

Computer Science

2528

Topics in Computational Linguistics
Winter 2004

Semantic spaces and contexonyms

Sabine Ploux and Hyungsuk Ji
Institut des Science Cognitives, CNRS, Lyon

Interpreted by
Graeme Hirst
January 2004

- Sabine Ploux and Hyungsuk Ji.
“A model for matching semantic maps between languages (French / English, English / French).”
Computational Linguistics, 29(2), June 2003, 155–178.
- Hyungsuk Ji, Sabine Ploux, and Eric Wehrli.
“Lexical knowledge representation with contexonyms.”
Proceedings, Machine Translation Summit IX, September 2003, New Orleans, 194–201.
- Hyungsuk Ji and Sabine Ploux.
“A mental lexicon organization model.”
Proceedings, Joint International Conference on Cognitive Science, Sydney, July 2003, 240–245.

3

Background:

Semantic distance

- Words may be “close” or “distant” in meaning.
 - *talk, speak*
 - *doctor, nurse*
 - *dentist, barber*
 - *telephone, theorem*

4

Factors in semantic distance

- Denote overlapping (or identical) classes.
- Are (relatively large) subclasses of same class.
- Tend to co-occur (in a salient way).
- Have many (salient) properties in common.

5

Near-synonymy

- Very close in meaning.
 - *Talk, speak; seep, drip.*
- May differ in connotation, implication, style:
 - *Slim, slender, skinny.*
 - *Enemy, foe.*
 - *Died, snuffed it.*

6

Lexical choice

- In language generation, selecting a word to denote a part of the intended meaning.
- Problems:
 - Covering the meaning without redundancy.
John removed the cork from the bottle.
John uncorked the bottle.
**John uncorked the cork from the bottle.*
 - Well-formedness in covering a sentence.
John made { an error | a mistake }.
*John { erred | *mistaked }.*

7

Lexical choice (cont.)

- Problems (cont.)
 - Fitting the intended meaning as closely as possible without omissions or unwanted extras.
John erred ≠ John blundered
- Especially in MT:
 - Target language often gives no precise equivalent to word in source language.

8

Prior research at UofT

- Semantic distance: Jane Morris, Steve Green, David St-Onge, Alex Budanitsky.
 - Use links in WordNet or thesauri to determine distance or relatedness.
- Near synonyms: Chrysanne DiMarco (UW), Phil Edmonds, Diana Inkpen.
 - Representations of differences; derived from “synonym” dictionaries.
 - Lexical choice — monolingual or in MT.

9

Prior research (cont.)

- Lexical choice to cover meaning: Manfred Stede.
 - Representations of components of meaning within word.

10

“Components of meaning”

- Much research on “primitive units of meaning”, decomposition of words into primitives.
- In AI and CL: Wilks, Schank.
In linguistics: Wierzbicka, Jackendoff.

11

Examples

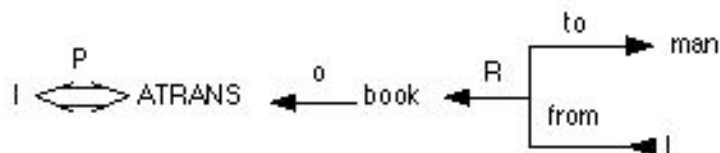
- Jackendoff

Laura released the bird from the cage.

[LET (LAURA, [GO (BIRD, [FROM (CAGE)])])]

- Schank

I gave the man a book.



www.umiacs.umd.edu/~bonnie/courses/cmsc723-03/lecture-notes/Lecture11.ppt
www.cogs.susx.ac.uk/lab/nlp/gazdar/teach/atc/1999/web/beateg/

12

Examples

- Wierzbicka and Goddard

X feels *happy* =

X feels something

sometimes a person thinks something like this:

something good happened to me

I wanted this to happen

I don't want anything else now

because of this, this person feels something good

X feels like this

www-personal.une.edu.au/~cgoddard/

13

Semantic primitives

- Abstractions?
 - Similar to mathematical objects.
- Psychological constructs?
 - Something in your head.
- What primitives are there? How many?
- A related question: What are word senses?

14

Views of word meaning

- Monolithic versus compositional.
- Monosemous versus polysemous.
- Discrete versus continuous.

Key point:

- There are more meanings than words.
- There are more meanings than word senses.

15

Some metaphors for word meaning

- Meanings are structures of primitives.
(A word is the “name” of a structure.)
 - Arbitrary structures are possible, but are not necessarily coherent meanings.
- Meanings are points in n -space whose dimensions are primitives.
 - Any (convex?) region is therefore also a coherent meaning.

16

Ploux & Ji: Mixed metaphors

- A word characterizes a meaning.
- Two or more words may (with redundancy) characterize a single meaning.
 - *fast, firm, lasting, secure, tight*
- Each word adds something to the structure.
- Each word is a nudge in n -space.

17

“Sense clustering” (1)

- Discovering the (presumed distinct) senses of a word by clustering its contexts of use by similarity (for use in IR, QA, etc).
 - Ted Pedersen, Hinrich Schütze, Dekang Lin & Patrick Pantel, ...
 - *heart:*
 - {*kidney, bone marrow, liver*}
 - {*psyche, consciousness, soul, mind*}
 - Problem: Labelling the clusters.

“Sense clustering” (2)

- Combining unnecessarily fine-grained lexicographic sense distinctions.
 - In WordNet, *bank* has 9 senses relating to repositories, inclines, etc.
- Both lead to much the same result.

19

Cliques of **synonyms**
and their semantic space

Cliques

- A **clique** is a complete graph whose nodes are words and whose edges are **synonymy relations** in lexicographic resources.
- Clique represents a more-precise meaning than any of the words in it.
 - \approx fuzzy intersection.
- Cliques are *found*, not artificially built.

21

Example

- Some cliques of good:
 - dependable, good, reliable, safe, secure.*
 - dependable, good, reliable, solid, sound.*
 - fair, good, honest, honourable, just, right, upright.*
 - able, capable, clever, expert, good, skilful.*

22

Example (screen)

You made the following query: **screen**
42 related words and 28 cliques were found.

Related words:

screen: CRT screen, blind, block out, camouflage, cloak, conceal, concealment, cover, covert, defend, disguise, filter, guard, hide, insulate, obscure, offer protection, partition, preserve, projection screen, protect, protection, riddle, safeguard, save, screen door, screen out, secure, shade, shelter, shield, shroud, shutter, sieve, sift, silver screen, sort, strain, test, veil, wall, web

Cliques:

- 1: CRT screen, screen
- 2: blind, camouflage, cloak, cover, screen
- 3: blind, cover, screen, shade
- 4: block out, screen
- 5: camouflage, cloak, conceal, cover, disguise, hide, screen, veil
- 6: camouflage, concealment, cover, disguise, screen
- 7: conceal, cover, disguise, hide, obscure, screen, veil
- 8: conceal, cover, hide, obscure, screen, shade, veil
- 9: conceal, cover, hide, screen, shroud, veil
- 10: concealment, cover, covert, screen
- 11: concealment, cover, screen, shelter
- 12: cover, defend, guard, protect, screen, shelter, shield
- 13: cover, guard, protection, screen, shelter, shield
- 14: cover, hide, screen, shade, shelter
- 15: cover, hide, screen, shelter, shroud
- 16: cover, protect, screen, shade, shelter
- 17: defend, guard, insulate, offer protection, preserve, protect, safeguard, save, screen, secure, shelter, shield
- 18: filter, screen, sieve, sift, strain

23

Gradation of meaning

- “Chains” of cliques show transition from one (sub-)sense to another.
- Each member of chain shares one or more words (in addition to headword) with previous.

24

Finding cliques

- **Lexicographic resources** (dictionaries, thesauri, ...) give evidence for cliques.
- Add edge whenever some resource lists two words as “synonymous”.
 - Nodes are words, not word-senses.
 - Could be within or across languages.
 - But (sub-)senses become apparent in the cliques.

25

Semantic space

- Construct a “**semantic space**” from cliques.
- A clique is a boolean vector of terms $\langle t_1, t_2, \dots \rangle$ where $t_i \in \{0, 1\}$ depends on whether the i^{th} term is in the clique; e.g., $\langle 0, 1, 0, 0, 1, \dots \rangle$
- Measure distance (really, correlation) between cliques in vector space by (ill-explained) factor-analysis (dimensionality reduction) method.
- Project onto Euclidean 2-space.
- Get regions that represent broader meanings.

26

Example (screen)

Component Selection for screen

Check desired component(s) and click on View.

View

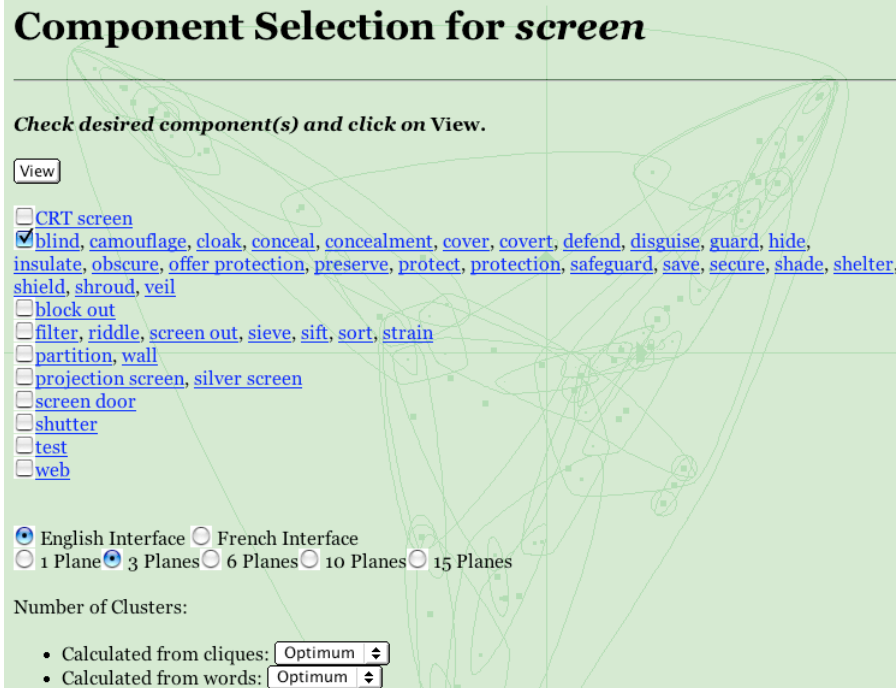
- CRT screen
- blind, camouflage, cloak, conceal, concealment, cover, covert, defend, disguise, guard, hide, insulate, obscure, offer protection, preserve, protect, protection, safeguard, save, secure, shade, shelter, shield, shroud, veil
- block out
- filter, riddle, screen out, sieve, sift, sort, strain
- partition, wall
- projection screen, silver screen
- screen door
- shutter
- test
- web

English Interface French Interface

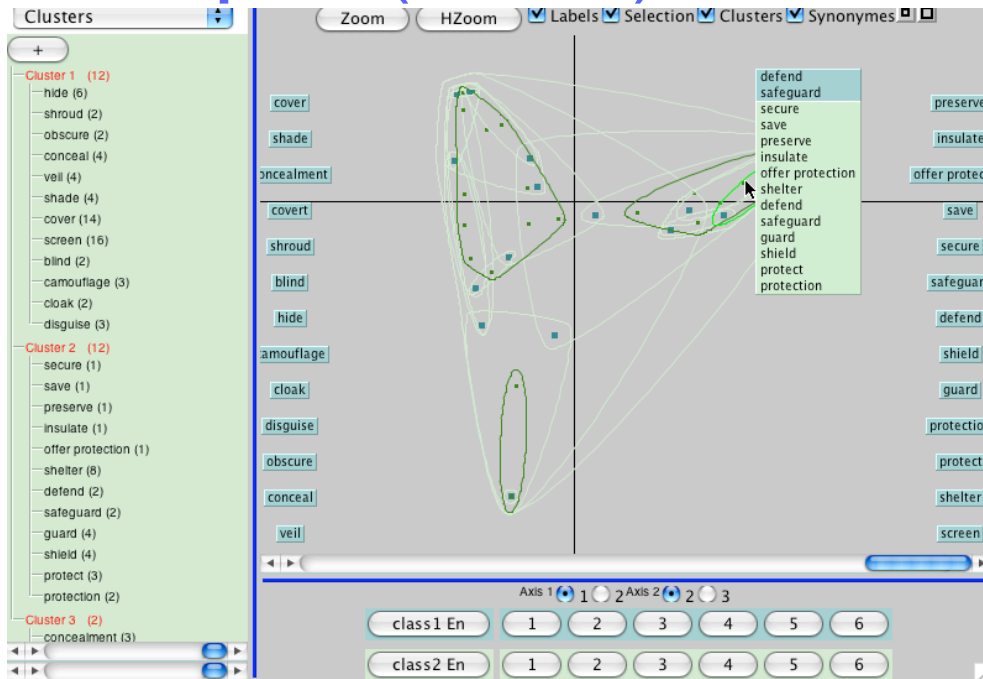
1 Plane 3 Planes 6 Planes 10 Planes 15 Planes

Number of Clusters:

- Calculated from cliques: Optimum
- Calculated from words: Optimum



Example (screen)



Clusters

Zoom HZoom Labels Selection Clusters Synonyms

Cluster 1 (12)

- hide (6)
- shroud (2)
- obscure (2)
- conceal (4)
- veil (4)
- shade (4)
- cover (14)
- screen (16)
- blind (2)
- camouflage (3)
- cloak (2)
- disguise (3)

Cluster 2 (12)

- secure (1)
- save (1)
- preserve (1)
- insulate (1)
- offer protection (1)
- shelter (8)
- defend (2)
- safeguard (2)
- guard (4)
- shield (4)
- protect (3)
- protection (2)

Cluster 3 (2)

- concealment (3)

cover

shade

concealment

covert

shroud

blind

hide

camouflage

cloak

disguise

obscure

conceal

veil

defend

safeguard

secure

save

preserve

insulate

offer protection

shelter

defend

safeguard

guard

shield

protection

protect

shelter

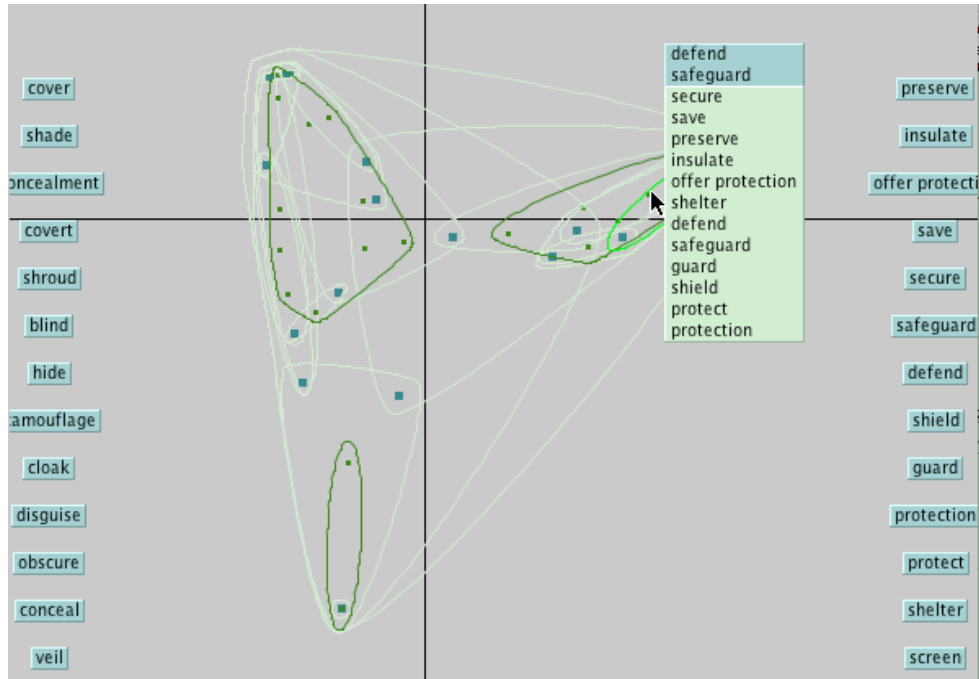
screen

Axis 1 1 2 Axis 2 2 3

class1 En 1 2 3 4 5 6

class2 En 1 2 3 4 5 6

Example (screen)



29

Notes

- Regions represent (sub-)senses.
- Clustering into regions is by unspecified process.
- Labels are derived from terms in cliques in the region.
- “Central” senses connect the others.

30

Matching semantic maps

- Do lexical choice in translation by “matching semantic maps” across languages.
- More precisely: Project cliques from both languages onto the same map to determine equivalences.

31

Method (1)

- Choose a term S in the source language. Compute cliques for S and its synonyms.*
- Find candidate translations T of these words in the target language (e.g., from bilingual lexicon). Compute their cliques.
- Retain the target-language cliques that are most **relevant** to the source-language cliques ...

*NP-complete.

32

Method (2)

- Regard each clique S, T as a vector of terms. Compute the boolean matrix M such that $M_{ij} = 1$ iff T_j translates S_i (or vice versa?).
- T is **relevant** to S if $\text{rank}(M) \geq 3$ (or 2).
 - i.e., at least 3 independent non-zero rows or columns.
 - Anomaly? Why require independence? Non-zero is what matters. Many 1s should mean high relevance even if low rank.

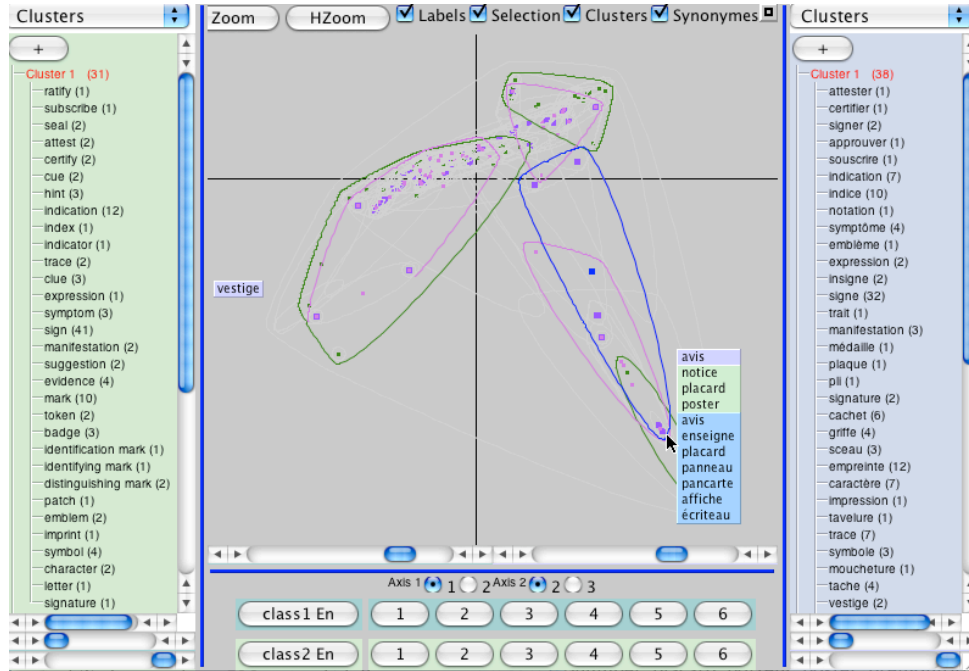
33

Method (3)

- Use the factor-analysis method to create a semantic space for the cliques from both languages.
 - But now use a matrix that takes translation into account.
(But this matrix is different: clique-to-clique, not clique-to-term. So how does this work?)
- Get distances between SL and TL cliques; make map of semantic space.

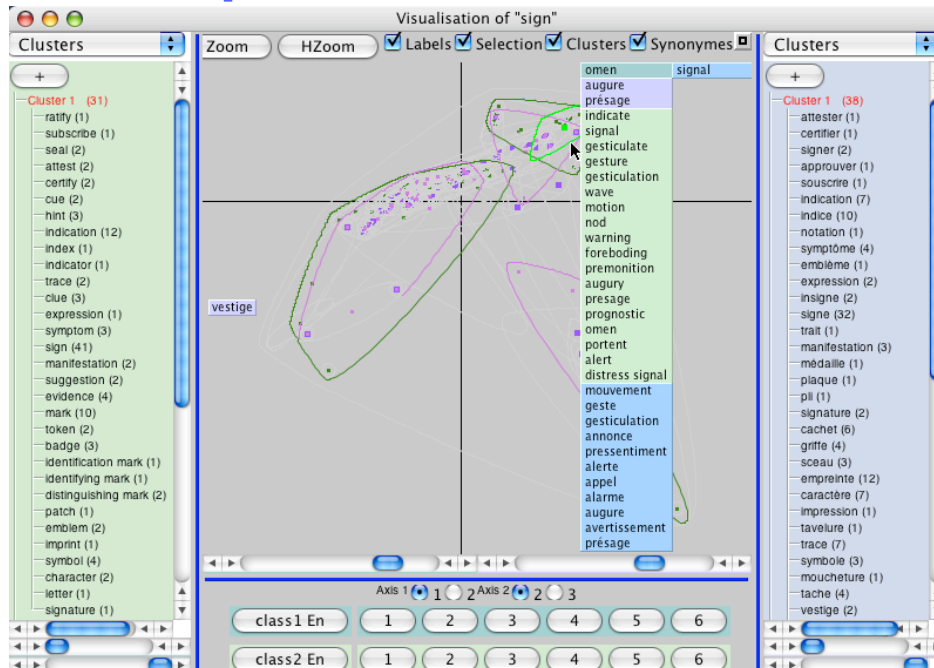
34

Example (sign EN) (1)



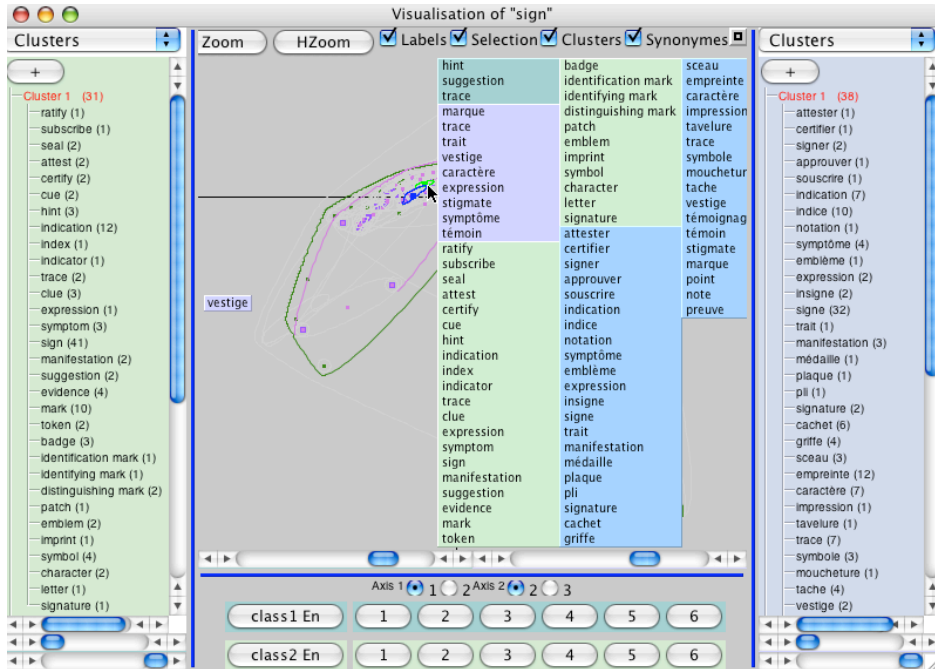
35

Example (sign EN) (2)



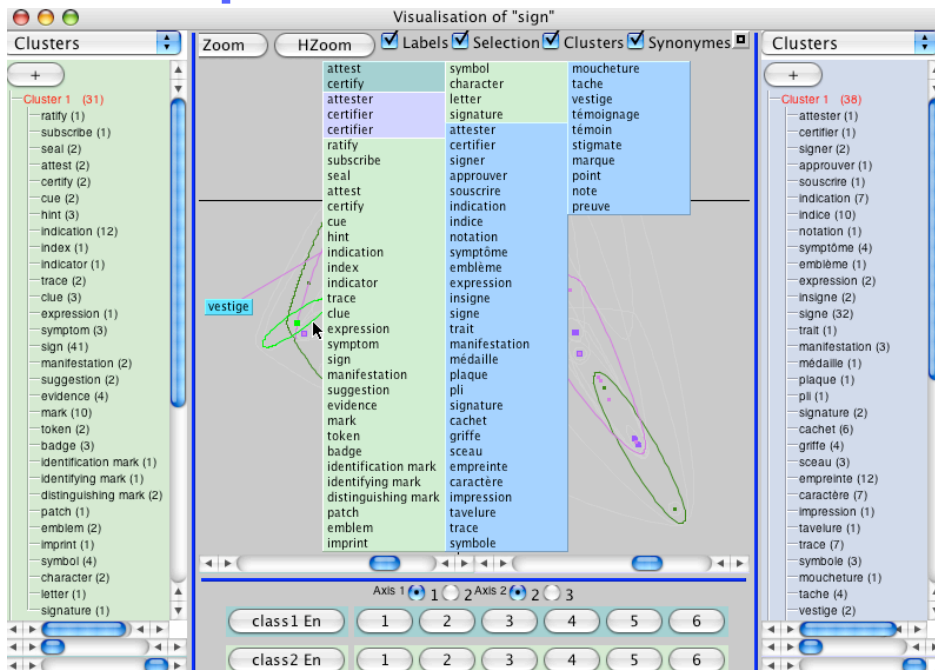
36

Example (sign EN) (3)



37

Example (sign EN) (4)



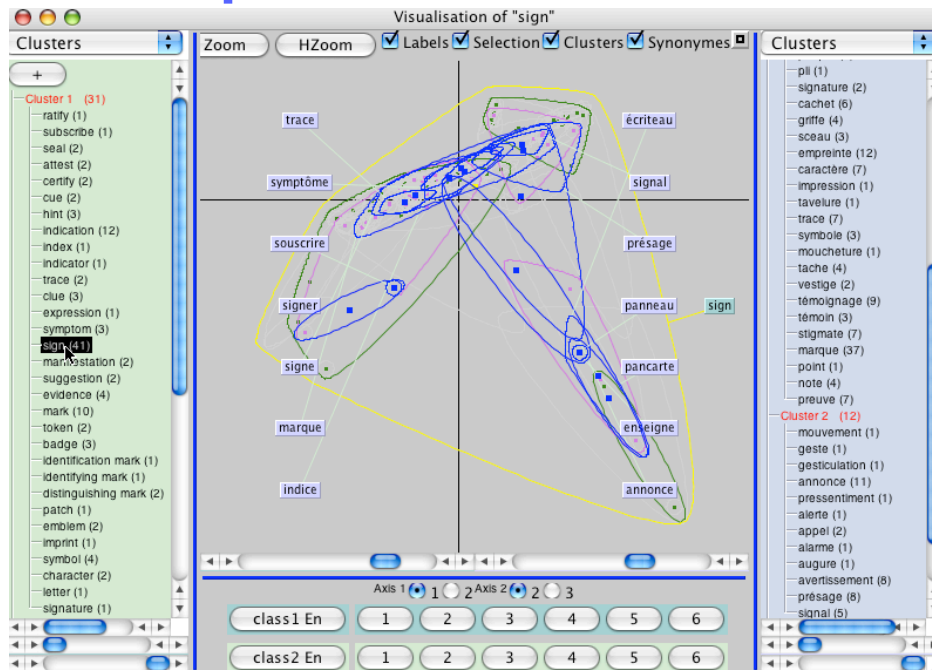
38

How does this help?

- Maps and interactive system as resource for translators – help with lexical choice.
- Better than translation database because it finds more terms. [Is that good?]
 - Avg no of terms: 14.1 vs 92.9.

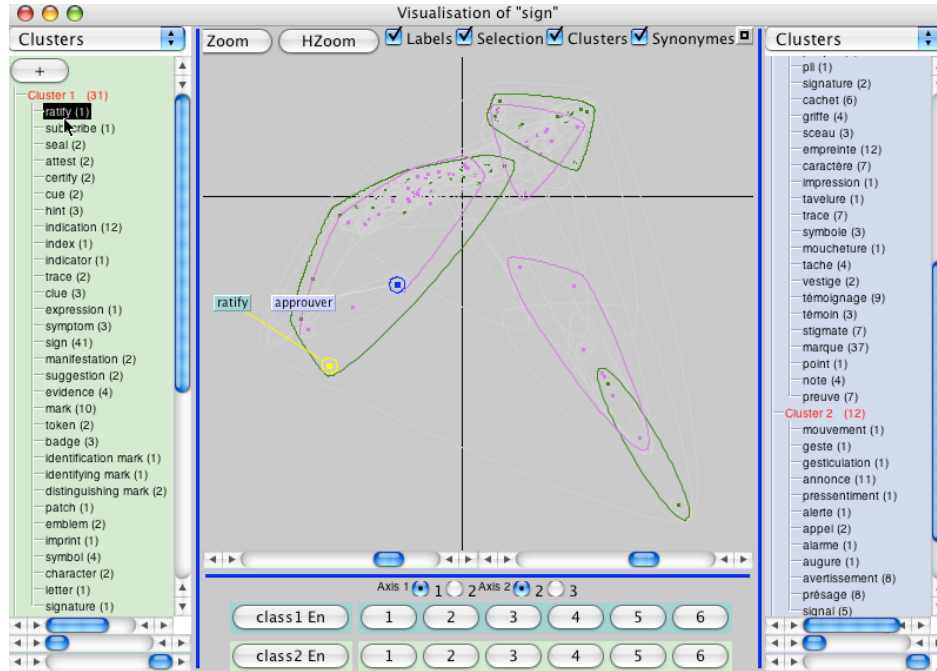
39

Example (*sign* EN) (5)



40

Example (sign EN) (6)



41

Use in MT?

- No claim re use in automatic lexical choice.
 - How could it be applied?
 - Requires knowing clique (precise intended sense) in TL.
 - “But the future looks bright for using this in automatic systems.” [Paraphrase]

42

Cliques of **contexonyms** and their semantic space

“Contexonyms”

- **Contexonyms** of W are the words that tend to co-occur with it.
 - Not symmetric or transitive.
 - [Poor term: Empirical, pretending to be lexical-semantic.]
- It's a relation between words. Hence can compute cliques with it.

Finding contexonyms

- For chosen word W , unspecified window size for co-occurrence, suitable corpus, stop-list, build [second-order] word-association table:
 - Take the k “children” that most frequently co-occur with W (in decreasing order).
 - Take the most frequent children of those children.
 - (Optional:) Discard any children whose children don't include W .
(i.e., require mutual frequent co-occurrence).

45

Finding contexonyms (2)

- Contexonyms are the remaining children of W , or are those plus *their* remaining children.
[Paper seems to say both of these.]

46

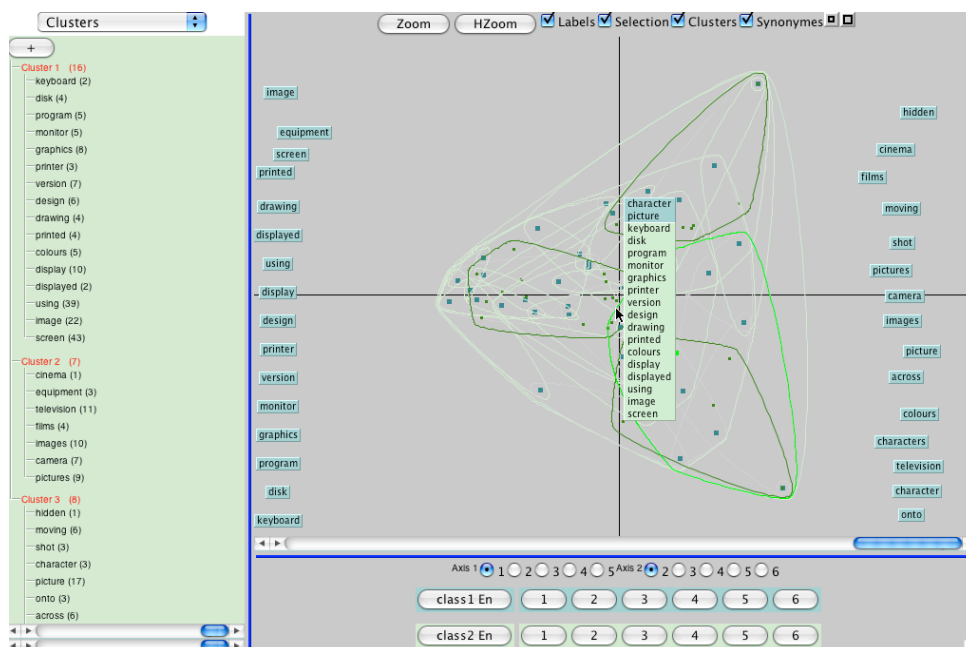
Cliques and semantic maps of contexonyms

- Compute cliques as before.
- Compute distances between cliques, project onto Euclidean space, find clusters of cliques (i.e., regions), label them.*

*Or, in this paper, cluster the contexonyms themselves (what does that mean?).

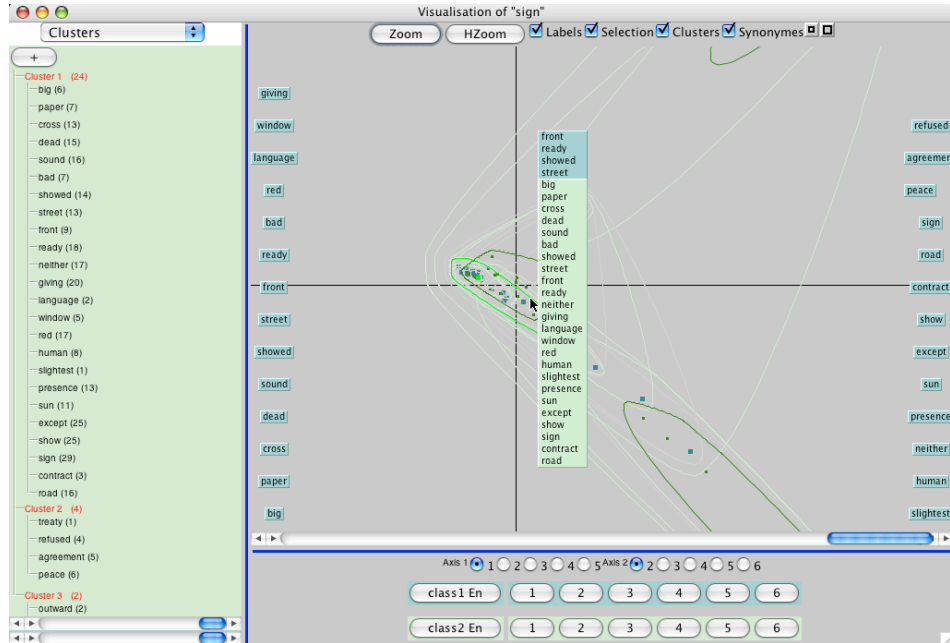
47

Example (screen)



48

Example (sign)



49

What does it mean?

- *Much* previous work on co-occurrence and on sense clustering; what's new here?
- Automatic determination of contextual cues for sense discrimination.
- Cues for lexical choice, e.g., in MT ...

50

Discover cues to near-synonym distinctions

- Find contexonyms for each member of set of near-synonyms.
 - Distinguishing nuances of near-synonyms may be reflected in contexonyms of each.
 - Also, some collocational information.
- E.g., *blunder* → *stupid, fatal, gross, commit, ...*
 - But only if parameters are set just right. With defaults, *blunder* → *commit* [only]

51

How does this help?

- Use in lexical choice:
 - Prefer words that are close (in this space) to other words in the text.
- Do we really need this whole apparatus for that?
 - Would frequent co-occurrence be sufficient?

52

Summary (1)

- Use (a) lexicographic resources and (b) raw text to find relations between words.
- Build cliques from relations, semantic space from cliques.
- Regions in this space reflects sub-senses.
- Distance in this space reflects near-synonymy and closely-related words.
 - Jane Morris: “Non-classical” lexical relations.

53

Summary (2)

- Regions are labelled; relations are not.
- Unclear what value is added over thesaurus, over other methods.
- Unclear how to apply in NLP applications or MT.

54