Onomasiological and Semasiological Lexicography: Past, Present, and Future

Patrick Hanks

Masaryk University, Brno

U. Bratislava, March 2007

Outline of the talk

1. The onomasiological approach: thesauruses
   – From Wilkins (1668) and Roget (1852) to WordNet
   – classifying words in a top-down ontology
   – moving towards hypertext
2. The semasiological approach: dictionaries
   – lexicography and technology
   – The dictionary as an inventory
3. The dictionary of the future
   – FrameNet and CPA
   – focus on words in context: phraseology
   – corpus analysis, syntagmatics, hypertext

The Onomasiological Tradition

- Onomasiology = the classification of concepts in a taxonomy or ontological structure.

Milestones:
   – 1668: John Wilkins: *Essay towards a Real Character and Philosophical Language*
   – 1852: Peter Mark Roget: *Thesaurus*
   – 1993: G. Miller, C. Fellbaum et al.: *WordNet*
   – 1998: P. Vossen, K. Pala, et al.: *EuroWordNet*
     - *Balkanet*

John Wilkins: *Essay towards a Real Character and a Philosophical Language*

- "The most complete project for a universal and artificial philosophical language that the 17th century was ever to produce."
- "As a preliminary step, a vast review of all knowledge, to establish [the] notions held in common by all rational beings."
   – U. Eco (1995), *The Search for the Perfect Language*

Wilkins and taxonomies

- Wilkins is sometimes called "the father of modern taxonomies"
- But W. did not distinguish scientific taxonomy from folk taxonomy, e.g.
   – *onions* and *garlic* = foodstuffs
   – *lilies* = flowers
     - Scientifically speaking, onions, garlic, and lilies are all Liliaceae.
     - Wilkins lived a century before Linnaeus (the true father of modern botanical taxonomy)

Wilkins and hypertext

- Wilkins was "groping towards the modern notion of *hypertext*" (Eco)
- He tried to say something about the conceptual relations between entities and actions, for example, as well as merely classifying entities.
- E.g. "*dogs* bark,bay, or yelp ... *wolves* howl".
   – 350 years later we are still groping towards hypertext

Wilkins on dogs (1)

- "RAPACIOUS Beasts of the DOG-Kind ... distinguished as:
  - European, Terrestrial ...
    - Bigger: ...noted for *tameness* and *docility* OR for *wildness* and *enmity to sheep*: **1a. DOG, Bitch, Puppy, Whelp** *(bark, bay, yelp).*
    - **1b. WOLF** *(howl).*
    - *Lesser; living usually in holes in the ground: ...* **2. FOX, BADGER**
  - Amphibious: 3. ... **SEAL**
  - Exotic, being noted for gregariousness: 4. **JACKALL ...**Feeding on ants: **5. ANT-BEARE, ARMADILLO**
  - **6. CARAGUYA"** [EH?!]

Wilkins on dogs (2)

- They [dogs] are chiefly distinguishable from the uses which men employ them about:
  - Delight: LAPDOGS
  - Companying: (...) CURRS
  - Custody of places or things: MASTIFS
  - Hunting, either by sight: GASE-HOUNDS
    - Smell, of *birds*: SPANIELS
    - of *beasts*: HOUNDS, ... BEAGLES
  - Swiftness: GREYHOUNDS, LURCHERS
  - Play: TUMBLERS

A Modern Hypertext Database

- Every node would be linked to numerous other nodes. E.g. for the concept *dog*
  - **Taxa:** *dog > canine > animal ... < wolves; cats*
  - **Breeds of dog:** *mastiff, beagle, labrador, corgi, etc.*
  - **Dog parts:** *body, head, tail, paws, teeth, hair, ...*
  - **Dog behaviour:** *bark, bite (as an act of aggression), lick, sniff, wag tail, hunt, guard, chase cats, ...*
  - **Dogs in history and culture***: symbiosis with man since the Neolithic; dogs in medieval castles; in paintings; culture of the lapdog; Hundewiese*
  - **and many other conceptual links**

WordNet on dogs

- **hypernym tree 1:** *canine > carnivore > placental mammal > mammal > vertebrate > chordate > animal > organism > living thing > whole > object > physical entity > entity*
- *hypernym tree 2: domestic animal > animal*
- **hyponym tree: 186 hyponyms** (!)
  - mostly breeds, e.g. *spaniel, beagle, dachshund*
  - but also designations: e.g. *puppy, pooch, cur, mongrel, lapdog*

Syntagmatics in Ontologies

- Italian Wordnet s.v. *cane* 'dog':
  - [role_agent] [v] *abbaiare* '*bark*'
  - [role_agent] [v] *braccare* 'hunt'

- Czech Wordnet s.v. *pes* 'dog':
  - link to Verbalex argument structures, e.g.
    - *pes štěká* 'a dog is barking'

Classification and Definition

- A classification is in no way equivalent to a definition.
- We are all prisoners of our time and our language.
  - Wilkins thought he was classifying universal concepts.
  - He was actually classifying the 17th-century Oxford Anglocentric world view.

Borges' warning to ontologists

In the *Celestial Emporium of Benevolent Knowledge*, animals are classified as:

a) those that belong to the Emperor,
b) embalmed ones,
c) those that are trained,
d) suckling pigs,
e) mermaids,
f) fabulous ones,
g) stray dogs,
h) those included in the present classification,
i) those that tremble as if they were mad, j) innumerable ones,
k) those drawn with a very fine camelhair brush,
l) others,
m) those that have just broken a flower vase,
n) those that resemble flies from a distance.

13

The semasiological approach to language

- **Dictionaries:**
  – concerned with words and word use, rather than with the classification of concepts
  – information on orthography, pronunciation, inflections, syntactic class, etymology, as well as meaning and other information
  – a **bottom-up** approach (without much "**up**")
- **Lexicographers**: glorified inventory clerks

14

The development of lexicography is dependent on development in technology

- 1450: Gutenberg invents the printing press
- c. 1498: Henri Estienne, printer in Paris
- 1532: Robert Estienne, "printer to the king" publishes a Latin-French dictionary
- 1474: William Caxton, printer in London
- 1530: Palsgrave, tutor to Princess Mary, publishes an English-French dictionary

15

From Johnson 1755

ALCOHOL,

An Arabick term used by chymists for a high rectified dephlegmated spirit of wine, or for anything reduced into an impalpable powder.
Quincy.

If the same salt shall be reduced into *alcohol*, as the chymists speak, or an impalpable powder, the particles and intercepted spaces will be extremely lessened.
Boyle.

Sal volatile oleosum will coagulate the serum on account of the *alcahol*, or rectified spirit, which it contains.

Arbuthnot.

16

Principles of Definition Writing

- **Three main choices:**
- Perfunctory, minimal hints (Duden, Chambers)
- Leibnizian substitutability (OED, Collins, ODE)
  – *"Eadem sunt quorum unum potest substitui alteri salva veritate."*
- Define words in context (Cobuild)
  – Phraseology disambiguates.

17

Random House Dictionary (1966)

alcohol …**1.** … a colorless, limpid, volatile, flammable, water-miscible liquid, C_2H_5OH, having an etherlike odor and pungent, burning taste, the intoxicating principle of fermented liquors, produced by yeast fermentation of certain carbohydrates, as grains, molasses, starch, or sugar, or obtained synthetically by hydration of ethylene or as a by-product of certain syntheses: used chiefly as a solvent in the extraction of specific substances, in beverages, medicines, organic synthesis, lotions, tonics, colognes, rubbing compounds, as an automobile radiator antifreeze, and as a rocket fuel. **2.** whiskey, gin, vodka, or any other intoxicating liquor containing this liquid. **3.** *Chem.* Any of a class of compounds …having the general formula ROH, where R represents an alkyl group and –OH a hydroxyl group…

18

Oxford Advanced Learner's Dictionary, 6th edition (2000)

alcohol *noun* [U] **1** drinks such as beer, wine, etc., that can make people drunk: *He never drinks alcohol.* | *Alcohol abuse* **2** the colourless liquid that is found in drinks such as beer, wine, etc., and is used in medicines, cleaning products, etc.: *Wine usually contains about 10% alcohol* | *levels of alcohol in the blood* | *He pleaded guilty to driving with excess alcohol* | *Low-alcohol beer* | *Choose an alcohol-free skin toner if you have dry skin*

- *This slide and the previous slide show Leibnizian, Aristotelian definitions.*

Cobuild (1987): 1

- When a horse **gallops**, it runs very fast so that all four legs are off the ground at the same time.
- If something **displeases** you, it makes you feel annoyed or rather angry.
- The **monsoon** is the season in southern Asia when there is a lot of very heavy rain.
  - Here the definienda are phraseological, not words in isolation

Cobuild (1987): 2

alcohol
- Drinks that can make people drunk, such as beer, wine, and whisky, can be referred to as **alcohol.** EG *Do either of you smoke cigarettes or drink alcohol? ... No alcohol is allowed on the premises.*
- **Alcohol** is a colourless liquid that is found in drinks such as beer, wine, and whisky. It is also used in products such as perfumes and cleaning fluids. EG *a low-alcohol beer ... Products for dry skin have little or no alcohol.*

Exploiting the potential of phraseological definitions

- If you say that someone speaks in **monosyllables**, you mean that they say very little.
- If you call a woman a **bitch**, you are saying in a very rude way that she is an unpleasant person.
- If you describe a situation as a **bitch**, you mean that it is difficult to deal with.
- A **bitch** is female dog.

  – adapted from Cobuild

The Word Sketch Engine

alcohol *(as modifier)*	**BNC freq.**	**Salience score**
alcohol consumption	*131*	*34.0*
alcohol abuse	*114*	*31.3*
alcohol intake	*53*	*18.2*
alcohol misuse	*23*	*17.7*
alcohol content	*35*	*15.3*
alcohol problem	*38*	*11.3*
alcohol dependency	*5*	*10.1*
alcohol dependence	*7*	*9.2*

Some issues in definition writing

- A verbless phrase or a full sentence (as in *Cobuild*)?
- Defining the 'essence' of something, or characterizing what it **typically** is?
- How much technical detail to include?
- How to represent normal phraseology and collocations?
- What is the role of examples? How chosen?
- How to express changing beliefs, scepticism (e.g. about *alchemy*)?
- How to relate figurative extensions to the 'literal' meaning (e.g. *alchemy* = a former science, but also apparent magic; alcohol = strong drink as well as a chemical)?

Deciding what is a sense

complain *v.i.* **1.** express dissatisfaction or annoyance about a state of affairs or an event. **2.** (**complain of**) state that one is suffering from (a pain): *he began to complain of headaches.*

- This could be: "express dissatisfaction about a state of affairs or a pain that one is suffering from" BUT:

complaint *n.* **1.** A statement that a situation is unsatis-factory or unacceptable. **2.** An illness or medical condition: *a skin complaint.*

- Words and senses interact with one another. Dictionaries have their own kind of textual discourse structure. They are not just lists.

25

How much grammatical information should a dictionary include?

- Problem: the (English) public are resistant to grammar.
- ODE uses [with obj.], [no obj.], [with adverbial], with obj. and adverbial], etc. (It dared not go much farther)
- ODE uses various other devices, including highlighting **common phraseology** in examples.
- A hypertext dictionary can go into much more detail.
- A new, lexically based grammar of English (especially of adverbials) is needed.

26

Differences between FrameNet and the CPA Pattern Dictionary

- FrameNet (http://framenet.icsi.berkeley.edu/) explores semantics of frames in depth.
- FrameNet focuses on the various roles in events
- CPA focuses on (more or less shallow) links between word meaning and word use
- CPA gives criteria for distinguishing one sense of a word from another, using a shallow ontology
- FrameNet proceeds frame by frame; CPA proceeds verb by verb.
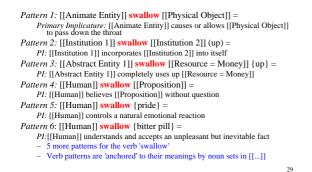- In due course CPA and FrameNet will complement each other.

27

Lexicography of the Future

- Demonstration: *English Pattern Dictionary*
  - *eat, drink, chew, swallow*
  - *calm*
    - *Note that verb arguments are expressed as onomasiological classifications.*
    - *Computer technology will enable a new, hypertext approach to dictionary making, using both onomasiological and semasiological techniques*

28

CPA entry for 'swallow'

Pattern 1: [[Animate Entity]] **swallow** [[Physical Object]] =
  Primary Implicature: [[Animate Entity]] causes or allows [[Physical Object]] to pass down the throat
Pattern 2: [[Institution 1]] **swallow** [[Institution 2]] (up) =
  PI: [[Institution 1]] incorporates [[Institution 2]] into itself
Pattern 3: [[Abstract Entity 1]] **swallow** [[Resource = Money]] {up} =
  PI: [[Abstract Entity 1]] completely uses up [[Resource = Money]]
Pattern 4: [[Human]] **swallow** [[Proposition]] =
  PI: [[Human]] believes [[Proposition]] without question
Pattern 5: [[Human]] **swallow** {pride} =
  PI: [[Human]] controls a natural emotional reaction
Pattern 6: [[Human]] **swallow** {bitter pill} =
  PI:[[Human]] understands and accepts an unpleasant but inevitable fact
  – 5 more patterns for the verb 'swallow'
  – Verb patterns are 'anchored' to their meanings by noun sets in [[...]]

29

Aspects of lexicography of the future

- Move on from Aristotelian-Leibnizian preconceptions. Analyse the evidence from large corpora and find out what's really going on in a language.
- First establish the normal context(s), then define the meaning. Don't try to define words in isolation.
- Get the argument structure right, plus correct levels of generalization – and noun sets will "fall into place", e.g.
  - [[Human]] hazard {guess}
  - [[Event]] {take place}
  - [[Horse]] gallop [NO OBJ]
- We need a lexically based grammar, especially of adverbials (the third argument), e.g.
  - [[Human]] put [[Physical Object]] [Adv[Location]]

30