Phraseology, meaning, and the future of lexicography

Patrick Hanks
Professor in Lexicography
University of Wolverhampton
Talk outline

• Dictionaries are dead.
  – The business model on which funding was based (anticipation of sales of printed products) is defunct.
  – Anyway, meanings are context-dependent

• But there is still a role for lexicographers.
  – Discovering and explaining word meaning
  – Cannot be left to (speculative) linguists
  – Nor to Wiktionary

• Lexicographers of the future will necessarily focus on discovering and explaining phraseological patterns.

• Question: Why focus on phraseology?
  – In order to process meaning.
Philosophy and meaning

• Grice (1957) posited that meanings are not just in the head.
  – They are events; interactions between people:
  – between speaker (S) and hearer (H);
  – (and with displacement in time) between writer and reader

• For this to work, S and H must share a body of linguistic conventions having the same meanings.

• Neither Grice nor anyone else has specified what these conventions are.

• We need to explore how to specify, for every language, the shared linguistic conventions on which meaningful communication depends.
Do Words have meaning?

• Can we get good evidence for meaning and phraseology by consulting our intuitions?
• Let’s think of a word.

• What’s the meaning of blow?
The meaning potential of a word

• What’s the meaning of **blow**? --
  – What the wind does? A disappointment? Something you do with your fist? With your nose? Or with a whistle?
  – Or is it a verb?
• What’s the meaning of **blow up**?
  – Destroy a building? What you do to a balloon? Lose your temper?

All of these things and more! Words are hopelessly ambiguous.

But put a word in context, and the ambiguity is reduced or eliminated.

Strictly speaking, words in isolation don’t have meaning; they only have **meaning potential**.

Different aspects of a word’s meaning potential are activated in different contexts.
The idiomaticity of words

- What are the meaning potentials of blow up?
- To answer, we need to sort the phraseology into patterns:
  - They blew up [bridges/buildings/tanks]. [a normal, idiomatic p.v.]
  - She blew up some balloons. [a normal, idiomatic p.v.]
  - You can blow up your vest once you are outside the aircraft. [Aeroflot announcement; grammatical and logical, but not natural]
- Is it idiomatic to say, “A wind blew up”?
  - ANSWER: Yes, but it’s rare.
  - “A hurricane blew up” is a bit more idiomatic.
  - “A light breeze blew up” is not idiomatic.
  - Advice to students: it’s OK to talk about “blowing things up” in a military context, but better to avoid talking about any kind of wind “blowing up”.
Prototypical patterns for *blow*, verb

The *Pattern Dictionary of English Verbs (PDEV)* has 77 patterns for the verb *blow*. Most frequent—showing SPOCA interdependencies (clause roles) and some stereotypical arguments—are:

- 12% [the wind/S] *blows/P* [~/O] ([Direction/A])
- 6% [the wind or an explosion/S] *blows/P* [something/O] [somewhere/A]
- 14% [a bomb (or a person using explosive)/S] *blows/P* [something/O] [up/A]
- 4% [a building/vehicle/ship/S] *blew/P* [~/O] [up/A]
- 3% [a disagreement/S] *blew/P* [~/O] [up/A]
- 4% [the wind (or an explosion)/S] *blew/P* [something/O] [off/A]
- 2% [an explosion/S] *blew/P* [the windows/O] [out/A]
Idioms are patterns

An idiom is a pattern at least one of whose arguments is populated by a very small lexical set

- *Something blew the project off course* [= wrecked it]
- *This move will blow the cobwebs away* [= get rid of old ideas]
- *He likes to blow his own trumpet* [= boast]
- *To blow the whistle on the government* [= expose wrongdoing]
- *She was blowing hot and cold* [= was indecisive]
- *He blew his top* [= lost his temper]
- *He blew a lot of his money on gambling* [= spent]
- *Lawrence blew my cover* [= he revealed the truth about me]

There are many more.
Semantic invariants? Necessary and sufficient conditions?

- Wierzbicka’s “Natural Semantic Metalanguage” (NSM) postulates that the core meaning of each word is defined by a semantic invariant.
  - What is the semantic invariant of the verb *blow*?
  - Does such a thing exist?

- What are the necessary and sufficient conditions (N&SCs) that define a “blowing” event?

- Is this even the right sort of question to ask?
Instead of seeking “the invariant”

• Don’t assume that all uses of a word must have something in common, but *look and see* how words are used. – L. Wittgenstein

• We build concepts around prototypes, not by defining conditions. – E. Rosch

• The stereotypical tiger is a fierce wild four-legged animal with black stripes. But a tame three-legged albino tiger is still a tiger. – H. Putnam

• Many if not most meanings require the presence of more than one word for their normal realization. – J. M. Sinclair
What’s the use of semantic invariants and N&SCs?

• Scientific and technical concepts need to be defined (i.e. their meaning must be stipulated). How is this done?
  – By using the ordinary words of natural language in their most ordinary senses.
  – But those ordinary natural meanings are different in kind from defined terminology in technical domains.

• There has been colossal confusion for over 300 years between the meaning of technical terms and meaning in natural language. We now know that:
  – To understand your meaning of an ordinary word when you use it, I need to know about its relationships with other words, with the world, and with other language users.
The need for a new kind of resource

• Trying to account for all possible uses of a word such as blow is impossible.

• But accounting for the normal phraseology of a word (and building from there) is quite possible.
  – Basic norms (patterns) can be collected, creating a corpus-driven dictionary of phraseology and collocations.
  – such a dictionary does not yet exist
  – In Wolverhampton, we are building one (www.pdev.org)

• Language learners and computer programs alike need to learn these basic patterns ("norms"), but they also need to know how the norms are exploited creatively.
Where to start?

- Start with verbs
  - and predicative adjectives (e.g. *I am happy to see you*)
- The verb is the pivot of the clause
  - We make conversation by using clauses
- Nouns are different
  - nouns need a different kind of analytic mechanism
  - Bilingual dictionaries are useful in helping learners or translators find the right noun, getting the gender and spelling right, etc.
  - Adjectives are also different (not part of this talk).
Corpus Pattern Analysis (CPA)

• We need not just a dictionary with word meanings, but also:
  – an inventory of normal contexts for each word;
  – A set of rules stating how each context a) is used normally or b) can be exploited creatively.

• CPA aims, by careful analysis of data, to establish:
  – An inventory of normal phraseological conventions
  – The meaning (semantics and pragmatics) associated with each phraseological norm.

• Out of this arises a new theoretical approach – the Theory of Norms and Exploitations (TNE)
Semantic Types

• Understanding text meaning depends on analysis of collocations and their variants
  – Groups and sets of collocates [example from R. Moon]:
    • shivering in her shoes / quaking in his boots / shaking in their sandals

• Lexical sets are grouped according to semantic type
  – In this example, the noun semantic type is [[Footwear]]
The CPA “Ontology”

A hierarchical inventory of 220 semantic types. Top types:

- [[Entity]]
  - [[Physical Object]]
    - [[Human]]
    - [[Animal]]
    - [[Artefact]]
  - [[Abstract Entity]]
    - etc.

- [[Eventuality]]
  - [[Event]]
    - [[State of Affairs]]
      - etc.

The semantic types of nouns disambiguate the verbs with which they are used.
Some implications of all this (1)

- **Nouns** (typically) are referring expressions.
  - They represent concepts (and the world).
  - They ‘plug into’ verbs.
- **Verbs** are ‘power sockets’:
- Plug a noun (or 2, or 3) into a verb, and you can make a meaning, i.e.
  - construct a proposition
  - ask a question
  - interact socially
  - etc.
Some implications of all this (2)

- Meanings in language are associated with words in prototypical phraseological patterns (not only with words in isolation).
- Meanings in text are interpreted by pattern matching
  - i.e. mapping bits of text onto the patterns in our heads
  - The patterns in our heads come from ‘lexical priming’ (Hoey 2005)
  - Members of a language community share primed patterns
- Some uses match well onto patterns; these are ‘norms’
- Some uses seem surprising; these are ‘exploitations of norms’ [but some are mistakes].
- For each language, a corpus-driven lexical database will identify the normal phraseology associated with each word
- A set of exploitation rules is needed to explain creative usage.
A “double-helix” theory of meaning in language

• A human language is a system of rule-governed behaviour
  – But not one, monolithic rule system.

• Rather, it is two interlinked systems of rules:
  – 1) Rules governing normal usage
  – 2) Rules governing exploitation of norms.

• The two systems interact, producing new norms:
  – Today’s exploitation may be tomorrow’s norm.
What are meanings?

HYPOTHESIS: Meanings are events.
- Meanings are created and understood by **pattern matching:**
  - i.e., people are constantly subconsciously matching word uses in texts and conversations (*parole*) with patterns of word use that are present in the language at large (*langue*)
  - In order to make sense of what they read or hear.

Pattern matching is going on in your head all the time, while you speak and write, or listen and read.

- But we don’t (yet) know what the patterns are!
Meaning in langue and parole

• Monolingual lexicographers have at least some of the skills needed to compile an inventory of conventional phraseological patterns in any langue.

• But, to achieve this, they need to approach the problem indirectly
  – i.e. to find recurrent phraseological patterns of any word, and
  – to associate a meaning with each phraseological pattern
  – Not just with each word in isolation.
A plethora of phraseological dictionaries?

Provided that sufficient corpus evidence is available, the corpus-driven lexicographical techniques described in this paper can be applied to:

– Pedagogical dictionaries

– Bilingual dictionaries
  • E.g. There is no French or Italian equivalent of *file* as a verb of movement, so how to translate “the jury filed back into court”?

– Period dictionaries (historical dictionaries)
  • E.g. People would like to know whether a particular phrase used by Shakespeare was coined by him or was part of the general convention of Tudor English

– Domain-specific dictionaries

– Single-author dictionaries