

**Notes on TNC data:**

1. All type and lemma lists are based on TNC written component.
2. Type lists are based on unique running words, be bare or inflected. All possible collocates and resulting multi-word units are excluded.
3. All ambiguities and homographs are preserved in the lemma lists. Frequency values of these lemmas are of all possible homographs and functions included in the bare or inflected forms of the given entry.
4. Proper Nouns, abbreviations, onomatopoeia, digits, interjections (e.g. fixed expressions) and foreign words are not included in any of the lists.
5. All lists are ranked by frequency and published under 5 groups, namely type, lemma, POS based lemma, thematic lists and suffixes (including verbal/nominal paradigms and suffix combination rankings).