Thanasis Georgakopoulos
(University of Liège)

Eitan Grossman
(Hebrew University of Jerusalem)

Stéphane Polis
(F.R.S.-FNRS / University of Liège)

in collaboration with Dmitry Nikolaev
(Hebrew University of Jerusalem)

ALT12– Canberra – 11.12.2017 – 15.12.2017

Weighted lexical semantic maps for areal lexical typology

Verbs of perception and cognition as a case study

Outline of the talk

- Semantic maps
  - What are semantic maps?
  - Why use semantic maps for areal lexical typology?
- Case study: verbs of perception and cognition
  - Why is this semantic field interesting?
  - The datasets
    - CLICS
    - Vanhove (2008)
    - WORDNET
- Areal patterns and general discussion

Introduction

Semantic maps

+ Semantic maps

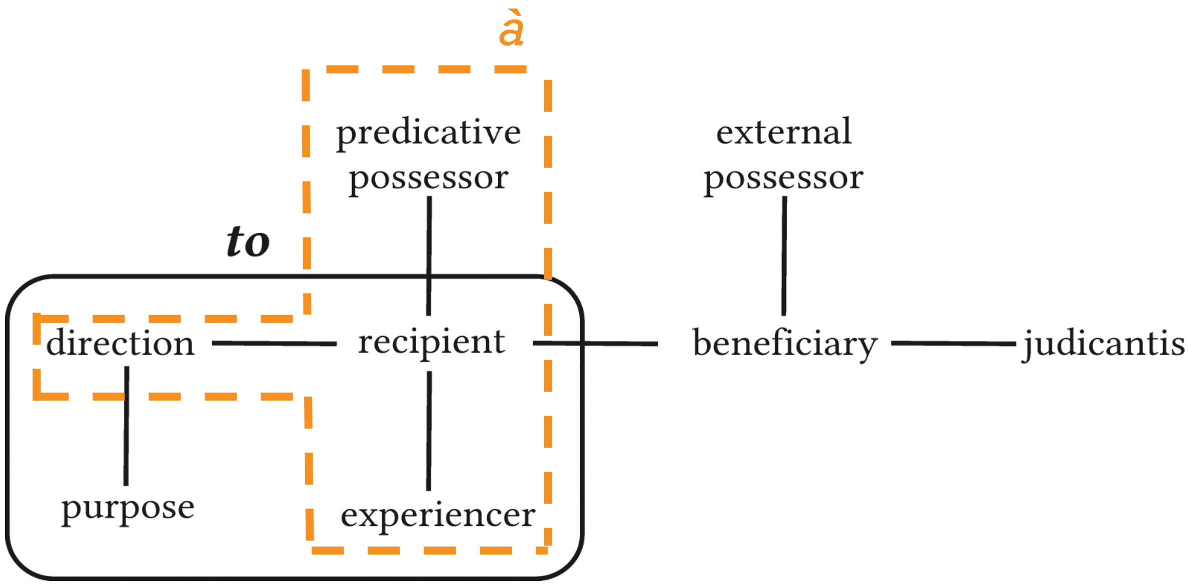


Figure 1. A semantic map of typical dative functions / the boundaries of English *to* and French *à*
(based on Haspelmath 2003: 213, 215)

Semantic maps

Weighted semantic maps

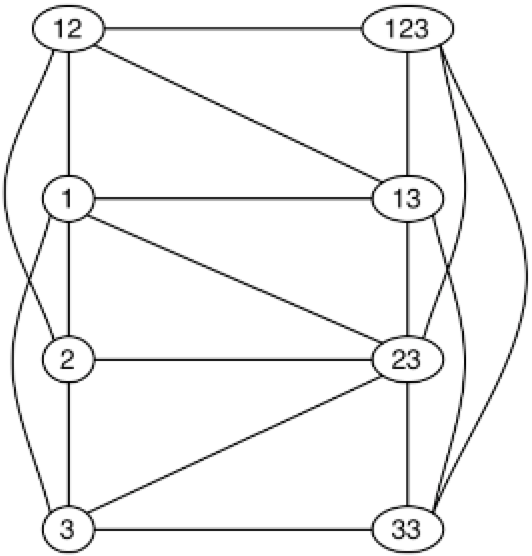


Figure 2a. A simple semantic map of person marking (Cysouw 2007: 231)

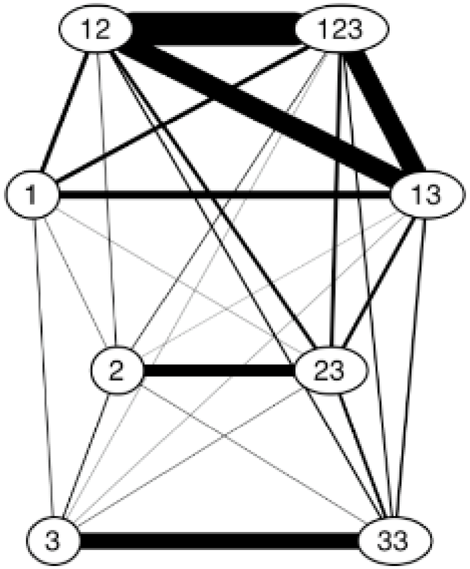


Figure 2b. A weighted semantic map of person marking (Cysouw 2007: 233)

Semantic maps

Lexical semantic maps

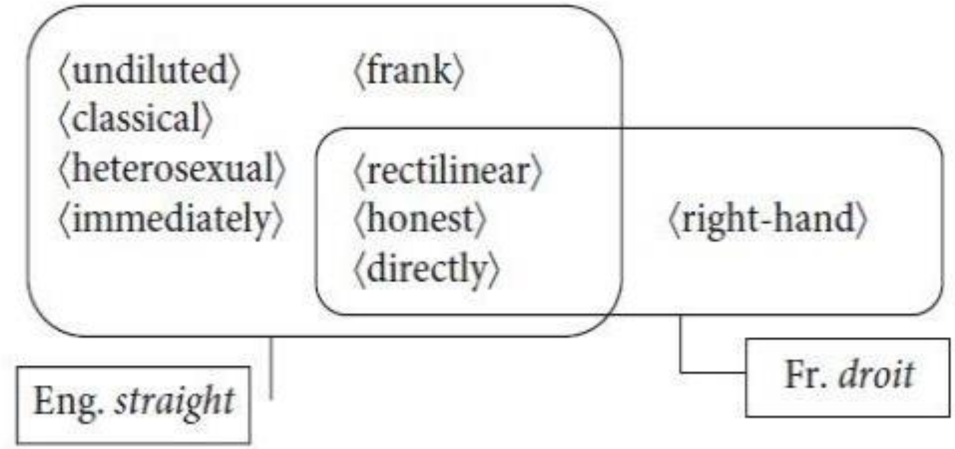


Figure 3. Overlapping polysemies:
Eng. straight *vs*. Fr. droit
(François 2008: 167)

Colexification = multifunctionality

Languages differ as to which senses they colexify

+ Semantic maps

Weighted lexical semantic maps

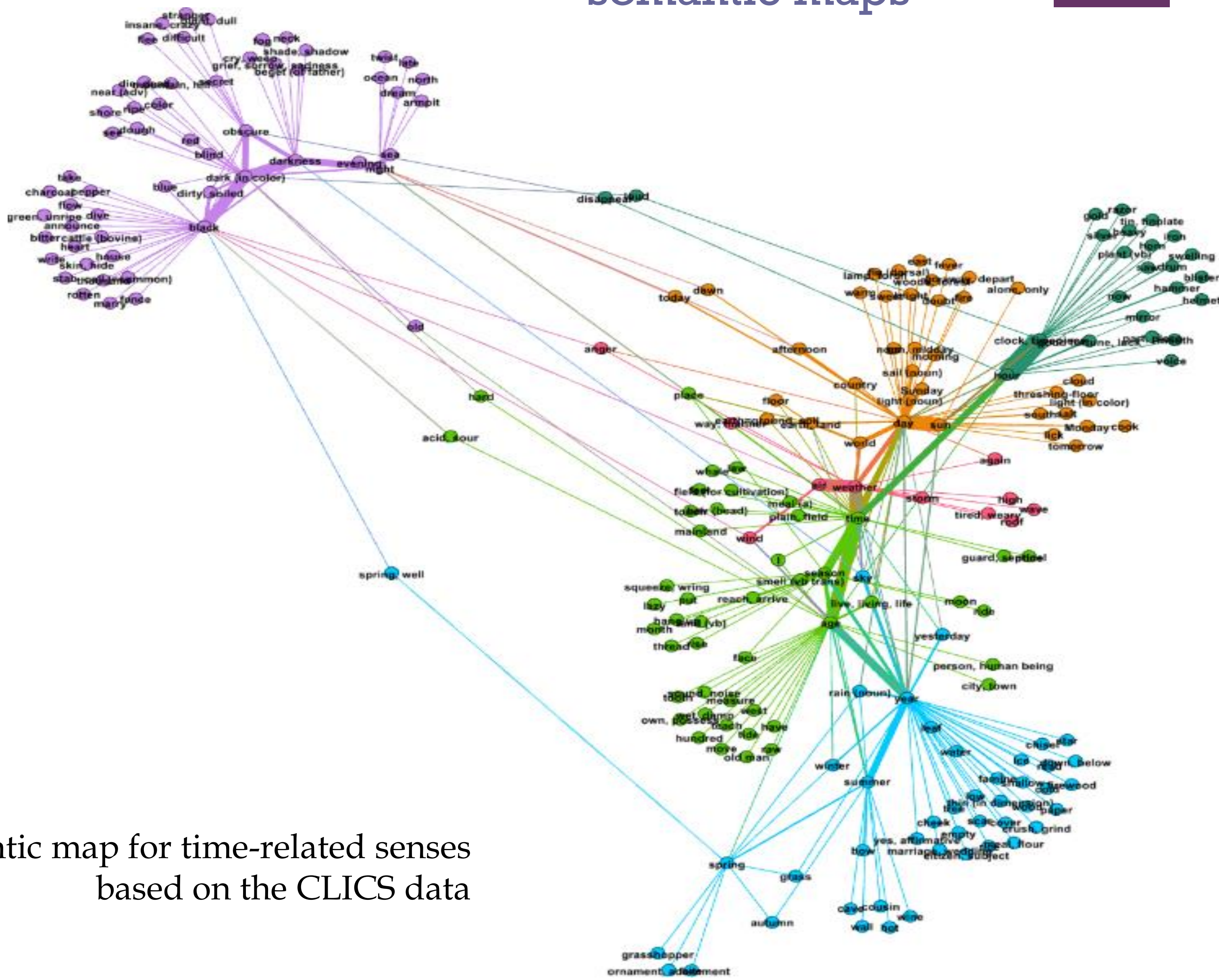


Figure 4. Semantic map for time-related senses based on the CLICS data

Semantic maps

Why use semantic maps for areal lexical typology?

- The map makes universal claims
  - Frequency: attested *vs.* non-attested and frequent *vs.* rare
  - Types of polysemy: possible *vs.* impossible
  - Implicational hierarchies (unlike other colexification networks)
- The mapping of language specific items allows
  - studying genetic, areal, and culture specific patterns,
  - but it also shows how hard it is to reach statistically significant results

Verbs of perception and cognition

A case study

+ Why perception & cognition?

- Perception and cognition are among the basic concepts that are lexicalized in the languages of the world (e.g. Swadesh 1952)
- Well-studied domain: our results can be compared (e.g. Sweetser 1990; Evans & Wilkins 2000; Vanhove 2008)
- The relevant literature has revealed both universal and culture-specific patterns

Verbs of perception & cognition

Semantic extensions

Intrafield (= *Intradomain*)
(senses: same semantic field)

Interfield (= *Interdomain/ Transfield*)
(senses: different semantic field)

(based on Wilkins 1996: 274; cf. Matisoff 1978)

Verbs of perception & cognition

Intrafield extensions

sight > hearing > touch > {smell / taste}

Figure 4. Viberg's sense modality hierarchy for semantic extensions and polysemies of perception verbs
(Viberg 1984: 136)

Table. Inventories of the verbs of perception
(Viberg 1984: 140)

Walbiri (West Australia) Source: Hale 1971: 478		Djaru (West Australia) Tasaku 1981: 418		Lesghian (East Caucasus) Dixon 1979: note 54	
nja-	'to see'	nyang-	'see/ look'	akun	'see/look'
puḍa-nja-	'to hear, to feel'	pura-nyang-	'hear/ listen'	van akun	'hear/listen'
paṇṭi-nja-	'to smell'				

+ Verbs of perception & cognition

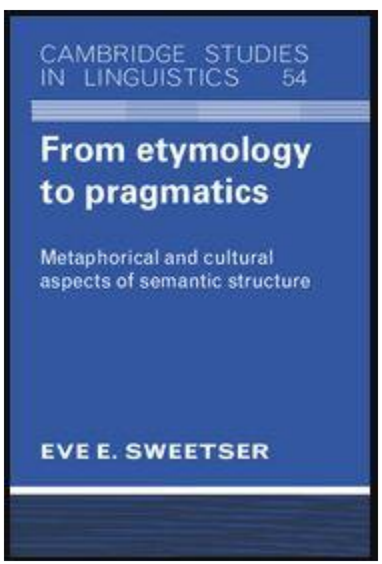
Interfield extensions

Mind-as-body-Metaphor:

The internal self is understood in terms of the bodily external self (Sweetser 1990: 45)

→

- **Common cross-linguistically (if not universal)**: the connection between VISION and KNOWLEDGE (Sweetser 1990: 45)



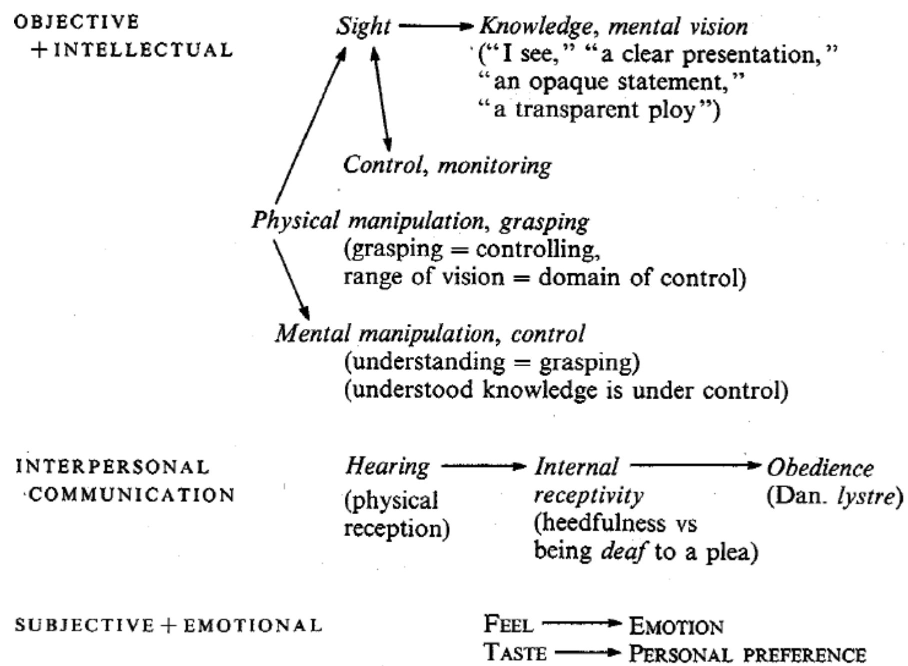


Figure 6. The structure of our metaphors of perception (Sweetser 1990: 38)

Verbs of perception & cognition

In Australian languages:
(Evans & Wilkins 2000)

- Cognitive verbs > 'hear'

(cf. intrafield extensions confirm the prevalence of vision)

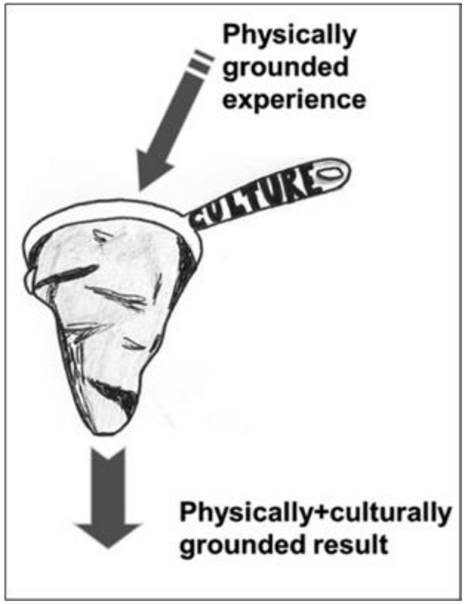


Figure 7. The culture sieve (Ibarretxe-Antuñano 2013)

Extensions in cognitive verbs:

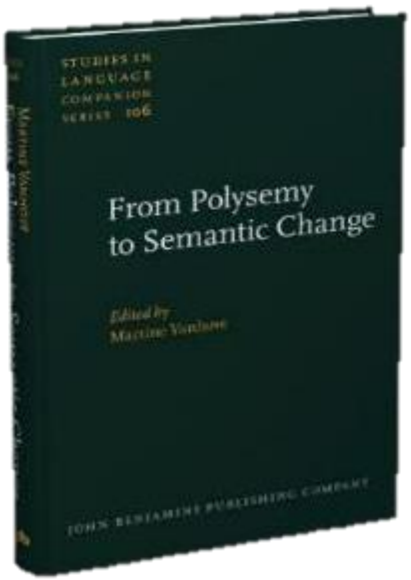
- A foot in culture: a relativistic aspect (cf. Sweetser 1990)
- A foot in nature: a universal aspect (Evans & Wilkins 2000)

The culture sieve:

- "filters" those elements that are in accordance with the premises of a given culture
- "impregnates" the mapping with touches of a culture in contrast with other cultural and social systems

(Ibarretxe-Antuñano 2013: 324)

More recent accounts



Vanhove 2008

- Sample of 25 languages (8 phyla); mostly African
  - Intrafield: vision prevails
  - Transfield: the auditory modality prevails
    - Stronger semantic association of hearing and mental perception
  - Implicational universal:
    - Hearing > vision > prehension

[no distinction between *controlled activity* (e.g. listen) vs. *non-controlled experience* (e.g. hear); cf. Viberg 1984; 2001]

More recent accounts

Wälchli 2016

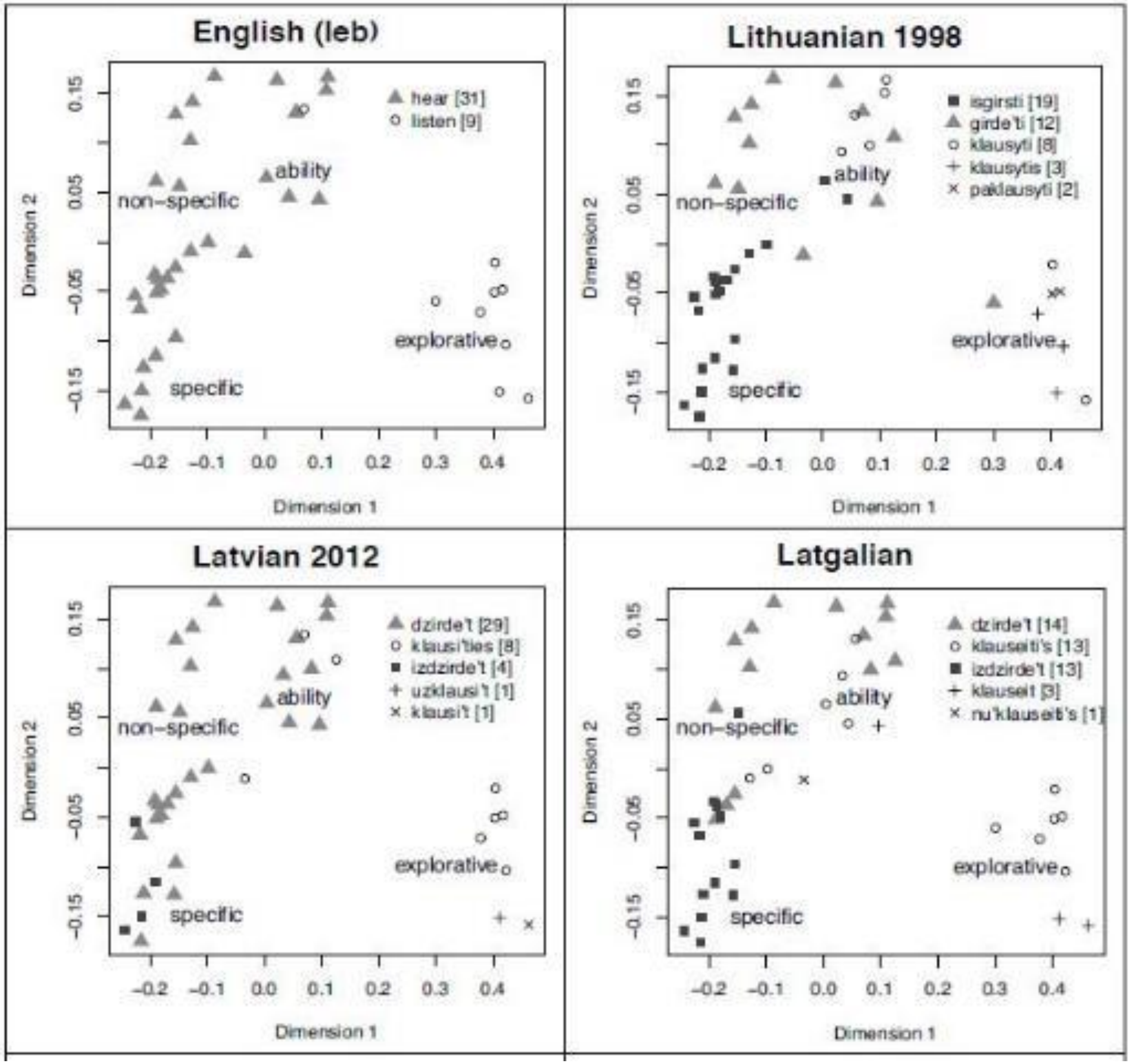
- Convenience sample: Central, East and North European languages
- Case study: Auditory and visual perception
  - *Explorative perception verbs* = *controlled activity* (e.g. listen)
  - *Opportunistic perception verbs* = *non-controlled experience* (e.g. hear)
    - *Specific perception verbs*: subtype of opportunistic perception verbs

- Goal: how the encoding of a specificity distinction may differ cross-linguistically.
  - Particular areal feature for Baltic languages

- Method: probabilistic semantic maps based on parallel corpora

More recent accounts

■ = specific 'hear', ▲ = non-specific 'hear', ○ = 'listen'



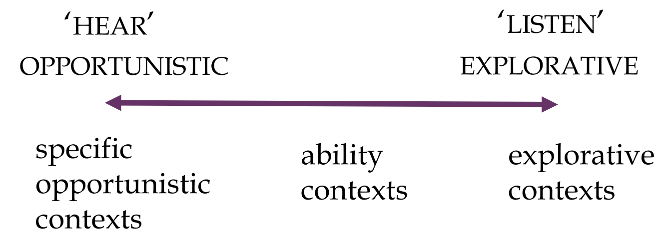
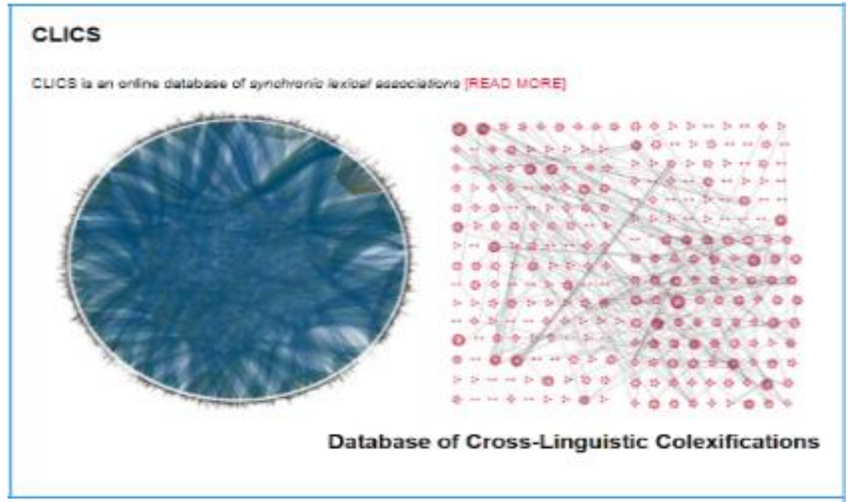


Figure 8. Probabilistic semantic map of 44 auditory contexts in Mark based on 64 doculects in English (leb), Lithuanian (1998), Latgalian and Latvian (2012) (Wälchli 2016: 77)

Datasets for building lexical semantic maps

Perception and cognition

CLICS



- *N* of lgs: 221
- *N* of lg families: 64
- *N* of concepts: 1280

Meaning 1	Meaning 2	*N of language*	*N of forms*	language:form
see	know	5	6	aro_std:[ba]//ayo_std:[iˈmoʔ]//haw_std:[ʔike]//mcq_std:[ɓanahe]//mri_std:[kitea]//tel_std:[aarayu]//tel_std:[arayu]
see	find	15	23	agr_std:[wainat]//arn_std:[pe]//con_std:[ˈatʰeye]//cwg_std:[yow]//emp_std:[uˈnu]//kgp_std:[we]//kpv_std:[addzɪnɪ]//kyh_std:[mah]//mca_std:[wen]//mri_std:[kitea]//oym_std:[ɛsa]//pbb_std:[uy]//plt_std:[mahìta]//pui_std:[duk]//ray_std:[tikeʔa]//rtm_std:[ræe]//sap_Enlhet:[neŋwetayˀ]//sei_std:[aʔo]//shb_std:[taa]//sja_std:[unu]//swh_std:[ona]//tbc_std:[le]//yag_std:[tiki]
see	get, obtain	6	6	kgp_std:[we]//mbc_std:[eraʔma]//pbb_std:[uy]//sap_Standard:[akwitayi]//srq_std:[tea]//udi_std:[акъсун]

Polysemy data from CLiCs (http://clics.lingpy.org/download.php)

(List et al. 2014)

+ CLICS

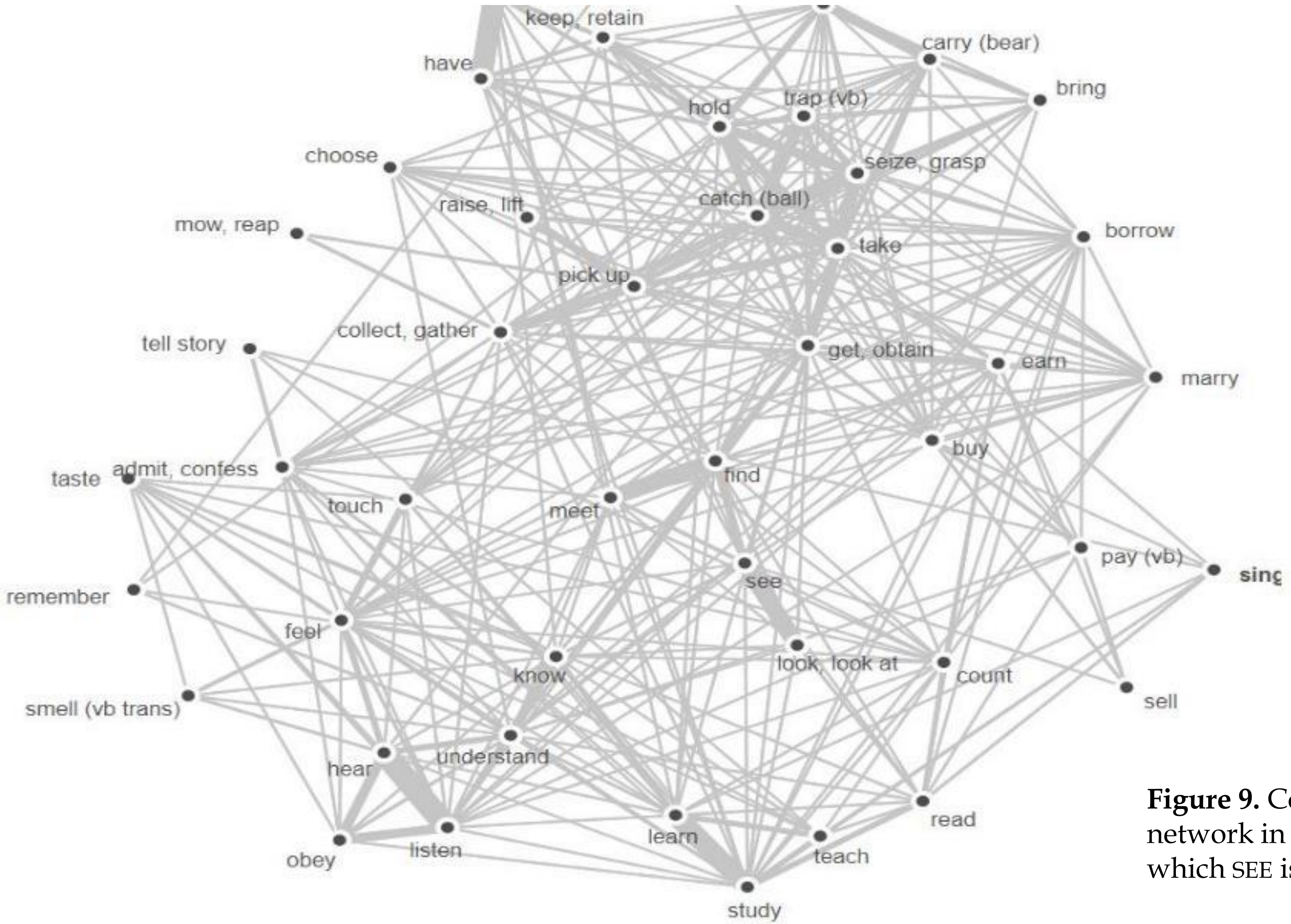


Figure 9. Complete network in CLICS of which SEE is part

CLICS

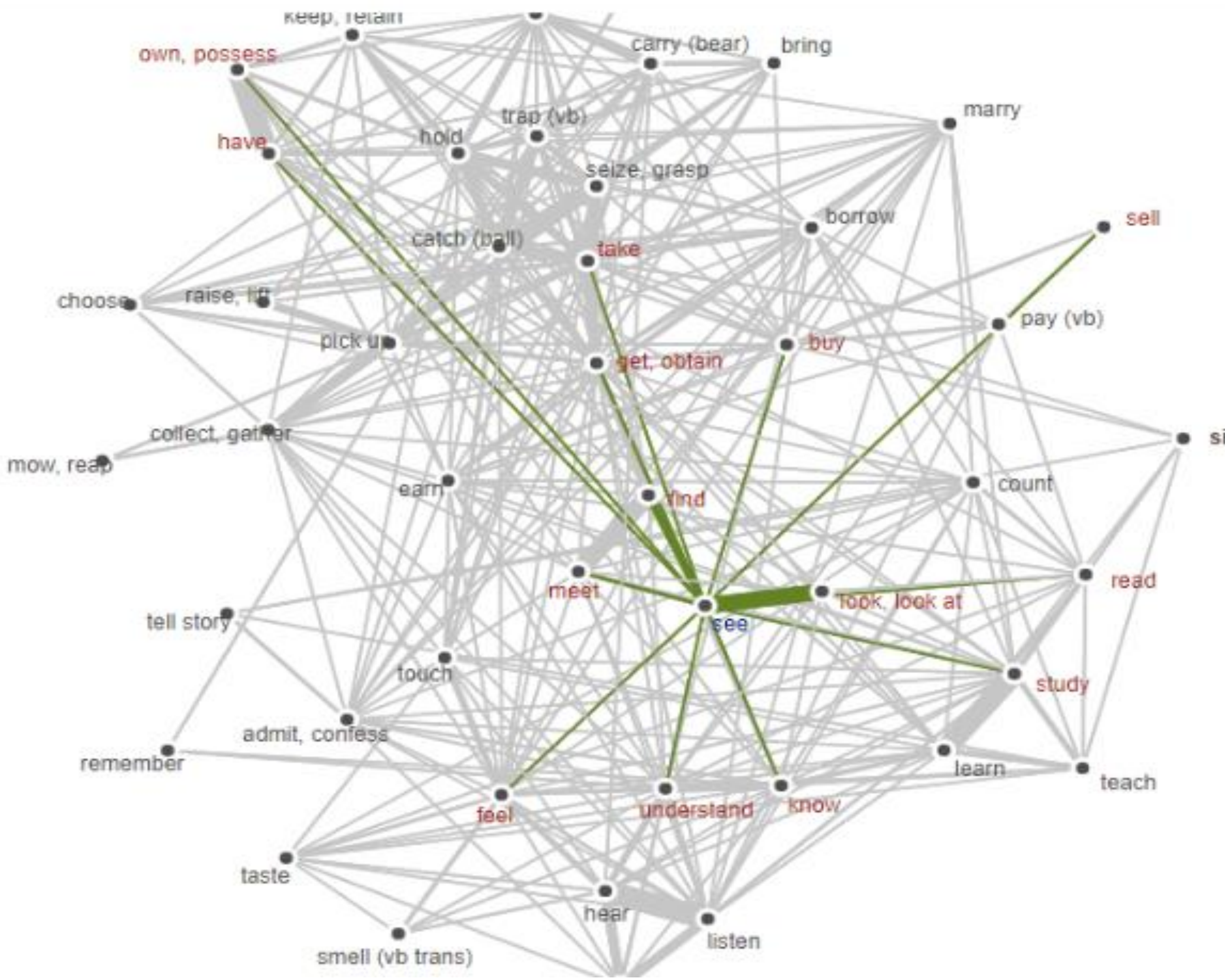
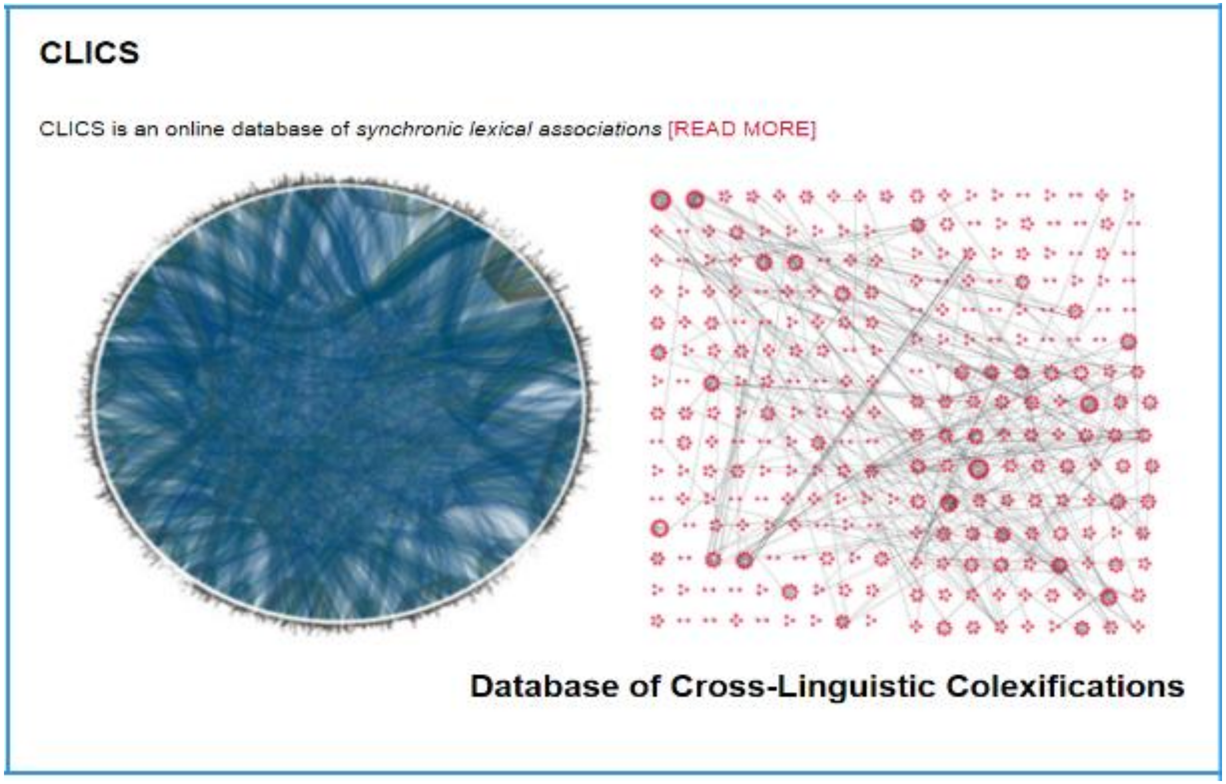


Figure 10. Complete network in CLICS with SEE as a pivot

CLICS



Waiting for CLICS 2.0 …
(List 2017)

- Increased quantity of data
- Increased quality of data (e.g., links to the Concepticon)
- Include partial colexifications
- Normalize the data which is analysed by CLICS

From CLICS to a more economical map

The economy principle

Given three meanings (Meaning$_A$, Meaning$_B$, Meaning$_C$), if the linguistic items expressing Meaning$_A$ and Meaning$_C$ always express Meaning$_B$, there is no need to draw an edge between Meaning$_A$ and Meaning$_C$ (the resulting map will not be triangular, i.e. a vacuous semantic map, with all the meanings connected).

(Georgakopoulos & Polis *forthcoming*)

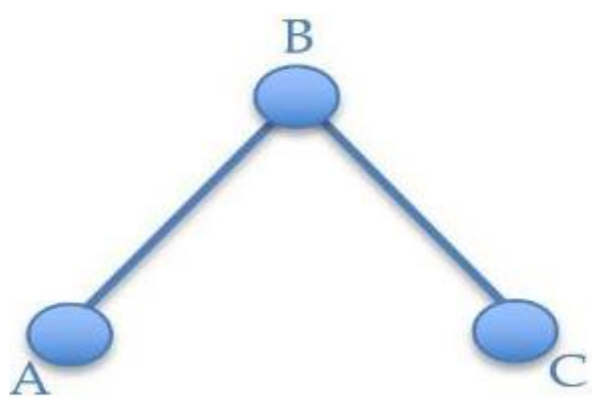


Figure 11. An abstract semantic map

From CLICS to a more economical map

- The synchronic polysemy patterns are converted into a lexical matrix

```
Tmap = [Tsenses]
for t in Tclean:
        split_langWord = t[2].split('//')
        for couple in split_langWord:
                langWord = couple.split(':')
                line = [langWord[0], langWord[1]]
                for i in range (2,len(Tsenses)):
                        line.append('0')
                line[Tsenses.index(t[0])] = '1'
                line[Tsenses.index(t[1])] = '1'
                Tmap.append(line)
```

Python script α

Languages   Forms

Source of constraint	Constraint name	Meaning 1 'SEE	Meaning 2 'KNOW'	Meaning 3 'GET, OBTAIN
Araona	ba	1	1	0
Ayoreo	iˈmoʔ	1	1	0
Hawaiian	ʔike	1	1	0
Ese	ɓanahe	1	1	0
Maori	kitea	1	1	0
Telugu	aarayu	1	1	0
Kaingang	we	1	0	1
Macushi	eraʔma	1	0	1
Páez	uy	1	0	1
Sanapaná (Standard)	akwitayi	1	0	1
Sirionó	tea	1	0	1
Udi	акъсун	1	0	1

From CLICS to a more economical map

Source of constraint (Languages)	Constraint name (Forms)	Meaning 1 'SEE	Meaning 2 'KNOW'	Meaning 3 'GET, OBTAIN
Araona	ba	1	1	0
Ayoreo	i'moʔ	1	1	0
Hawaiian	ʔike	1	1	0
Ese	ɓanahe	1	1	0
Maori	kitea	1	1	0
Telugu	aarayu	1	1	0
Kaingang	we	1	0	1
Macushi	eraʔma	1	0	1
Páez	uy	1	0	1
Sanapaná (Standard)	akwitayi	1	0	1
Sirionó	tea	1	0	1
Udi	акъсун	1	0	1

Weighted semantic map based on an adapted version of the algorithm suggested by Regier et al. (2013)

```
# CREATE INITIAL GRAPH
# graph G: add each term's nodes, no edges in graph yet.
G = nx.Graph()  # create empty graph (undirected)
PossE = []      # list of possible edges, filled below
for t in T:
        # add all nodes in t, if not already in graph
        for n in t:
                if (not G.has_node(n)):
                        G.add_node(n)
        # add to PossE a link between each pair of nodes in t
        # adding a link between every node in G is needless and slower
        for pair in allpairs(t):
                u = pair[0]
                v = pair[1]
                if (not (((u,v) in PossE) or ((v,u) in PossE))):
                        PossE.append((u,v))
```

Python script β

+ From CLICS to a more economical map

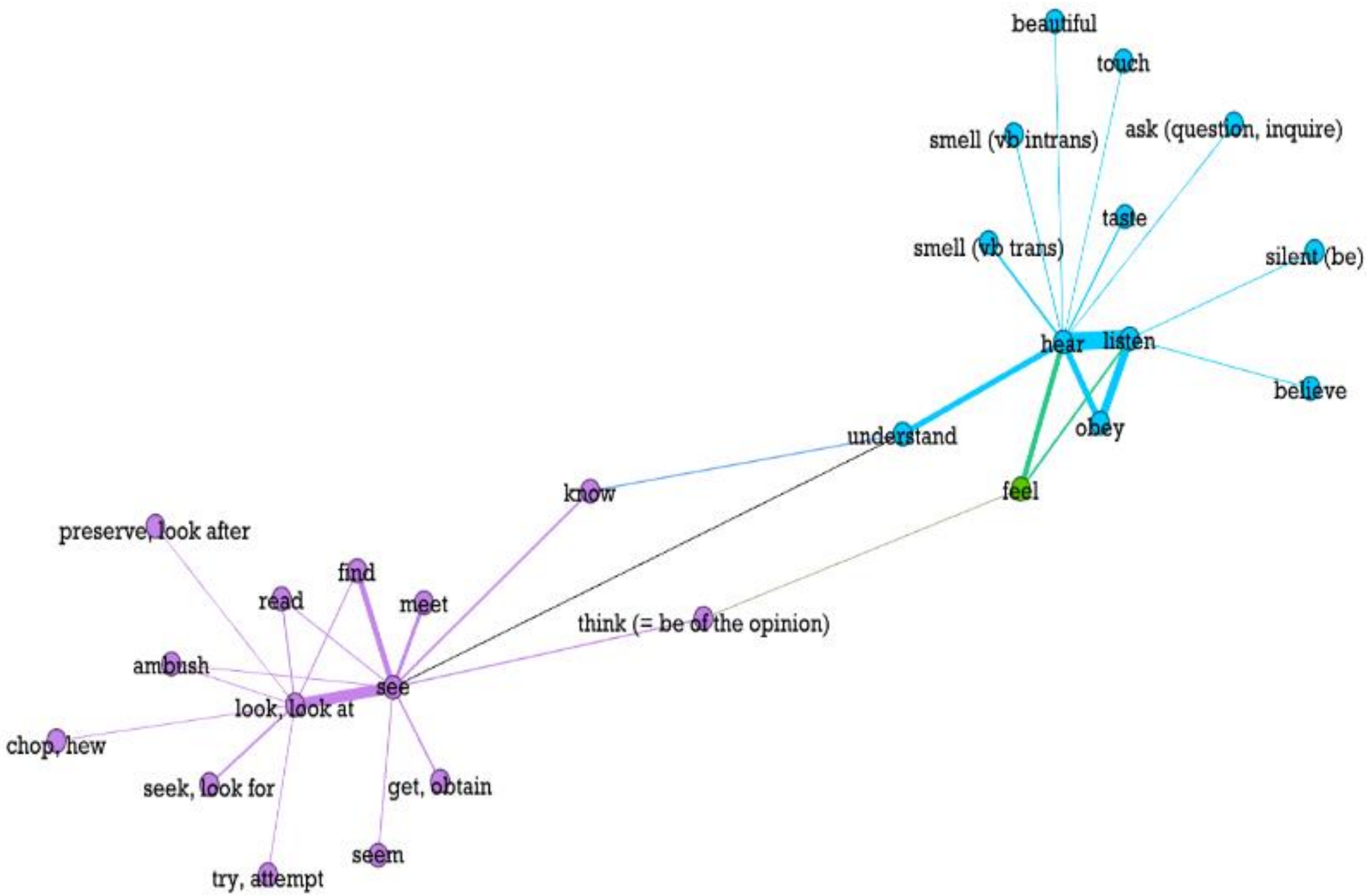


Figure 12. Full semantic map for the cognition-perception domain, visualized with modularity analysis* (Blondel et al. 2008) in *Gephi*

* A method to extract the community structure of large networks. Here, the different colors point to modules (also called clusters or communities) with dense connections between the nodes within the network.

+ From CLICS to a more economical map

- Direct edge between perception verbs denoting *non-controlled experience* (e.g., HEAR, SEE) and cognitive verbs (e.g., UNDERSTAND)

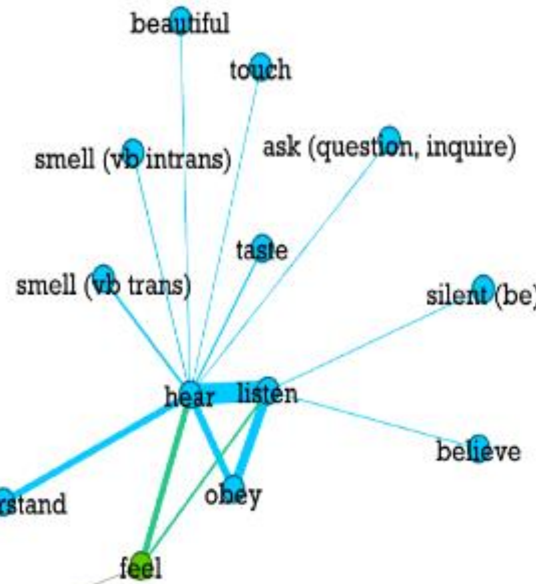
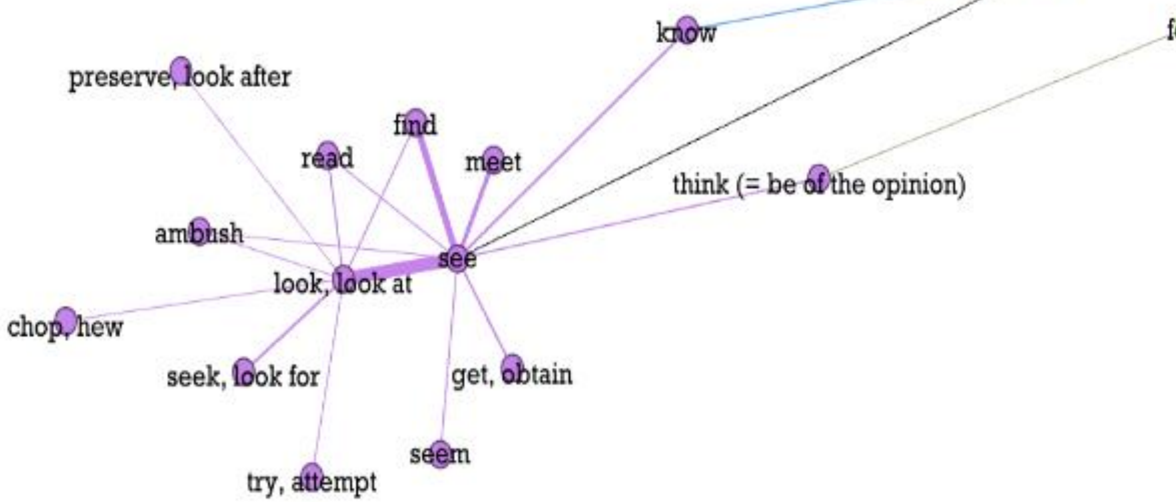



Figure 12. Full semantic map for the cognition-perception domain, visualized with modularity analysis* (Blondel et al. 2008) in *Gephi*

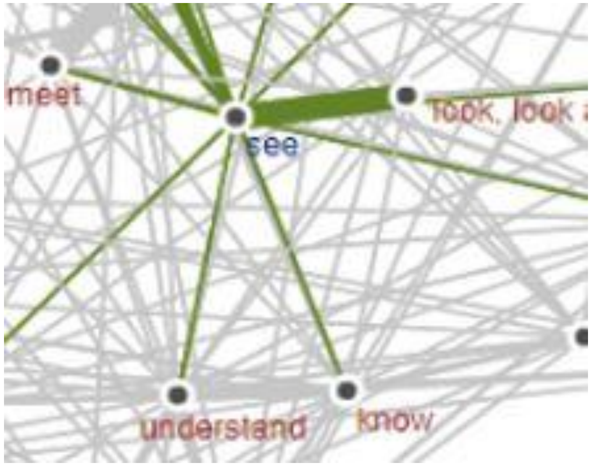


Figure 13. Snapshot from CLICS with SEE as a pivot

+ From CLICS to a more economical map

- Direct edge between perception verbs denoting *non-controlled experience* (e.g., HEAR, SEE) and cognitive verbs (e.g., UNDERSTAND)

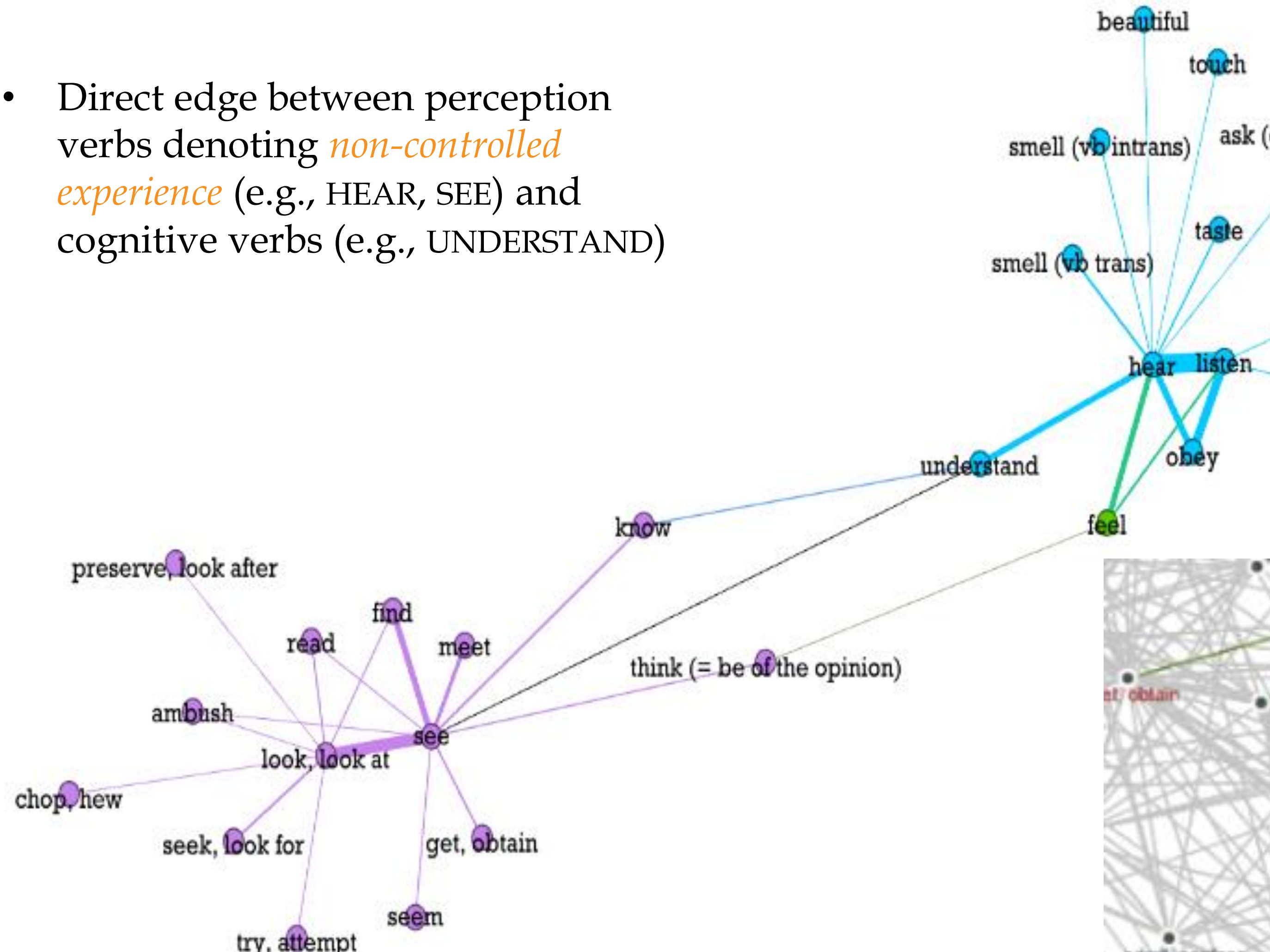


Figure 12. Full semantic map for the cognition-perception domain, visualized with modularity analysis* (Blondel et al. 2008) in *Gephi*

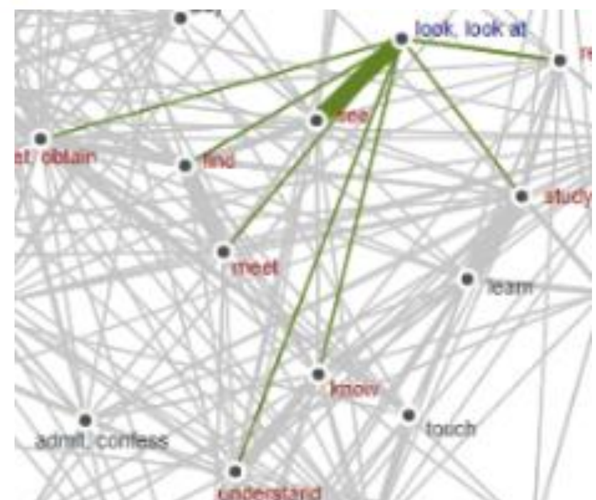


Figure 14. Snapshot from CLICS with LOOK as a pivot

+ From CLICS to a more economical map

- Direct edge between perception verbs denoting *non-controlled experience* (e.g., HEAR, SEE) and cognitive verbs (e.g., UNDERSTAND)
- No intrafield extension between SEE and HEAR, without going through *interfield meanings*

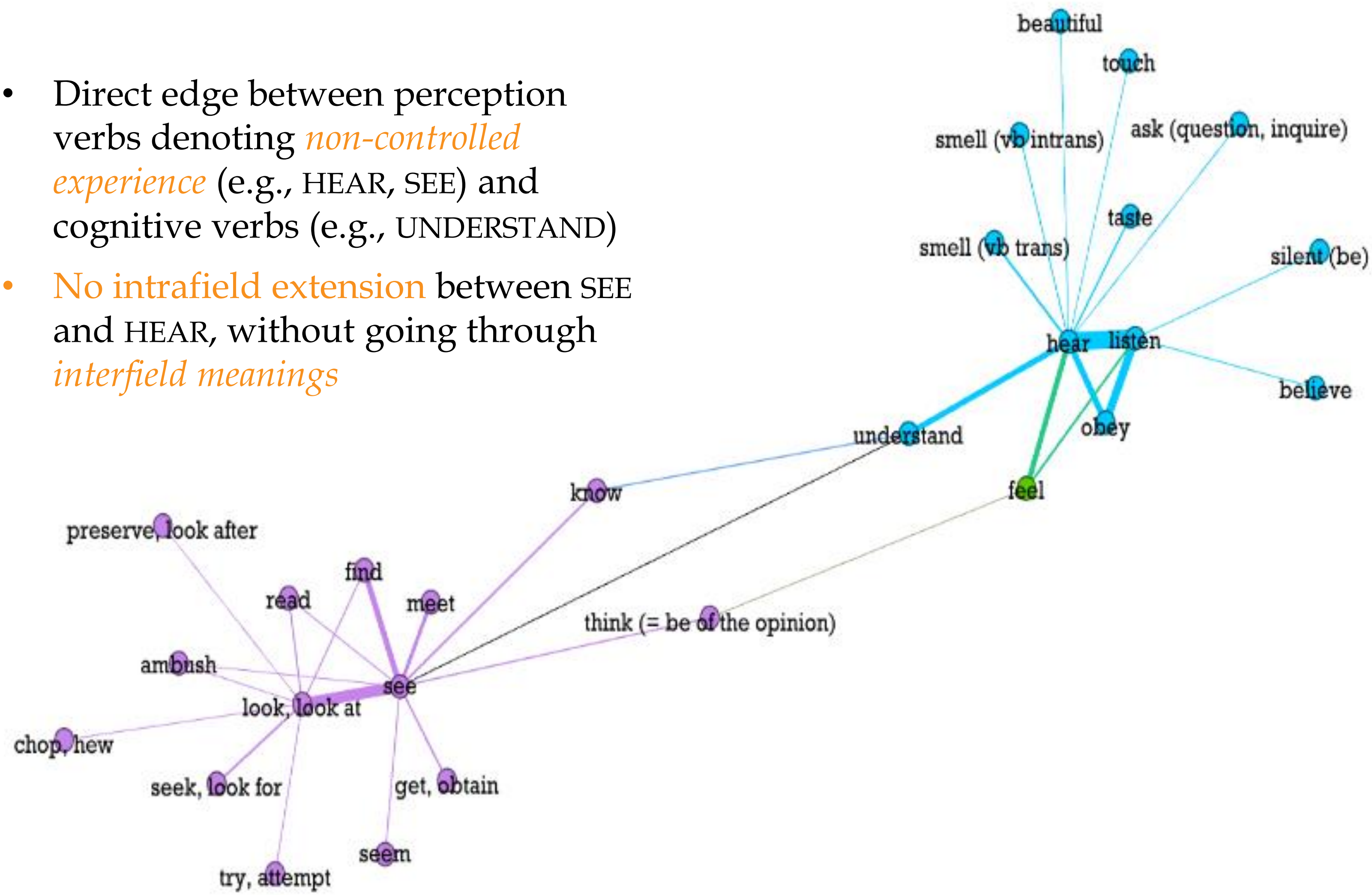


Figure 12. Full semantic map for the cognition-perception domain, visualized with modularity analysis* (Blondel et al. 2008) in *Gephi*

Mapping Vanhove's data

- Visualization of frequency of polysemy patterns

- Implicational hierarchies:
  - If THINK and SEE, then KNOW
  - If HEAR and LEARN, then KNOW

- The map predicts more than the attested data
  - If REMEMBER and SEE, then UNDERSTAND
  - "[A] good model always predicts a few things not yet encountered" (Cysouw 2007: 233)

- HEAR, KNOW, and UNDERSTAND are the most important nodes in the map (articulation points)

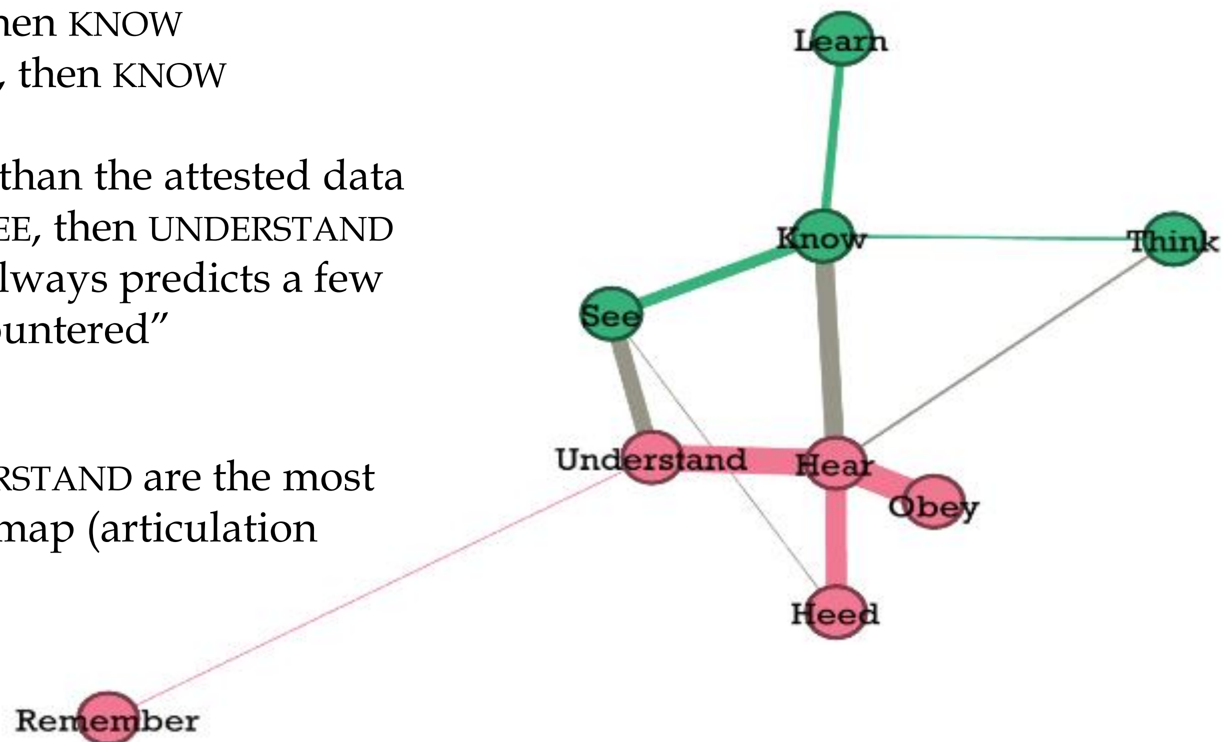


Figure 15. Semantic map for the cognition-perception domain based on Vanhove's (2008) data

+ Mapping Vanhove's data

- Visualization of frequency of polysemy patterns

- Implicational hierarchies:
  - If THINK and SEE, then KNOW
  - If HEAR and LEARN, then KNOW

- The map predicts more than the attested data
  - If REMEMBER and SEE, then UNDERSTAND
  - "[A] good model always predicts a few things not yet encountered" (Cysouw 2007: 233)

- HEAR, KNOW, AND UNDERSTAND are the most important nodes in the map (articulation points)

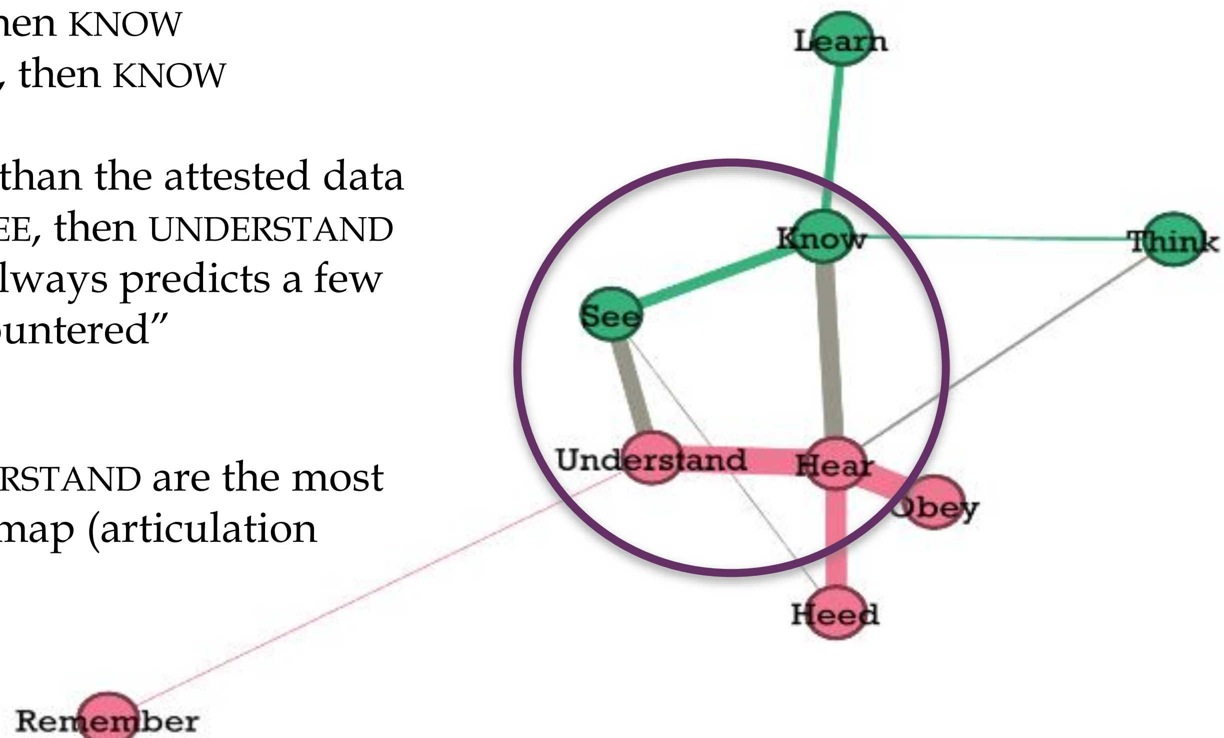


Figure 15. Semantic map for the cognition-perception domain based on Vanhove's (2008) data

- Again, no intrafield polysemy is allowed without the intervention of an interfield polysemy
  - If SEE and HEAR, then either KNOW or UNDERSTAND

Wordnet

A database of words that are linked together by their semantic relationships

- *N* of lgs: 25

Global WordNet Association

A free, public and non-commercial organization that provides a platform for discussing, sharing and connecting wordnets for all languages in the world.

More info GWA

Global WordNet Association

Core concept

Words are grouped together as sets of synonyms (Fellbaum 1998: 72ff.)

Synset: A synonym set; a set of words that are roughly synonymous in a given context

A prerequisite for the representation of meanings in a lexical matrix (Miller et al. 1993)

Wordnet

A database of words that are linked together by their semantic relationships

- *N of lgs*: 25

Global WordNet Association

A free, public and non-commercial organization that provides a platform for discussing, sharing and connecting wordnets for all languages in the world.

More info GWA

Global WordNet Association

Core concept

Words are grouped together as sets of synonyms (Fellbaum 1998: 72ff.)

Synset: A synonym set; a set of words that are roughly synonymous in a given context

⇓

A prerequisite for the representation of meanings in a lexical matrix (Miller et al. 1993)

Method

1. Choose four basic senses:
   a) SEE, HEAR (non-controlled experience / opportunistic perception verbs)
   b) LOOK; LISTEN (controlled activity / explorative perception verbs)
2. Collect the forms that lexicalize these 4 senses
3. Retrieve the list of all the senses of these forms (the total of the synsets)

+ Wordnet

Method

4. For each form, check whether the senses collected are among its senses
5. Generate a polysemy matrix

These five steps are implemented in a Python script that uses the Wordnet module of the Natural Language Toolkit (NLTK)

- S: (v) **see** (perceive by sight or have the power to perceive by sight) *"You have to be a good observer to see all the details"; "Can you see the bird in that tree?"; "He is blind--he cannot see"*
- S: (v) understand, realize, realise, **see** (perceive (an idea or situation) mentally) *"Now I see!"; "I just can't see your point"; "Does she realize how important this decision is?"; "I don't understand the idea"*
- S: (v) witness, find, **see** (perceive or be contemporaneous with) *"We found Republicans winning the offices"; "You'll see a lot of cheating in this school"; "The 1960's saw the rebellion of the younger generation against established traditions"; "I want to see results"*
- S: (v) visualize, visualise, envision, project, fancy, **see**, figure, picture, image (imagine; conceive of; see in one's mind) *"I can't see him on horseback!"; "I can see what will happen"; "I can see a risk in this strategy"*
- S: (v) **see**, consider, reckon, view, regard (deem to be) *"She views this quite differently from me"; "I consider her to be shallow"; "I don't see the situation quite as negatively as you do"*
- S: (v) learn, hear, get word, get wind, pick up, find out, get a line, discover, **see** (get to know or become aware of, usually accidentally) *"I learned that she has two grown-up children"; "I see that you have been promoted"*
- S: (v) watch, view, **see**, catch, take in (see or watch) *"view a show on television"; "This program will be seen all over the world"; "view an exhibition"; "Catch a show on Broadway"; "see a movie"*

Image 1. A snapshot of Wordnet's synsets of the verb see in English

Language	Form	Sense					
		SEE	UNDERSTAND	WITNESS	CONSIDER	LOOK	WATCH
English	see	1	1	1	0	0	1
French	regarder	0	0	0	1	1	1
Spanish	mirar	0	0	0	0	1	1
Spanish	observar	0	0	0	0	1	1
Spanish	ver	1	0	1	0	1	1

Wordnet

- Direct edges between perception verbs denoting *non-controlled experience* (e.g., HEAR) and cognitive verbs (e.g., UNDERSTAND)

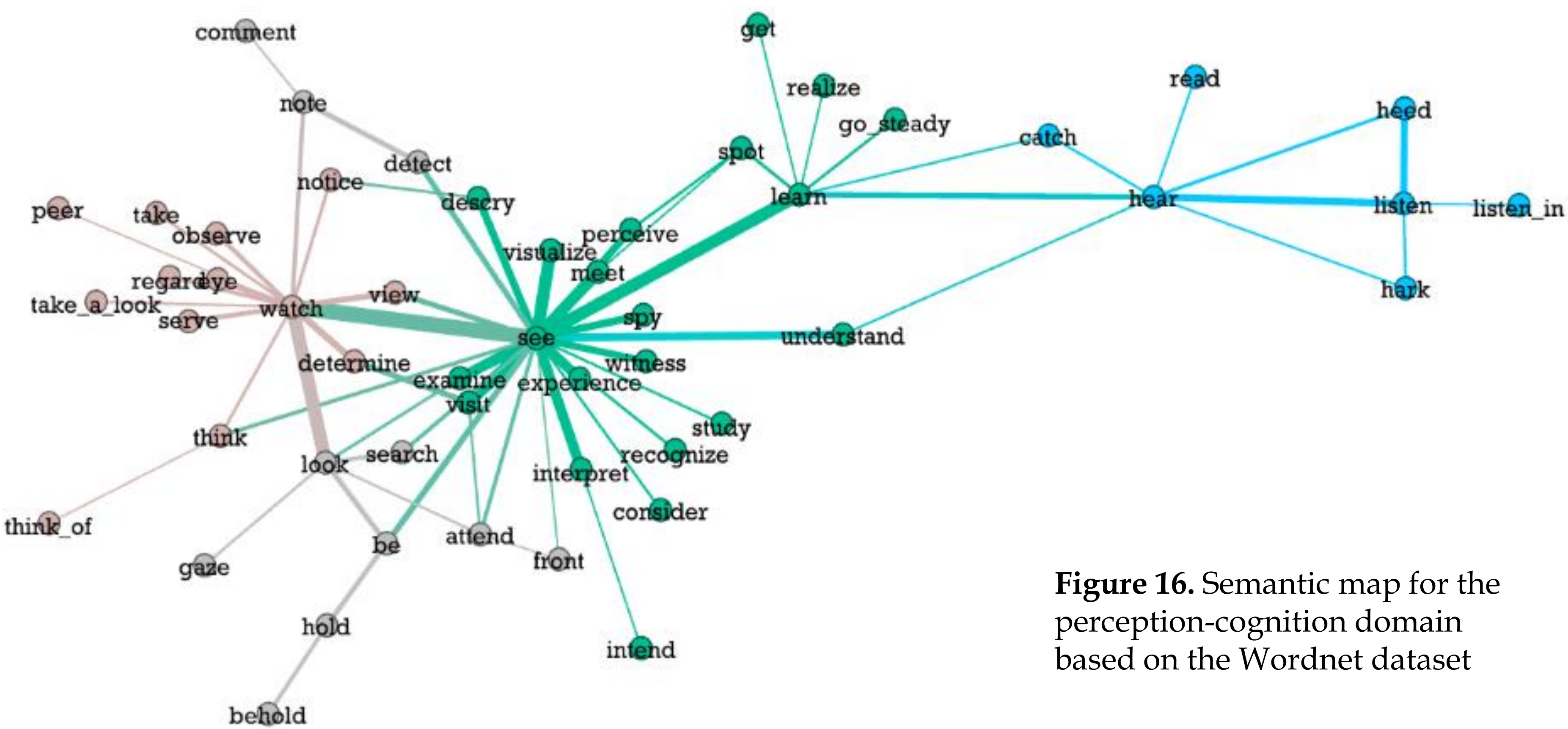


Figure 16. Semantic map for the perception-cognition domain based on the Wordnet dataset

Discussion

I. Areal patterns in CLICS

II. General discussion

Discussion

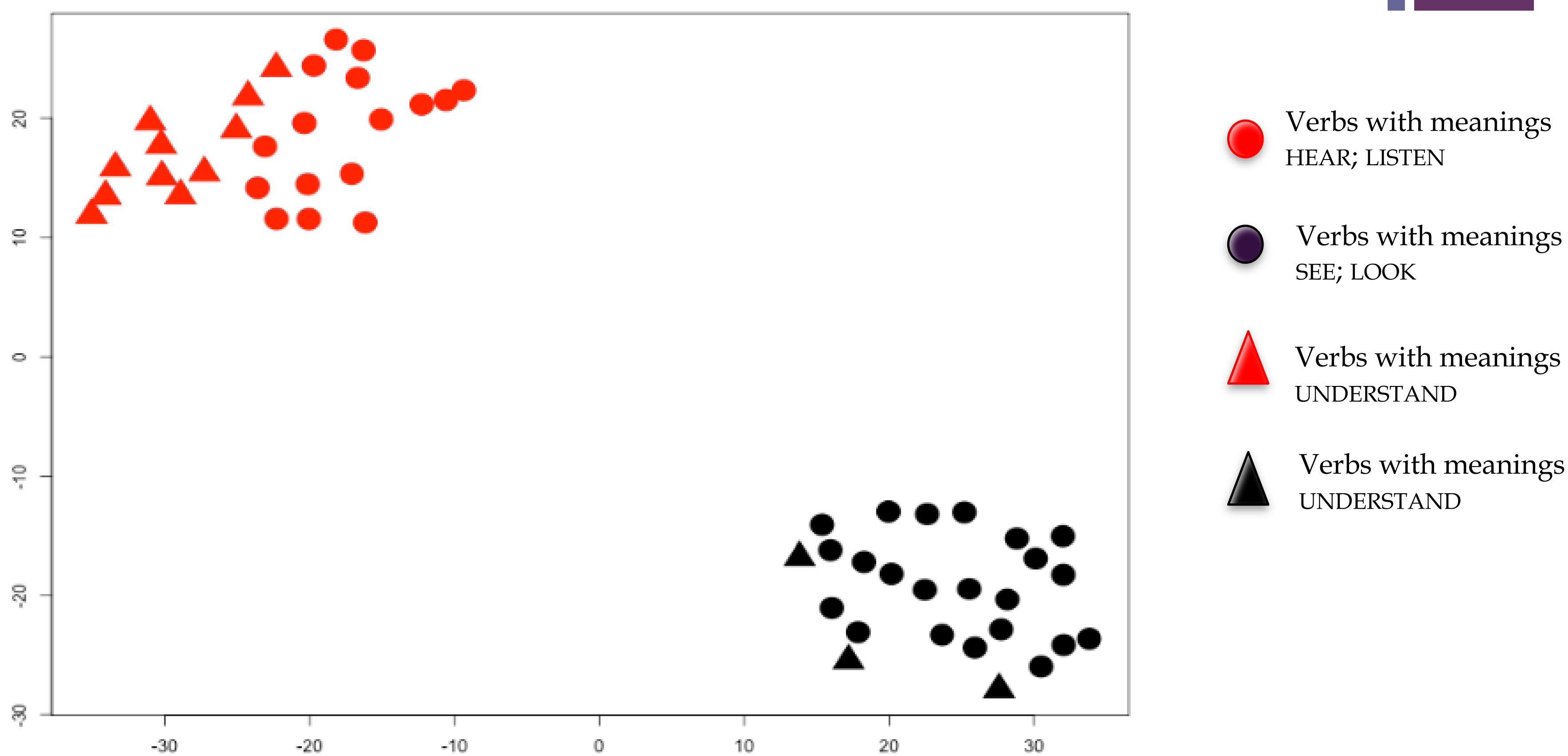


Figure 17. HEAR/ LISTEN *vs.* SEE/ LOOK: A 2D t-SNE projection (van der Maaten & Hinton 2008) of CLICS polysemy data

+ Discussion: Areal patterns in CLICS

Macro-areas in CLICS

Africa
Australia
Eurasia
North America
Papua
South America

- Coverage of the world's languages in CLICS is biased towards certain regions of the world (South American languages, languages of the Caucasus, languages of Europe figure particularly prominently).

(List et al. 2014)

Discussion: Areal patterns in CLICS

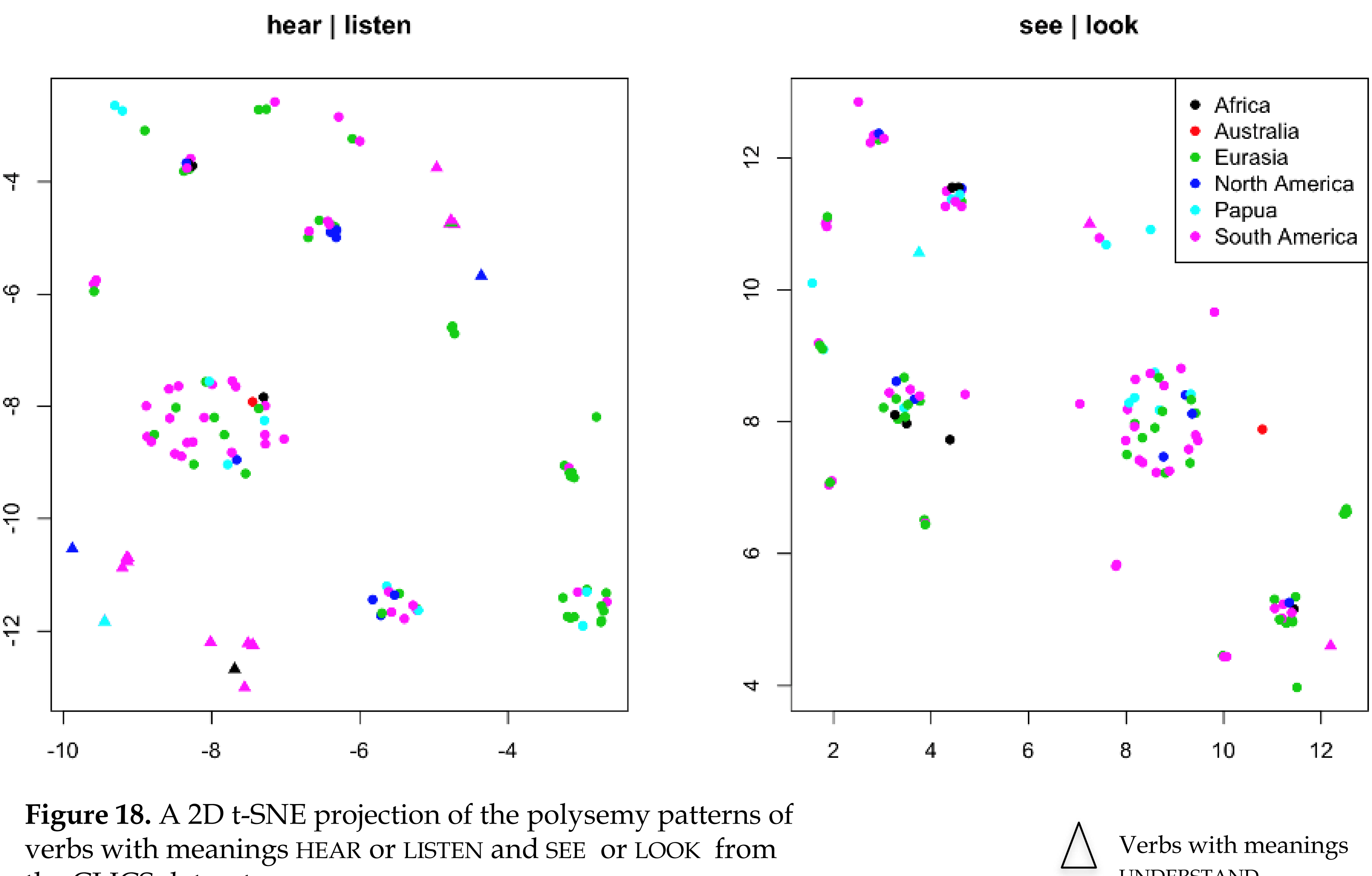


Figure 18. A 2D t-SNE projection of the polysemy patterns of verbs with meanings HEAR or LISTEN and SEE or LOOK from the CLICS dataset

+ Discussion: Areal patterns in CLICS

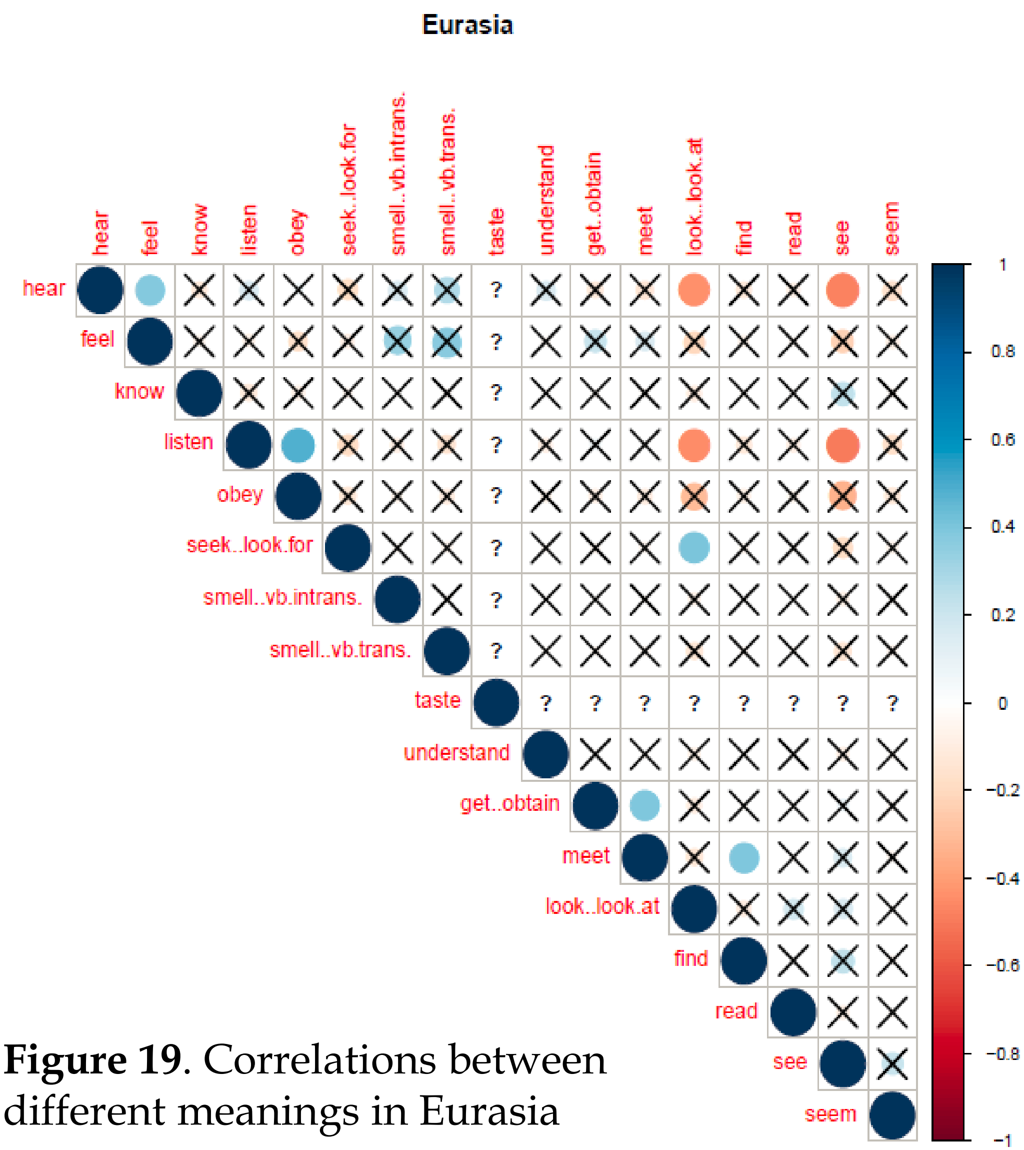


Figure 19. Correlations between different meanings in Eurasia

Discussion: Areal patterns in CLICS

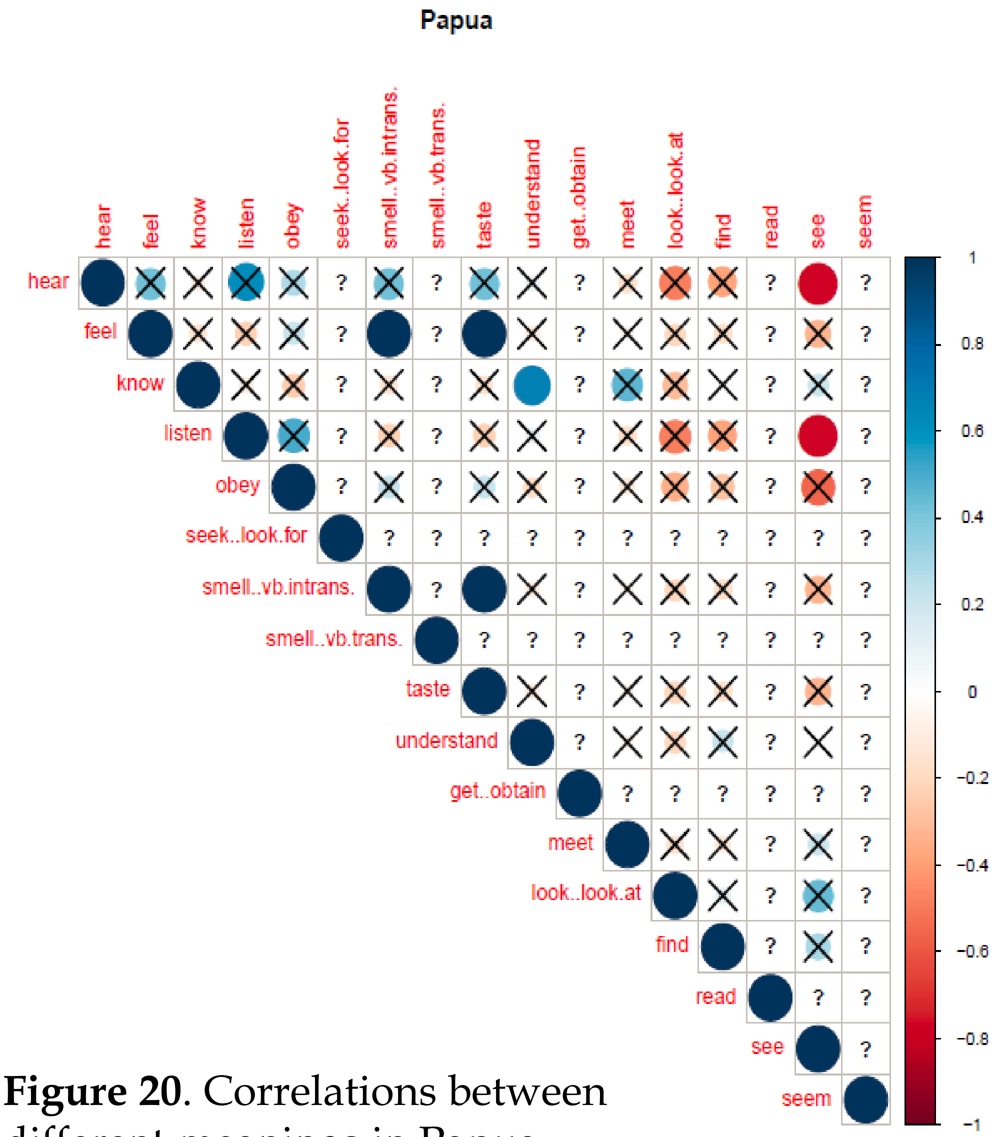


Figure 20. Correlations between different meanings in Papua

Discussion: Areal patterns in CLICS

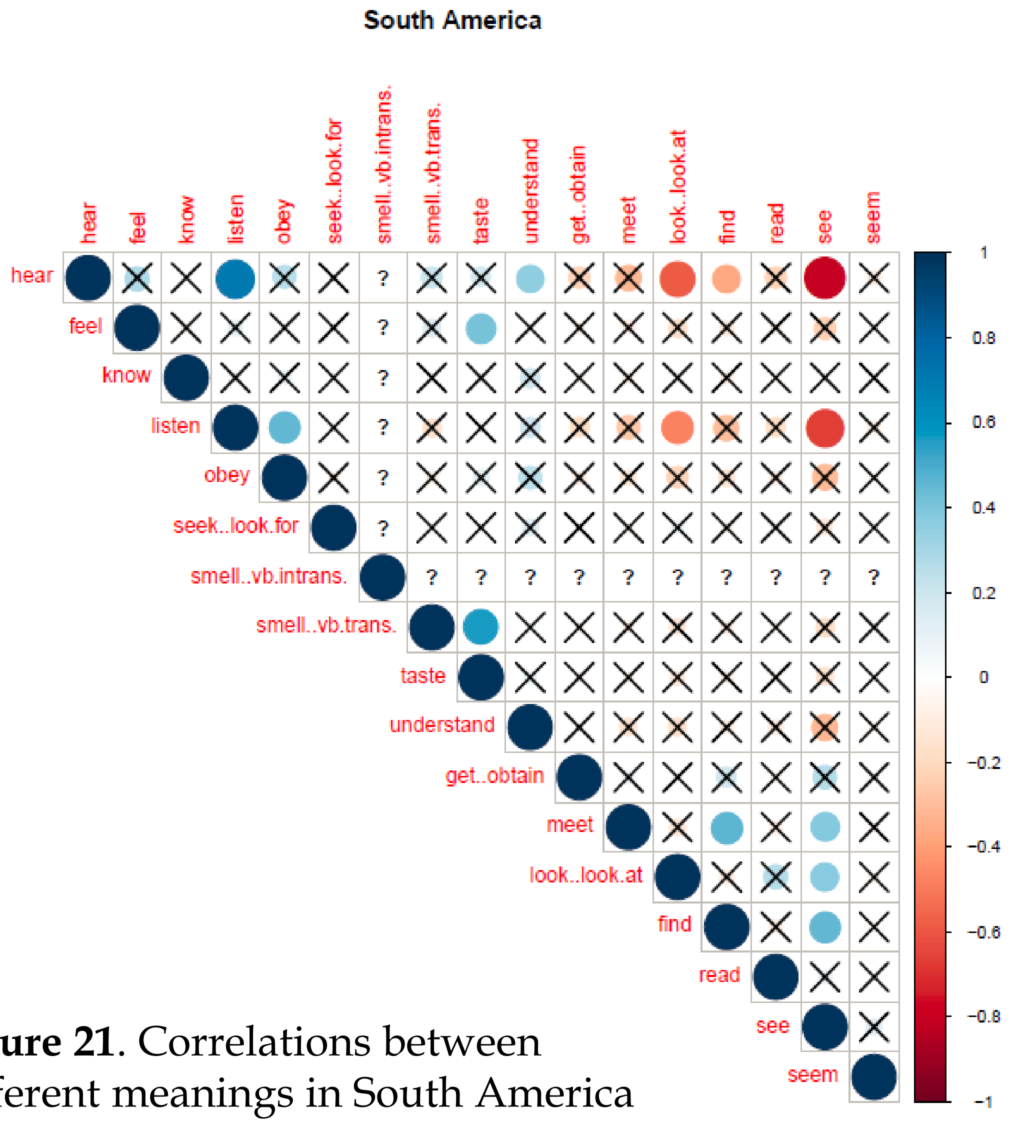


Figure 21. Correlations between different meanings in South America

+ General discussion

- The colexification patterns presented here are typical Greenbergian implicational universals.
- The three samples show some stable patterns
  - The indirect connection between SEE and HEAR that are mediated by cognition verbs
  - The direct connection between perception verbs denoting non-controlled experience and cognitive verbs
- The areal impact is difficult to establish besides some limited cases (cf. SEE)
  - Smaller areas might provide more insightful results (provided that we have an adequate sample).
  - Statistical significance is difficult to reach with the 'small' samples at our disposal

+ General discussion

- A sample of areally related, but genetically diverse languages (with enough languages in each family in order to reach statistical significance) would be the way to go in order to investigate further these questions (i.e., beyond semantic factors).

+ Selected references

Cysouw, M. (2007). Building semantic maps: The case of person marking. In B. Wälchli & M. Miestamo (Eds.), *New challenges in typology* (pp. 225–248). Berlin/New York: Mouton de Gruyter.

François, A. (2008). Semantic Maps and the Typology of Colexification: Intertwining Polysemous Networks across Languages. In: M. Vanhove (Ed.), *From Polysemy to Semantic Change. Towards a Typology of Lexical Semantic Associations* (pp. 163–215). Amsterdam/Philadelphia: John Benjamins.

Georgakopoulos T. & Polis, S.. *Forthcoming.* The semantic map model: State of the art and future avenues for linguistic research". *Language & Linguistics Compass*.

Haspelmath, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In: M. Tomasello (Ed.), *The new psychology of language*, Vol. 2 (pp. 211–242). Mahwah/ New Jersey: Lawrence Erlbaum Associates.

List, J.-M., Mayer, T., Terhalle, A., & Urban, M. (2014). *CLICS: Database of Cross-Linguistic Colexifications*. Marburg: Forschungs-zentrum Deutscher Sprachatlas (Version 1.0, online available at http://CLICS.lingpy.org, accessed on 2016-27-10).

Regier, T., Khetarpal, N., & Majid, A. (2013). Inferring semantic maps*, in: Linguistic Typology*, *17*, 89–105.

Swadesh, M. (1952). Lexicostatistic Dating of Prehistoric Ethnic Contacts. *Proceedings of the American Philosophical Society*, *96*, 452–463.

Sweetser, E. (1990). *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure.* Cambridge: Cambridge University Press.

Viberg, Å. (1984). The verbs of perception: a typological study. *Linguistics, 2*, 123–162.

Wälchli, B. (2016). Non-specific, specific and obscured perception verbs in Baltic languages. *Baltic Linguistics, 7,* 53–135.

Acknowledgments

http://web.philo.ulg.ac.be/lediasema/

Université de Liège

BeIPD-COFUND