Danish orthography for folks and robots

Peter Juel Henrichsen
Det Danske Sprognævn (DSN)

In November 2024, the Danish Language Council (DSN) released the 5th edition of Retskrivningsordbogen (RO), the lexicon defining the official Danish orthography. RO has two volumes, a dictionary (64.000 lemmas) and a manual (principles of interpunction, abbreviation, listing, numbering, emphasis, and more). As a novelty, the dictionary part of RO5 is now available as a database called COR1 (Det Centrale Ordregister, level 1). In COR1, each lemma and each wordform carry an id, as shown in these samples:

COR1-id	form	inflextion (RO terms)
COR.73562.110	sky	sb.fk.sg.ubes
COR.22093.302	sky	adj.sg.best
COR.32735.200	sky	vb.inf.akt
COR.86598.110	klor	sb.fk.sg.ubest
COR.86598.120	klor	sb.itk.sg.ubest
COR.86598.111	kloren	sb.fk.sg.best
COR.86598.121	kloret	sb.itk.sg.best

Notice that "sky" is ambiguous in Danish between a noun (*cloud*, with lemma index 73562 and inflexion index 110), an adjective (*shy*, 22093.302) and a verb (*avoid*, 32735.200), but effectively disambiguated by the COR-id.

COR1 is easy to import in programs (far easier than a book), but even more importantly, the COR framework also accommodates other dictionaries allowing lexical information to flow freely among all. DSN and the National Committee of Chemical Nomenclature are currently working on COR.TALE, a COR-compatible dictionary covering the Danish chemical terms, which in certain cases diverge from the RO spelling.

COR.KEMI.086598.120	chlor	sb.itk.sg.ubest
COR.KEMI.086598.121	chloret	sb.itk.sg.best

The element *chlorine* is spelled "klor" in RO while as a chemical term it is "chlor" (compare COR.86598 and COR.KEMI.086598 in the examples). The RO lemma accepts common gender (111) as well as neuter gender (121) while the professional jargon accepts neuter only (121). The fine-structure of the COR-ids in COR1 and COR.KEMI allows a COR-based text tagger or a grammar checker (or any other language technological appliation) to combine the relevant pieces of lexical information in the inter-linked dictionaries.

DSN is currently working on a COR-based application for automatic text annotation called CLINK (short for COR-linker). We conclude the presentation with a (very) short introduction.

COR1 is available at ordregister.dk (under licence CC0, i.e. free for all).