



Dialexification

A tool for assessing trends in semantic change

Alex François

LATTICE – CNRS, ENS, P3 – ANU

Siva Kalyan

University Queensland – ANU

From colex to dialex

- Looking back

- a concept:

- Colexification**

- (François 2008)

- an online resource:

- CLiCS**

- (List & al. 2018; Rzymiski & al. 2020)

comparing languages
in their
synchronic structures

- Looking forward

- a concept:

- Dialexification**

- (François & Kalyan in prep.)

- an online resource:

- EvoSem**

- (François, Kalyan, Dehouck,
Pastor, Kletz, in prep.)

comparing languages
in their
diachronic evolutions

Colexification

François, Alexandre. 2008. **Semantic maps and the typology of colexification: Intertwining polysemous networks across languages.** In Martine Vanhove (ed.), *From Polysemy to Semantic change: Towards a Typology of Lexical Semantic Associations*. Benjamins, 163–215.

- Colexification is a structural property of a language, in synchrony:

(1) A language **colexifies** two distinct senses
iff it can express them using the same lexical form.

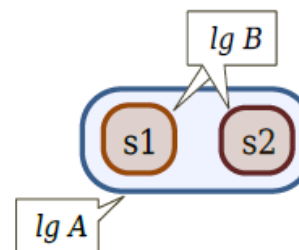
- Italian **sentire** colexifies ‘hear’ and ‘feel’
 - *Non t’avevo sentito* ‘I hadn’t heard you.’
 - *Sentiva dolore* ‘He was feeling pain.’
- Mwotlap **yoñteg** (Oceanic, Vanuatu) shows the same *colex*
 - *Nok et yoñteg te nēk* ‘I hadn’t heard you.’
 - *Kē ni-yoñteg ne-memeh.* ‘He was feeling pain.’
- English **dislexifies** (=must distinguish) the two meanings
 - **I hadn’t felt you* -- **He heard pain*

cf. **con**-junct, **dis**-junct
con-cord, **dis**-cord
con-verge, **di**-verge

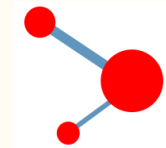
(2) A language **dislexifies** two senses
iff it must express them using different lexical forms.

- Two meanings s1 and s2 are either colexified or dislexified.
 - Can be represented as a **semantic map** or **(hyper)graph**.
 - Senses are **comparative concepts** (Haspelmath 2010), lg-independent.
 - The **onomasiological perspective** (going from meaning to form) is what makes language comparison possible.

Figure 1 - Two senses s1 and s2 may be colexified (language A) or dislexified (language B)

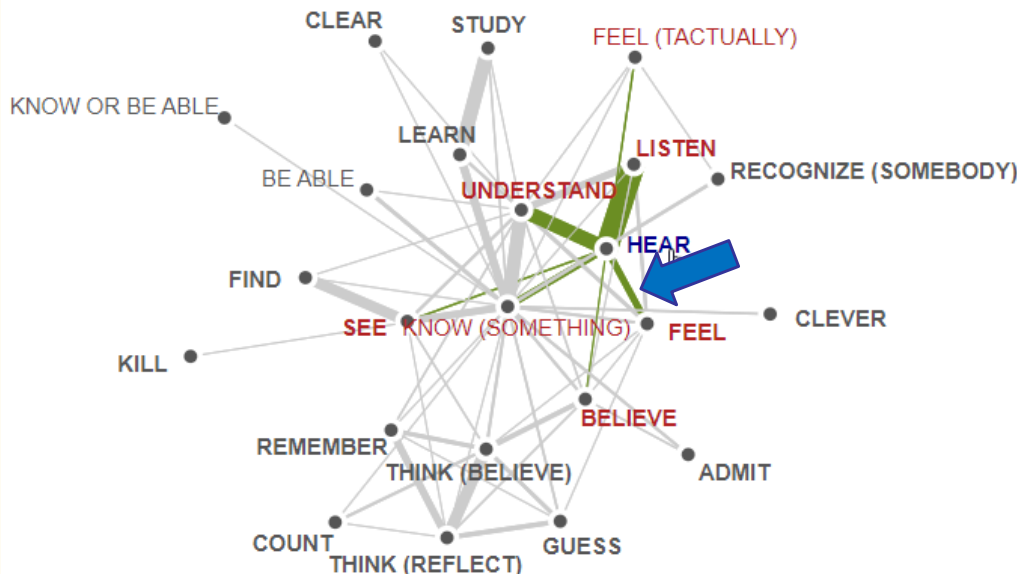


Typology of colexification



■ CLiCS: Cross-Linguistic Colexifications — clics.clld.org

- List, J.-M., Mayer, T., Terhalle, A., & Urban, M. (2014). *CLiCS: Database of Cross-Linguistic Colexifications*. Marburg: Forschungszentrum Deutscher Sprachatlas.
- [1.0] 1280 meanings, 221 language varieties
- [2.0] 1521 meanings, 1156 language varieties
- [3.0] **2919** meanings, **3156** language varieties



25 colexifications for "FEEL" and "HEAR":

Language	Family	Form
Hausa	Afro-Asiatic	ji
Hausa	Afro-Asiatic	ji
Polci	Afro-Asiatic	wom
Ignaciano	Arawakan	sama
Wapishana	Arawakan	abathan
Waurá	Arawakan	eteme
Cuoi	Austroasiatic	nge33co31
Prai	Austroasiatic	mecnang
Maori	Austronesian	rongohia
Maori	Austronesian	rongo
White Hmong	Hmong-Mien	hnov
Albanian (Tosk variety)	Indo-European	ndien
Catalan	Indo-European	sentir
Italian	Indo-European	sentire

CLiCS turned my programmatic idea (2008)
into a powerful resource for all linguists to use.



Typology of colexification

DE GRUYTER MOUTON

Linguistic Typology 2021; aop



Baring the bones: the lexico-semantic association of bone with strength in Melanesia and the study of colexification

Antoinette Schapper*

<https://doi.org/10.1515/lingty-2021-2082>
Received March 22, 2021; accepted March 3, 2021; published online July 27, 2021

Abstract: In this article I demonstrate that there is a pervasive lexico-semantic association BONES ARE STRENGTH in the languages of Melanesia, but that its linguistic expression is highly varied; languages are scattered along a lexical-to-clausal cline in their expression of the association between bone and strength, with a large number of language-specific idioms based on the association to be observed in Melanesia.

NEW TRENDS IN INDO-EUROPEAN LINGUISTICS Number 2 | 2 October 2020

Colexification Patterns of Proto-Indo-European Root “Speaking” Nouns, and Noun Classes

By Roland A. Pooth

Once again thanks to the great approval: “A spectre is haunting Indo-European linguistics, the spectre of the novel *Templatic Model* of Proto-Indo-European morphosyntax.”

Abstract

Cognition 226 (2022) 105179

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognition



When do languages use the same word for different meanings? The Goldilocks principle in colexification

Thomas Brochhagen^{a,*}, Gemma Boleda^{a,b}

^a Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat, 130, 08010 Barcelona, Spain
^b Catalan Institution for Research and Advanced Studies (ICREA), Pg. Lluís Companys, 23, 08010 Barcelona, Spain

ABSTRACT

Lexical ambiguity is pervasive in languages, and often systematic. For instance, a toe or a finger, that is, these two meanings coexist in Spanish, and they do so in many languages. Previous work shows that related meanings are more likely to colexify. We hypothesize that colexification follows a Goldilocks principle: the interplay between this pressure and the competing pressure for language transfer. We hypothesize that colexification follows a Goldilocks principle: meanings are more likely to attach to the same word when they are related, but not too little. We find support for this principle in data from over 120 languages. We argue that forces shape the lexicons of natural languages evolve to

ARTICLE INFO

Keywords:
Language universals
Lexical ambiguity
Cognitive effort
Communication

COGNITIVE SCIENCE A Multidisciplinary Journal



Cognitive Science 45 (2021) e13035
© 2021 Cognitive Science Society LLC
ISSN: 1551-6709 online
DOI: 10.1111/cogs.13035

Conceptual Similarity and Communicative Need Shape Colexification: An Experimental Study

Andres Karius^{a,b,c}, Richard A. D. ...

Thanasis Georgakopoulos – Daniel A. Werning – Jörg Hartlieb – Tomoki Kitazumi – Lidewij E. van de Peut – Annette Sundermeyer – Gaëlle Chantraine

The Meaning of Ancient Words for ‘Earth’: An Exercise in Visualizing Colexification on a Semantic Map

This paper aims at investigating the polysemic patterns associated with the notion ‘soil/earth’ by using the semantic map model as a methodological tool. We focus on the applicability of the model to the lexicon, since most of past research has been devoted to the analysis of grammatical morphemes. The most concise result of our research is a diagrammatic visualization of the semantic spaces of twenty lexemes in nine different languages, belonging to the Indo-European and the Afro-Asiatic language families.

Affective Science
<https://doi.org/10.1007/s42761-021-00033-1>

RESEARCH ARTICLE

Colexification Networks Encode Affective Meaning

Anna Di Natale^{1,2,3}, Max Pellert^{1,2,3}, David Garcia^{1,2,3}

Received: 30 September 2020 / Accepted: 27 January 2021
© The Author(s) 2021

Abstract

Colexification is a linguistic phenomenon that occurs when multiple concepts word. Colexification patterns are frequently used to estimate the meaning of words that these are related to. This is still missing direct empirical validation at scale. Here, by colexification patterns capture similar affective meanings. Using pre-existing databases to cover much longer word lists. We achieve this with an unsupervised colexification network data to interpolate the affective ratings of words. We find positive correlations between network-based estimates and empirical affective ratings. We find that colexification networks contain information related to affective meanings. Finally, we use machine learning, trained on a large corpus, and show that our simple linguistic model can predict affective ratings with high explainability. These results show that it is possible to

Behavior Research Methods
<https://doi.org/10.3758/s13428-023-02063-y>

LEXpander: Applying colexification networks to automated lexicon expansion

Anna Di Natale^{1,2,3}, David Garcia^{1,2,3,4}

Accepted: 6 January 2023
© The Author(s) 2023

Abstract

Approaches to text analysis from social media and other corpora rely on word lists to detect topics, me

frontiers | Frontiers in Psychology

Check for updates

ACCESS
Carling
University Frankfurt, Germany

Wichmann,
University of Kiel, Germany

University of Science and Technology

Science
Paris Lodron
University of Salzburg

Inference of partial colexification from multilingual wordlists

Johann-Mattis List^{1,2,*}

¹Department of Linguistic and Cultural Evolution, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany; ²Chair of Multilingual Computational Linguistics, University of Passau, Passau, Germany

The past years have seen a drastic rise in studies devoted to the investigation of colexification patterns in individual languages families in particular. Languages of the world in specific. Specifically computational studies have enabled scholars to infer colexification patterns for large collections of languages. Studies devoted to partial colexifications—colexification pa

An Interactive Visualization of Crosslinguistic Colexification

Thomas Mayer¹, Johann-Mattis List¹, Anselm Terhalle², Matthias Urban³

¹Philipps-Universität Marburg; ²Heinrich-Heine-Universität Düsseldorf; ³Universität Leipzig
thomas.mayer@uni-marburg.de, mattis.list@uni-marburg.de,
terhalle@phil.uni-duesseldorf.de, m.urban@hum.leidenuniv.nl

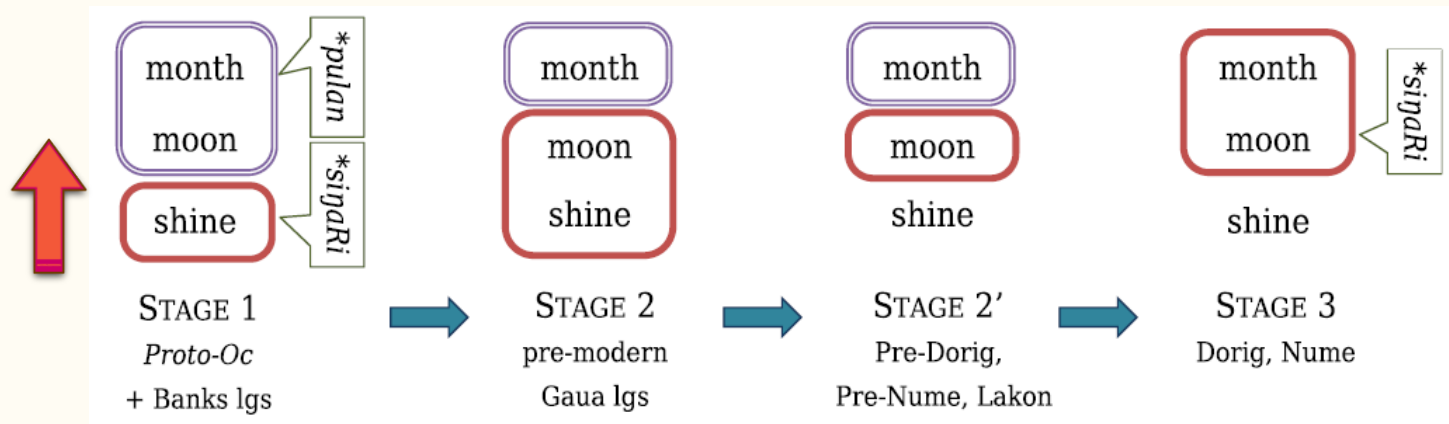
Abstract

In this paper, we present an interactive web-based visualization for the CLICS database, an online resource for crosslinguistic colexification patterns. The associations cover 1,288 concepts and represent for concepts to be expressed by the same words in the same languages and language varieties of the world. The network structure in the CLICS database calls for a visualization component that makes it easier for researchers to explore crosslinguistic colexifications. The network is represented as a force-directed graph and features a number of interactive features that allow the user to get an overview of the overall structure while at the same time providing an opportunity to explore colexification in more detail. An integral part of the visualization is an interactive listing of all languages that colexify a given concept. Each language in the list is thereby attributed a different color. This way, given associations can be inspected for genealogical or areal patterns.

Keywords: interactive visualization

What about semantic change?

- Most studies in colexification have been synchronic
 - cf. “**strict** colexification”: “2 senses encoded by a single lexeme in synchrony”
- But colexification can also be viewed as a **dynamic** process
 - cf. lexical change in languages of Vanuatu :



Colex links can appear [*semantic merger*] or disappear [*semantic split*] over time, as words move and migrate across semantic space (François 2022).

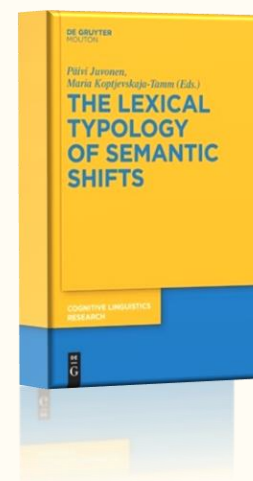
“lexical tectonics”

François, Alexandre. 2022. Lexical tectonics: Mapping structural change in patterns of lexification. In T. Georgakopoulos & S. Polis (eds), *The future of mapping: New avenues for semantic maps research*. Special issue of *Zeitschrift für Sprachwissenschaft*.

Lexical tectonics

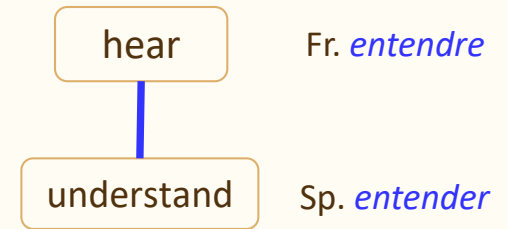
- Question
 - How does lexical change **propagate through conceptual space**?
 - Does *lexical tectonics* follow random paths?
 - Or are certain semantic changes **more likely** than others?
 - Can we identify the **most common paths of lexical change**?
 - cf. Kümmel 2007: most common paths of *sound change*
 - cf. Heine & Kuteva 2002: most frequent *paths of grammaticalisation*
 - Part of a larger effort on **lexical typology of linguistic change**
- We need a **typological database of lexical change**
 - Build a diachronic equivalent to *CLiCS*.

synchronic link		diachronic link
Colexification	→	Dialexification
CLiCS		<i>EvoSem</i>



Dialexification

Two senses X and Y are **dialexified**
 iff they are **the meanings of cognate forms**,
 = associated with **reflexes of a single etymon**.



cf. “false friends”

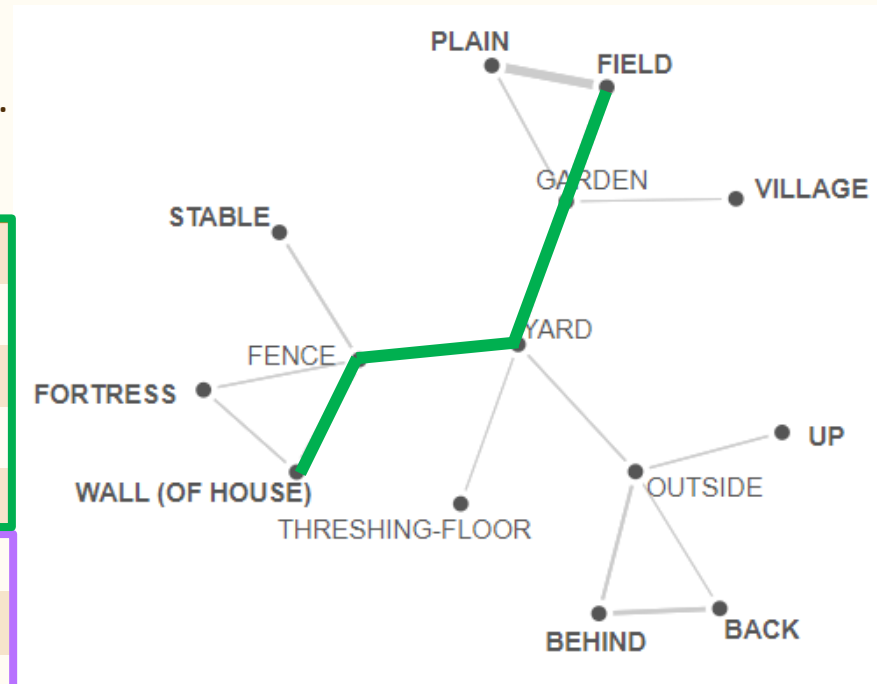
p-Indo-E root * <i>g^hord^h-</i>	Albanian	<i>gardh</i>
	Icelandic	<i>garður</i>
	Old English	<i>ġeard</i>
	French	<i>jardin</i>
	Irish	<i>garraí</i>
	Old Norse	<i>gjarð</i>
	Gothic	<i>ΓΑΡΔΣ</i> <i>gards</i>
	Czech	<i>hrad</i>
	Russian	<i>город</i> <i>gorod</i>
	Rromani	<i>kher</i>
	Bengali	<i>ঘর</i> <i>ghor</i>
	Sanskrit	<i>गृह</i> <i>grhá</i>

Dialexification

Two senses X and Y are **dialexified**
iff they are **the meanings of cognate forms**,
= associated with **reflexes of a single etymon**.

p-Indo-E
root
**g^hord^h-*

Albanian	<i>gardh</i>	'fence'
Icelandic	<i>garður</i>	'wall'
Old English	<i>ġeard</i>	'yard'
French	<i>jardin</i>	'garden'
Irish	<i>garraí</i>	'field'
Old Norse	<i>gǫrð</i>	'belt'
Gothic	<i>ΓΑΡΔΣ gards</i>	'court'
Czech	<i>hrad</i>	'castle'
Russian	<i>город gorod</i>	'city'
Rromani	<i>kher</i>	'house'
Bengali	<i>ঘর ghor</i>	'family'
Sanskrit	<i>गृह grhá</i>	'wife'



cf. CLiCS subgraph **YARD**

> Dialex° captures **more semantic connections** than just colexification
because it is not limited to synchrony of 1 lg

Dialexification

For each **pair of dialexified senses**

- Is that dialex **unique** to this etymon?
 - Is it attested with **other** etyma ?
 - Is it **common** ? rare ?
- Found across different lg families?

p-Indo-E
root
**g^hord^h-*
'enclose'?

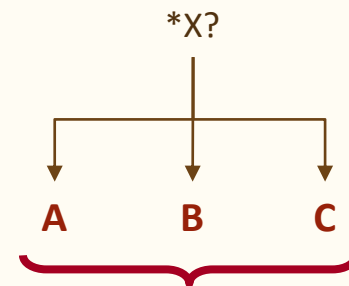
Albanian	<i>gardh</i>	'fence'
Icelandic	<i>garður</i>	'wall'
Old English	<i>ġeard</i>	'yard'
French	<i>jardin</i>	'garden'
Irish	<i>garraí</i>	'field'
Old Norse	<i>gǫrð</i>	'belt'
Gothic	<i>ΓΑΡΔΣ gards</i>	'court'
Czech	<i>hrad</i>	'castle'
Russian	<i>город gorod</i>	'city'
Rromani	<i>kher</i>	'house'
Bengali	<i>ঘর ghor</i>	'family'
Sanskrit	<i>गृह grhá</i>	'wife'

Reconstructed proto-meanings are **speculative**, so instead of studying

- paths of change *X → A

we choose **attested** data as a proxy:

- relations of **semantic cognacy**



Senses A, B, C are **dialexified**.

Dialexification

For each **pair of dialexified senses**

- Is that dialex **unique** to this etymon?
 - Is it attested with **other** etyma ?
 - Is it **common** ? rare ?
- Found across different lg families?



the **EvoSem** project

= which **semantic evolutions** are most frequent?

Empirical, quantitative approach to lexical change

p-Indo-E
root

***g^hord^h-**

Albanian	gardh	'fence'
Icelandic	garður	'wall'
Old English	ġeard	'yard'
French	jardin	'garden'
Irish	garraí	'field'
Old Norse	gǫrð	'belt'
Gothic	ΓΑΡΔΣ <i>gards</i>	'court'
Czech	hrad	'castle'
Russian	город <i>gorod</i>	'city'
Rromani	kher	'house'
Bengali	ঘর <i>ghor</i>	'family'
Sanskrit	गृह <i>gṛhá</i>	'wife'

The *EvoSem* database



cnrs



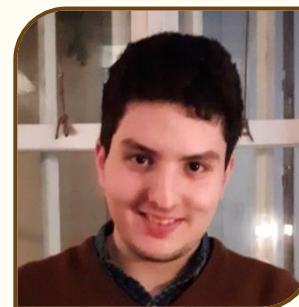
Siva Kalyan

Univ Queensland
– A.N.U. (Aus)



Alex François

LaTTiCe – CNRS, ENS –
A.N.U. (Aus)



Chams Bernard

Paris LabEx *Empirical
foundations of linguistics*



David Kletz

LaTTiCe – CNRS, ENS







Mathieu Dehouck

LaTTiCe – CNRS, ENS



Martial Pastor

LaTTiCe – CNRS, ENS

11 etyma dialectifying <i>see</i> — <i>understand</i>		
FAMILY	ETYMON	REFLEXES
PIE	*wéyǵ-e-ti	
Bantu	*bóna	 <i>language data</i>
PAN	*kita₂	 <i>language data</i>
PIE	*kri-né-h₁- *kri-n-h₂-	 <i>language data</i>

The *Ev*Sem database

tiny.cc/EvoSem_ICLC

▼ 50 meanings dialexified by ***g^hórd^h-os**

MEANINGS	REFLEXES	#
garden	► <i>language data</i>	10
yard	► <i>language data</i>	9
city	► <i>language data</i>	8
castle	► <i>language data</i>	6
town	► <i>language data</i>	6
fence	► <i>language data</i>	5
court	► <i>language data</i>	4
courtyard	► <i>language data</i>	4
dwelling	► <i>language data</i>	4
enclosure	► <i>language data</i>	4
farm	► <i>language data</i>	4
wall	► <i>language data</i>	3
estate	► <i>language data</i>	2
house	► <i>language data</i>	2
land	► <i>language data</i>	2
stone wall	► <i>language data</i>	2
stronghold	► <i>language data</i>	2

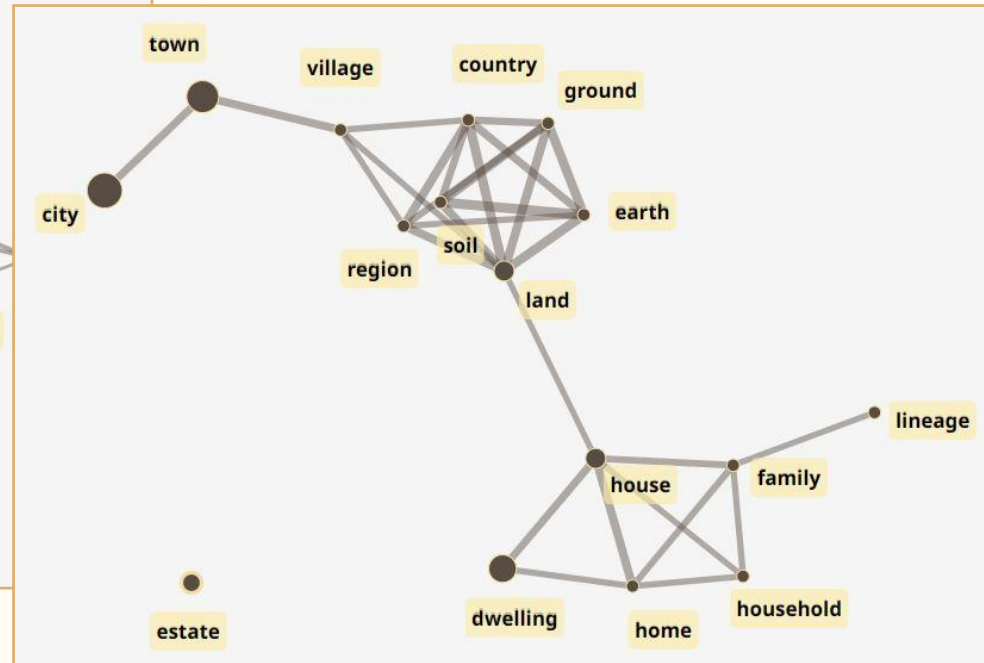
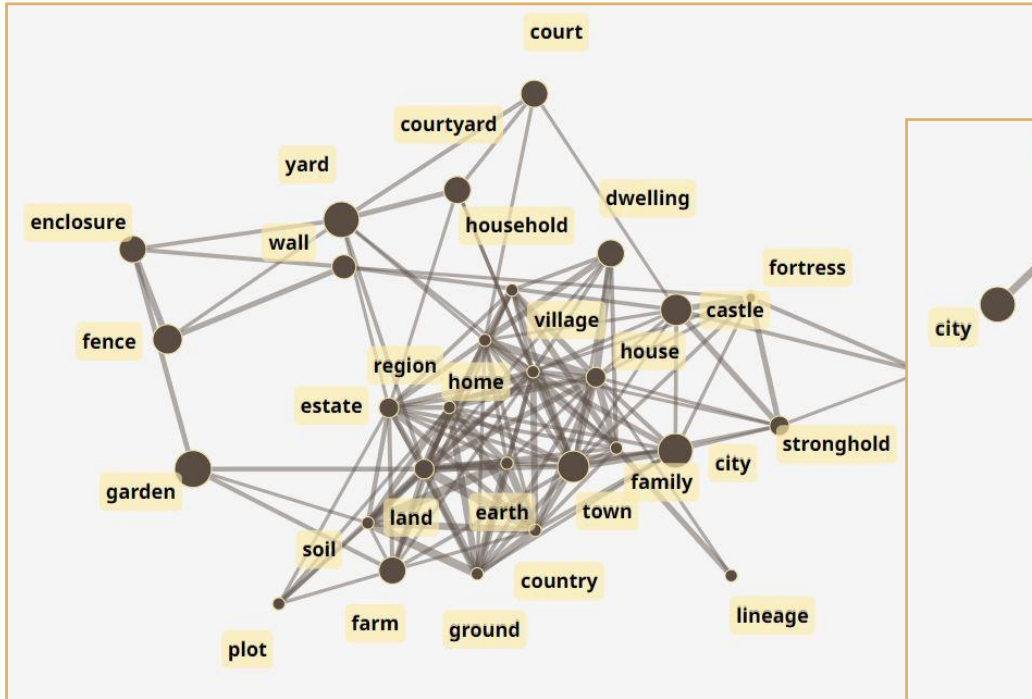
▼ 50 meanings dialexified by ***ghórdh-os**

MEANINGS	REFLEXES	#
garden	▼ <i>language data</i>	10
	<i>Albanian</i> gardh	
	<i>Dutch</i> gaard	
	<i>Faroese</i> garður	
	<i>Icelandic</i> garður	
	<i>Irish</i> garraí	
	<i>Old French</i> jart	
	<i>Old Saxon</i> gard	
	<i>Scots</i> yaird	
	<i>Scottish Gaelic</i> gàrradh	
	<i>Welsh</i> gardd	
yard	► <i>language data</i>	9
city	▼ <i>language data</i>	8
	<i>Belarusian</i> гóрад - hórad	
	<i>Bulgarian</i> град - grad	
	<i>Polish</i> gród	
	<i>Russian</i> góрод - górod	
	<i>Skolt Sami</i> gåårad	
	<i>Slovene</i> grād	
	<i>Ukrainian</i> góрод - hórod	
	<i>Yakut</i> кyорат - kuorat	

The *EvoSem* database

Etymograph of Indo-European **ghórdh-os*

● Node size shows frequency of each meaning among descendants of that specific etymon.
→ quick view of the cognate set's semantic profile



Threshold θ for minimum rate of dialexification (across the entire database)

Dialexification threshold: $\theta = 2, \lambda = 347$

2  50












Dialexification threshold: $\theta = 11, \lambda = 33$

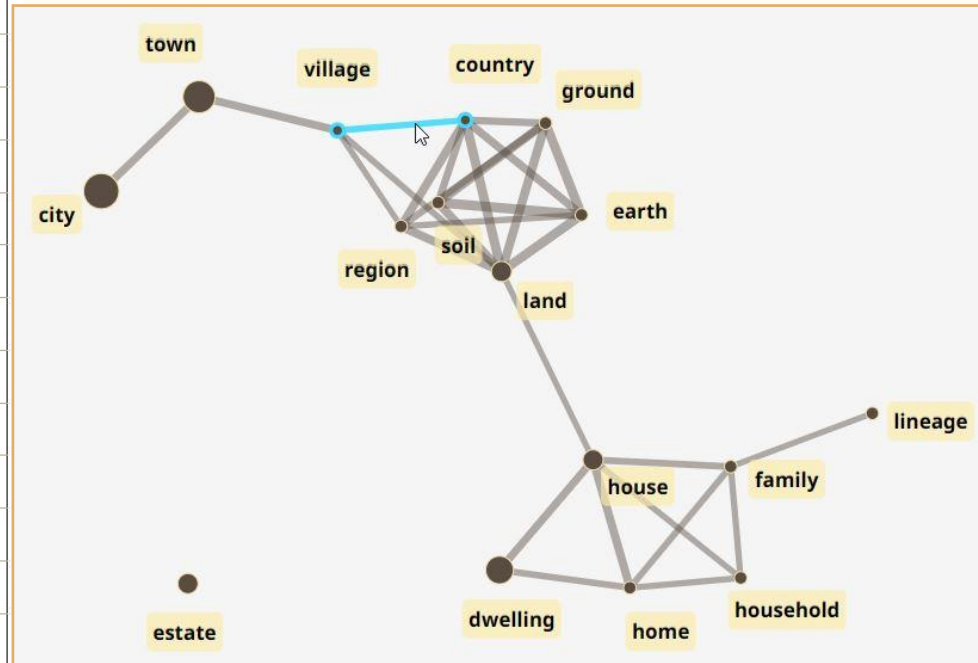
2  50

The *EvoSem* database

Etymograph of Indo-European **ghórdʰ-os*

13 etyma dialexifying *country* — *village*












FAMILY	ETYMON	► REFLEXES
Indo-European	<i>*ghórdʰ-os</i>	 ► <i>language data</i>
Austronesian	<i>*banua</i>	 ► <i>language data</i>
Austronesian	<i>*bawaŋ₁</i>	 ► <i>language data</i>
Austronesian	<i>*hili</i>	 ► <i>language data</i>
Austronesian	<i>*panua</i>	 ► <i>language data</i>
Basque	<i>*(h)eRi</i>	► <i>language data</i>
Dravidian	<i>*nāṭu</i>	► <i>language data</i>
Indo-European	<i>*bʰúH-m-i-s</i>	 ► <i>language data</i>
Indo-European	<i>*deyk-ó-s</i>	 ► <i>language data</i>
Indo-European	<i>*h₂ég-ro-s</i>	 ► <i>language data</i>
Indo-European	<i>*londʰ-om</i>	 ► <i>language data</i>
Indo-European	<i>*tewtéh₂</i>	 ► <i>language data</i>
Indo-European	<i>*wéyk-i-s</i>	 ► <i>language data</i>









The $\epsilon v \sigma$ Sem database

Etymograph of Indo-European ***gʰórdʰ-os**

13 etyma dialexifying **country** — **village**

FAMILY	ETYMON	► REFLEXES
Indo-European	*g^hórd^h-os	 ► <i>language data</i>
Austronesian	*banua	 ► <i>language data</i>
Austronesian	*bawaŋ₁	 ► <i>language data</i>
Austronesian	*hili	 ► <i>language data</i>
Austronesian	*panua	 ► <i>language data</i>
Basque	*(h)eRi	► <i>language data</i>
Dravidian	*nāṭu	► <i>language data</i>
Indo-European	*b^húH-m-i-s	 ► <i>language data</i>
Indo-European	*deyǵ-ó-s	 ► <i>language data</i>
Indo-European	*h₂éǵ-ro-s	 ► <i>language data</i>
Indo-European	*lond^h-om	 ► <i>language data</i>
Indo-European	*tewtéh₂	 ► <i>language data</i>
Indo-European	*wéyǵ-i-s	 ► <i>language data</i>

13 etyma dialexifying **country** — **village**

FAMILY	ETYMON	▼ REFLEXES												
Indo-European	*g^hórd^h-os	 ► <i>language data</i>												
Austronesian	*banua	 ► <i>language data</i>												
Austronesian	*bawaŋ₁	 ► <i>language data</i>												
Austronesian	*hili	 ► <i>language data</i>												
Austronesian	*panua	 ► <i>language data</i>												
Basque	*(h)eRi	► <i>language data</i>												
Dravidian	*nāṭu	► <i>language data</i>												
Indo-European	*bʰúH-m-i-s	 ► <i>language data</i>												
Indo-European	*deyǵ-ó-s	 ► <i>language data</i>												
Indo-European	*h₂ég-ro-s	<div> ▼ <i>language data</i><table><tr><td>Latin</td><td>ager</td></tr><tr><td>Mycenaean Greek</td><td>𐀓𐀆 - a-ko-ro</td></tr><tr><td>Georgian</td><td>აგარაკი - agaraḳi</td></tr><tr><td>Georgian</td><td>აგარაკი - agaraḳi</td></tr><tr><td>Old Armenian</td><td>ագարակ - agarak</td></tr><tr><td>Old Georgian</td><td>აგარაკი - agaraḳi</td></tr></table></div>	Latin	ager	Mycenaean Greek	𐀓𐀆 - a-ko-ro	Georgian	აგარაკი - agaraḳi	Georgian	აგარაკი - agaraḳi	Old Armenian	ագարակ - agarak	Old Georgian	აგარაკი - agaraḳi
Latin	ager													
Mycenaean Greek	𐀓𐀆 - a-ko-ro													
Georgian	აგარაკი - agaraḳi													
Georgian	აგარაკი - agaraḳi													
Old Armenian	ագարակ - agarak													
Old Georgian	აგარაკი - agaraḳi													
Indo-European	*lond^h-om	 ► <i>language data</i>												
Indo-European	*tewtéh₂	 ► <i>language data</i>												
Indo-European	*wéyǵ-i-s	 ► <i>language data</i>												

The *EvoSem* database



Colexification is
a special case of
Dialexification

Etymograph of Indo-European **ghórdʰ-os*

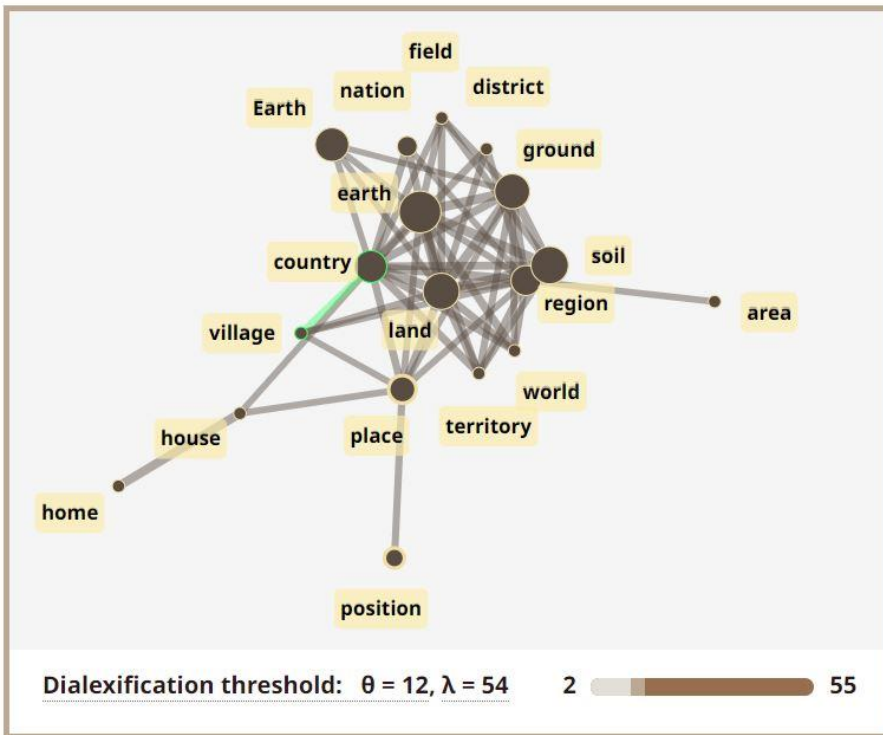
13 etyma dialexifying **country** — **village**






FAMILY	ETYMON	▼ REFLEXES
Indo-European	<i>*ghórdʰ-os</i>	 ► <i>language data</i>
Austronesian	<i>*banua</i>	 ► <i>language data</i>
Austronesian	<i>*bawaŋ₁</i>	 ► <i>language data</i>
Austronesian	<i>*hili</i>	 ► <i>language data</i>
Austronesian	<i>*panua</i>	 ► <i>language data</i>
Basque	<i>*(h)eRi</i>	 ► <i>language data</i> <div> <div>pure colex</div> <div> <div>Basque</div> <div>herri</div> </div> </div>
Dravidian	<i>*nāṭu</i>	► <i>language data</i>
Indo-European	<i>*bʰúH-m-i-s</i>	 ► <i>language data</i> <div> <div>colex + dialex</div> <div> <div>Hindi</div> <div>भूमि - bhūmi</div> </div> <div> <div>Middle Persian</div> <div>𐭠𐭣𐭥 - būm</div> </div> <div> <div>Middle Persian</div> <div>𐭠𐭣𐭥 - b'm</div> </div> <div> <div>Persian</div> <div>بوم - bum</div> </div> <div> <div>Sanskrit</div> <div>भूमि - bhūmi</div> </div> <div> <div>Khmer</div> <div>ភ្នំ - phuum</div> </div> </div>
Indo-European	<i>*deyǵ-ó-s</i>	 ► <i>language data</i>
Indo-European	<i>*h₂ég-ro-s</i>	 ► <i>language data</i>
Indo-European	<i>*londʰ-om</i>	 ► <i>language data</i>
Indo-European	<i>*tewtéh₂</i>	 ► <i>language data</i>

13 etyma dialexifying **country** — **village**

FAMILY	ETYMON	▼ REFLEXES
Indo-European	<i>*ghórdʰ-os</i>	 ► <i>language data</i>
Austronesian	<i>*banua</i>	 ► <i>language data</i>
Austronesian	<i>*bawaŋ₁</i>	 ► <i>language data</i>
Austronesian	<i>*hili</i>	 ► <i>language data</i>
Austronesian	<i>*panua</i>	 ► <i>language data</i>
Basque	<i>*(h)eRi</i>	► <i>language data</i>
Dravidian	<i>*nāṭu</i>	► <i>language data</i>
Indo-European	<i>*bʰúH-m-i-s</i>	 ► <i>language data</i>
Indo-European	<i>*deyǵ-ó-s</i>	 ► <i>language data</i>
Indo-European	<i>*h₂ég-ro-s</i>	 ► <i>language data</i> <div> <div>pure dialex</div> <div> <div>Latin</div> <div>ager</div> </div> <div> <div>Mycenaean Greek</div> <div>𐀀𐀃𐀆 - a-ko-ro</div> </div> <div> <div>Georgian</div> <div>აგარაკი - agaraḱi</div> </div> <div> <div>Georgian</div> <div>აგარაკი - agaraḱi</div> </div> <div> <div>Old Armenian</div> <div>ագարակ - agarak</div> </div> <div> <div>Old Georgian</div> <div>აგარაკი - agaraḱi</div> </div> </div>
Indo-European	<i>*londʰ-om</i>	 ► <i>language data</i>
Indo-European	<i>*tewtéh₂</i>	 ► <i>language data</i>
Indo-European	<i>*wéyǵ-i-s</i>	 ► <i>language data</i>

The *Ev*Sem database

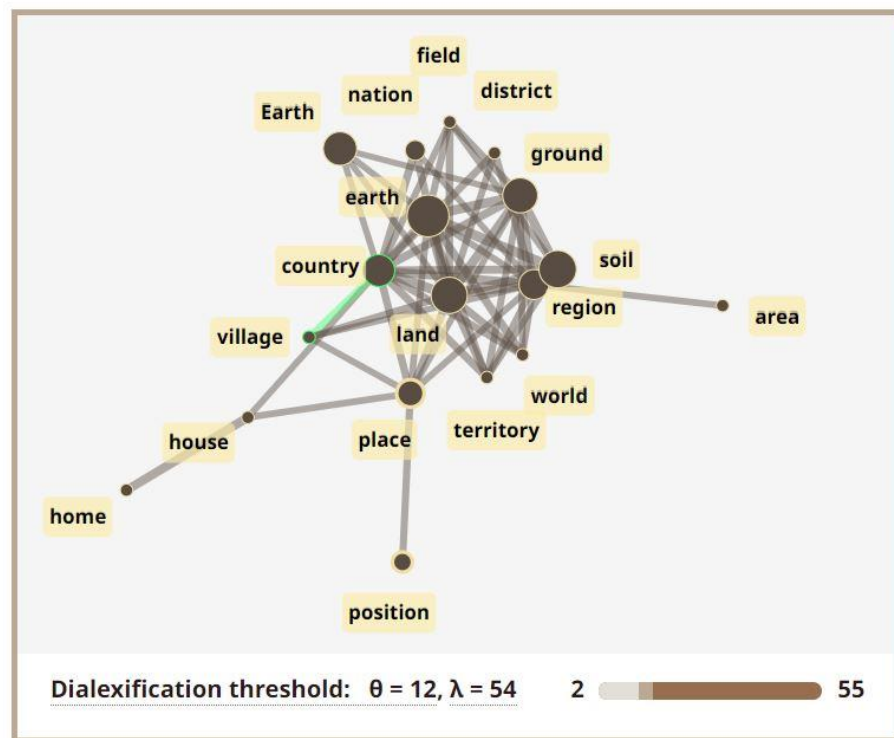
Etymograph of Indo-European ***bʰúH-m-i-s**

13 etyma dialexifying country — village														
FAMILY	ETYMON	▼ REFLEXES												
Indo-European	*gʰórdʰ-os	 ► <i>language data</i>												
Austronesian	*banua	 ► <i>language data</i>												
Austronesian	*bawaŋ₁	 ► <i>language data</i>												
Austronesian	*hili	 ► <i>language data</i>												
Austronesian	*panua	 ► <i>language data</i>												
Basque	*(h)eRi	► <i>language data</i>												
Dravidian	*nāṭu	► <i>language data</i>												
Indo-European	*bʰúH-m-i-s	 ► <i>language data</i>												
Indo-European	*deyk-ó-s	 ► <i>language data</i>												
Indo-European	*h₂ég-ro-s	<div> ▼ <i>language data</i><table><tr><td>Latin</td><td>ager</td></tr><tr><td>Mycenaean Greek</td><td>𐀀𐀃𐀆𐀃 - a-ko-ro</td></tr><tr><td>Georgian</td><td>აგარაკი - agaraki</td></tr><tr><td>Georgian</td><td>აგარაკი - agaraki</td></tr><tr><td>Old Armenian</td><td>ագարակ - agarak</td></tr><tr><td>Old Georgian</td><td>აგარაკი - agaraki</td></tr></table></div>	Latin	ager	Mycenaean Greek	𐀀𐀃𐀆𐀃 - a-ko-ro	Georgian	აგარაკი - agaraki	Georgian	აგარაკი - agaraki	Old Armenian	ագարակ - agarak	Old Georgian	აგარაკი - agaraki
Latin	ager													
Mycenaean Greek	𐀀𐀃𐀆𐀃 - a-ko-ro													
Georgian	აგარაკი - agaraki													
Georgian	აგარაკი - agaraki													
Old Armenian	ագարակ - agarak													
Old Georgian	აგარაკი - agaraki													
Indo-European	*londʰ-om	 ► <i>language data</i>												
Indo-European	*tewtéh₂	 ► <i>language data</i>												
Indo-European	*wéyk-i-s	 ► <i>language data</i>												

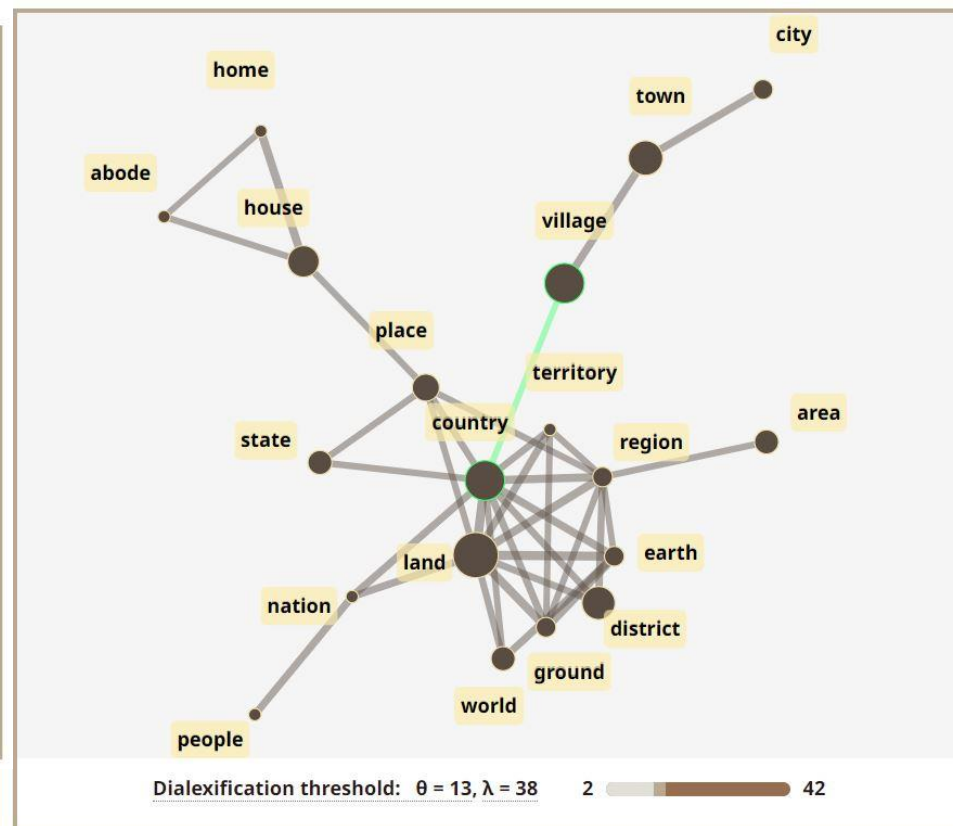
The *EvoSem* database

Each etymograph places a given dialex pair in the context of that particular etymon.

Etymograph of Indo-European ***b^húH-m-i-s**



Etymograph of Austronesian ***banua**



The *EvoSem* database

know

Search

Our search engine gives access to all concepts and their dialexifications (δ = # of dialex links).

Concepts	#neighbours
know	180 ▾
	understand ($\delta=25$)
	knowledge ($\delta=16$)
	recognize ($\delta=14$)
	perceive ($\delta=11$)
	see ($\delta=11$)
	learn ($\delta=10$)
	can ($\delta=9$)
	notice ($\delta=9$)
	feel ($\delta=9$)
	remember ($\delta=9$)
	sense ($\delta=8$)
	able ($\delta=7$)
	comprehend ($\delta=7$)
	consider ($\delta=7$)
	look ($\delta=6$)
	observe ($\delta=6$)
	ready ($\delta=5$)
	skilled ($\delta=5$)
	understanding ($\delta=5$)

Concepts	#neighbours
heavy	116 ▾
	weighty ($\delta=9$)
	burdensome ($\delta=8$)
	weight ($\delta=8$)
	severe ($\delta=8$)
	important ($\delta=7$)
	large ($\delta=7$)
	strong ($\delta=7$)
	difficult ($\delta=7$)
	big ($\delta=6$)
	dense ($\delta=6$)
	fat ($\delta=6$)
	great ($\delta=6$)
	harsh ($\delta=6$)
	grave ($\delta=6$)
	hard ($\delta=6$)
	major ($\delta=5$)
	thick ($\delta=5$)
	serious ($\delta=5$)
	dull ($\delta=5$)

Dialexification metrics:

- shed light on **semantic associations** among concepts
- can help historical linguists assess semantic hypotheses when **reconstructing lexical change**

The *EvoSem* database

know

Search

Our search engine gives access to all concepts and their dialexifications (δ = # of dialex links).

Concepts	#neighbours
know	180 ▾
	understand ($\delta=25$)
	knowledge ($\delta=16$)
	recognize ($\delta=14$)
	perceive ($\delta=11$)
	see ($\delta=11$)
	learn ($\delta=10$)
	can ($\delta=9$)
	notice ($\delta=9$)
	feel ($\delta=9$)
	remember ($\delta=9$)
	sense ($\delta=8$)
	able ($\delta=7$)
	comprehend ($\delta=7$)
	consider ($\delta=7$)
	look ($\delta=6$)
	observe ($\delta=6$)
	ready ($\delta=5$)
	skilled ($\delta=5$)
	understanding ($\delta=5$)

11 etyma dialexifying know — see

FAMILY	ETYMON	REFLEXES
Indo-European	*wéyd-e-ti	language data
Austronesian	*ila ₂	language data
Austronesian	*ilo ₂	language data
Austronesian	*kita	language data
Austronesian	*qiro	language data
Austronesian	*taqu-an	language data
Indo-European	*wóyde *widér	language data
Malayo-Polynesian	*kita	language data
PTB	*ka(ŋ)	language data
PTB	*m-kya(ŋ)	language data
PTB	*syey-s	language data

Access to sources

EvoSem sources

- A large etymological database: **Wiktionary**
 - data is sourced, reliable (for Indo-Eur.)
 - sources for lgs + reconstructions

argentum

References [edit]

- “[argentum](#)”, in Charlton T. Lewis and Charles Short (1879) *A Latin Dictionary*, Oxford: Clarendon Press
- “[argentum](#)”, in Charlton T. Lewis (1891) *An Elementary Latin Dictionary*, New York: Harper & Brothers
- [argentum](#) in Charles du Fresne du Cange’s *Glossarium Mediæ et Infimæ Latinitatis* (augmented edition with additions by D. P. Carpenterius, Adelungius and others, edited by Léopold Favre, 1883–1887)
- [argentum](#) in Gaffiot, Félix (1934) *Dictionnaire illustré latin-français*, Hachette
- Carl Meißner; Henry William Auden (1894) *Latin Phrase-Book*^[1], London: Macmillan and Co. [2 phrases]
- “[argentum](#)”, in Harry Thurston Peck, editor (1898) *Harper’s Dictionary of Classical Antiquities*, New York: Harper & Brothers

Reconstruction:Proto-Indo-European/h₂rǵntóm

- ¹ Mallory, James P. (1984), “Proto-Indo-European Silver”, in *Zeitschrift für vergleichende Sprachforschung*^[1], volume 1, issue 97, Göttingen: Vandenhoeck und Ruprecht, retrieved 7 May 2018, pages 1–12
- ² ⁰ ^{2.1} Kroonen, Guus (2013), “akra- 2”, in *Etymological Dictionary of Proto-Germanic* (Leiden Indo-European Etymological Dictionary Series; 11), Leiden, Boston: Brill, →ISBN, page 18: “h₂reǵ-nt-o-”
- ³ ⁰ ^{3.1} Delamarre, Xavier (2003), “*arganto-”, in *Dictionnaire de la langue gauloise: une approche linguistique du vieux-celtique continental* [Dictionary of the Gaulish language: A linguistic approach to Old Continental Celtic] (Collection des Hespérides; 9), 2nd edition, Éditions Errance, →ISBN, page 41: “h₂erǵnto-”
- ⁴ ⁰ ^{4.1} Martirosyan, Hrach (2010) *Etymological Dictionary of the Armenian Inherited Lexicon* (Leiden Indo-European Etymological Dictionary Series; 8), Leiden, Boston: Brill, pages 138–139
- ⁵ ⁰ Rastorgujeva, V. S.; Edel’man, D. I. (2000) *Etimologičeskij slovar’ iranskix jazykov* [Etymological Dictionary of Iranian Languages] (in Russian), volume I, Moscow: Vostochnaya Literatura, page 231
- ⁶ ⁰ , Kent, Ronald G; Old Persian, Grammer, Text, Lexicon; Univerity of pennsylvania: 1950 p:171.

1 root, 27 etyma
= 27 cognate sets

Proto-Indo-European [edit] [↕]

Root [edit] [↕]

*weyd- (stative)^[1]^[2]

1. to see

Derived terms [edit] [↕]

► Terms derived from the Proto-Indo-European root *weyd-

- *wéyd-e-ti (thematic root present)^[1]^[3]
 - Proto-Balto-Slavic: ***wei[?]detei**
 - ⇒ Proto-Balto-Slavic: ***wei[?]dē[?]tei**
 - Latvian: **viedēt**
 - Lithuanian: **veizdėti**
 - Proto-Slavic: ***viděti** (*see there for further descendants*)
 - Proto-Celtic: ***wēdeti** (“to tell, relate”) (*see there for further descendants*)
 - >² Proto-Germanic: ***wlītaną** (“to see, look”) (*#wl contamination perhaps from *wel- ~ *wl- (“to see”)^[4] (see there for further descendants)*)
 - ⇒ Proto-Germanic: ***waitōną** (“to search”) (*see there for further descendants*)
 - ⇒ Proto-Germanic: ***wiltiz** (“sight, appearance, face”) (*see there for further descendants*)
 - Proto-Germanic: ***wītaną** (“to direct the attention to, to scold”) (*see there for further descendants*)
 - Proto-Hellenic: ***wēidō**
 - Ancient Greek: **εἶδω** (eídō, “to be seen, appear”)

- *wéyd-se-ti (“to want to see”, desiderative)^[1]

- *wéyd-ti (athematic root present)
 - Proto-Indo-Iranian: ***wáy[?]ti** (“to know”) (*see there for further descendants*)

- *wid-é-t (thematic root aorist)
 - Armenian:
 - Old Armenian: **տգիւն** (egit, “to find”)
 - Proto-Germanic: ***wītaną** (“to know”) (*see there for further descendants*)
 - Proto-Hellenic: ***wīdon**
 - Ancient Greek: **εἶδον** (eídon)
 - Proto-Italic: ***wīdī**
 - Latin: **vidī**
 - ⇒ Proto-Italic: ***widēō^[1]** (*see there for further descendants*)
 - Proto-Indo-Iranian: ***Háwidat**
 - Proto-Indo-Aryan: ***Háwidat**
 - Sanskrit: **अविदत्** (ávidat)

EvoSem sources



STEDT

Sino-Tibetan Etymological Dictionary and Thesaurus



STEDT Etymon #2670

8.0 Abstract Nouns and Verbs, Psychological Verbs, Verbs of Utterance > 8.3 Psychological verbs > 8.3.1 Verbs of Feeling > 8.3.1.12 ? > 8.3.1.12.1 **Know, Believe**

#2670 PTB *syey-s KNOW / UNDERSTAND

Reconstructed mesoroots below:

- PKC *thay-I, thay?-II KNOW / ABLE / HEAR
- PCN *m-thət KNOW
- PTK *t^hej KNOW / SEE
- PNN *C_o-yæ:y KNOW
- PNN *C_o-syæ:y LEARN
- PLB *šey^{2/3} KNOW
- PL *si² KNOW

form	gloss	gfn	language	source	srcid	notes
0.1 Tibeto-Burman (previously published reconstructions)						
syey	know		*Tibeto-Burman	Chou 72	1257e	
syey	know	v.	*Tibeto-Burman	Weidert 87 T8To	325	
syey	know		*Tibeto-Burman	Matisoff 85 GSTC	048	
syey	know		*Tibeto-Burman	LaPolla 87	98	
syey	know		*Tibeto-Burman	Benedict 72 STC	182	
syey (*C)	know		*Tibeto-Burman	Coblin 86	101	
syey-s	know		*Tibeto-Burman	Matisoff 03 HPTB	614	
1 NE Indian Areal Group						

Introduction Index to Sets Cognate Sets Finderlist
Subgroups Languages Words Proto-form indexes
References+ Roots Loans Noise Formosan

THE AUSTRONESIAN COMPARATIVE DICTIONARY

web edition

by

Robert Blust



***banua** inhabited land, territory supporting the life of a community

PMP *banua inhabited land, territory supporting the life of a community [🔗](#)

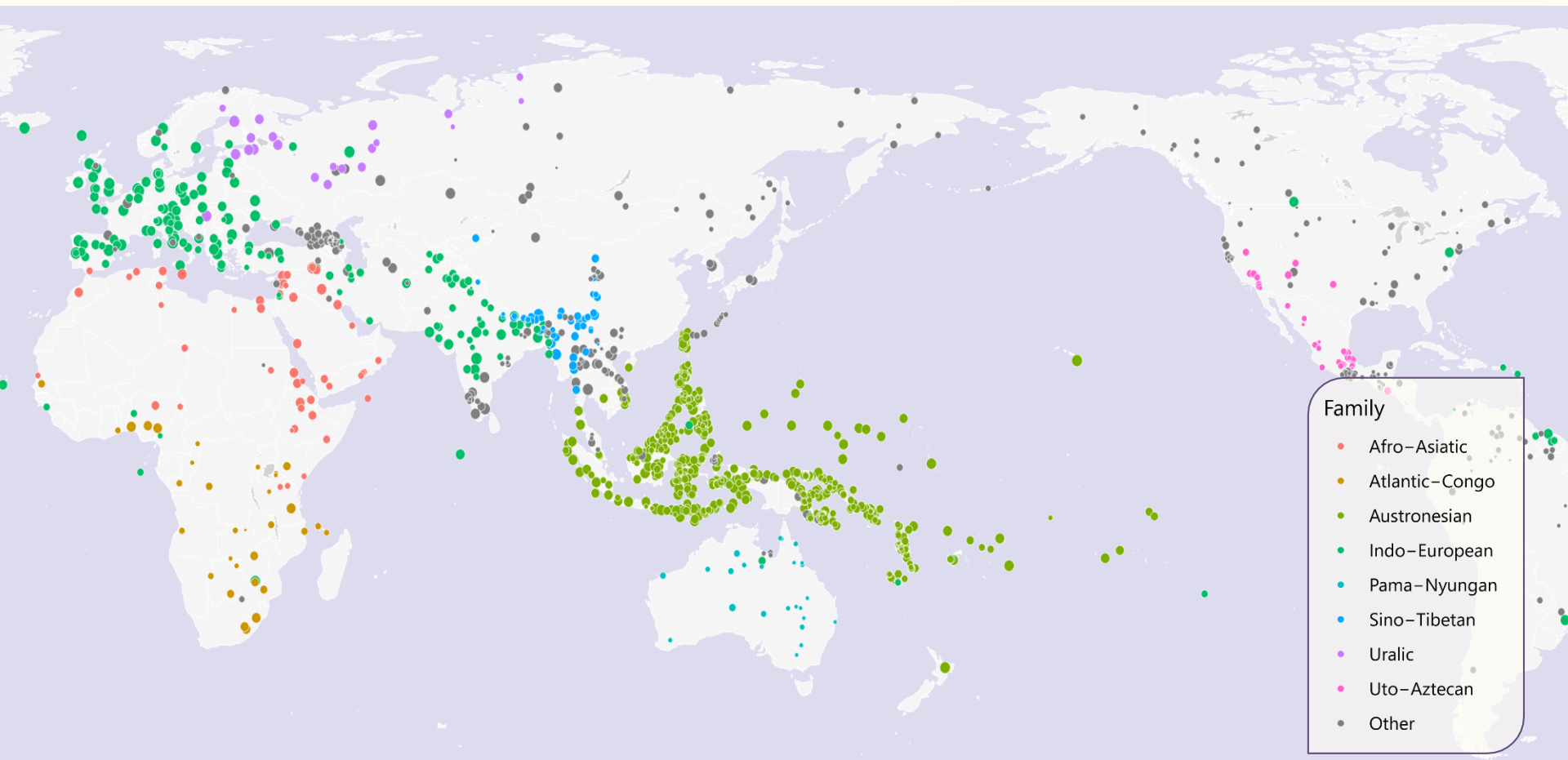
WMP

Itbayaten	vanua	landing place, port
Pangasinan	banwá	sun
Kapampangan	banwá	year; sky, heaven
Buhid	bánwa	sky
Aklanon	banwá	town, country
Hiligaynon	banwá	town, community; a compactly settled area usually larger than a village but smaller than a city
Cebuano	banwá	fatherland; town, village
Subanen/Subanun	meg-banua	live, dwell
Subanon	meg-banua	live, dwell
Palauan	belúu	country, village, place
Sangir	banua	land, district; people; state; sea; weather
Simalur	bano	land, place, district
	fano	land, place, district
Toba Batak	banua toru	underworld, world of the dead

EvoSem stats

EvoSem = largest **comparative etymological dictionary** ever compiled

Languages	2436
Top-level families	56
Proto-languages	109
Etyma	19,876
Reflexes (words)	187,396
Meanings	50,945



Conclusion

- **Dialexification**

- a tool for the historical linguist
 - a complement to *colexification* and CLiCS
- reveals deep-reaching semantic connections among concepts
- helps in modelling processes of lexical change (*lexical tectonics*)

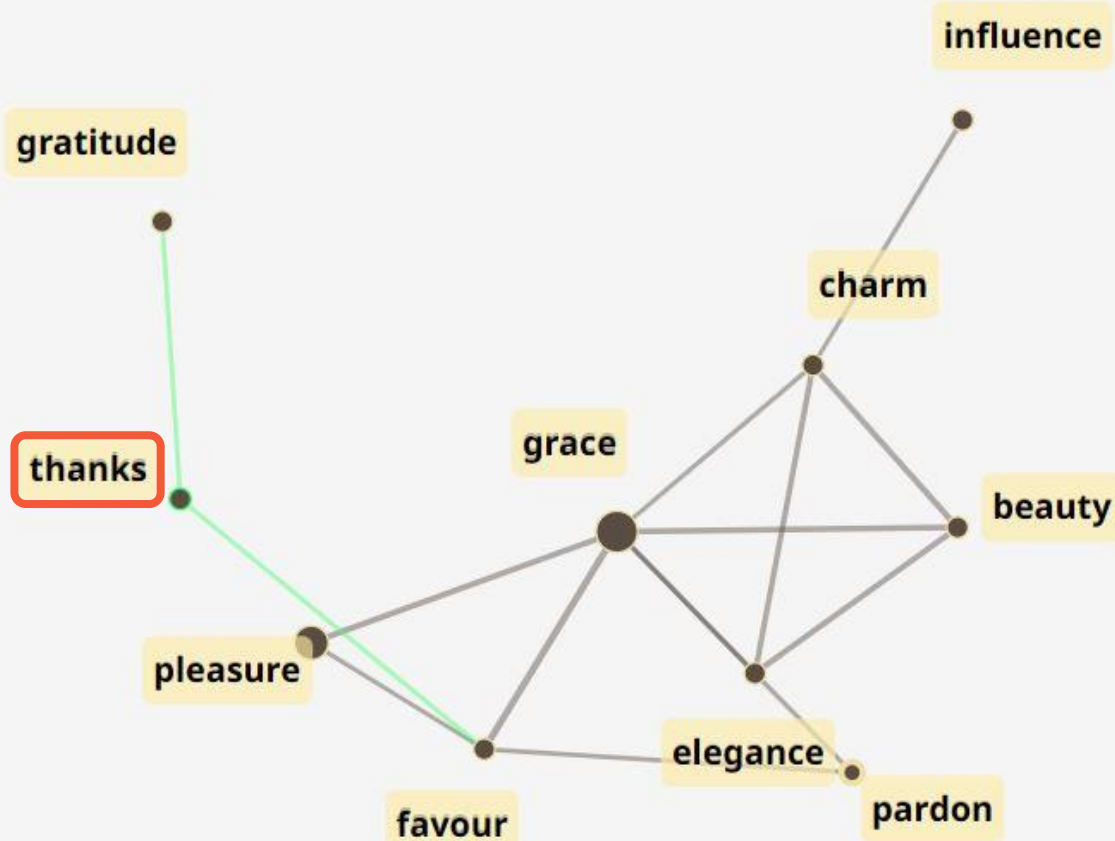
- Perspectives for *EvoSem*

- Making the most of 200 years of research on etymology
- Incorporate more families
 - based on available etymological databases
- Collective effort, for years to come
- Comments & suggestions welcome

Languages	2436
Top-level families	56
Proto-languages	109
Etyma	19,876
Reflexes (words)	187,396
Meanings	50,945



Thanks for your attention



Dialexification threshold: $\theta = 3, \lambda = 14$

tiny.cc/EvoSem_ICLC