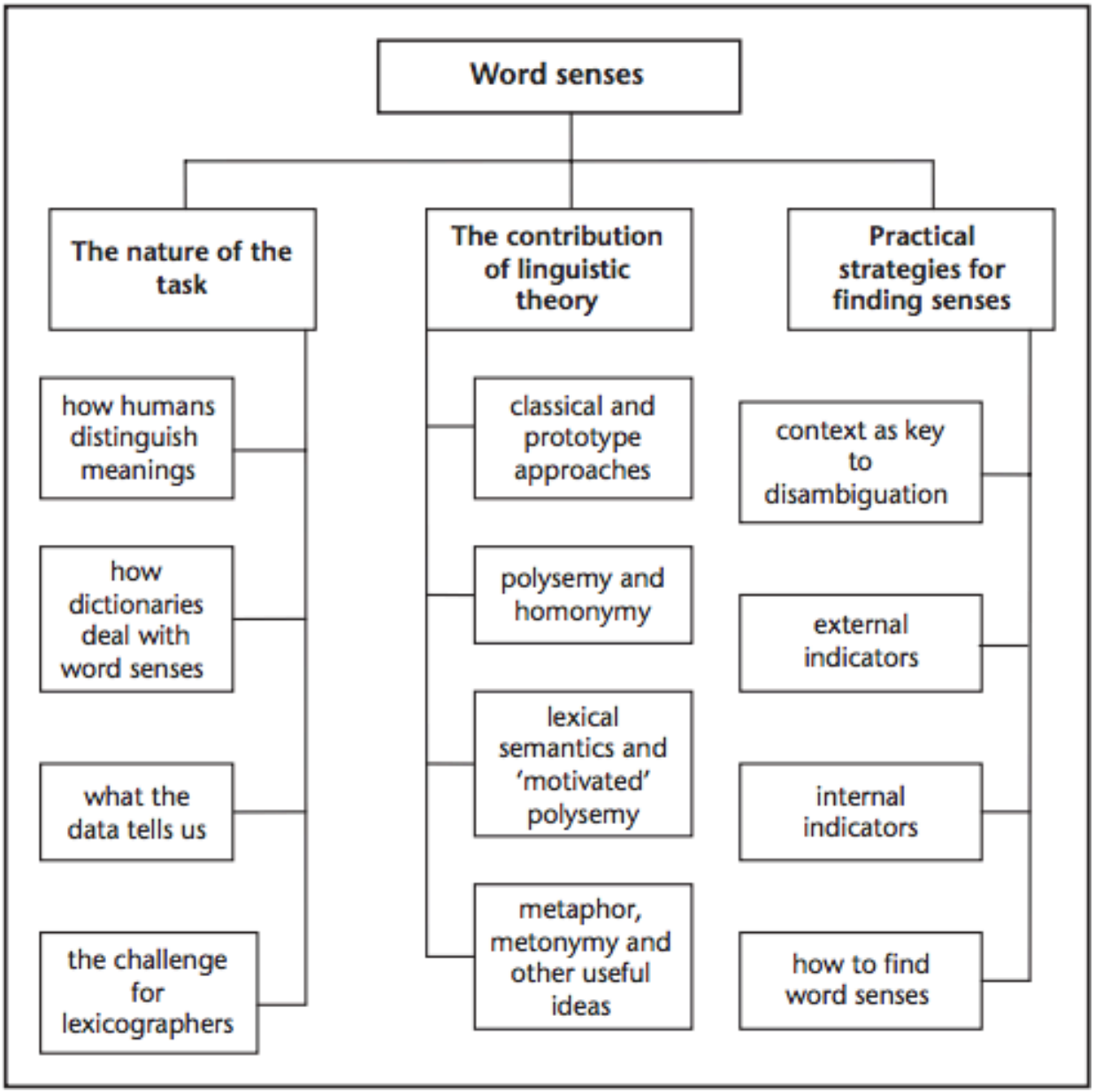
The Oxford Guide to Practical Lexicography
Building the database (1): word senses

Dodé Réka

2019. április 5.


Word senses
The nature of the task
The contribution of linguistic theory
Practical strategies for finding senses
how humans distinguish meanings
how dictionaries deal with word senses
what the data tells us
the challenge for lexicographers
classical and prototype approaches
polysemy and homonymy
lexical semantics and 'motivated' polysemy
metaphor, metonymy and other useful ideas
context as key to disambiguation
external indicators
internal indicators
how to find word senses

- dictionaries typically present words → word senses → dictionary senses
- identifying and describing word senses
- little agreement about what word senses are
- the hardest problems to describe the meaning of so-called polysemous words

- dictionary senses abstracted from raw language data
- Lexicographers undertake this task with reasonable confidence
- an inventory of senses for each headword and log them in the database
- most words have only one meaning
- there is only one possible ‘reading’ of the word when we encounter it in text

- the proportion of single-sense words is even higher in dictionaries with larger headword lists
- strong correlation between frequency and polysemy
- but the converse is also true

Two kinds of polysemy

- *Party*: five lexical units (LUs) – four main senses and one multiword expression
- high level of convergence among these accounts: rare
- dictionaries to divide up the various uses of a polysemous word in different ways
- *Overwhelm*: overwhelm in radically different ways, gives a somewhat different description
- described at various levels of granularity

overwhelm **1** if someone is overwhelmed by an emotion, they feel it so strongly that they cannot think clearly **2** if work or a problem overwhelms someone or something, it is too much or too difficult to deal with **3** to surprise someone very much, so that they do not know how to react **4** to defeat an army completely **5** if water overwhelms an area of land, it covers it completely and suddenly

LDOCE-4 (2003)

overwhelm **1** If you **are overwhelmed by** a feeling or event, it affects you very strongly and you do not know how to deal with it. **2** If a group of people **overwhelm** a place or another group, they gain complete control or victory over them.

COBUILD-5 (2006)

Approaches

1. try to account for the word's uses with just two or three broadbrush descriptors: **splitting**

2. in a more fine-grained analysis, we might identify as many as six or seven senses, each matching a precise context: **lumping**

- at the database stage, it is a good idea to split meanings fairly finely
- synthesis stage: it is easier to lump

- our first task is to analyse word forms into distinct meanings, or LUs → word sense disambiguation (WSD)
- nothing 'definitive' about the way we divide words into LUs
- lexical choices are intuitive, but rule-governed

The contribution of linguistic theory

- theoretical ideas especially
- identify any promising ways in which these ideas can inform the practical task of WSD
- understand the various mechanisms (meanings develop) → better position to tackle the practical task of finding dictionary senses

Classical approaches and prototype theory (Aristotle's Metaphysics)

- cluster of 'criterial features'
- the particular conjunction of **'necessary and sufficient conditions'** uniquely identify that meaning → straightforward binary choice
- *Bachelor*: an adult male, someone who has never married
    - (A man may be in a long-term relationship and may have children with his partner, but without actually being married: is he a bachelor?)

- the existence of borderline cases → classical theory begins to break down
- **many everyday words whose meanings cannot be captured by means of a set of necessary and sufficient features**

Wittgenstein: game

- criss-crossing network of similarities
- Membership of the category: 'family resemblances'
- understand and internalize what game means → extrapolating the meaning from all the exemplars they encounter

‘Prototype’ theory of meaning

- represents an ideal exemplar of a category (a ‘prototype’)
- a large extent shared by members of a speech community
- A prototype functions: ‘cognitive reference point’
- other entities seen as belonging to the same category provided they are **sufficiently similar** to the prototypical member
- ‘degrees of category membership’, so that some members are ‘better’ exemplars than others

Eleanor Rosch and ‘prototype theory’

- primary or ‘focal’ colours
- Subjects were asked to judge how well a given item represented a particular category

- We will approach the job in a pragmatic frame of mind, appreciating the limitations of WSD and recognizing that there are **no absolute truths awaiting discovery**
- We may also feel that conventional ways of presenting word senses in dictionaries **need to be modified to** reflect the way the language works.

The entry for *climb*

- *A car may climb up a steep hill.*
- *A plane may climb into the air after take-off.*
- *A column of smoke may also climb into the air.*
- *A plant may climb up a wall.*
- *A road or path may climb up the side of a hill.*
- *. . . and so on.*

- The entry opens with a broad definition of the prototypical 'ascend' meaning, and this is followed by a series of 'subsenses' describing recurrent uses which are both more specialized and less prototypical (like the ones listed above)

- two major advantages
  - It reflects the way people create meanings when they communicate, and thus it goes with the grain of the language, and accommodates creativity and fuzziness.
  - It makes the lexicographer's task more manageable, because it allows us to focus on the prototype and its common exploitations, rather than requiring us to predict and account for every possible instantiation of a meaning.
- provide a useful explanation both of the prototypical uses of *climb*, and of the most common variations on this prototype

Polysemy and homonymy

- Dictionaries have traditionally distinguished polysemous words from homonyms

- *Punch*

(1) *She gave him a punch in the stomach.* (a hard blow with the fist)
(2) *It lacks the emotional punch of French cinema.* (a forceful, memorable quality)
(3) *Glasses of punch were passed around.* (an alcoholic drink mixed from several ingredients)

second sentence can be interpreted as a metaphorical extension of the physical punch in sentence 1.

- Sentence 3: nothing in common with the other two: an entirely different semantic area,
- (accidental) fact of sharing the same orthographic form.
- This third use of *punch*: entered the language later and from a different source

Variables of same word-form

- Homonymy: do we separate *punch*-drink from *punch*-blow with fist?
- Wordclass: do we separate punch-noun (‘a punch in the stomach’) from punch-verb (‘punched me in the stomach’)?
- Homophony: do we separate *bow* (the weapon used with arrows) from *bow* (an act of bending the body forward)? The two words look identical but sound different.
- Capitalization: do we separate *swede* (a vegetable) from *Swede* (a person from Sweden)?

- homophones and capitalized forms are separate entries, or ‘homographs’
- Differences based on wordclass are variously handled: a single entry with subsections for each wordclass, separate homographs for each wordclass
- Is homonymy relevant? → depends on the intended uses (and target users) of the dictionary

- ‘on the tip of my tongue’ or ‘tip someone the wink’?
- it was decided to ignore homonymy completely’

Lexical semantics and 'motivated' polysemy

- books on lexical semantics, D. A. Cruse (1986, 2004) catalogues the varieties of polysemy
- Three concepts:
- (1) Contextual modulation:
    - My car has broken down.
    - I'm having the car resprayed
    - What a comfortable car!
    - Our car got a puncture.
    - I'm going to fill up the car.

different features of this concept are foregrounded in different situations

- a set of ‘meaning potentials’
- (2) Antagonism: cannot simultaneously interpret the word in two or more different ways, they compete with one another, and the best one can do is to switch rapidly from one to the other
- mutually exclusive → are thus the clearest manifestation of polysemy

- (3) Motivation
- linguistic behaviour is **motivated** rather than **arbitrary**
- reason why a word has more than one meaning, and the links between its various senses can be explained

1. Specialization

- a default reading: more general, more specialized
- 'specialization' – is especially common, and has a number of variations.
    - She raised the cup to her mouth and drank. (to swallow a liquid)
    - Jack doesn't drink, so I ordered him a Coke. (to drink alcohol)
    - She left her first husband because he drank. (to drink alcohol excessively)

- Sometimes is a function of a specialized domain, or sublanguage
- as forms of specialization: amelioration and pejoration (Thus notorious, at one time merely synonymous with ‘well-known’, has undergone pejoration and now combines the notions of ‘famous’ and ‘wicked)
- In hyperbole, similarly, we can observe specialization from a neutral to an emotionally charged sense (It’s boiling in here, I’ve been waiting for ages, or Those shoes cost a fortune)

2. Regular polysemy

- all the members of a particular semantic set behave in predictably similar ways
- classic example: words for trees typically function both as countable nouns (referring to the tree itself) and as mass nouns (referring to its wood), thus:
    - Is that an oak or a beech? [countable]
    - The doors are made of solid oak. [mass]

→ the opportunity of streamlining the editorial process

- ‘verb alternations’, where all the verbs in a particular semantic set exhibit predictable variations in syntactic behaviour
- adjectives that are applied to people for describing how they feel – a very large class of words – also regularly describe people’s actions and responses

angry

- people: crowd, mob, man, demonstrators, fans, residents
- actions and responses: voice, response, retort, face, outburst, letter.

- It makes obvious sense that any adjective belonging to this broad semantic set should be handled in the same way in a dictionary

3. Figurative extensions: metaphor and metonymy

- While others are too marginal and idiosyncratic to qualify for sensehood-- > at these outer edges, the boundaries can never be quite clear, and a word's lexicographic treatment will depend as much on the size and scope of the dictionary as on the weight of corpus evidence.
- In a high proportion of cases, the journey from original meaning to 'new' senses involves figurative uses of language, where there is an implied resemblance between primary and secondary meanings.

- The most obvious manifestation: a literal or physical meaning (they climbed the mountain) underpins a non-literal or abstract one

(1) Figurative uses in general: *haunt*, *sacrifice*, and *broadcast*.

haunt

- The default reading of haunt is that it is what ghosts do when they visit a place. This sense incorporates a number of features or images:
  - recurrence (the ghost keeps coming back, haunting isn't something that happens just once)
  - a sense of anxiety or menace (the ghost's appearance is unsettling to those who witness it)
  - a gloomy or ethereal quality (ghosts don't generally appear in broad daylight).

(2) Metaphorical sets

- a whole set of specialized vocabulary may be generated through metaphor

Computer terminology

  - the computer is thought of as a person
  - the computer screen is thought of as an office

(3) Conceptual metaphor

a fundamental cognitive process that shapes the way we form concepts and give them names

- basic spatial concepts (like up/down, in/out)
- ILLNESS IS WAR – the idea that illnesses 'attack' the body and must be 'resisted'.
- helping us to perceive underlying systems in the language → better equipped to make sense of language data

(4) Metonymy

- Kuwait was invaded by Iraq in 1990.
- Blair's reputation was profoundly damaged by Iraq

- two basic mechanisms is at work:
- a focus on one attribute or aspect of something when your intention is to **refer to the thing as a whole** (He wanted everyone to admire his new wheels. (= his new car))
- a reference to **something as a whole**, when your intention **is to focus on one of its attributes or aspects** (Would you mind if I used your bathroom? (= the toilet in your bathroom))

how far a dictionary entry can or should account for such alternations in meaning

- no simple answer, ourdecision will be guided by much the same criteria as we apply to other aspects of the WSD problem:
- **Systematicity**: does the metonymy instantiate a recurrent, wellestablished pattern?
- **Frequency**: is the extended meaning a common usage?
- **Longevity**: is the extended use ephemeral or likely to endure?
    - the Pentagon, the Kremlin, Capitol Hill, Downing Street, and so on

Linguistic theory and its relevance

- recognize that polysemy comes in many forms and arises through many mechanisms
- tackle the job with greater confidence if understand the underlying systems, will be better equipped to make good judgments in the more marginal, less clear-cut cases.

Word senses and corpus patterns: context disambiguates

- **lexical** and **syntactic environment** in which a **word appears:** the **most reliable indicator** of the meaning it conveys in any particular instance
- the minimal contexts available that enabled lexicographers to decide which meaning of a polysemous word was being used'
- two examples of the verb operate:
    - Human beings will simply be unable to operate them.
    - They operated but it was too late.

- With corpus-querying software: have the option of retrieving more of the original context – but rarely need to do this
- more prompts the question: how do we do this?

issue

- *Typically half of the 90 children in Gardner-Betts have mental health issues.*
- *He wasn't hateful, but I could tell he had some issues.*
- *...a supplement to support bladder control and health in dogs especially those with incontinence issues*
- *Issues around sexuality can be deeply threatening for young people.*
- *Due to his emotionally chaotic upbringing, he likely does have significant intimacy issues.*
- *It helps to understand . . . how issues around gender, dependency, daily routine, and staff responsibility impinge on the environment.*

- how this use of issue diverges from its familiar sense of ‘an important topic or problem for debate or discussion
- plenty of borderline cases
- But corpus data supports the view that a specialization of the basic meaning
    - n (issues) informal personal problems or difficulties: emotions and intimacy issues that were largely dealt with through alcohol.

- Our first observation relates to the domain: social science (in fields such as psychology, counselling, criminology, and childcare).
- Apart from this, the clues are in the context:
    - It is always (in this use) pluralized, as the ODE entry acknowledges.
    - It is rarely sentence-initial or clause-initial, but usually occurs in the patterns have + issues or with + issues.
    - It is often followed by around (this use of around to mean 'concerning' is itself quite recent, and characteristic of the same discourse types as issues in this meaning).
    - It is often premodified by another noun (as in intimacy issues or incontinence issues).

a combination of subjective judgments and objective observations

Practical strategies for successful WSD

- 'context disambiguates': a given meaning typically occurs ('external' indicators) and immediate lexico-grammatical environment typically associated with a particular meaning ('internal' indicators)

External indicators: domain, dialect, and setting

- a lexicographic corpus is diversity: the widest possible range of sources, well-balanced corpus

1. Domain: the domain (or subject matter) of a text determines the meanings that certain polysemous words

  - *Once the substrate is bound to the surface of the enzyme, covalent bonds may be formed or broken.*
  - *Japan's leading brokerages agreed to stop issuing new shares and convertible bonds for at least a month.*
  - *A dominant form of family life began to take shape . . . characterized by close emotional bonds between parents and children.*

2. Regional dialect when American speakers say they are going to *wash up* they mean they are going to wash their hands and face, whereas the same verb in British English refers to the activity of washing plates and cutlery after a meal.

3. Time some of the mechanisms by which words can acquire new meanings in the course of time. Sometimes, a new meaning completely replaces an older one, and in such cases time is the best indicator of intended meaning.

- dictionary providing **a diachronic description** of the language, these older uses have to be accounted for: thus in ODE, the entry for enthusiasm includes a sense (labelled 'archaic, derogatory') which is defined as 'religious fervour'.

4. Subcultures Any speech community includes numerous subcultures (any group whose characteristic preoccupations or lifestyles differentiate it from mainstream culture)

Examples

- recent instance of ‘youthspeak’ is the idiosyncratic use of the adjective *random*. (2001–2006)
- **our (subjective) intuitions** and knowledge of the world interact with **the (objective) linguistic data**.
- well-balanced and well-annotated corpus: clear uses
- Strategy: mirrors what readers and listeners do when they encounter such uses in real life

Internal indicators

- we have to rely on internal evidence – specifically on evidence of a word's . . .
  - syntactic and lexico-grammatical behaviour
  - collocational features and selectional restrictions
  - colligational preferences.

typically cluster together on the basis of several shared features

Syntax and lexicogrammar

Friendly

- (A) The enquirer was a friendly, bubbly girl of about twenty-three.
- One more advantage of living in Taiwan is that people here are polite and friendly towards foreigners.
- She met me with a friendly smile, shook my hand and introduced me to the class.
- The Samos is an established hotel which is well run and has a friendly, relaxed atmosphere.
- (B) It was known, also, that the old lady had been friendly with his mother.
- I had become friendly with Vivien Fish, whose father was Churchill's dentist.
- He said he was still friendly with the princess but would do nothing to embarrass her.

- A: people's behaviour or gestures
- B: a relationship of friendship

- key indicator here is the preposition *with*

- **Countability**: in many cases, countable/uncountable alternations reflect types of regular polysemy, such as the 'generic'/'item of' alternation
    - She received the sort of welcome normally reserved for royalty (= members of a royal family)
    - The author's royalties will all go to charity (= income received from a book)

- **Transitivity**: when a verb is polysemous, transitivity is often an indicator of meaning.
  - *No drink at all should be the rule, particularly when driving, operating machinery or taking certain medicines (object obligatory)*
  - *Surgeons were forced to operate after an infection set in (no object).*

  In its 'surgical' use, *operate* is always intransitive, and is used either on its own or in a PP with on.

- I borrowed $5 from him
- Some firms had to borrow in order to stay in business.

- *borrow* normally takes an object, and the object specifies the thing being borrowed (money, someone's car, a book, etc.). But when the object is omitted – as in the second example here – the only possible reading is a specialization of the general sense of borrow:
- the thing being borrowed is money (it can't be anything else)
- the lender is a financial institution (not, say, a friend of the borrower)
- the borrower has undertaken a programme of repayments.

Syntactic patterns

- the verb *remember* has a number of uses, and each is typically associated with a particular syntactic pattern

Example	Pattern	Meaning
I remember sitting alone in the cafeteria, slowly drinking my cup of coffee	+ *-ing* form	call to mind a past experience or event
Then she remembered why Nahum had come home in a terrible temper	+ *wh*-clause	call to mind information you knew before
But just remember that you are being judged even before you speak a word	+ *that*-clause	often imperative: keep in mind a relevant fact
He only hoped Jane had remembered to leave the window open	+ *to*-INFIN	do something you undertook to do
Feynman, who died in 1988, is remembered for his many contributions to theoretical physics	+ PP/*for*	usually passive: be known or celebrated for a particular achievement
Please remember me to your mother. I trust that she is well.	+ PP/*to*	usually imperative: convey greetings to

Collocation and selectional restrictions

- an observable tendency of certain words to occur frequently with certain other words
- 'selectional restrictions': the general semantic category of items that typically appear as the subjects or objects of a verb, or as the complements of an adjective.
- A collocation is a recurrent combination of words, where one specific lexical item (the 'node') has an observable tendency to occur with another (the 'collocate'), with a frequency far greater than chance.

‘selectional restrictions’ of *forge*

- the verb *forge*: forge is used in a simple V + O pattern
- A Word Sketch for forge lists the following words as typical objects: *alliance, banknotes, bonds, friendship, links, metalwork, painting, partnership, passport, relationship, signature, sword, ties, unity*

Collocation

- Collocation: **less open-ended** – more arbitrary – phenomenon. We talk about people committing crimes: *Commit a crime*

- 'words that often co-occur' provide **valuable evidence for meaning differences**
- Most adjectives are light on syntax and can be used both **attributively** (an excellent performance) and **predicatively** (her performance was excellent).

- For the most part, though, syntax doesn't help us to disambiguate the uses of this adjective. But the word's typical complements point to a number of clear selectional restrictions

- corpus data **confirms our intuitions** and **gives them objective support**.
- Sometimes, however, the **corpus reveals important distinctions** which **introspection alone is unlikely** to provide

build up

- build up can be used with or without an object: you can build something up, or something can build up – build up (intransitive) are overwhelmingly negative and undesirable, whereas things that people build up (transitive) are almost always positive

1	thirties, has spent the last five years	building up	a successful business. However, she recen
2	and Lee Grant also scored as Colts	built up	a 6-0 lead. Sam Reed had set them a diffi
3	her own mark and has been steadily	building up	a quality client list. Spence, who made the
4	tive enterprise; over time, researchers	build up	a body of wisdom which tells them which
5	eel for the language. This in turn will	build up	their confidence in English, which they
6	n exposed to the infectious agent and	built up	an immunity to it. It is only in remote

Fig 8.6 Concordances for the transitive use of *build up*

1	aotic conditions. Massive bottlenecks	built up	in the early spring on the railway network
2	cerned that inflationary pressures are	building up	in the region's economy, foreshadowing
3	t followed incident, tension had been	building up	to a new peak. Raymond was accused of
4	the kidneys stop working and poisons	build up	in the blood. The patient will experience
5	we realise that a situation could be	building up	which could lead to the ultimate defeat of
6	ash in 1987. Huge backlogs of work	built up	in the security dealers' back offices, and

Fig 8.7 Concordances for the intransitive use of *build up*

Colligational preferences

- ‘a midway relation between grammar and collocation
- some words have to favour (or avoid) particular forms or positions
- a particular noun is **almost always plural**, and is **never sentence-initial**, then we have a prima facie case of colligation – an observable preference for a subset of the available grammatical options.

- colligation can include any of the following:
- in verbs, a marked preference for one particular form or use, such as the imperative, passive, reflexive, or progressive (-ing form)
- in nouns, a marked preference for either the singular or plural form, or for modifying other nouns
- in adjectives, a marked preference for either attributive or predicative position, or for comparative or superlative forms
- in any wordclass, a marked preference for one particular position within the sentence or clause
- in any wordclass, a marked preference for appearing in negative (or 'broad negative') constructions: think of words like compunction, remotely, afford, tenable, or budge
- a strong tendency to avoid any of these forms, structures, or positions.

Relevance of colligation to the task of identifying LUs

- (1) Adjectives: many adjectives can be used both as 'classifiers' (or 'pertainyms') and as descriptive words.
- (2) Nouns: (*trial*): refers to a legal process presided over by a judge or magistrate. When used in this way, the noun has no obvious preferences for singular or plural form or as subject or object
- (3) Verbs: the two main senses of acquit show marked colligational features
- in its 'legal' use it has a strong preference for the passive
- its other meaning is only invoked when the verb is used reflexively:
  - The former Scarborough goalkeeper certainly acquitted himself well on his debut.
- In its most usual ('applaud') meaning, cheer can be either transitive or intransitive. In a less frequent use, meaning 'to encourage', the verb is almost always passive:
  - I was much cheered by the fact that he expressed unqualified approval for it.

- the colligational features of the two meanings of consequence (‘result’ and ‘importance’):
- ‘result’ use is at least ten times more frequent than the ‘importance’ use, and shows a marked preference for being pluralized. In its ‘importance’ use, consequence has a number of clear preferences (this is a brief summary of a detailed study):
- It almost always appears in a PP (of great/little consequence).
- It is never the subject of a verb.
- It never occurs with a specific deictic (like these or the), favouring instead words like some, no, or any.
- It has a strong tendency for ‘denial’, that is for saying that something is not important

- colligation makes an important contribution to the task of identifying senses.
- The accumulated evidence from all these 'internal indicators' complements our intuitions about meaning and underpins an analysis which is as objective and 'scientific' as it reasonably can be

- LU-1 ‘quarrel, dispute’ (don’t argue with her)
- LU-2 ‘maintain, make a case for’ (he argued for a change in tactics)
- LU-3 ‘indicate, constitute evidence for’ (this argues a lack of support)
- LU-4 ‘persuade’ (she argued them out of going)

LU/meaning	corpus examples	linguistic features
LU-1 quarrel	*The teachers and medics were arguing about who has which square inch of my time.* *We spent most of our time in cafes, arguing and holding hands.* *The platoon commander was arguing with a gang of Christian Phalangists.*	– no object, no passive – often in *-ing* form – allows reciprocal use (with two or more subjects) – you argue *with* a person, and *about* (sometimes *over*) an issue
LU-2 maintain, make a case (for)	*Employers in the industry argued strongly for the retention of a statutory levy.* *She argues against the radical feminist view of 'male violence in the hands of the state'.* *He argues the need for a written constitution which is compatible with the rule of law.* *Headland has persuasively argued that there was just not enough food for such groups in the forest itself.* *Of course, it can be argued that readers get the paper that they deserve.* *Originally, France had argued for these plans to be confirmed by popular referendum.*	– usually with a PP (you argue *for* or *against* something) or with a *that*-clause – often modified by an adverb (*persuasively*, *cogently*, *convincingly*, etc.); though some of these collocates are shared with LU-3 – modality is common, in patterns like *It can/could be argued…*, *One could argue…* – occasionally with a simple noun object – occasionally in the pattern *argue* + *for* + *to*-infinitive – rarely in *-ing* form
LU-3 indicate, constitute evidence for	*The congestion on our roads argues that a serious vehicle tax should be levied.* *These features argue for a local origin.* *This lack is a key factor arguing against the existence of such a relationship.*	– non-human subject – usually with a PP (facts argue *for* or *against* something) or with a *that*-clause – rarely in *-ing* form, typically in simple present
LU-4 persuade	*Don't try to argue him out of it now – it's too late.* *Better not tell her the truth. Better just argue her into going back where she belonged.*	– obligatory PP: you either argue someone *into* something or *out of* it – often in infinitive, after words like *tried to, managed to*

- it’s unlikely that listeners will have any problem decoding the speakers’ meaning
- the language system allows us to generate novel meanings (or ‘stretch’ existing ones), without compromising intelligibility
- true ambiguity is exceptionally rare

- Rather than being a weakness, polysemy could be seen as an elegant design feature which, with maximum economy, enables language to respond to new situations while keeping to a minimum its demands on our short-term memories and processing capacity.
- dictionary senses: We do this by making generalizations (or abstractions) from the mass of available language data.
- These generalizations aim to make explicit the meaning distinctions
- their purpose is to enable dictionary users to associate what they have encountered in a specific context with a particular area of meaning.

How to find word senses

- (1) analyses instances of usage, typically in concordances or lexical profiles
- (2) provisionally identifies different word senses (this is the subjective, intuitive part)
- (3) collects good, typical corpus examples for each of these provisional senses.
- (4) analysing each cluster of examples in turn, the lexicographer identifies the features that are typically associated with it (and that distinguish it from all the other clusters);
- (5) finally, our inventory of senses is refined if necessary (which may involve further splitting, or conversely, lumping of closely related clusters) so that all uses of the word that occur frequently in text are fully accounted for.

- A **usage becomes a norm** – deserves to be described in a dictionary – when it is judged to be ‘part of the language’.
- ‘sensehood’ on those uses which can be observed to recur independently in a number of different texts
- the dictionary’s job is to deal with **‘the probable, not the possible**

- The various indicators we have described throw up clusters of examples which all behave in much the same way – and from these clusters we abstract our dictionary senses.