Computational Lexicography: Mapping Meaning onto Use

Course: Lexicology: words and Meanings

Why is Sinclair important?

.

Why is Sinclair important? (1)

- "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." (Sinclair 1998 'The Lexical Item', page 4)

Why is Sinclair important? (2)

The **idiom principle** (also known as *the phraseological tendency*) vs. **the open-choice principle**:

 "The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments." (Sinclair 1991. *Corpus, Concordance, Collocation*, p. 110)

 "Tending towards open choice is what we can dub the terminological tendency, which is the tendency for a word to have a fixed meaning in reference to the world. ... tending towards idiomaticity is the phraseological tendency, where words tend to go together and make meanings by their combinations." (Sinclair 2004. *Trust the Text*, p. 29)

Two senses of "computational lexicography"

1. Exploiting published dictionaries for use in new computer programs

2. Using computer programs to create new dictionaries

Using dictionaries for computational purposes

- Inventory of the words of a language
  + tokenization, lemmatization
- Word class recognition (noun vs. verb vs. adj.)
  - but dictionaries don't give comparative frequencies
  - **see, sees** *n.* district of a bishop: 136 in BNC.
  - **see, sees** *vb.* perceive: 118,500 in BNC.
- Word sense disambiguation
  - assumes that dictionary sense distinctions are reliable.
  - dictionaries don't give comparative frequencies!

Word Sense Disambiguation

Lesk (1986): *'How to tell a pine cone from an ice cream cone'*, using OALD definitions:

pine 1. kind of <u>evergreen</u> <u>tree</u> with needle-shaped leaves. **2.** waste away through sorrow or illness.

cone 1. a solid object with a round flat base and sides that slope up to a point… **2.** something of this shape whether solid or hollow. **3.** a piece of thin crisp biscuit shaped like a cone, which you can put ice cream in to eat it. **4.** the fruit of certain <u>evergreen</u> <u>trees</u>.

Some problems

- There is no general agreement on what counts as a word sense

- No clear criteria are given in dictionaries for distinguishing one sense from another

- There is very little syntagmatic information in dictionaries

Lumping and splitting

Most dictionaries are splitters. E.g. why did OALD 1963 make these two senses **(cone)**?

- **1.** a solid object with a round flat base and sides that slope up to a point… **2.** something of this shape whether solid or hollow.

Why not:

- a solid or hollow object with a round flat base and sides that slope up to a point

This problem is endlessly multiplied in entry after entry.

Implicatures: taking prototypes amd domain seriously

If someone **files a lawsuit**, they activate a procedure asking a court for justice.

When a pilot **files a flight plan**, he or she informs ground control of the intended route and obtains permission to begin flying. …

When a group of people **file into a room** or other place, they walk in one behind the other.

(12 more such definitions of *file*, verb.)

The problem: deciding relevant context

- Peter treated Mary.
- Peter treated Mary for her asthma.
- Peter treated Mary badly.
- Peter treated Mary with respect.
- Peter treated Mary with antibiotics.
- Peter treated Mary to lunch.
- Peter treated Mary to his views on George W. Bush
- Peter treated the woodwork with creosote.

The CPA method

- CPA: Corpus Pattern Analysis (based on TNE: the Theory of Norms and Exploitations).

1. Create a sample concordance (KWIC index):
   - from a 'balanced' corpus (i.e. general language): BNC50
   - 250 examples of actual uses of the word to start with

2. Classify *every* line in the sample, by context.

3. Take further samples if necessary.

4. Use introspection to interpret data, but not to create data.

Sample from a concordance

```
       incessant noise and bustle had abated. It seemed everyone was up
        after dawn the storm suddenly abated. Ruth was there waiting when
               Thankfully, the storm had abated, at least for the moment, and
     storm outside was beginning to abate, but the sky was still ominous
Fortunately, much of the fuss has abated, but not before hundreds of
       , after the shock had begun to abate, the vision of Benedict's
been arrested and street violence abated, the ruling party stopped
  he declared the recession to be abating, only hours before the
'soft landing' in which inflation abates but growth continues moderate
 the threshold. The fearful noise abated in its intensity, trailed
ability. However, when the threat abated in 1989 with a ceasefire in
  bag to the ocean. The storm was abating rapidly, the evening sky
    ferocity of sectarian politics abated somewhat between 1931 and
    storm. By dawn the weather had abated though the sea was still angry
```

The Importance of Context

- "You shall know a word by the company it keeps" – J. R. Firth.

- Corpus analysis can show what company our words keep.

- Frequency alone is not enough: "*of the*" is a frequent collocation – but not interesting!

- "*storm abated*" is less frequent, but more interesting. Contrasted with "*threat abated*", it can give a different meaning to the verb *abate*.

- So we need a way of measuring the statistical significance of collocations.

Mutual information

- A way of computing the statistical significance of two words in collocation.

- Compares the actual co-occurrence of two words in a corpus with chance.

- Church and Hanks (1990): 'Word Association Norms, Mutual Information, and Lexicography' in *Computational Linguistics* 16:1.

- Kilgarriff, Rychlý et al. (2004): "The Sketch Engine", Proceedings of Euralex 2004. Lorient, France.

In CPA, every line in the sample must be classified

An important principle of statistical analysis.

The classes are:

- Norms
- Exploitations
- Alternations
- Names (*Midnight Storm:* name of a horse, not a storm)
- Mentions (to **mention** a word or phrase is not to **use** it)
- Errors (e.g. *learned* mistyped as *leaned*)
- Unassignables

Methodological precepts

- Focus on the probable. On the basis of what has happened, predict what is likely to happen.
- Don't look for necessary conditions for the meaning of a word. (There aren't any.)

  - "***This elephant is a mouse***" is an unlikely sentence of English – but <u>not</u> meaningless
- Don't try to account for ***<u>all</u>*** possibilities.
- Use prototype theory to account for ***<u>probable</u>*** meanings.
- Don't ever say "all and only".

Norms

- How the words are ***normally*** used.

- Descriptive (not prescriptive).

- Norms are discovered by systematic, empirical Corpus Pattern Analysis (CPA).

Exploitations

- People don't just say the same thing, using the same words repeatedly.
- They also *exploit* norms in order to say new things, or in order to say old things in new and interesting ways.
- Exploitations include metaphor, ellipsis, word creation, and other figures of speech.
- Exploitations are a form of creativity.

Example of a CPA verb norm

abate/V   BNC frequency: 185 in 100m.

1. [[Event = Storm]] abate [NO OBJ](11%)

2. [[Event = Flood]] abate [NO OBJ]   (4%)

3. [[Event = Fever]] abate [NO OBJ] (2%)

4. [[Event = Problem]] abate [NO OBJ] (44%)

5. [[Emotion = Negative]] abate [NO OBJ] (20%)

6. [[Person | Action]] abate [[State = Nuisance]] (19%)
   (Domain: Law)

[[Event = Storm]] abate [NO OBJ]

dry kit and go again.The **storm** **abates** a bit, and there is no problem in
ling.Thankfully, the **storm** had **abated**, at least for the moment, and the
sting his time until the **storm** **abated** but also endangering his life, Ge
storm outside was beginning to **abate**, but the sky was still ominously o
bag to the ocean.The **storm** was **abating** rapidly, the evening sky clearin
 after dawn the **storm** suddenly **abated**.Ruth was there waiting when the h
t he wait until the **rain storm** **abated**.She had her way and Corbett went
 storm.By dawn the **weather** had **abated** though the sea was still angry, i
lcolm White, and the **gales** had **abated**: Yachting World had performed the
he **rain**, *which* gave no sign of **abating**, knowing her options were limite
n became a **downpour** that never **abated** all day.My only protection was
ned away, **the roar of the wind** **abating** as he drew the hatch closed behi

[[Event = Problem]] abate [NO OBJ]

'soft landing' in which **inflation abates** but growth continues modera
Fortunately, **much of the fuss** has **abated**, but not before hundreds of
 the threshold. The **fearful noise abated** in its intensity, trailed
   incessant **noise and bustle** had **abated**. It seemed everyone was up
ability. However, when **the threat abated** in 1989 with a ceasefire in
the Intifada shows little sign of **abating**. It is a cliche to say that
h he declared **the recession** to be **abating**, only hours before the pub
he ferocity of sectarian politics abated somewhat between 1931 and 1
been arrested and **street violence abated**, the ruling party stopped b
the **dispute** showed no sign of **abating** yesterday. Crews in

[[Emotion = Negative]] abate [NO OBJ] (selected lines)

```
ript on the table and his anxiety abated a little.This talented, if
 that her initial awkwardness had abated # for she had never seen a
es if some inner pressure doesn't abate.He wanted to play at the fun
Baker in the foyer and my anxiety abated.He seemed disappointed and
hained at the time.When the agony abated he was prepared to laugh wi
self; the pain gradually began to abate spontaneously, a great relie
ght, after the shock had begun to abate, the vision of Benedict's sn
y calm, control it!) The fear was abating, the trembling beginning t
 his dark eyes. That fear did not abate when, briefly, he halted. For
```

AN EXPLOITATION OF THIS NORM:

```
isapproval, his kindlier feelings abated, to be replaced by a resurg
```

("kindlier feelings" are normally positive, not negative.)

Part of the lexical set [[Event = Problem]] as subject of 'abate'

From BNC: {fuss, **problem, tensions, fighting,** price war, hysterical media clap-trap, disruption, slump, inflation, recession, the Mozart frenzy, working-class militancy, hostility, intimidation, ferocity of sectarian politics, diplomatic isolation, dispute, …}

From AP: {threat, crisis, **fighting,** hijackings, protests, **tensions,** violence, bloodshed, **problem**, crime, guerrilla attacks, turmoil, shelling, shooting, artillery duels, fire-code violations, unrest, inflationary **pressures,** layoffs, bloodletting, revolution, murder of foreigners, public furor, eruptions, bad publicity, outbreak, jeering, criticism, infighting, risk, crisis, …}

(All these are kinds of **problem**.)

Part of the lexical set [[Emotion = Negative]] as subject of 'abate'

From BNC: {*anxiety,* fear, emotion, **rage, anger,** fury, pain, agony, feelings,…}

From AP: {**rage, anger**, panic, animosity, concern, …}

A domain-specific norm:
[[Person | Action]] abate [[Nuisance]]
(**DOMAIN: Law. Register: Jargon**)

```
o undertake further measures to abate the odour, and in Attorney Ge
us methods were contemplated to abate the odour from a maggot farm
s specified are insufficient to abate the odour then in any further
as the inspector is striving to abate the odour, no action will be
t practicable means be taken to abate any existing odour nuisance,
ll equipment to prevent, and or abate odour pollution would probabl
rmation alleging the failure to abate a statutory nuisance without
 t I would urge you at least to abate the nuisance of bugles forthw
 way that the nuisance could be abated, but the decision is the dec
otherwise the nuisance is to be abated.They have full jurisdiction
ion, or the local authority may abate the nuisance and do whatever
```

Lexical sets are contrastive

- Different lexical sets generate different meanings.
- Lexical sets are not like syntactic structures.
- In principle, lexical sets are open-ended, but most have high-value **best examples**.
- In practice, a lexical set may have only 1 or 2 members, e.g. *take a {look | glance}*.
- No certainties in word meaning; only probabilities.
- … but probabilities can be measured.

A more complicated verb: 'take'

- **61** *phrasal verb patterns, e.g.*

  [[Person]] **take** [[Garment]] **off**

  [[Plane]] **take off**

  [[Human Group]] **take** [[Business]] **over**

- **105** *light verb uses (with specific objects), e.g.*

  [[Event]] **take place**

  [[Person]] **take {photograph | photo | snaps | picture}**

  [[Person]] **take {the plunge}**

- **18** *'heavy verb' uses, e.g.*

  [[Person]] **take** [[PhysObj]] [Adv[Direction]]

- **13** *adverbial patterns, e.g.*

  [[Person]] **take** [[TopType]] **seriously**

  [[Human Group]] **take** [[Child]] **{into care}**

- TOTAL: 204, and growing (but slowly)

A fine distinction: 'take + place'

- [[Event]] take {place}: **A meeting took place**.
- [[Person 1]] take {[[Person 2]]'s place}:
  - **George took Bill's place**.
- [[Person]] take {[COREF POSDET] place}: **Wilkinson took his place among the greats of the game**.
- [[Person=Competitor]] take {[ORDINAL] place}: **The Germans took first place**.

Noun norms

- Norms for nouns are different in kind from norms for verbs.

    - Adjectives and prepositions are more like verbs.

- A different analytical apparatus is required for nouns.

- Prototype statements for each true noun can be *derived from a corpus*.

- Examples for the noun '*storm'* follow.

Storm (literal meaning) (1)

WHAT DO STORMS DO?

- Storms *blow*.

- Storms *rage*.

- Storms *lash* coastlines.

- Storms *batter* ships and places.

- Storms *hit* ships and places.

- Storms *ravage* coastlines and other places.

Storm (literal meaning) (2)

BEGINNING OF A STORM:

- Before it begins, a storm is ***brewing, gathering,*** or ***impending***.

- There is often a ***calm*** or a ***lull before*** a storm.

- Storms last for a certain period of time.

- Storms ***break***.

END OF A STORM:

- Storms ***abate***.

- Storms ***subside***.

- Storms ***pass.***

Storm (literal meaning) (3)

WHAT HAPPENS TO PEOPLE IN A STORM?

- People can *weather, survive,* or *ride (out)* a storm.
- Ships and people may get *caught in* a storm.

Storm (literal meaning) (4)

WHAT KINDS OF STORMS ARE THERE?

- There are *thunder storms, electrical storms, rain storms, hail storms, snow storms, winter storms, dust storms, sand storms, tropical storms…*

- Storms are *violent, severe, raging, howling, terrible, disastrous, fearful, ferocious…*

Storm (literal meaning) (5)

OTHER ASSOCIATIONS OF 'STORM':

- Storms, especially snow storms, may be *heavy*.

- An unexpected storm is a *freak* storm.

- The centre of a storm is called the *eye of the storm*.

- A major storm is remembered as *the great storm* (of [[Year]]).

- **STORMS ARE ASSOCIATED WITH** *rain, wind, hurricanes, gales,* and *floods*.

Conclusions

- Meanings are best associated with normal *contexts*, rather than words in isolation.

- Normal contexts correlate statistically significant collocations in different clause roles.

- The whole language system is probabilistic and preferential.

- The probabilities can be analysed in a new kind of dictionary – a syntagmatic dictionary.