category in the *Semantyka* [Semantics] tab (Figure 2).

Słownik pojęciowy języka staropolskiego

O słowniku Słownik Wyszukiwarka

artykuł hasłowy: dom

Jednostki:	Szczegóły:
(1) 1. 'budynek mieszkalny (często wraz z mieszkańcami i przynależnym gospodarstwem), samo mieszkanie, miejsce stałego pobytu, aedificium, domus (saepe una cum habitatoribus et re familiar), habitanti locus stabilis, cubiculum, interna domus spatia'.	Typ Jednostka pochodzi ze Słownika staropolskiego
(2) ~	Rodzaj autosemantyczna
(3) ~ 'kryjówka'.	Numer (1) 1.
(4) ~ dubium.	Definicja 'budynek mieszkalny (często wraz z mieszkańcami i przynależnym gospodarstwem), samo mieszkanie, miejsce stałego pobytu, aedificium, domus (saepe una cum habitatoribus et re familiar), habitanti locus stabilis, cubiculum, interna domus spatia'.
(5) 2. 'świątynia, templum'.	Gramatyka rzeczownik
(6) ~	Semantyka 2.4.6.4. BUDOWNICTWO. ZAMIESZKIWANIE
(7) ~	2.4.6.4.1. BUDOWLA. BUDYNEK. DOM etc.
(8) ~	Przykład w transliteracji Oradował 'iżm do gich domy, by penędze brali, a ony gich nechceli wżacz 1395
(9) ~	Przykład w transkrypcji Orędownał j->śm do jich domu, by pieniądze brali, a oni jich nie chcieli wziąć 1395
(10) ~	Lokalizacja Leksz II nr 1773
(11) ~	Uwagi [Cytat został uzupełniony o fragment a ony gich nechceli wżacz.]
(12) 3. 'zakład, instytucja, aedificia, vario publicoque apta usu': dom kupiecki 'sklep, bazar, taberna mercatoris'.	
(13) ~ 'skład, sklep z lekami, korzeniami i kosmetykami, apotheca, emporium'.	
(14) ~ 'dom noclegowy, hospitium'.	
(15) ~ 'domus conductitia'.	
(16) ~ 'dom publiczny, lupanar'.	
(17) 4. 'ród, rodzina, gens, familia'.	

Figure 2: Entry *dom* in the SPJS (1st subentry 'building')

From the side of single entries, such a classification scheme facilitates the semantic characteristics of a given word, and it may emphasise its polysemousness when particular meanings obtain different classifications (partially or fully), which is often the case. Such information is valuable for people who are learning a new language, but it is also good for other users (Batko-Tokarz 2018: 82-83).⁹ This was especially important in the SPJS due to the nature of the vocabulary that was collected in the dictionary — and assignments to particular fields is a valuable supplementation of a definition. The dictionary user also has access to definitions and other elements of the microstructure, for example in the WSJP PAN an entry includes large collections of various semantic, grammatical, and etymological information, as well as collocations and full sentence quotations (Koziół-Chrzanowska 2017: 204; Żmigrodzki 2018: 211). In the SPJS, grammatical characteristics as well as examples of word use in medieval texts are provided (Sieradzka-Baziur 2020: 59-63).

If dictionary users are interested in a given meaning, they can move to the field into which it was classified to see the units that are thematically related to this topic. In both dictionaries, users can search for entries according to a hierarchical thematic index that gives users access to the linguistic material in the fields, even though it looks slightly different in terms of the graphics in each dictionary. In the SPJS, the whole long and complex structure becomes visible in the bookmark *Schemat klasyfikacyjny SPJS* [SPJS Classification scheme]. Within the *Słownik tematyczny* [Thematic dictionary] section of the WSJP PAN, users

are presented with an expandable hierarchical structure of thematic scheme (Figure 3). Clicking the question mark icon on this page provides instructions on how to use the search engine and the thematic classification principles used in this dictionary.



Figure 3: The *Słownik tematyczny* [Thematic dictionary] in the WSJP PAN

If users search for words using the classification scheme, they can see lexical units assigned to various thematic areas, fields and especially subfields (in alphabetical arrangement). After choosing the unit, they can quickly move to a description that is typical of a semasiological dictionary (Figures 4 and 5).



Figure 4: Units classified into the subfield *Dom/inne miejsca zamieszkania i ich otoczenie* [Home/other places of residence and their surroundings] in the WSJP PAN (fragment view; the full list contains 702 entries)

2.4.6.4.1. BUDOWLA. BUDYNEK. DOM etc. - jednostki:
baszta (1); bóżnica (1); buda (1); buda (2); budowanie (2); cerkiew, cerki, cyrkiew (1); cerkiew, cerki, cyrkiew (2); chałupa (1); chlew (6); chycz, częściej pl. tantum chycze (1); chyczec a. chyczek (1); chyż, chyża (1); chyżek (1); chyżka (1); chyżka (2); ciemnica (1); ciemniczny (1); ciemność (3); cieplica, cieplice (2); ciosanie (2); ćma (4); dom (1); dom (3); dom (5); dom (12); dom (14); dom (15); dom (16); doma (1); domek (1); domek (2); (S) domostwo, domowstwo (1); domowny (1); domowy (4); dworowy (1); dworzysko (3); dwór (1); dwór (3); dwór (5); dwór (7); dymowe (1); folwark, folwerk, forwak, forwalk, forwark, wolwark (1); folwarkowy, forwarkowy (1); gałka (2); glinianka (1); gmach (1); gnojny (1); gołębiniec (1); gospoda (1); gospoda (2); gospodarz (5); gospodarz (6); gospodarzyć (1); gospodka (1); gospodyni (2); gospodzić (1); gościniec (3); gościnny, gościni, gościny (2); gość (3); grodziec (1); gród (1); izba, histba, istba, izdba (1); jata (1); kamienica (1); kaplica (1); karczemnikow(y) (1); karczemny (1); karczma, kaczma (1); karczmisko (1); kąt (3); kiczka (2); kikieć (?) (1); klasztor (1); klatka, kletka (1); kleć (1); kluza (1); kmiećnica (?) (1); komnata, kownata (1); komora (1); komora (5); komórka (1); konica (1); kościół (1); kościół (4); kościółek, kościółek (1); kram, krom (1); kramnica (1); kramny (1); kramny (3); kruchta, krukta (1); kruchta, krukta (2); kruchta, krukta (3); kryłos (1); kuczka (1); 1. kupiecki, kupiecki (1); kupiectwo (1); kupny (6); łakwa (1); lekarski (1); lonar (1); ława (2); łęczysko (2); łęczyszcze (2); modlebnia (1); modlebnica (1); modlitwa, moglitwa (4); murowanie (1); murowany (1); namiot (1); nasuć (4); nieczysty (5); niepoczesny, niepoczesny (2); noc (3); noc (4); noc (5); nocleg (1); nocować (1); obcowanie, hobcowanie (5); obora (1); obręba (2); odpoczynienie, otpoczynienie (4); odpoczywać, odpoczywać (2); okół (2); oratarz (1); pacisz (?) (1); pałac (1); pałacowy (1); piekuta (1); plac, plec (1); płot (6); pobudowanie (1); podniebie (2);

Figure 5: Units classified into the category *Budowla. Budynek. Dom* etc. [Structure. Building. House etc.] in the SPJS (fragment view; the full list contains 249 entries)

The electronic form of a dictionary helps lexicographers overcome the faults of the thematic order that are visible in the aforementioned printed Polish thematic dictionaries, and this seems to be the solution to problems of metaideography (Burkhanov 1999: 144-147). Such faults may include a frequent lack of definitions of thematically classified units, which results in certain doubts that are related to the reasons for their classification. When classification in printed dictionaries was carried out at the level of individual meanings, the various senses of a single word were distributed across different sections of the scheme. Problems also include the arbitrariness and non-intuitiveness of a given classification or its assignments — especially when there are no definitions — as well as related difficulties when searching for particular units by the thematic scheme. Researchers have mentioned that those units that are thematically non-classifiable or hard to classify have been omitted in such dictionaries (cf. Batko-Tokarz 2019: 114-122). Until now, the lexicographers have combined various kinds of access to the dictionary material through thematic or alphabetical indices in order to remove the faults, and they have used the advantages of both arrangements of a dictionary macrostructure (Burkhanov 2010: 108-109). This was dealt with in the above-mentioned Polish dictionaries (including the historical ones), but they have not solved all of the problems (Batko-Tokarz 2019: 122-124).

4.2 Abundant and rich linguistic material in thematic fields and various opportunities for its characteristics

Both of the above-mentioned Polish dictionaries follow the assumptions that thematic classification should be made on the level of meanings separated within an entry, and that there can be several possible classifications. Due to these assumptions, thematic classification was able to include a considerable number of lexical units. First, in the dictionaries of general Polish, entry articles often have several or even more than a dozen meanings. Second, many of them have been classified several times. According to the recent official statistics of the WSJP PAN¹⁰ of as many as 38 215 units were simultaneously classified into two fields, and 4 001 units were classified into three or more fields (Żmigrodzki 2022: 156).

The electronic form of a dictionary definitely facilitates broadly understood thematic classification, as there are no limitations to the volume of such a lexicon. Because of this, the WSJP PAN is the largest thematic dictionary of contemporary Polish at the moment and outrivals the other Polish thematic dictionaries mentioned above. According to the status as of August 2021, it included 141 561 meanings that were classified in a thematic manner. If multiple classifications are, however, taken into account, it turns out that the total number of classifications to thematic categories is as many as 188 037 (Żmigrodzki 2022: 155, 160). Although it refers to the early stage of the development of Polish, the SPJS also includes 80 594 thematic classifications, while 44 177 units were classified thematically into at least one (and often more) categories (Sieradzka-Baziur 2020: 90). A similar number of units are included in the largest translation dictionaries that are arranged thematically and include Polish, while the above-mentioned contemporary Polish glottodidactical thematic dictionaries (Kita and Polański 2002; Kurzowa 2020) or phraseological dictionaries (Lebda 2009) count from several to more than ten thousand entries (Żmigrodzki 2022: 155).

All of this influences the richness of the material in particular thematic categories and the completion of them with new units (Batko-Tokarz 2018: 87-88), because the WSJP PAN (as the dictionary of general language) collects lexical units with various features (e.g., scientific, colloquial, expressive, and official features) that will all be included in a given category if they are connected by the same topic. In both dictionaries, not only single words but also discontinuous units (phrasemes and proverbs) are subject to classification. Due to this, the user obtains simultaneous access to both of those groups in the fields. Thematic fields that are created in this way may be useful for specialised users of a language, such as translators, journalists, or authors of various texts who want to extend their vocabulary on a given topic. Additionally, such fields may be useful for both native and non-native language users. Each of them may adjust the access to the dictionary to their needs, and each user may select the linguistic material in a variety of ways. This is enabled by the advanced search options.

Such a dictionary and its material selection options may be very useful for the researchers of a language (Sieradzka-Baziur 2020: 90-96). With the use of

electronic devices, it is easy to carry out varied characteristics of particular categories and their contents and structure as well as quantitative analyses. For example, the field *Życie społeczne* [Social life] dominates in the most developed group in SPJS (*Człowiek* [Human]), which counts as many as 166 categories. The next two places are occupied by those aspects that are related to psychological and spiritual life (Sieradzka-Baziur 2020: 85-86). In the WSJP PAN, the two groups with the largest number of thematically classified units were *Człowiek jako istota psychiczna* [Man as a psychological being] (23,6%) and *Człowiek w społeczeństwie* [Man in society] (22,18% of the units) (Żmigrodzki 2022: 160). Based on such quantitative analyses, interesting conclusions can be drawn on the thematic structure of the vocabulary and comparative analyses can be carried out. It is possible to see which thematic fields or subfields are the most or least numerous (Sieradzka-Baziur 2020: 86-87; Żmigrodzki 2022: 160-164).

4.3 Advanced search

Electronic dictionaries offer various kinds of searching in which modern researchers are increasingly more interested, including various forms of thematic searching (Pastor and Alcina 2010; Camacho-Niño 2021).

A bookmark *Wyszukiwanie zaawansowane* [Advanced search] in the WSJP PAN merits special attention for its exceptional appeal (Figure 6). This is because various search filters may be combined with searching by thematic categories, which gives users an almost unlimited opportunity to arrange thematic fields and search for the material that interests them. The option of searching by thematic criteria in an advanced search looks similar to that on the main page of this dictionary (Option 3: *Wyszukaj według przynależności do kategorii tematycznej* [Search by Thematic Category]) (Żmigrodzki 2022: 158-159).

Searching by a specific thematic category can be combined with other options, such as with the option to search by the dictionary's rich set of chronological, expressive, normative, and specialist qualifiers, as well as qualifiers that characterise the vocabulary in terms of the geographical, stylistic, and environmental ranges. Such a searching method can further be combined with searching by the language of origin of a given unit, which enables the semantic analysis of borrowings. Another method of searching is possible by grammar characteristics (i.e., by the types of entries, such as phraseological expressions) or by specified parts of speech. In this way, users may gain access to only the colloquial and expressive vocabulary that is typical of a given specialisation, for example, or to nouns or phraseological expressions that belong to a given thematic category (Batko-Tokarz 2018, 89-91; Żmigrodzki 2022: 158-159). This is a valuable supplementation of the above-described thematic dictionaries of various kinds.

The WSJP PAN has recently added another search option based on the date of the appearance of a given form in Polish (based on the chronologising of dictionary entries in the WSJP PAN). This makes it possible to see, for example, how the vocabulary of a specific stage (such as the beginning of the 21st century) is classified thematically and, thus, in which fields the greatest changes

have taken place in this respect. The advanced search, which allows one to search thematic fields and classify the vocabulary that is collected in them according to various criteria, seems to be an invaluable tool that allows one to tailor the thematic dictionary to the diverse needs of its users and, in fact, transform it into the various types of thematic dictionaries that were described above.

Ustawienia
zaawansowanego wyszukiwania

Wpisz szukane słowo... **Szukaj** ↻

1a Wyszukaj hasła (znaczenia) zawierające podany ciąg znaków:
☒ w formie hasłowej ☐ w definicji ☐ w cytacie ☐ w połączeniach

1b Wyszukaj hasła (znaczenia) zawierające dowolne formy fleksyjne podanych wyrazów (zwrotów):
☒ nie uwzględniaj znaków przestankowych i nawiasów ☐ uwzględniaj znaki przestankowe i nawiasy

2 Wyszukiwanie według kwalifikatorów (kwalifikatory): ▼

3 Wyszukaj według przynależności do kategorii tematycznej (Kategoria tematyczna)
[dowolna kategoria] [dowolna podkategoria] [dowolna pozycja]

4 Wyszukiwanie według języka pochodzenia (Pochodzenie): ▼

5 Wyszukiwanie według charakterystyki gramatycznej ▼

6 Wyszukiwanie według chronologii
Stulecie: ☒ nie określono ☐ określono
Rok: Od: [] Do: []

Figure 6: Advanced search in the WSJP PAN

5. Summary

Researchers often indicate the centuries-long domination of the alphabetical arrangement in lexicography (McArthur 1998: 150; Kawalya and De Schryver 2013: 174-177; Kay 2015: 54). Although the same is true from Polish lexicography, the results that were collected in this article show that, in previous centuries as well as in recent decades, many thematic dictionaries have emerged in Polish. Polish scientific research is increasingly focusing on onomasiological and thematic lexicography, which is related to the growing interest in these types of dictionaries in world linguistics as well (Kawalya and De Schryver 2013: 178-180). This phenomenon is reflected in Polish electronic dictionaries that enable the new and definitely more effective functioning of this method and overcoming its limitations. Such dictionaries seem to be the lexicons of the future — especially because they are very useful and their functioning in the electronic form is very promising. In Slavic lexicography, a revival of the research on thematic dictionaries can also be observed in the electronic format; this is clearly related to new lexicographic projects in this area. This is evident from a thesaurus of the Czech language that was published at the beginning of the 21st century, which maintained the format of Roget's thesaurus (Klégr 2007). In the following years, a Czech online onomasiological dictionary and database (<https://beta.najdislovo.cz/>) was developed (Obstová et al. 2022) on the basis of this thesaurus and another Czech thematic dictionary (Haller 1969-1986). Also in Croatian lexicography, the Croatian thesaurus project was described (cf. Nikolić-Hoyt 2004), and now it is a project of the somatic thesaurus for this language (Nikolić-Hoyt 2020). Furthermore, one can find descriptions of projects focused on the compilation of thematic dictionaries for various types of vocabulary (cf. Bihday 2018; Nenasheva 2024).

Endnotes

1. See Batko-Tokarz (2015; 2019: 21-32) for a detailed description of the perception of these terms in European and Polish lexicography.
2. In a dictionary of Slavic linguistic terminology (Jedlička 1977: 222-223), the term *thematic* dominates, but other terms (sometimes several of them) also appear for a given language. The word *ideographical* is popular — especially in the Russian tradition (Burkhanov 1999). In older works, the following names were used: *subject dictionary* — *słownik rzeczowy* (Polish), and *věcný slovník* (Czech). In newer ones terms such as *conceptual*: *pojęciowy* (Pol.) and *pojmový* (Cz.) are also found.
3. In a narrow understanding, a thematic dictionary is identified only with pedagogical dictionaries that are to support language learning (Burkhanov 2010: 242-243).
4. All of those dictionaries have been characterised in detail against the broader background of the Polish linguistic thought (Batko-Tokarz 2019: 39-127).
5. This was prepared in the late 1960s, but communist censorship prevented its publication because the topics reflected a student perspective that differed from the official state propaganda. The dictionary was only published after the fall of communism in Poland.

6. The one part of the dictionary, arranged in alphabetical order, has been provided with subject indexes modeled on McArthur's (1981) dictionary.
7. In the middle of 2022, however, one of the publishing houses made it possible for users to download the electronic version of the *Słownik tematyczny polsko-ukraiński* [Polish-Ukrainian Thematic Dictionary] (Kononenko, Mytnik and Wasiak 2015). The dictionary was dedicated to Ukrainian refugees who were forced to leave their country and move to Poland because of the war.
8. So far, they do not include image or sound elements. This is one of the challenges that Polish electronic lexicography face, and Polish researchers are thinking about ways of implementing them. For example, Biesaga (2017) analysed which thematic fields were illustrated most often and least often in selected dictionaries by using the thematic scheme applied in the WSJP PAN. Furthermore, she pointed out that it was justified to introduce images that were connected with the thematic division of the vocabulary.
9. This was described in the example of the entry *zdechnąć* [die]. In the WSJP PAN, seven meanings are separated (<https://wsjp.pl/haslo/podglad/35274/zdechnac>) with references to (1) the deaths of animals, and (2) the deaths of human beings, (3) to the functioning of man in the physical aspect (tiredness), and (4) in the psychological aspect (boredom), but also, respectively, with reference (5) to plants, when they wither, (6) to machines and devices when they stop working, and (7) to non-material phenomena that which disappear. Thematic classification emphasises this polysemy as each meaning has been classified in a different way. This is helpful especially to less proficient users of Polish, especially because meanings 3 through 7 are colloquial. In the SPJS, two meanings are distinguished in the entry *zdechnąć*, which receive different thematic classifications (death of people, and death of animals) (<https://spjs.ijppan.pl/haslo/index/21523/9978>).
10. Work on the WSJP PAN is still ongoing, and these numbers should be higher now.

References

Printed dictionaries

- Anusiewicz, J. and J. Skawiński. 1996. *Słownik polszczyzny potocznej*. Warszawa: Wydawnictwo PWN.
- Artomiusz, P. 1591. *Nomenclator selectissimas rerum appellationes tribus linguis, Latina, Germanica, Polonica explicatas indicans*. Kraków.
- Bartmiński, J. and S. Niebrzegowska-Bartmińska (Eds.). 1996–2024. *Słownik stereotypów i symboli ludowych*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Haller, J. (Ed.). 1969–1986. *Český slovník věcný a synonymický 1–3 and Rejstřík*. Praha: Státní Pedagogické Nakladatelství.
- Hallig, R. and W. von Wartburg. 1952. *Begriffssystem als Grundlage für die Lexikographie*. Berlin: Akademie Verlag.
- Kaczmarek, L., T. Skubalanka and S. Grabias. 1994. *Słownik gwary studenckiej*. Lublin: Wydawnictwo UMCS.
- Kita, M. and E. Polański. 2002. *Słownik tematyczny języka polskiego*. Łódź: Wydawnictwo Literatura.
- Klégr, A. 2007. *Tezaurus jazyka českého. Slovník českých slov a frází souznačných, blízkých a příbuzných*. Praha: Nakladatelství lidové noviny.

- Kononenko, I., I. Mytnik and E. Wasiak.** 2015. *Słownik tematyczny polsko-ukraiński*. Warszawa: Wydawnictwo PWN.
- Kucała, M.** 1957. *Porównawczy słownik trzech wsi małopolskich*. Wrocław: Zakład Narodowy im. Ossolińskich.
- Kurzowa, Z.** 2020. *Ilustrowany słownik podstawowy języka polskiego wraz z indeksem pojęciowym wyrazów i ich znaczeń*. Kraków: Universitas.
- Lebda, R.** 2009. *Wielki słownik frazeologiczny*. Kraków: Krakowskie Wydawnictwo Naukowe.
- McArthur, T.** 1981. *Longman Lexicon of Contemporary English*. Essex: Longman.
- Murmeliusz, J.** 1526. *Dictionarius Ioannis Murmellii variarum rerum*. Kraków.
- Mymer, F.** 1528. *Dictionarium trium linguarum: Latinae: Teutonicae et Polonicae* Kraków.
- Pelcowa, H.** 2012–2024. *Słownik gwar Lubelszczyzny*. Lublin: Wydawnictwo UMCS.
- Seretny, A.** 2008. *A co to takiego? Obrazkowy słownik języka polskiego*. Kraków: Universitas.
- Słowarz** (so-called). Anonymous author. 1532. *Dictionarius seu nomenclatura quatuor linguarum*. Kraków.
- Zawiliński, R.** 1926. *Dobór wyrazów. Słownik wyrazów bliskoznacznych i jednoznacznych. Do praktycznego użytku ułożył Roman Zawiliński*. Kraków: "Poradnik Językowy". Accessed at: <https://rcin.org.pl/dlibra/show-content/publication/edition/36954?id=36954> (23 April 2005)

Online dictionaries

- SPJS** — Sieradzka-Baziur, B. (Ed.). *Słownik pojęciowy języka staropolskiego*. Accessed at: <https://spjs.ijp.pan.pl/spjs/strona/kartaTytulowa> (16 November 2025)
- WSJP PAN** — Żmigrodzki, P. (Ed.). 2006–. *Wielki słownik języka polskiego PAN*. Accessed at: <https://wsjp.pl/> (16 November 2025)

Other literature

- Balalaieva, O.** 2021. Класифікації Електронних Словників [Classifications of Electronic Dictionaries], *Міжнародний Філологічний Часопис* [International Philological Journal] 12(2): 60–63. <https://doi.org/10.31548/philolog2021.02.060>
- Batko-Tokarz, B.** 2008. Tematyczny podział słownictwa w Wielkim słowniku języka polskiego. Żmigrodzki, P. and R. Przybylska (Eds.). 2008. *Nowe studia leksykograficzne* 2: 31–48. Kraków: Lexis. Accessed at: <http://ijp.pan.pl/publikacje-elektroniczne/nowe-studia-leksykograficzne-t-2> (1 April 2025)
- Batko-Tokarz, B.** 2015. Słowniki o układzie hierarchicznym — próba uporządkowania pojęć. *Język Polski* 95(5): 406–417. <https://doi.org/10.31286/JP.95.5.2>
- Batko-Tokarz, B.** 2018. Klasyfikacja tematyczna w WSJP PAN. Żmigrodzki, P., M. Bańko, B. Batko-Tokarz, J. Bobrowski, A. Czelakowska, M. Grochowski, R. Przybylska, J. Waniakowa and K. Wegrzynek (Eds.). 2018. *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*: 79–92. Kraków: Libron. Accessed at: <https://rcin.org.pl/dlibra/publication/edition/68631?id=68631&from=publication> (10 April 2025)
- Batko-Tokarz, B.** 2019. *Tematyczny podział słownictwa współczesnego języka polskiego — teoria, praktyka, leksykografia*. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego. Accessed at: <https://ruj.uj.edu.pl/entities/publication/f52bfbba-4e31-4246-bd1b-f222891e8d06> (12 April 2025)

- Biesaga, M.** 2017. Dictionary Tradition vs. Pictorial Corpora: Which Vocabulary Thematic Fields Should Be Illustrated? *Lexikos* 27: 132–151.
<https://doi.org/10.5788/27-1-1397>
- Bihday, M.** 2018. Representation of Lexical-Semantic Classification of the Ukrainian Verbs in the Form of Ideographic Dictionary. *Linguistic Studies* 36: 44–49.
<https://doi.org/10.31558/1815-3070.2018.36.6>
- Burkhanov, I.** 1999. *Linguistic Foundations of Ideography: Semantic Analysis and Ideographic Dictionaries*. Rzeszów: Wydawnictwo Wyższej Szkoły Pedagogicznej.
- Burkhanov, I.** 2010. *Lexicography: A Dictionary of Basic Terminology*. Rzeszów: Wydawnictwo Uniwersytetu Rzeszowskiego.
- Camacho-Niño, J.** 2021. Estructuras de acceso y remisión en los diccionarios del siglo XXI. Moreno Moreno, M.A. and M. Torres Martínez (Eds.). 2021. *Estudios del léxico en el ámbito universitario del siglo XXI*: 125–148. Barcelona: Editorial Octaedro S.L. Accessed at:
https://www.researchgate.net/publication/352996840_Estructuras_de_acceso_y_remision_e_n_los_diccionarios_del_siglo_XXI (1 April 2025)
- Daković, S.** 2011. Słowniki tematyczne w Polsce, Serbii i Chorwacji. *Südslavistik online — zeitschrift für südslavische Sprachen und Literaturen und Kulturen* 3. Dvojezična leksikografija bosanskog/hrvatskog/srpskog jezika: 33–52. Accessed at:
<http://www.suedslavistik-online.de/03/03.pdf> (1 April 2025).
- Ferrett, E. and S. Dollinger.** 2021. Is Digital Always Better? Comparing Two English Print Dictionaries with Their Digital Counterparts. *International Journal of Lexicography* 34(1): 66–91.
<https://doi.org/10.1093/ijl/ecaa016>
- França, P.C.** 2010. *Onomasiological Dictionaries and Ontologies*. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010*: 1291–1298. Ljouwert: Fryske Akademy. Accessed at:
http://www.euralex.org/elx_proceedings/Euralex2010/121_Euralex_2010_9_FRANCA_Onomasiological%20dictionaries%20and%20ontologies.pdf (8 April 2025)
- Granger, S. and M. Paquot (Eds.).** 2012. *Electronic Lexicography*. Oxford: Oxford University Press.
- Greń, Z.** 2023. Klasyfikacja tematyczna słownictwa w leksykografii czeskiej jako źródło do badania zmian w językowym obrazie świata. Falski, M. and E. Kaczmarska-Zglejszewska (Eds.). 2023. *Czechy przez naukę języka do rozumienia kultury. Księga dedykowana Teresie Piotrowskiej-Malek*: 15–43. Kraków: Libron.
- Gruszczyński, W. and Z. Saloni.** 2013. From Multilingual to Monolingual Dictionaries. A Historical Overview of Polish Lexicography. *Studies in Polish Linguistics* 8(4): 205–227.
<https://doi.org/10.4467/23005920SPL.13.010.1691>
- Hartmann, R.R.K.** 2005. Onomasiological Dictionaries in 20th-Century Europe. *Lexicographica. International Annual for Lexicography* 21: 6–19.
<https://doi.org/10.1515/9783484604742.6>
- Hartmann, R.R.K. and G. James.** 2002. *Dictionary of Lexicography*. London/New York: Routledge.
- Hüllen, W.** 2006. *English Dictionaries, 800–1700. The Topical Tradition*. Oxford: Oxford University Press.
- Jedlička, A. (Ed.).** 1977. *Slovník slovanské lingvistické terminologie = Словарь славянской лингвистической терминологии = Dictionary of Slavonic Linguistic Terminology*. Praha: Academia nakladatelství Československé akademie věd Praha.

- Kawalya, D. and G.-M. de Schryver.** 2013. Introducing a New Lexicographical Model: AlphaConceptual⁺ (and How it Could Be Applied to Dictionaries for Luganda). *Lexikos* 23: 172-200.
<https://doi.org/10.5788/23-1-1210>
- Kononenko, I. and I. Mytnik.** 2017. Potencjał teoretyczny i praktyczny dwujęzycznego słownika tematycznego. Borys, O., M. Jeż, A. Samadowna and M. Saniewska (Eds.). 2017. *Szkice językowe i literacko-kulturowe*: 40-57. Warszawa: Katedra Ukrainistyki, Uniwersytet Warszawski. Accessed at: <https://open.icm.edu.pl/handle/123456789/19237> (10 April 2025)
- Koutny, I.** 2005. A Corpora Assisted Multilingual Thematic Dictionary. *Archives of Control Sciences* 15(4): 573-582.
- Koutny, I.** 2012. From Semantic Networks to Dictionary Structure. *Język. Komunikacja. Informacja. Language. Communication. Information* 7: 115-128. Accessed at: <https://repozytorium.amu.edu.pl/server/api/core/bitstreams/61eb6b33-8db3-4f17-8023-25e811b2968d/content>
- Kozioł-Chrzanowska, E.** 2017. What Do Users of General Electronic Monolingual Dictionaries Search For? The Most Popular Entries in the *Polish Academy of Sciences Great Dictionary of Polish*. Kosem, I., C. Tiberius, M. Jakubiček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century. Proceedings of eLex 2017 Conference, Leiden, the Netherlands, 19–21 September 2017*: 202-220. Brno: Publisher: Lexical Computing CZ s.r.o., Brno, Czech Republic. Accessed at: <https://elex.link/elex2017/wp-content/uploads/2017/09/paper12.pdf>
- McArthur, T.** 1998. *Living Words: Language, Lexicography and the Knowledge Revolution*. Exeter: University of Exeter Press.
- Nenasheva, L.V.** 2024. «Тематический словарь архангельских говоров» и электронная система словаря ["Thematic Dictionary of Arkhangelsk Dialects" and the Electronic Dictionary System]. *Philology. Theory & Practice* 17(9): 3376-3383.
<https://doi.org/10.30853/phil20240478>
- Nikolić-Hoyt, A.** 2004. *Konceptualna leksikografija: prema tezaurusu hrvatskoga jezika*. Zagreb: Hrvatska sveučilišna naklada.
- Nikolić-Hoyt, A.** 2020. Iz riznice Benešićeva rječnika i Somatskoga tezaurusa hrvatskoga jezika. *Filologija* 75: 47-65.
<https://dx.doi.org/10.21857/y54jofkgqm>
- Obstová, Z., O. Tichý and A. Klégr.** 2022. Projekt Lexikálně-sémantické databáze češtiny (LSD-Czech): uživatelský pojmový slovník a online databáze. *Studie z Aplikované Lingvistiky* 13: 86-98. Accessed at: https://studiezaplikovanelingvistiky.ff.cuni.cz/wp-content/uploads/sites/19/2023/01/Zora_Obstova_-_Ondrej_Tichy_-_Ales_Klegre_86-98.pdf (28 April 2025)
- Pastor, V. and A. Alcina.** 2010. Search Techniques in Electronic Dictionaries: A Classification for Translators. *International Journal of Lexicography* 23(3): 307-354.
<https://doi.org/10.1093/ijl/ecq015>
- Rembiszewska, D.K. (Ed.).** 2016. *Słowańskie słowniki gwarowe*. Warszawa-Łomża: Łomżyńskie Towarzystwo Naukowe im. Wagów.
- Sękowska, E.** 2020. A Thematic Dictionary as a Source of Cultural Competence (Using the Example of a Chinese-Polish Dictionary). *Acta Universitatis Lodzensis. Kształcenie Polonistyczne Cudzoziemców* 27: 339-348.
<https://doi.org/10.18778/0860-6587.27.19>

- Sieradzka-Baziur, B.** 2020. *Onomazjologiczny słownik online. Metodologia, zawartość, wykorzystanie w badaniach na przykładzie "Słownika pojęciowego języka staropolskiego"*. Kraków: Instytut Języka Polskiego Polskiej Akademii Nauk. Accessed at:
<https://rcin.org.pl/dlibra/publication/270658/edition/233563> (12 April 2025)
- Sierra, G.** 2000. *The Onomasiological Dictionary: A Gap in Lexicography*. Heid, U., S. Evert, E. Lehmann and C. Rohrer (Eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany, August 8th–12th, 2000*: 223-235. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart. Accessed at:
http://www.euralex.org/elx_proceedings/Euralex2000/027_Gerardo%20SIERRA_The%20onomasiological%20dictionary_a%20gap%20in%20lexicography.pdf (12 April 2025)
- Żmigrodzki, P.** 2008. *Słowo — słownik — rzeczywistość. Z problemów leksykografii i metaleksykografii*. Kraków: Lexis.
- Żmigrodzki, P.** 2018. Methodological Issues of the Compilation of the *Polish Academy of Sciences Great Dictionary of Polish*. Čibej, J., V. Gorjanc, I. Kosem and S. Krek (Eds.). 2018. *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, 17–21 July 2018, Ljubljana, Slovenia*: 209-219. Ljubljana: Ljubljana University Press, Faculty of Arts. Accessed at:
<https://euralex.org/publications/methodological-issues-of-the-compilation-of-the-polish-academy-of-sciences-great-dictionary-of-polish/>
- Żmigrodzki, P.** 2022. *Wielki słownik języka polskiego PAN jako słownik tematyczny. Studia Językoznawcze. Synchroniczne i diachroniczne aspekty badań polszczyzny* 21: 153-166. Accessed at:
<https://wnus.edu.pl/sj/pl/issue/1252/article/19943/>