



New developments in Elexifinder, a discovery portal for lexicographic literature

Iztok Kosem, David Lindemann
Jožef Stefan Institute, Ljubljana



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Overview

- Introduction
- Elexifinder version 2
 - Interface and search functions
 - *Concepts and Categories*
 - New item collections
- LexBib bibliographical data
 - LexBib Zotero group
 - Sources and workflows
 - Author disambiguation
- LexVoc vocabulary of lexicographic terms
 - Sources, structure, and applications
 - LexVoc terms translation
- Outlook

Introduction

- Elexifinder version 1, introduced in Kosem & Krek (2019)
- LexBib, introduced by Lindemann, Kliche & Heid (2018)
- Overlapping interests: Bibliodata meeting at eLex Sintra
- 2019 to 2020:
 - Existing datasets merged
 - More data included
 - Creation of RDF database as central data repository
 - Author disambiguation
- LexVoc vocabulary
 - Subject indexation of articles
 - Subject categories for Elexifinder

Elexifinder Interface: Live Demo

The screenshot displays the @lexiFinder web interface. At the top, the @lexis logo is on the left, and the @lexiFinder logo is in the center. Below the logo is a search bar containing the query 'B. T. Sue Atkins AND B Bilingual' and a 'SEARCH' button. To the right of the search bar are filter options: 'Filter by: Locations Sources Categories Time of interest Any language'. Below the search bar, the text 'List of publications (13 results found)' is displayed. On the left side, there is a sidebar with navigation links: 'List of publications', 'Top Concepts', 'Languages', 'Tag Cloud', 'Timeline', 'First Author Locations', 'Sources', 'Publication authors', and 'Concept Graph'. The main content area shows a list of publications. The first publication is 'Lexicographic Profiling: an Aid to Consistency in Dictionary Entry Design' by B. T. Sue Atkins et al., published in EURALEX (2006) on Sun, 01 Jan 2006. The second publication is 'The Contribution of Framenet to Practical Lexicography' by Michael Rundell et al., published in IJL 16/3 (2003) on Mon, 01 Sep 2003. The third publication is 'Then and Now: Competence and Performance in 35 Years of' published in EURALEX.

Elexifinder version 2: contents

Collections

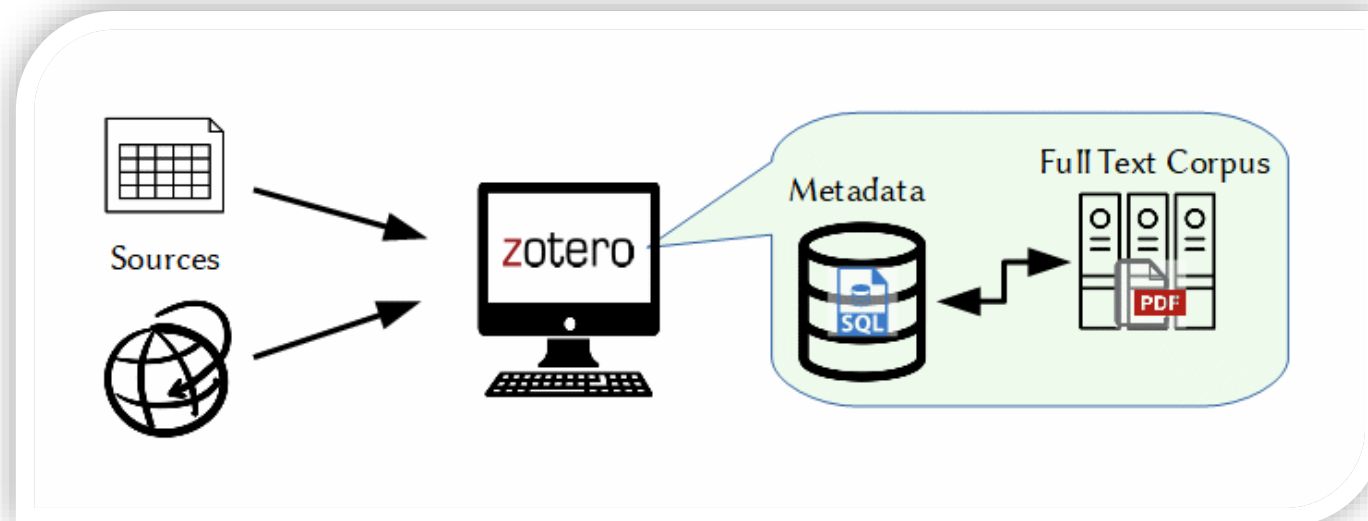
- Conference proceedings
 - 2,018 items: eLex, Euralex, Globalex
- Journals
 - 4,145 items: 9 journals, complete
- Collective vols. and handbooks
 - 319 items: 12 volumes, complete
- Videos
 - 86 items

Top 10 publication languages

- English: 4,853
- German: 910
- Danish: 436
- Spanish: 395
- Swedish: 369
- French: 227
- Bokmål: 179
- Afrikaans: 135
- Slovene: 81
- Italian: 44
- (others: 83 items in 10 languages)

LexBib bibliographical data

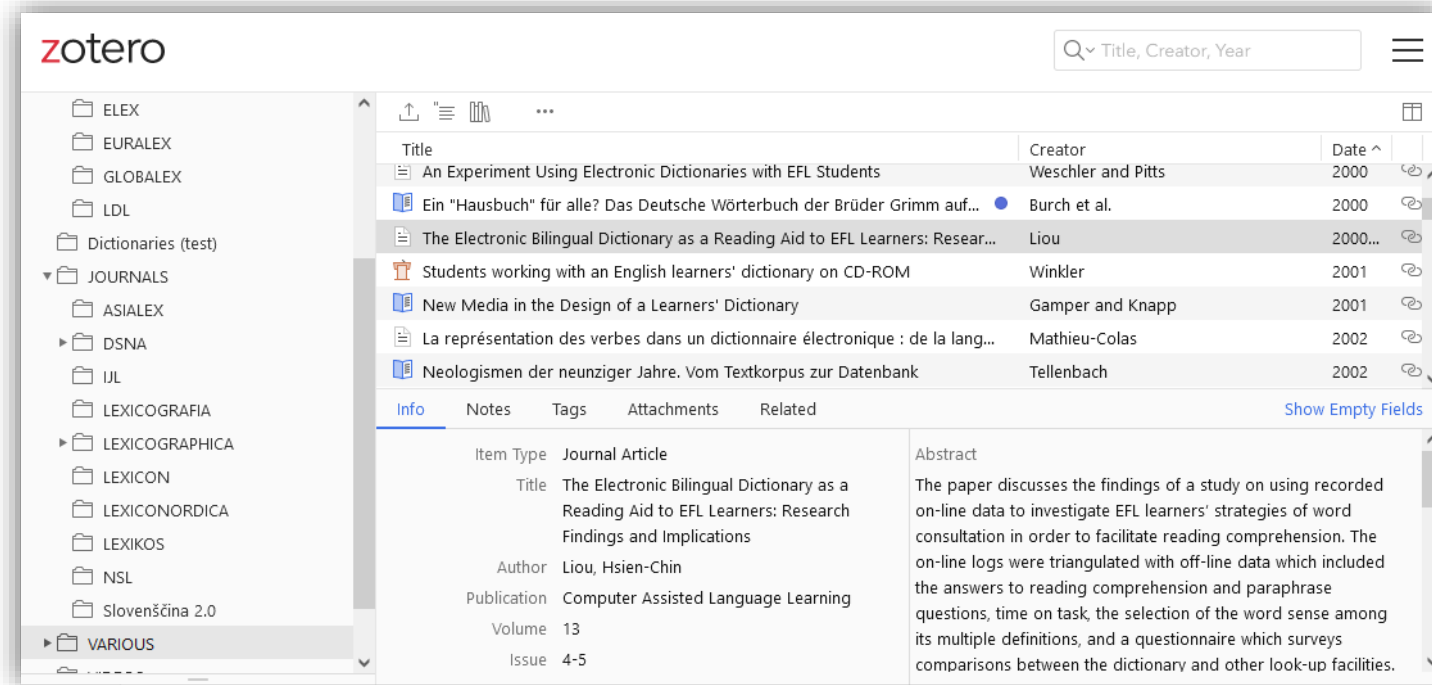
- Zotero
 - Recording of publication metadata, and full texts
 - In tabular format, from existing bibliographies (EURALEX, OBELEX-meta)
 - Using Zotero web scraping tool: From web-repositories
 - Manual validation of metadata records (team uses „Zotero groups“)
 - Always with full text



LexBib Zotero Group

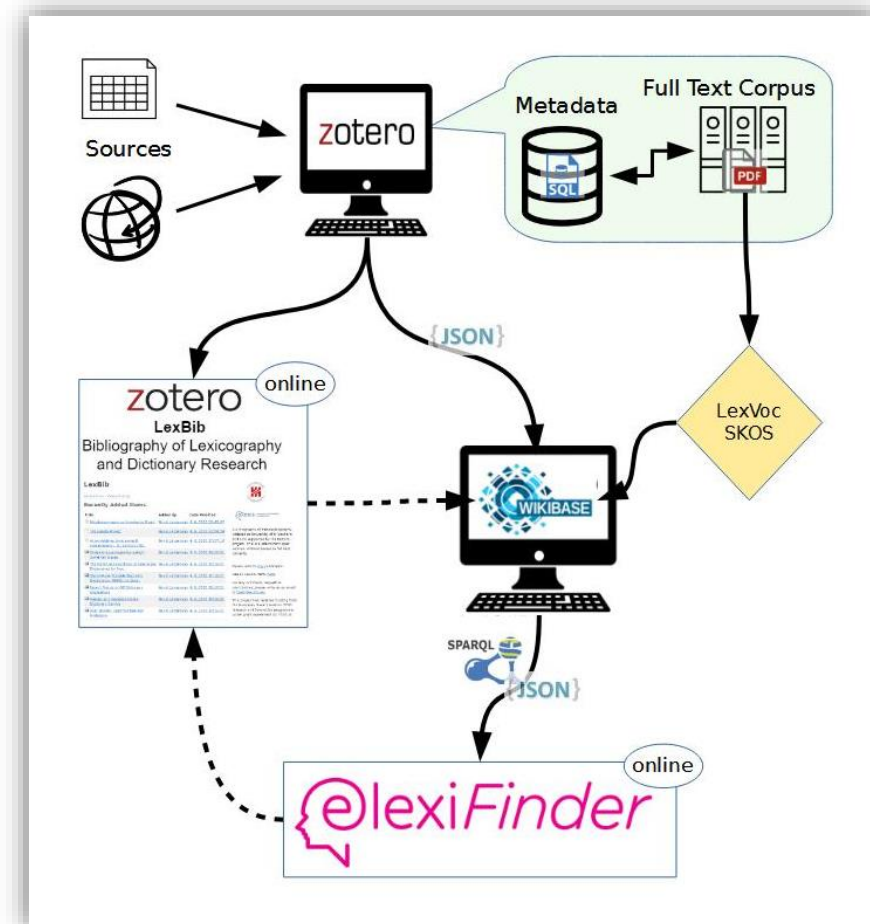
LexBib Zotero group:
<http://lexbib.org/zotero>

- Browse by collection, search for publication metadata
- Metadata download in structured formats (bibtex, etc.)
- Useful for import to own citation manager



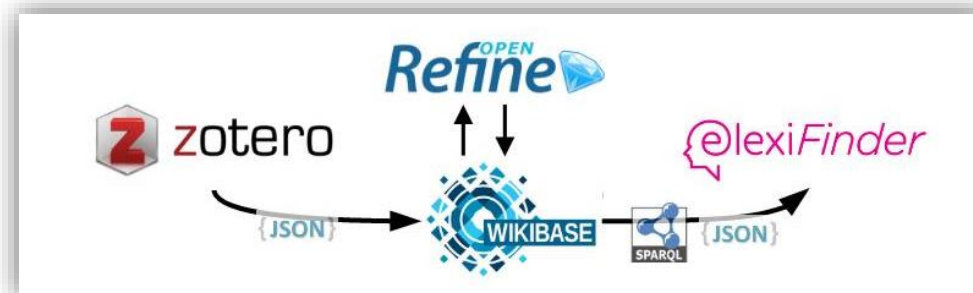
LexBib bibliographical data

- Export from Zotero to LexBib Wikibase
- Mapping literals to ontology items:
 - Creators (authors, editors)
 - Containers (journal issues, etc.)
 - Events (conferences, workshops)
 - Languages (publication language)
- Export from Wikibase to Elexifinder discovery portal
 - Disambiguated entities
 - Full text body (GROBID, manual cleaning)
- Full Texts: upload to SketchEngine
 - Custom subcorpora (by event, time, etc.)



LexBib bibliographical data

- Author disambiguation using Open Refine
 - Initial dataset (Klaes 2021)
 - 18,000 creator statements in Zotero
 - 5,000 different names
 - 4,000 different persons
 - Updates since Aug. 2021
 - Reconciliation of Zotero name literals against LexBib Wikibase



120 matching rows (425 total)

Show as: rows records Show: 5 10 25 50 rows

	bibItem	statement_id	coll	creator
236.	http://data.lexbib.org/entity/Q22805	Q22805-43E59BC9-D28F-4CCF-8A80-3E86B6F64DA7	26	B. T. S. Atkins
237.	http://data.lexbib.org/entity/Q22805			Anna Wierzbicka
238.	http://data.lexbib.org/entity/Q22805			Jurij D. Апресян
239.	http://data.lexbib.org/entity/Q22805			Sue Atkins
240.	http://data.lexbib.org/entity/Q22805			Sidney I. Lan
241.	http://data.lexbib.org/entity/Q22805			

Search for match

Search for "B. T. S. Atkins"

☒ Match other cells with same content

☐ Match this cell only

Select an item from the list:

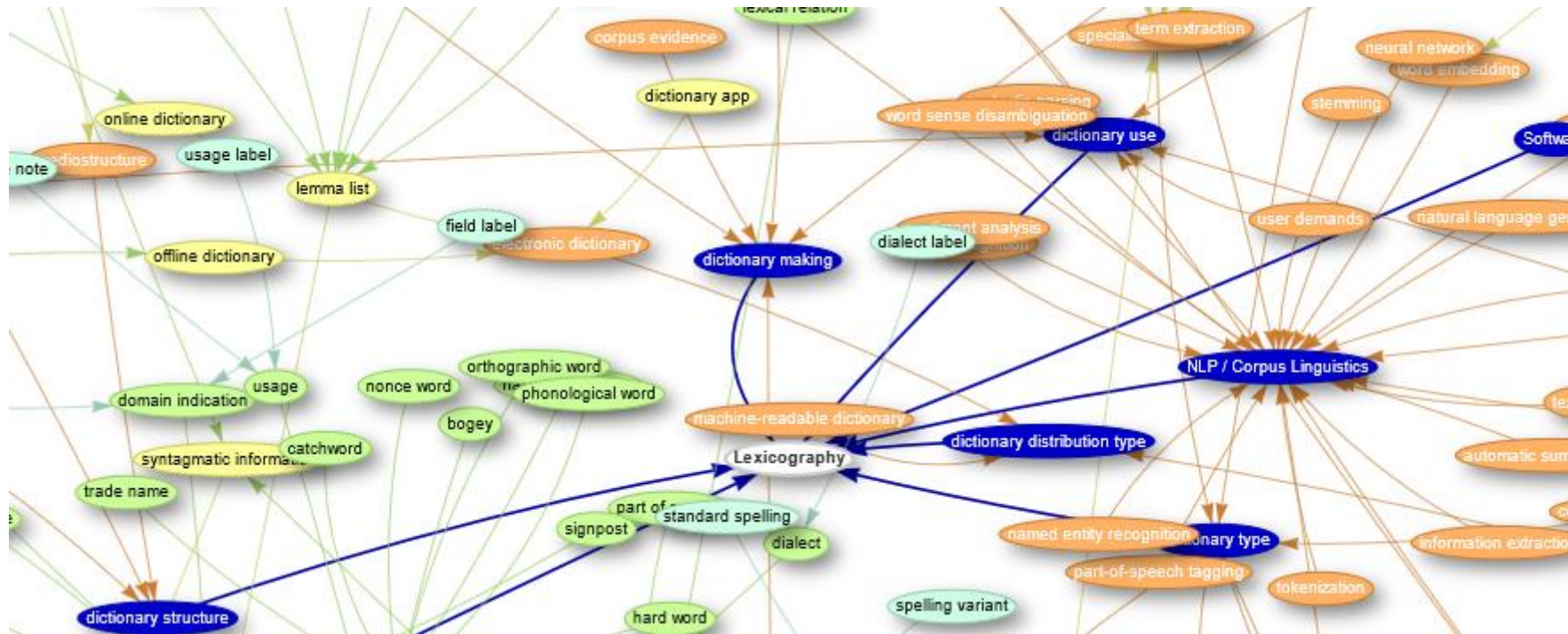
- Sue Atkins Q20324
- a Person

Match New Item Don't Reconcile Cell Cancel

LexVoc Vocabulary of Lexicographic Terms

- Sources
 - Updated and extended version of “Bibliografía Temática de la Lexicografía” subject index (Córdoba Rodríguez 2003), translated to English
 - The typology of dictionaries by Engelberg and Storrer (2016)
 - The glossary of lexicographic terms by Kipfer (2013)
 - Index of the volume “Using Online Dictionaries” (Müller-Spitzer 2014)
 - “Linguistic Property” branch of the GOLD ontology (linguistics-ontology.org)
 - Term extraction: Elexifinder version 2 English corpus (SketchEngine)
- Today, around 500 validated terms in SKOS tree
 - Validation: Attestation of term lexicalizations in LexBib English corpus

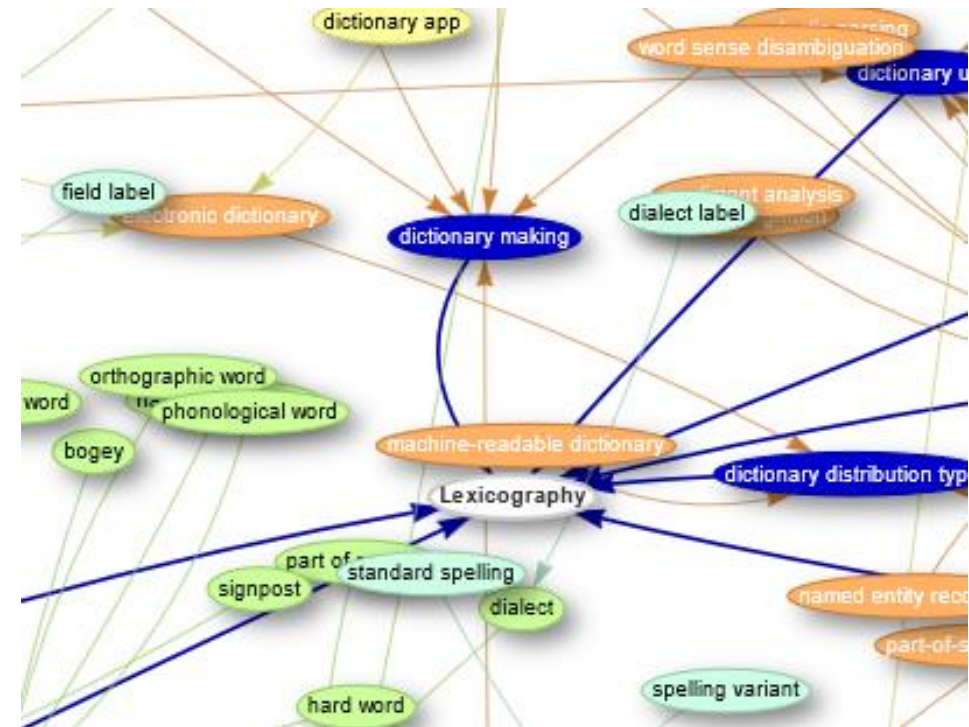
LexVoc Vocabulary of Lexicographic Terms



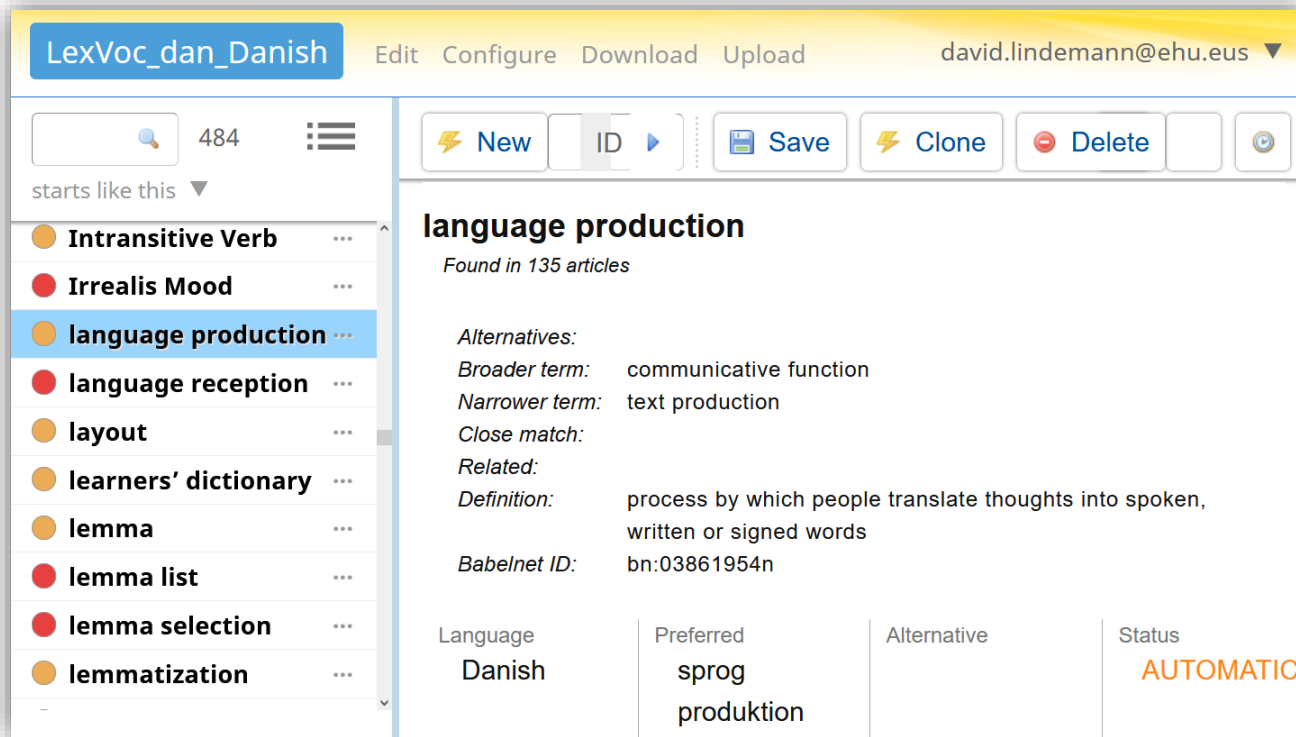
- SKOS vocabulary: Tree structure with “Lexicography” as root
- Description: <https://lexbib.elex.is/wiki/LexVoc>

LexVoc: Structure

- Root “Lexicography”
- Main branches
 - Dictionary Structure
 - Dictionary Type
 - Dictionary Making
 - Dictionary Use
 - Dictionary Function
 - Dictionary Distribution Type
 - Linguistic Property
 - NLP / Corpus Linguistics
- Discussion, see: <https://meet.elex.is/groups/lexicographic-concepts-vocabulary/>



LexVoc: Translation



The screenshot shows the LexVoc interface for the 'LexVoc_dan_Danish' project. The top bar includes 'Edit', 'Configure', 'Download', and 'Upload' buttons, along with the user email 'david.lindemann@ehu.eus'. A search bar on the left shows '484' results. A sidebar on the left lists various linguistic categories, with 'language production' selected. The main area displays details for 'language production', including its definition, alternatives, and a table of translations.

language production
Found in 135 articles

Alternatives:
Broader term: communicative function
Narrower term: text production
Close match:
Related:
Definition: process by which people translate thoughts into spoken, written or signed words
Babelnet ID: bn:03861954n

Language	Preferred	Alternative	Status
Danish	sprog produktion		AUTOMATIC

- 39 languages (Elexis languages)
- Call for collaboration
- Drafted translations from BabelNet, Wikidata
- Translation editing process: Lexonomy app at Elexis
- 38 bilingual dictionaries, see https://lexbib.elex.is/wiki/LexVoc_Lexonomy
- Discussion, see <https://meet.elex.is/groups/lexvoc-translation/>

Summary, Outlook

- Elexifinder / LexBib as a service to the community
 - Elexifinder: <https://finder.elex.is>
 - LexBib: <https://lexbib.elex.is>
- Open calls for collaboration
 - Proposals for more publications to include <https://meet.elex.is/discussions/topic/elexifinder-new-research-publications-suggestions/>
 - Discussion of LexVoc structure <https://meet.elex.is/groups/lexicographic-concepts-vocabulary/>
 - LexVoc translation <https://meet.elex.is/groups/lexvoc-translation/>

