

## CELEX

Dit expertisecentrum heeft lexicale databestanden ontwikkeld voor het Nederlands, het Engels en het Duits. Deze bestanden bevatten uitgebreide informatie over de orthografie, fonologie, morfologie, syntaxis en frequentie van woorden, maar bijvoorbeeld geen enkele informatie over hun betekenis. De CELEX data kunnen worden gebruikt in verschillende typen taalkundig onderzoek en taalkundige experimenten.

When starting to use any of the three databases, the user first has to choose between three so-called 'lexicon types':

- lemma lexicon
- wordform lexicon
- corpus type lexicon

Each lexicon type uses a specific kind of entry. The CELEX lemma lexicon is the one most similar to an ordinary dictionary since every entry in this lexicon represents a set of related inflected words. In a lexicon, a lemma can be represented by using a headword (cf. traditional dictionary entries) such as, for example, 'call' or 'cat'. The wordform lexicon yields all possible inflected words: every entry in the lexicon is an inflectional variant of the related headword or stem. So, a wordform lexicon contains words like 'call', 'calls', 'calling', 'called', 'cat', 'cats' and so on. A corpus type lexicon, on the other hand, simply gives you an ordered list of all alphanumeric strings found in the corpus with raw string counts, undisambiguated for relations to either lemmas or wordforms.

For all types of lexicons, the user may subsequently select any number of columns -- from a total of approximately 950(!) database columns -- combining information on the orthography, phonology, morphology, syntax and frequency of the entries.

- Orthography
  - with or without diacritics
  - with or without word division positions
  - alternative spellings
  - number of letters/syllables
- Phonology
  - phonetic transcriptions (using SAM-PA notation or Computer Phonetic Alphabet (CPA) notation) with:
    - syllable boundaries
    - primary and secondary stress markers
  - consonant-vowel patterns
  - number of phonemes/syllables
  - alternative pronunciations
  - frequency per phonetic syllable within words
- Morphology
  - Derivational/compositional:
    - division into stems and affixes
    - flat or hierarchical representations
  - Inflectional:
    - stems and their inflections
- Syntax
  - word class
  - subcategorisations per word class
- Frequency
  - disambiguated for homographic lemmata

An arbitrary query using a small English lemma lexicon (that is, one with very few columns) might yield the following result:

SHOW					
Headword	PhonStrsCPA	MorphStructure	MorphC	Cl	Freq
celebrant	˘sE. 11. br@nt	((celebrate), (ant))	Vx	N	6
celebrated	˘sE. 11. bre/. tId	((celebrated))	V	A	158
celebration	"sE. 11. ˘bre/. Sn,	((celebrate), (ion))	Vx	N	221
celibacy	˘sE. 11. b@. sI	((celibate), (cy))	Ax	N	13
celibate	˘sE. 11. b@t	((celibate))	A	N	2
cell	˘sE. 1	(cell)	N	N	1216
cellar	˘sE. 1@r*	(cellar)	N	N	225
cellarage	˘sE. 1@. rIJ/	((cellar), (age))	Nx	N	0
cellist	˘T/E. 1Ist	((cello), (ist))	Nx	N	6
cello	˘T/E. 10/	(cello)	N	N	36

START GOTO ZOOM HIDE COUNT PRINT SAVE QUERY

Page: 1 (2) Columns: 6 (6) Tempo: 10 Count: ^

An example selection from a small English wordform lexicon, showing the inflectional variants of the headwords given in the previous example, is presented in the next table:

SHOW						
Word	WordSyll	PhonStrsCPA	Cl	File	TransIn	Freq
celebrant	cel-e-brant	˘sE. 11. br@nt	N	S	@	2
celebrants	cel-e-brants	˘sE. 11. br@nts	N	P	@+s	4
celebrate	cel-e-brate	˘sE. 11. bre/t	V	eP	@	48
celebrated	cel-e-brat-ed	˘sE. 11. bre/. tId	A	b	@	158
celebrated	cel-e-brat-ed	˘sE. 11. bre/. tId	V	aP	@+d	24
celebrated	cel-e-brat-ed	˘sE. 11. bre/. tId	V	pa	@+d	24
celebrates	cel-e-brates	˘sE. 11. bre/ts	V	e3S	@+s	21
celebrating	cel-e-brat-ing	˘sE. 11. bre/. tIN	V	pe	@-e+ing	76
celebration	cel-e-bra-tion	"sE. 11. ˘bre/. Sn,	N	S	@	156
celebrations	cel-e-bra-tions	"sE. 11. ˘bre/. Sn, z	N	P	@+s	65

START GOTO ZOOM HIDE COUNT PRINT SAVE QUERY

Page: 1 (2) Columns: 7 (7) Tempo: 10 Count: ^

Bron: [http://www.kun.nl/celex/subsecs/section\\_rep.html](http://www.kun.nl/celex/subsecs/section_rep.html) [20-11-2001]