ISOcat
www.isocat.org

Prologue:
Corpus Spoken Dutch (CGN)

CGN?

CGN *: **C**orpus **G**esproken **N**ederlands*

(Corpus of Spoken Dutch)

- Also used for written texts (with a few additional tags)
- Dutch: language spoken in the Netherlands and northern part of Belgium (Flanders)

Today: **Part of Speech** (PoS) annotation

(verb, noun, adjective, …)

Over the years: **CGN** = *de facto* standard for Dutch

Tag sets for Dutch

Another tag set for Dutch: Parole

Different theoretical 'frames', short vs full version

N(x,y,z) or NC / NP

common noun

proper noun

N

Question: how to relate such tag sets?

ISOcat to the rescue?

ISOcat
www.isocat.org

Complex tags

- PAROLE
  - VMIP3S0

  (Verb Modal Indicative Present Third_person Singular)

  hij kan. Kan hij?
- CGN
  - WW(pv,tgw,met-t)

  (verb, finite, present,has-t)

  hij kom**t** kom**t** hij

  jij kom**t**

  jij kunt

  hij kan. Kan hij?

  kom jij?

  kun jij?

ISOcat
www.isocat.org

Some characteristics of CGN

- Rather elaborated tags (over 320)
  - Not just: Noun (or N)
  - But: **N(soort,mv,dim)**
    - soort meaning ‘common’
    - mv meaning ‘plural’
    - dim meaning ‘diminutive’

stoeltjes	small chairs
mannetjes	little men
baby’tjes	tiny babies

ISOcat
www.isocat.org

Characteristics CGN 2

- *"The old are outnumbering the young, " so he said*
- *"Those books are very old"*

What is the PoS of *old* and *young* ?

In CGN these are considered specific forms of **adjectives** : ADJ(nom) and ADJ(vrij)

[meaning: the adjectives appear in the position where one would expect a noun or an adverb]

ISOcat
www.isocat.org

form vs function

- *"The old are outnumbering the young, " so he said*
- *"Those books are very old"*

In many other tag sets the words would be considered noun and adverbs, respectively!

What is decisive in a specific tagset? The FORM of the FUNCTION ?

Consequences for ISOcat

- What is a noun?
- What is in a specific definition meant with this notion?
- And when should this be made clear?
  - When it is the notion to be defined?
  - When the notion is used in a definition?
- Can all DCs within a domain in principle be combined?

ISOcat
www.isocat.org

How to …

►in general

- What is to be taken into account when creating a DC (Data Category)?
- What is to be defined? A full tag or its ‘building blocks’?

ISOcat
www.isocat.org

CGN Showcase

- Later this morning:

- Some guidelines: what to do (or NOT to do)
- Some examples, showing how we tried to overcome all kinds of issues