Norms and Exploitations in Word Use: Corpus Pattern Analysis

Patrick Hanks
Research Institute of Information and Language Processing

University of the West of England, Bristol

patrick.w.hanks@gmail.com
Talk outline

• Background questions
  • What is meaning? How does language work?
  • Do words have meanings?
  • They have patterns – quantifiable patterns of use

• We propose a new approach to processing lexis and meaning, based on corpus pattern analysis (CPA):
  – First find the patterns in which each word is embedded
  – Attach meanings to patterns (constructions), not words

Creating a resource for use in:

• processing text by matching occurrences to patterns:
  – Crudely: best match wins
Challenging basic assumptions:
Do words have meaning?

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? Your nose? Or a whistle? Spend a lot of money? …

• What’s the meaning of *blow up*?
  - Destroying a building? What you do to a balloon? Lose your temper? Start to become publicly notorious? …

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

• They have meaning potential but not meaning.
  - The meaning potential of many words is a big messy soup of components and connections, not something neat and precise.

But put a word in context, and its ambiguity is reduced or eliminated.
Prototypical patterns for *blow*, verb

[54 patterns for *blow*, verb.] The main ones are:

- 12% the wind [breeze, gale, …] blows (+ direction)
- 6% the wind, or an explosion, blows something somewhere
- 14% a bomb or a person using explosive blows something up
- 4% the ship (house, tin, etc.) blew up
- 3% a disagreement blew up (between them)
- 4% the wind (or an explosion) blew something off
- 2% an explosion blew the windows out
Some idioms for *blow*, verb

- *Something blew the project off course* [wrecked it]
- *This will blow the cobwebs away* [get rid of useless old ideas]
- *He likes to blow his own trumpet* [boast]
- *She felt she had a duty to blow the whistle on the government* [expose wrongdoing]
- *He blew his brains out* [killed himself]
- *She was blowing hot and cold* [was indecisive]
- *Lawrence blew my cover* [revealed]

- Idiom meanings are not compositional
- The meanings of most clauses are quasi-idiomatic
Corpus Pattern Analysis (CPA)

• To understand the meaning of texts, 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 is either a) used normally or b) exploited to make metaphors etc.

• 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)
Patterns in Corpora

• When you first open a concordance for a lexical item, very often some patterns of use leap out at you.
  – Collocations make patterns: one word goes with another
  – in structures (constructions, valencies)
  – To see how words make meanings, we need to analyse contexts: valencies and collocations

• The more you look, the more patterns you see.

• BUT THEN

• When you try to formalize the patterns, you start to see more and more exceptions.

• Fuzzy boundaries between patterns

• How to make sense of the data?
Collocations:

• “Many, if not most meanings, require the presence of more than one word for their normal realization. ...

  “Patterns of co-selection among words, which are much stronger than any description has yet allowed for, have a direct connection with meaning.”

Concept meanings vs. word and phrase meanings

• Understanding text meaning depends on analysis of collocations.

• Collocations are variable.

  – [corpus-derived examples, from R. Moon]:
    • shivering in her shoes /
    quaking in his boots /
    shaking in their sandals
Semantic Types

- **Lexical sets** are lexical items grouped according to a shared semantic type.
  
  - *Boots, shoes, sandals, etc.*, share a semantic type.
  - Call it [[Footwear]]

Procedure for CPA

STEP 1: Take a sample concordance for each word
   – 250-500-1000 examples

STEP 2: Identify statistically salient collocates using the Sketch Engine (Kilgarriff 2004)

STEP 3: Group collocates into lexical sets

• Use introspection to interpret data, but not to create data.
GROUP 1:
It is hard to believe that bull-leapers grasped the horns and relied on the tossing movement to get them over the bull’s head. Ursula leaned slowly back against the window-sill, one hand grasping the edge tightly while the other held her cigarette. He grasped the handle of the door in one hand and the spoon in the other. He reached out wildly, trying to grasp the creature, but it had moved away. Benjamin stretched across and grasped the man’s hand. Laura grasped Maggie by the arm.

GROUP 2:
In the end we will grasp the truth. I was too intelligent not to be already grasping the rules of the game we played. After fifteen minutes, Julia thought that she had grasped most of the story. Teachers should grasp the fact that the DES can lay down details of a policy but that the Department of Employment funds it. He could never grasp the essentials, the obligations of living in a western society. He had not grasped that Ruby worked that day with a mere photograph. She grasped what was happening.
GROUP 3:
Lawrence hoped his players would grasp the chance of cup glory.
The Prime Minister failed to grasp that opportunity.
Kylie, singing like she had never before, grasped the moment.

GROUP 4:
Ian Corner, David Chell and their staff are bravely grasping the nettle of recession.
The Labour Party has failed to grasp the nettle in Monklands.
That's what the GMB need to do, to grasp the nettle, to move forward.

GROUP 5:
Theda had gone paler than usual, and she grasped at the bedpost for support.
The child was still crying as Alan sat down with him, but he grasped greedily for the milk.

GROUP 5a:
Nadirpur's eyes widened. He was grasping at straws.
Patterson’s eyes flickered as if I’d given him a straw to grasp.
Classes used in CPA

• **Norms** (normal uses in normal contexts)
• **Exploitations** (e.g. coercions and ad-hoc metaphors)
• **Alternations**
  • e.g. [[Doctor]] treat [[Patient]] --> [[Medicine]] treat [[Illness]]
  - Names (*Midnight Storm*: name of a horse, not a kind of storm)
  - Mentions (to mention a word or phrase is not to use it)
  - Errors
  - Unassignables

Every line in the sample must be classified
Norms and Exploitations

- In order to understand meaning in language, it is essential to distinguish between:
  - norms (the basic shared conventions that S and H mutually rely on), and
  - exploitations (freshly created metaphors and other tropes, unusual phrasing, etc.)

- Two different rule systems
- The two rule systems interact.
Exploitations: what to ignore when writing a dictionary

• Exploitations are unusual uses of words, coined for rhetorical effect, economy of space, etc.
• Exploitations are deliberate.
• They are ways of creating new meanings, ad hoc.
• Exploitations are among the most interesting uses of words in a language.
• Lexicographers have a duty to ignore them.
Exploitation rule 1: ellipsis

- I hazarded various Stuartesque destinations such as Bali and Istanbul.
  - Julian Barnes
Stuart needlessly scraped a fetid plastic comb over his cranium.
‘Where are you going? [I asked.] You know, just in case I need to get in touch.’
‘State secret. Even Gillie doesn’t know. Just told her to take light clothes.’

He was still smirking, so I presumed that some juvenile guessing game was required of me. I hazarded various Stuartesque destinations like Florida, Bali, Crete and Western Turkey, each of which was greeted by a smug nod of negativity.
Exploitation Rule 2: Anomalous argument

Another example:

- Always vacuum your moose from the snout up, and brush your pheasant with freshly baked bread, torn not sliced.

  —said to be from *The Massachusetts Journal of Taxidermy*, 1986 (per Associated Press newswire)

- Can you vacuum a moose? ... Is it normal?
- “Can you say X in English? – the wrong question to ask. Ask instead, “Is it normal?”
Exploitation rule 3: metaphor

- Stoke Mandeville station is a little oasis; clean and bright and friendly.
- New Town Hotel -- a relaxing oasis for professional and business men.
- Driffield, which was a pleasant oasis in the East Riding of Yorkshire.
- The planned open-cast site was a pleasant oasis in a decaying industrial landscape.
- She regards her job as an oasis in a desert of coping with Harry’s illness.
- … an oasis in the midst of this desert of feuding.

An oasis in English (and other European languages) is prototypically pleasant, relaxing, calm, and surrounded by barren, nasty desert. (The reality may be very different.)
Salient collocates for ‘oasis’ (SkE)

BNC frequency for ‘oasis’: 307

<table>
<thead>
<tr>
<th>Collocate</th>
<th>Co-occurrences</th>
<th>Salience score</th>
</tr>
</thead>
<tbody>
<tr>
<td>greenery</td>
<td>3</td>
<td>8.11</td>
</tr>
<tr>
<td>serenity</td>
<td>2</td>
<td>7.53</td>
</tr>
<tr>
<td>desert</td>
<td>12</td>
<td>7.07</td>
</tr>
<tr>
<td>calm</td>
<td>7</td>
<td>7.28</td>
</tr>
<tr>
<td>lush</td>
<td>2</td>
<td>6.82</td>
</tr>
<tr>
<td>tranquillity</td>
<td>2</td>
<td>6.76</td>
</tr>
<tr>
<td>peaceful</td>
<td>3</td>
<td>5.75</td>
</tr>
<tr>
<td>welcome</td>
<td>4</td>
<td>5.68</td>
</tr>
<tr>
<td>pleasant</td>
<td>3</td>
<td>5.12</td>
</tr>
<tr>
<td>tropical</td>
<td>4</td>
<td>5.07</td>
</tr>
</tbody>
</table>
Regular and irregular linguistic performance

- Norms are first-order regularities of linguistic behaviour
- Alternations are second-order regularities of linguistic behaviour
- Exploitations are irregularities, deliberately chosen by a speaker or writer for rhetorical or literary effect
- Mistakes are irregularities that occur accidentally, not deliberately
A “double-helix” theory of 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:
  – Rules governing normal usage
  – Rules governing exploitation of norms
• The two systems interact, producing new norms:
  – Today’s exploitation may be tomorrow’s norm!
Browse it for yourself

- **A Pattern Dictionary of English Verbs (PDEV)**
- Currently being created by CPA:
  - Related projects are starting up for:
    - Italian (Elisabetta Jezek; *Universita degli Studi, Pavia*)
    - Spanish (Janet de Cesaris, Paz Battaner, and others, *Universitat Pompeu Fabra, Barcelona*)