Workshop on corpus Pattern Analysis

Patrick Hanks
Institute of Formal and Applied Linguistics,
Charles University in Prague, Czech Republic

Patterns

• When you first open a concordance, very often some patterns of use leap out at you.
  – Collocations make patterns: one word goes with another
  – To see how words make meanings, we need to analyse collocations
• The more you look, the more patterns you see.
• BUT
• When you try to formalize the patterns, you start to see more and more exceptions.
• The boundaries are fuzzy and there are many outlying cases.

Analysis of Meaning in Language

• Analysis based on predicate logic is doomed to failure:
  – Words are NOT building blocks in a ‘Lego set’
  – A word does NOT denote ‘all and only’ members of a set
  – Word meaning is NOT determined by necessary and sufficient conditions for set membership
• Instead, a prototype-based approach to the lexicon is necessary:
  – mapping prototypical interpretations onto prototypical phraseology

The linguistic ‘double-helix’ hypothesis

• A language is a system of rule-governed behaviour.
• Not one, but TWO (interlinked) sets of rules:
  1. Rules governing the normal uses of words to make meanings
  2. Rules governing the exploitation of norms

Exploitations

• People exploit the rules of normal usage for various purposes:
  • For **economy** and **speed**:
    – Conversation is quick
    – Listeners (and readers) get bored easily
    – Words that are ‘obvious’ can sometimes be omitted
  • To **say new things** (reporting discoveries)
  • To **say old things in new ways**
  • For rhetoric, humour, poetry, politics …

Lexicon and prototypes

• Each word is typically used in one or more patterns of usage (valency + collocations)
• Each pattern is associated with a meaning:
  – a meaning is a set of prototypical beliefs
  – In CPA, meanings are expressed as ‘anchored implicatures’.
  – Few patterns are associated with more than one meaning.
• Corpus data enables us to discover the patterns that are associated with each word.
What is a pattern? (1)

• The verb is the pivot of the clause.
• A pattern is a statement of the clause structure (valency) associated with a meaning of a verb, together with the typical semantic values of each argument.
• Different semantic values of arguments activate different meanings of the verb.

What is a pattern? (2)

• [Human] fire [Firearm]
• [Human] fire [Projectile]
• [Human 1] fire [Human 2]
• [Anything] fire [Human] {with enthusiasm}
• [Human] fire [NO OBJ]
• Etc.

Semantic Types and Ontology

• Items in double square brackets are semantic types.
• Semantic types are being gathered together into a shallow ontology.
  – (This is work in progress in the current CPA project)
• Each type in the ontology will (eventually) be populated with a set of lexical items on the basis of what’s in the corpus under each relevant pattern.

Shimmering lexical sets

• Lexical sets are not stable – not „all and only‟.
• Example:
  – [Human] attend [Event]
  – [Event] = meeting, wedding, funeral, etc.
  – But not thunderstorm, suicide.

Meanings and boundaries

• Boundaries of all linguistic and lexical categories are fuzzy.
  – There are many borderline cases.
• Instead of fussing about boundaries, we should focus instead on identifying prototypes.
• Then we can decide what goes with what
  – Many decision will be obvious.
  – Some decisions – especially about boundary cases – will be arbitrary.

The Idiom Principle (Sinclair)

• In word use, there is tension between the „terminological tendency” and the „phraseological tendency”:
  – The terminological tendency: the tendency for words to have meaning in isolation
  – The phraseological tendency: the tendency for the meaning of a word to be activated by the context in which it is used.