

# The role of homosemasy for constructing comparative concepts in cross-linguistic studies

**Martin Haspelmath**

*Max-Planck-Institut für evolutionäre Anthropologie, Leipzig*

Re-thinking synonymy, University of Helsinki, 2010 October 28-30



MAX-PLANCK-GESELLSCHAFT

# homosemasy

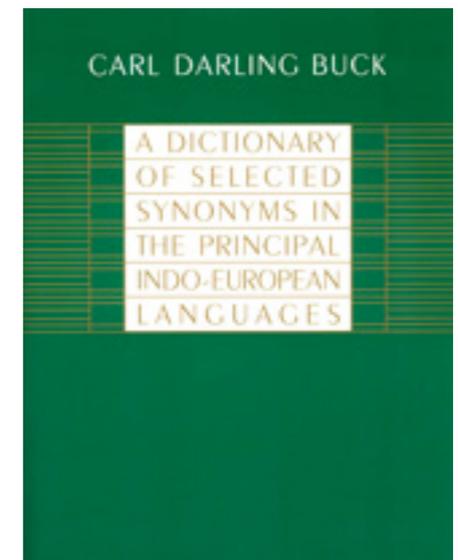
(“cross-linguistic synonymy”)

- **synonymy** = identity of meaning of two forms of the same language
- **homosemasy** = identity of meaning of two forms of two different languages
- (Greek **ομο** ‘equal’ + **σημασία** ‘meaning’)
- (other possibilities: *homosemy*, *synsemy*)

# homosemasy

(“cross-linguistic synonymy”)

- *synonymy* may be used as a synonym of homosemasy
- **cf.** Buck, Carl Darling. 1949. *A dictionary of selected synonyms of the principal Indo-European languages*. Chicago: University of Chicago Press.



# homosemasy

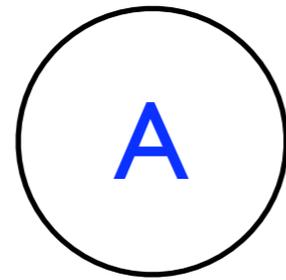
(“cross-linguistic synonymy”)

- homosemasy  $\neq$  translational equivalence
- because translation is an activity, a practical application of linguistics
- homosemasy is a theoretical notion – ideally translational equivalents should be based on homosemasy, not the other way round

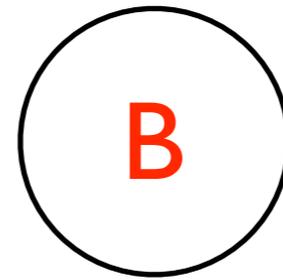
# theoretical language comparison

- pre-modern: based on **equation** of categories across languages
- modern: based on linking categories to a special ***tertium comparationis***, a kind of technical measure (to be called *comparative concept* in the following)

# equation of categories



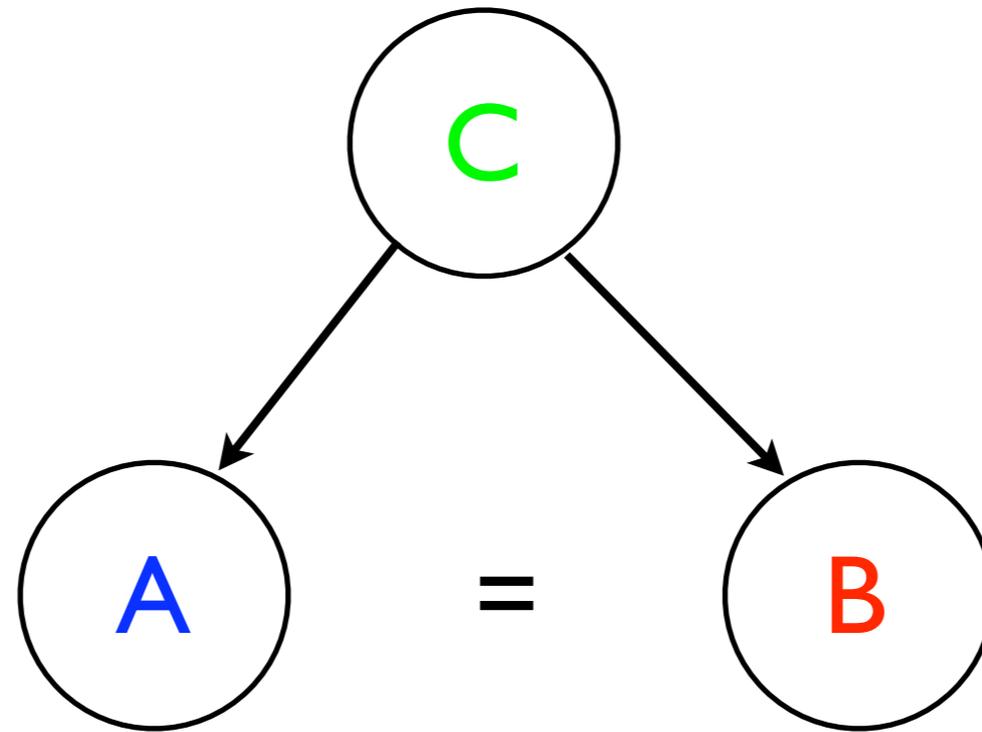
=



language 1

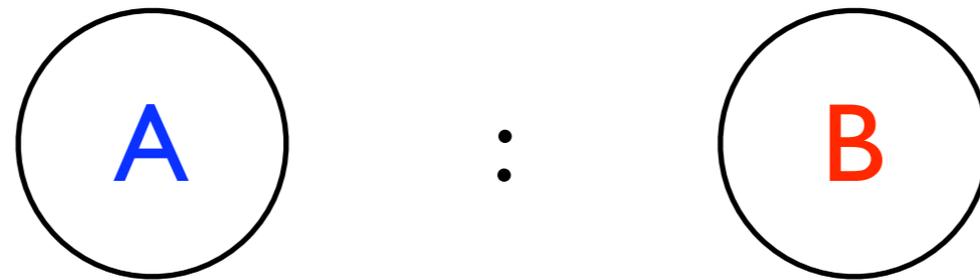
language 2

# equation of categories



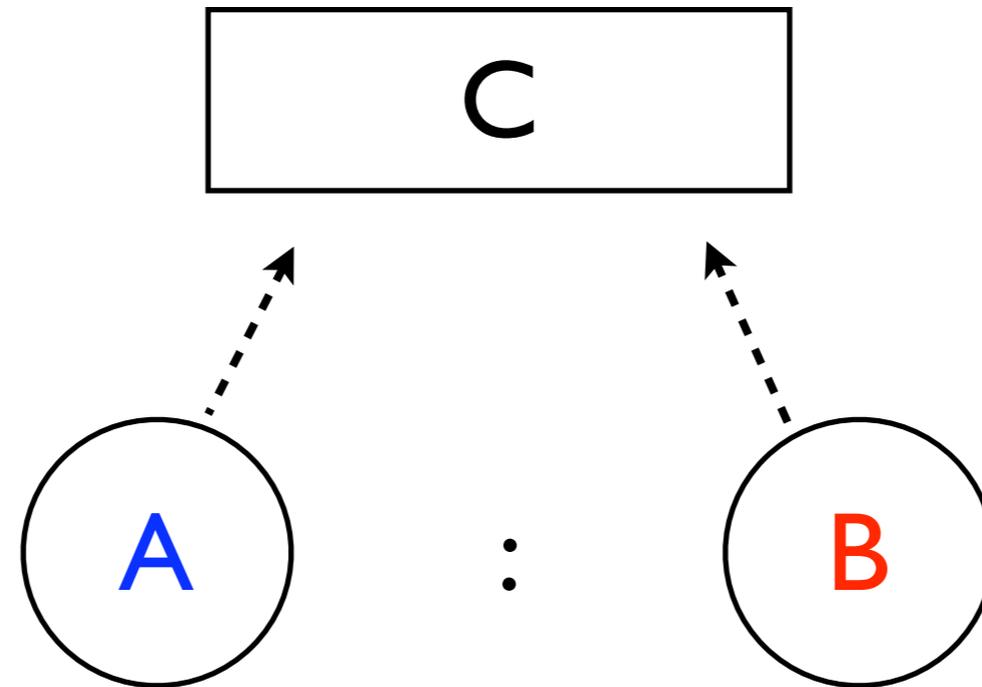
A and B are **equated**, and both **instantiate** the universal/cross-linguistic category C.

# modern comparison of categories



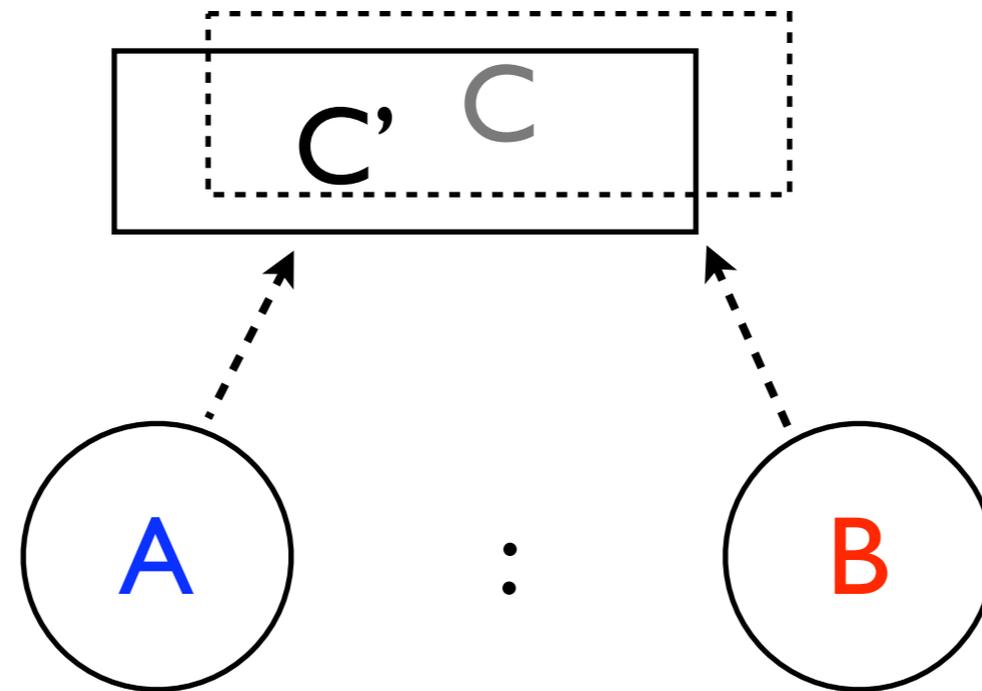
A and B are compared, based on some similarity.

# modern comparison of categories



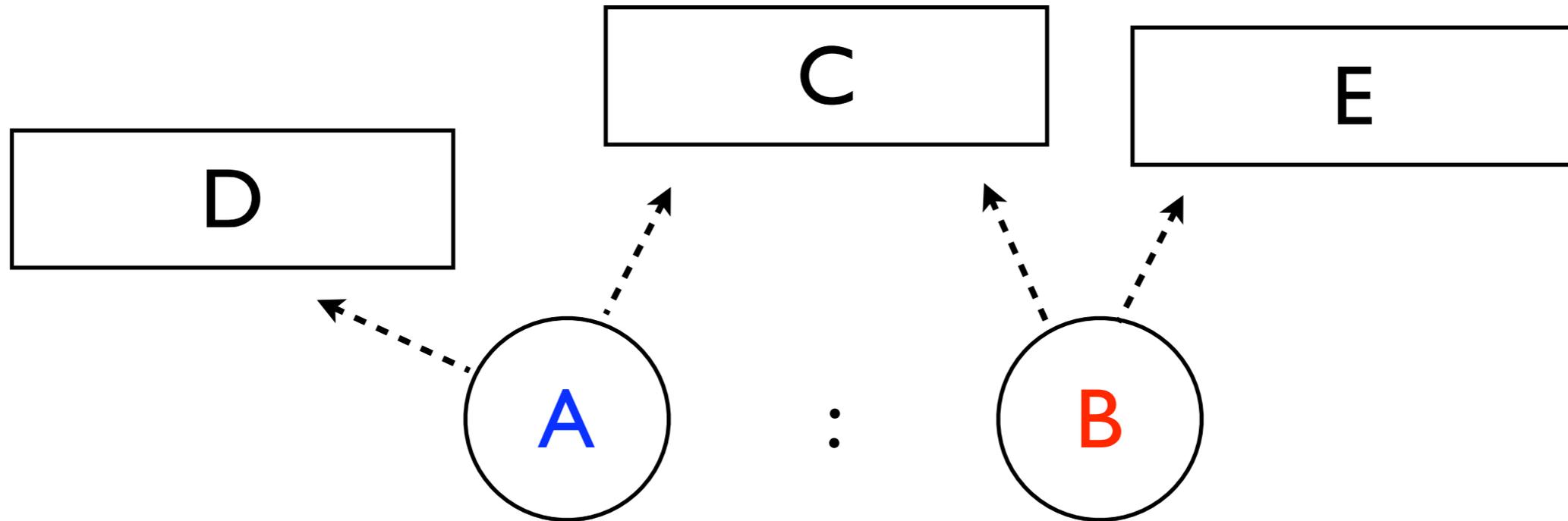
A and B are related to a *tertium comparationis* C (which is not a category, but a **comparative concept**).

# modern comparison of categories



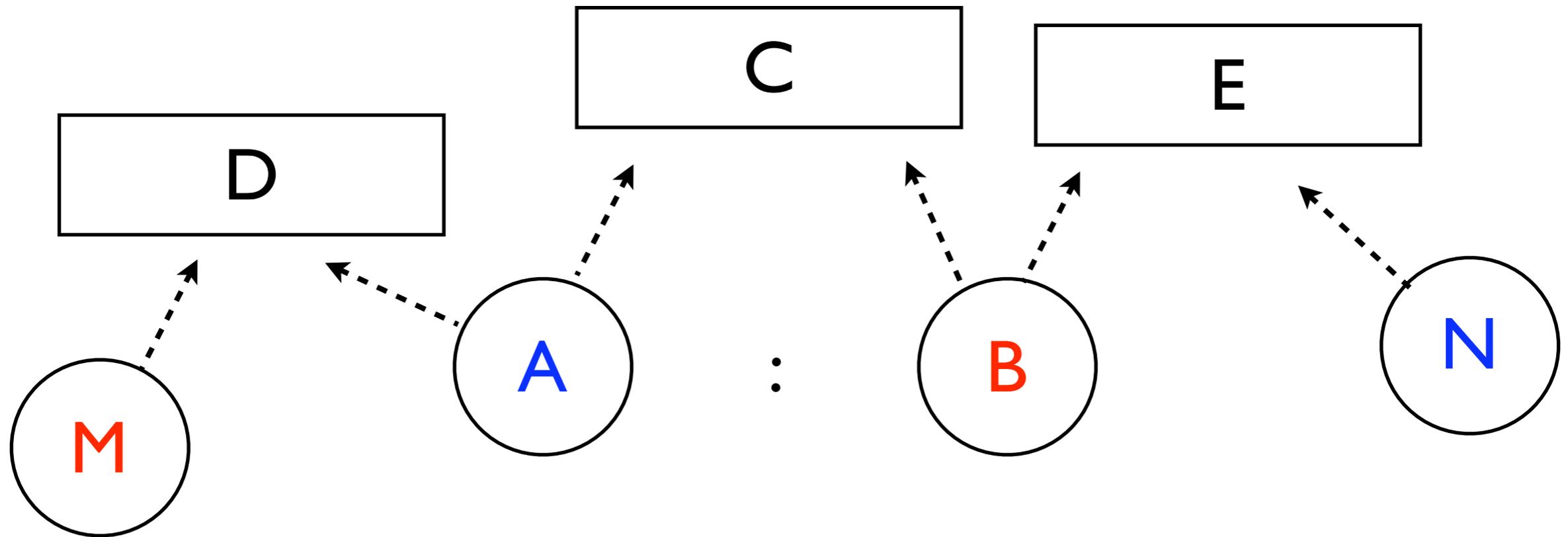
A and B could also be related to an alternative comparative concept C'.

# modern comparison of categories



Or A and B could be related to several comparative concepts simultaneously.

# modern comparison of categories



Or A and B could be related to several comparative concepts simultaneously.

# pre-modern equation of categories

Latin cases equated with Greek cases:

nominative	sacerdos	ιερευς
dative	sacerdoti	ιερει
accusative	sacerdotem	ιερα

cf. Roman gods equated with Greek gods:

Artemis	Diana
Hera	Juno
Hermes	Mercurius

# pre-modern equation of categories

- semantic categories (homosemasy, synsemy)
  - Latin venio = Greek έρχομαι ('come')
- phonological categories (“symphony”)
  - Latin /p/ = Greek /p/ (π)

# pre-modern equation of words in cross-linguistic studies

- word meanings are routinely equated across languages
  - both in historical-comparative linguistics
  - and in general-comparative (“typological”) linguistics

# equating words

## in historical-comparative linguistics

### Swadesh list (in seven languages, partial):

No ☑	English ☑	French ☑	German ☑	Italian ☑	Spanish ☑	Dutch ☑	Swedish ☑
207	name *	nom	Name	nome	nombre	naam	namn
206	because	parce que	weil	perché	porque	omdat	eftersom, ty
205	if	si	wenn, falls, ob	se	si	als, indien	om
204	and	et	und	e	y	en	och
203	with	avec	mit	con	con	met	med
202	in	dans	in	in	en	in	i
201	at	à	bei, an	a	a, en, ante	aan, te, bij	hos, vid
200	left	à gauche	links	sinistra	izquierda	links	vänster
199	right	à droite	rechts	destra	derecha	rechts	höger
198	far	loin	weit, fern	lontano	lejos	ver	långt bort, fjärran
197	near	proche	nah, nahe	vicino	cerca	naar	nära
196	correct	juste, correct	richtig	corretto	correcto	richting, correct	rätt, riktig
195	dry *	sec	trocken	asciutto, secco	seco	droog	torr
194	wet	mouillé	nass, feucht	bagnato	mojado	nat	våt, blöt
193	smooth	lisse	glatt	liscio	suave, liso	glad	len, slät

# equating words in historical-comparative linguistics

language.psy.auckland.ac.nz : Austronesian : Crows : Bantu :

## Austronesian Basic Vocabulary Database

**Word: belly**

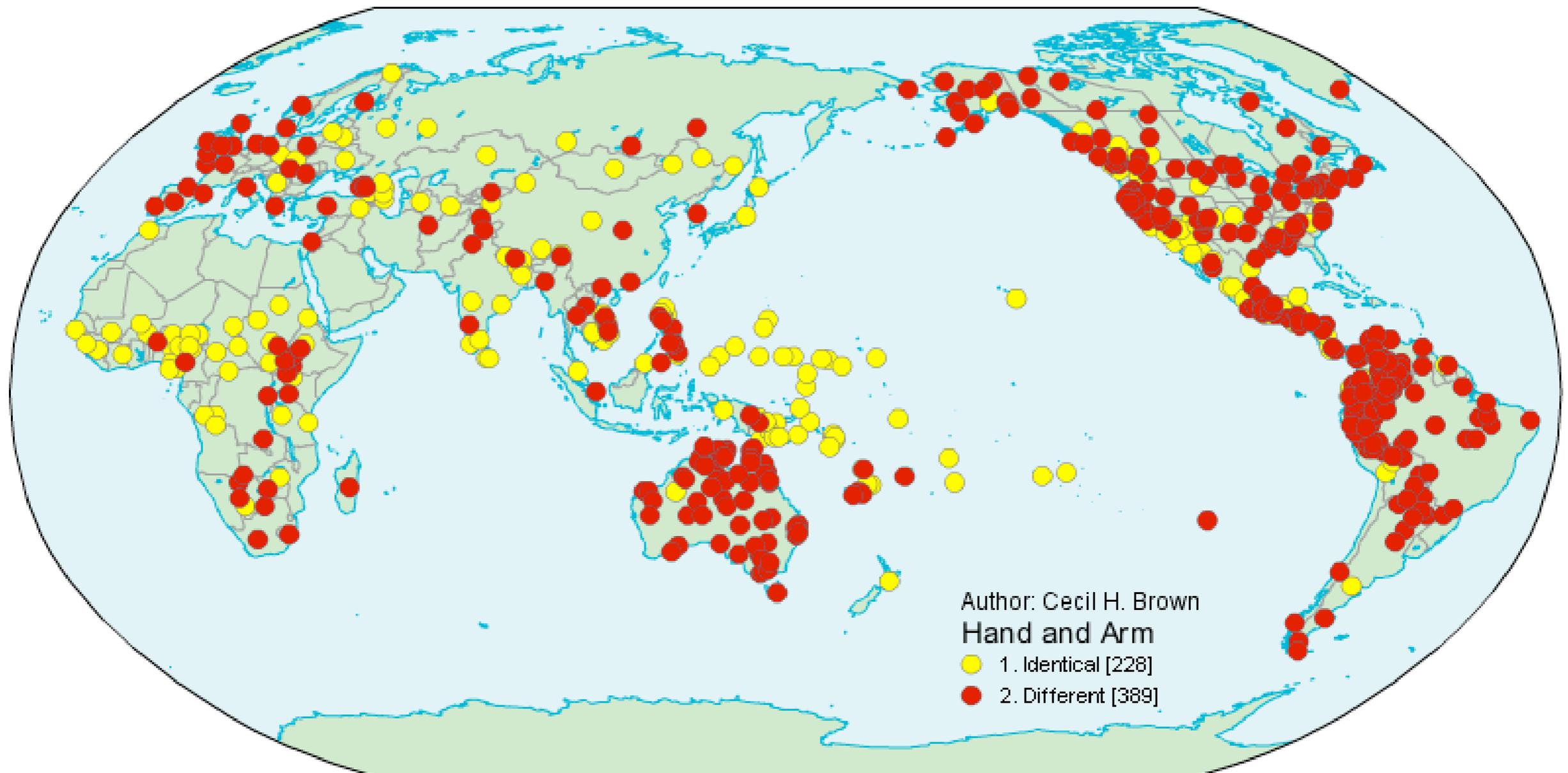
**Entries for "belly":**

ID	Language	Item	Annotation	Cognacy	Classification
172592.	<a href="#">Kalaw Kawaw Ya</a>	maitha			<a href="#">A:P</a>
175281.	<a href="#">Khmer</a>	pəəh			<a href="#">A:M:E</a>
175071.	<a href="#">Vietnamese</a>	bụng			<a href="#">A:M:V</a>
4921.	<a href="#">Proto-Austronesian</a>	*tiaN		1	<a href="#">A</a>
4896.	<a href="#">Squiliq Atayal</a>	ktu?			<a href="#">A:F</a>
4897.	<a href="#">Ci' uli Atayal</a>	labos			<a href="#">A:F</a>
4916.	<a href="#">Sediq</a>	búyas			<a href="#">A:F</a>
72069.	<a href="#">Bunun</a>	ti'án		1	<a href="#">A:F</a>
72070.	<a href="#">Bunun</a>	tteyan		1	<a href="#">A:F</a>
4991.	<a href="#">Central Amis</a>	tiaL	belly, abdomen	1	<a href="#">A:F:E</a>
90215.	<a href="#">Central Amis</a>	ti(j)aL		1	<a href="#">A:F:E</a>
72083.	<a href="#">Kavalan</a>	pələś			<a href="#">A:F:E</a>
72084.	<a href="#">Kavalan</a>	bəʔəś			<a href="#">A:F:E</a>
72085.	<a href="#">Kavalan</a>	tian		1	<a href="#">A:F:E</a>
72086.	<a href="#">Kavalan</a>	tteyan		1	<a href="#">A:F:E</a>
87375.	<a href="#">Basai</a>	tijan		1	<a href="#">A:F:E</a>
72091.	<a href="#">Siraya</a>	vouil			<a href="#">A:F:E</a>

[http://  
language.psy.auckland.  
ac.nz/](http://language.psy.auckland.ac.nz/)  
R. Blust, R. Gray, S.  
Greenhill

# equating words in general-comparative linguistics

Languages with 'hand' words and 'hand/arm' words:



Cecil H. Brown. 2005. In: *The World Atlas of Language Structures*

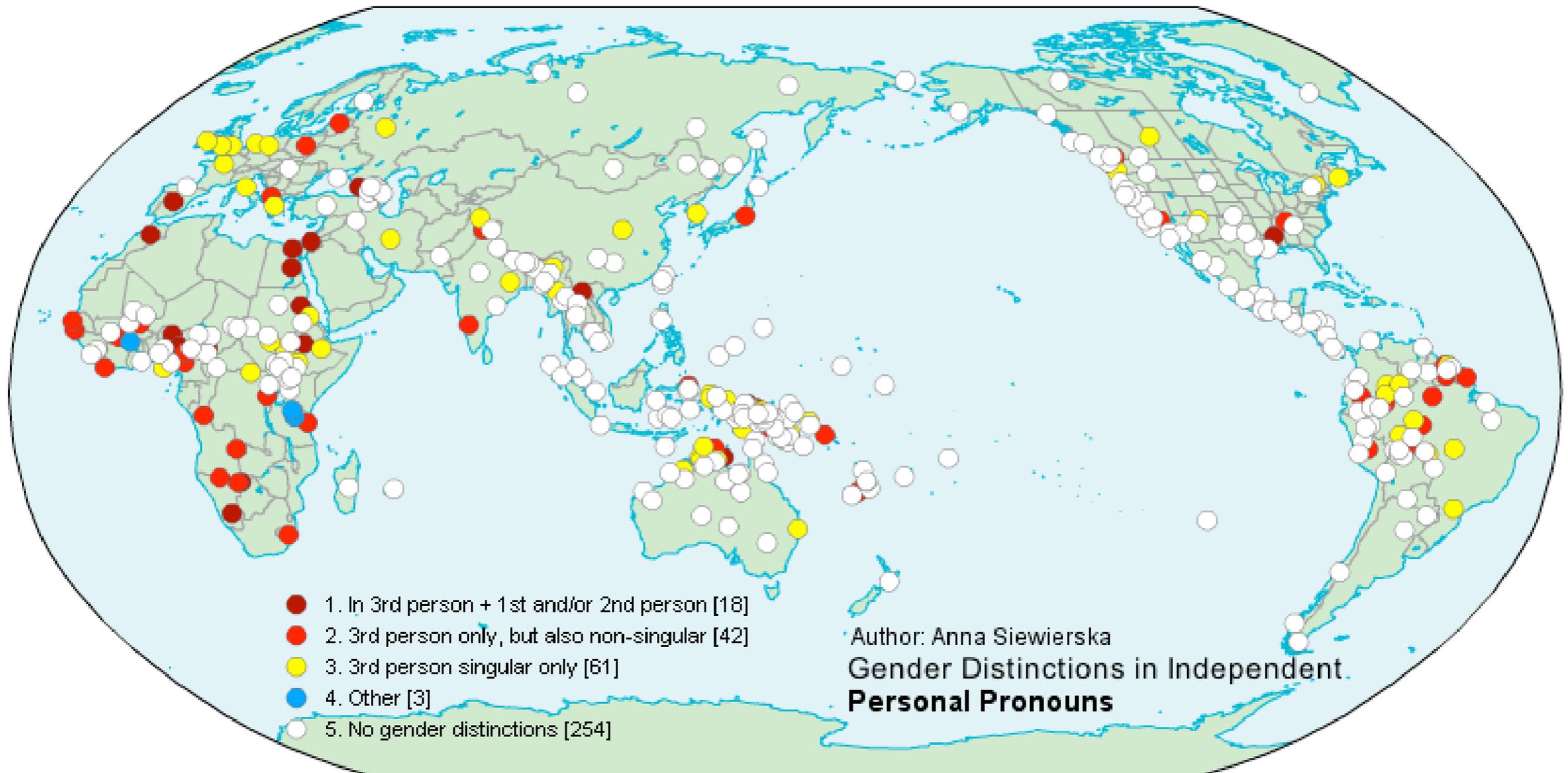
# equating grammatical forms in historical-comparative linguistics

## Independent personal pronouns in Semitic languages:

		<i>Independent Personal Pronouns</i>								
		Egyptian	Tuareg	Rendille	*Pr.-Sem.	P.Syr.	O.Bab.	Ugaritic	Hebrew	Aramaic
Sing.										
1	<i>in-k</i>	<i>n-ək</i>		<i>an(i)</i>	<i>'an-a</i>	<i>'anna</i>	<i>anāku</i>	<i>'an, 'ank</i>	<i>'ānī, 'ānōkī</i>	<i>'ānā</i>
2 m.	<i>nt-k</i>	<i>kay</i>	}	<i>at(i)</i>	<i>'an-ta/ka<sup>(?)</sup></i>	<i>'anta</i>	<i>atta</i>	<i>'at</i>	<i>'attā</i>	<i>'anta, 'att</i>
f.	<i>nt-t</i>	<i>kəm</i>			<i>'an-ti/ki<sup>(?)</sup></i>	?	<i>atti</i>	<i>'at</i>	<i>'att</i>	<i>'anti, 'att</i>
3 m.	<i>nt-f</i>	}	<i>nt-a</i>	<i>us(u)</i>	<i>šū-wa</i>	<i>šūwa</i>	<i>šū</i>	<i>hw</i>	<i>hū'(a)</i>	<i>hū(')</i>
f.	<i>nt-š</i>			<i>iče</i>	<i>ši-ya</i>	<i>šiya</i>	<i>šī</i>	<i>hy</i>	<i>hī'(a)</i>	<i>hī(')</i>

# equating grammatical forms in general-comparative linguistics

Languages with gender in independent personal pronouns:



Siewierska, Anna. 2005. In: *The World Atlas of Language Structures*

# equating constructions in general-comparative linguistics

## Ergative constructions:

Lezgian (Haspelmath 1993)

*Ali-di kic' q'e-na.*

Ali-ERG dog kill-AOR

'Ali killed the dog.'

West Greenlandic Eskimo (Fortescue 1984:265)

*Inu-it nanuq-Ø taku-aat.*

people-ERG bear-ABS see-3PL>3SG.INDIC

'The people saw the polar bear.'

# What is the problem with equating categories?

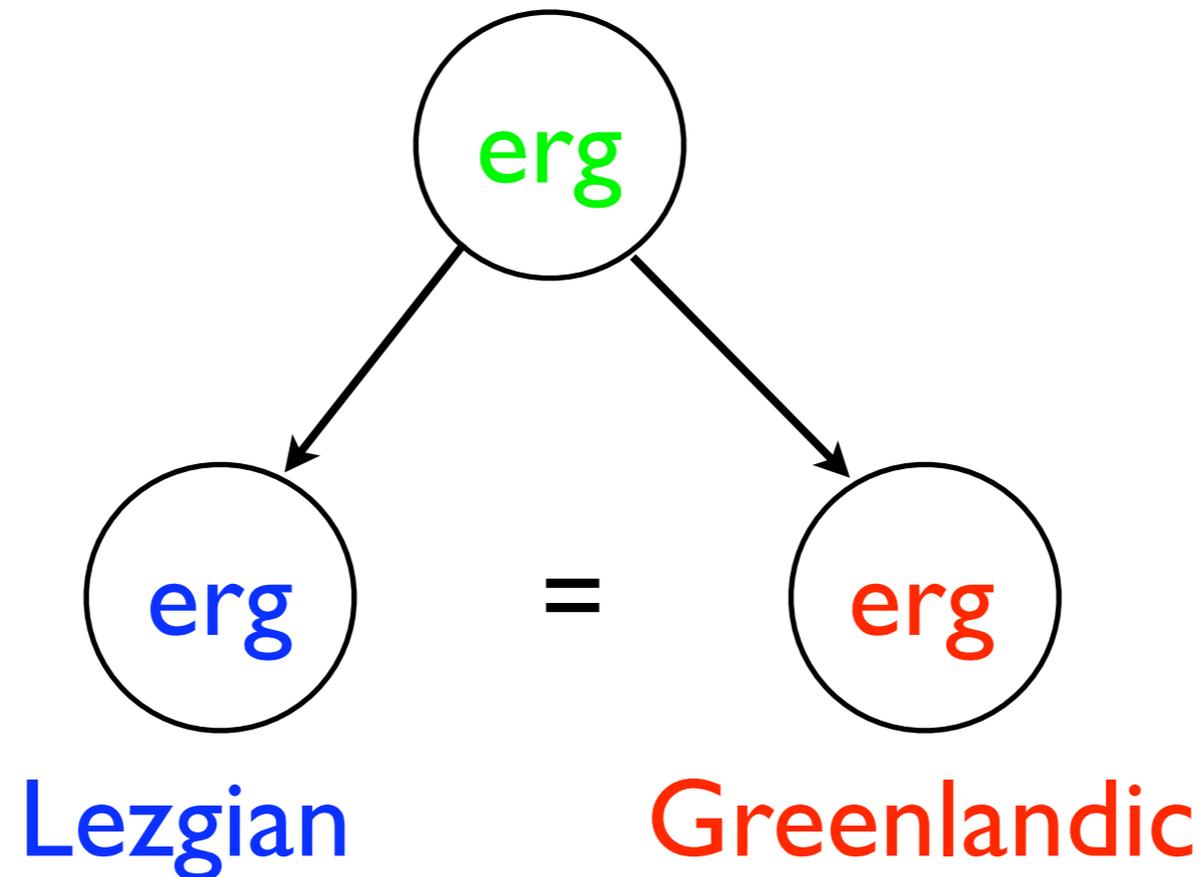
**each language has its own categories**

- grammar: Lezgian “ergative” cannot be equated with Greenlandic “ergative”



Franz Boas (1858-1942)

wrong:



Both ergative cases are **equated**, and both **instantiate** the cross-linguistic category “ergative”.

# What is the problem with equating categories?

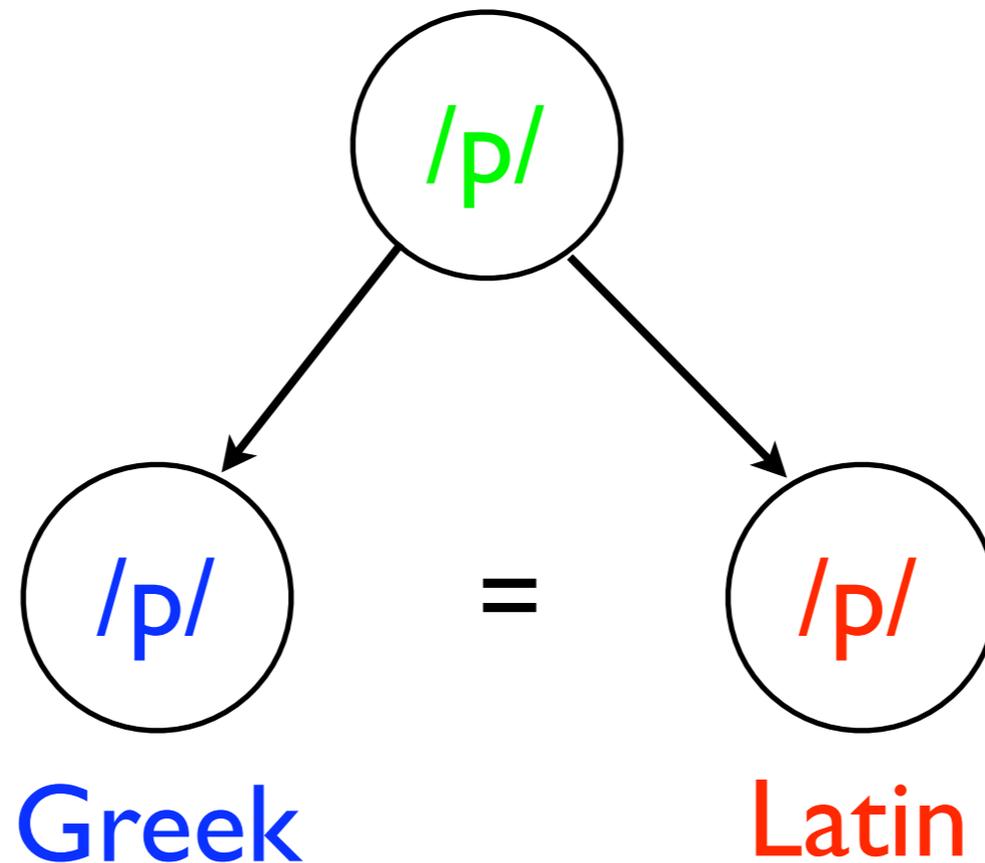
## **each language has its own categories**

- grammar: Lezgian “ergative” cannot be equated with Greenlandic “ergative”
- phonology: Greek /p/ cannot be equated with Latin /p/

Nikolai S. Trubetzkoy (1890-1938)



wrong:



Greek /p/ and Latin /p/ are **equated**, and both **instantiate** the cross-linguistic category /p/.

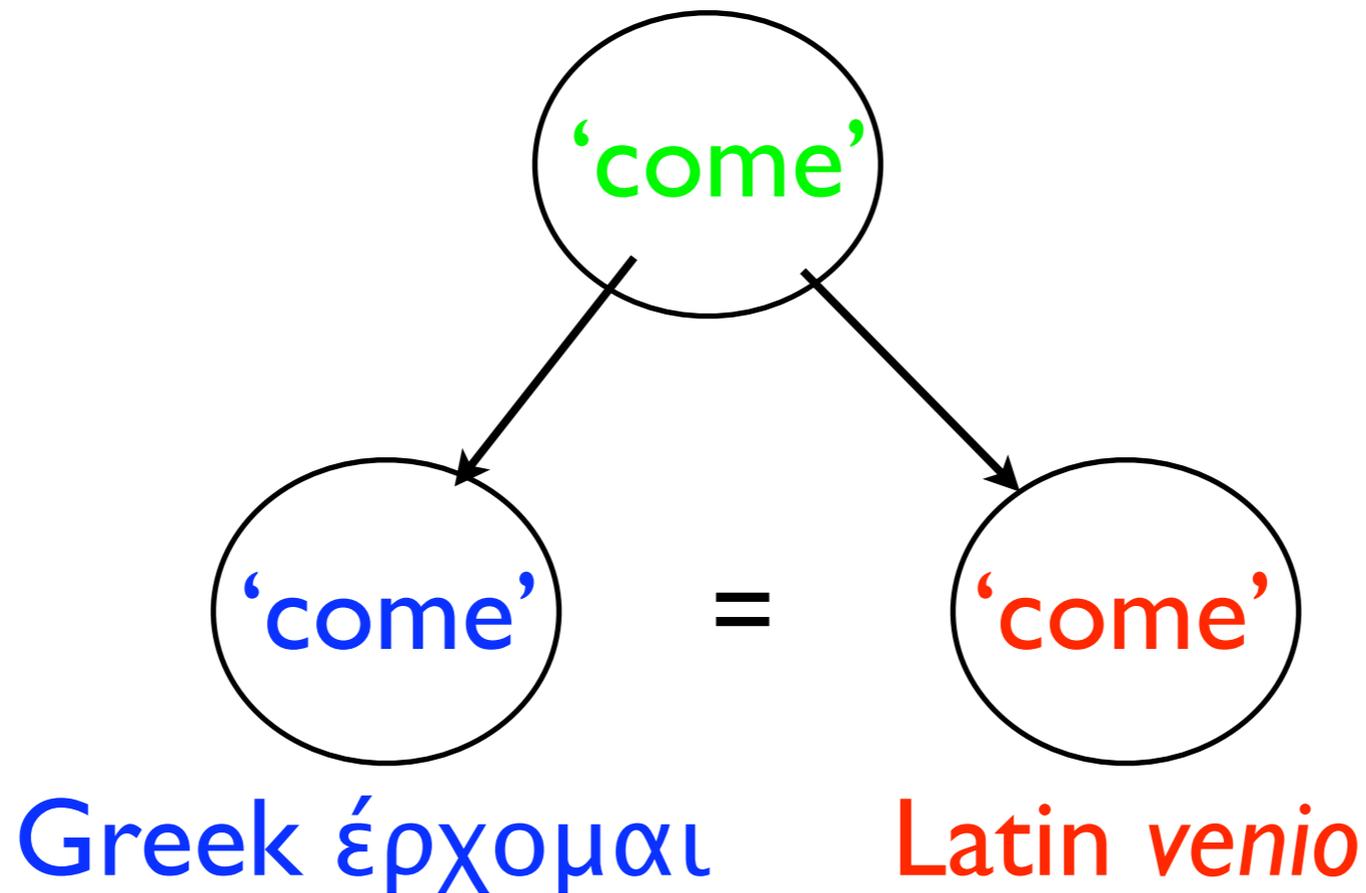
# What is the problem with equating categories?

## **each language has its own categories**

- grammar: Lezgian “ergative” cannot be equated with Greenlandic “ergative”
- phonology: Greek /p/ cannot be equated with Latin /p/
- word meaning: Greek “come” cannot be equated with Latin “come”

**There is no true homosemasy**

wrong:



Greek 'come' and Latin 'come' are **equated**, and both **instantiate** the cross-linguistic semantic category 'come'.

# homosemasy

- **there is no true homosemasy:**
- meanings are often similar across languages, but they almost never match perfectly; cross-linguistic semantic variation is rampant
- cf. Hjelmslev (1943) on ‘tree’ ‘wood’ ‘forest’:

English	German	Danish	French	Spanish
tree	Baum	træ	arbre	árbol
wood	Holz		bois	leña
woods	Wald	skov	forêt	madera
forrest				bosque
				selva

cf. Finnish *metsä*

# homosemasy

Evans (2011):

	♂ referent		♀ referent		
	♀ speaker	♀ speaker	♀ speaker	♀ speaker	
(elder)	1	3	5	7	
(younger)	2	4	6	8	
(elder)	brother		sister		English
(younger)					
(elder)	kakang				Indonesian
(younger)	adik				
(elder)	ani		otōto		Japanese
(younger)	ane		imōto		
(elder)	thabuju	kularrind	kularrind	yakukathu	Kayardild
(younger)	duujind	kularrind	kularrind	duujind	

Evans, Nicholas. 2011. Semantic typology. In: Song, Jae Jung (ed.) 2011. *The Oxford handbook of linguistic typology*. Oxford: OUP, 504-533.

# how to compare languages

- **How do we compare languages? Three approaches:**
- strict universalism (meanings are identical across languages)
- low-level universalism (semantic primitives are identical across languages)
- particularism (languages have different meanings)

# strict universalism

- There is a substantial set of **universal/cross-linguistic** meanings (perhaps innately given) that can be instantiated in all languages.

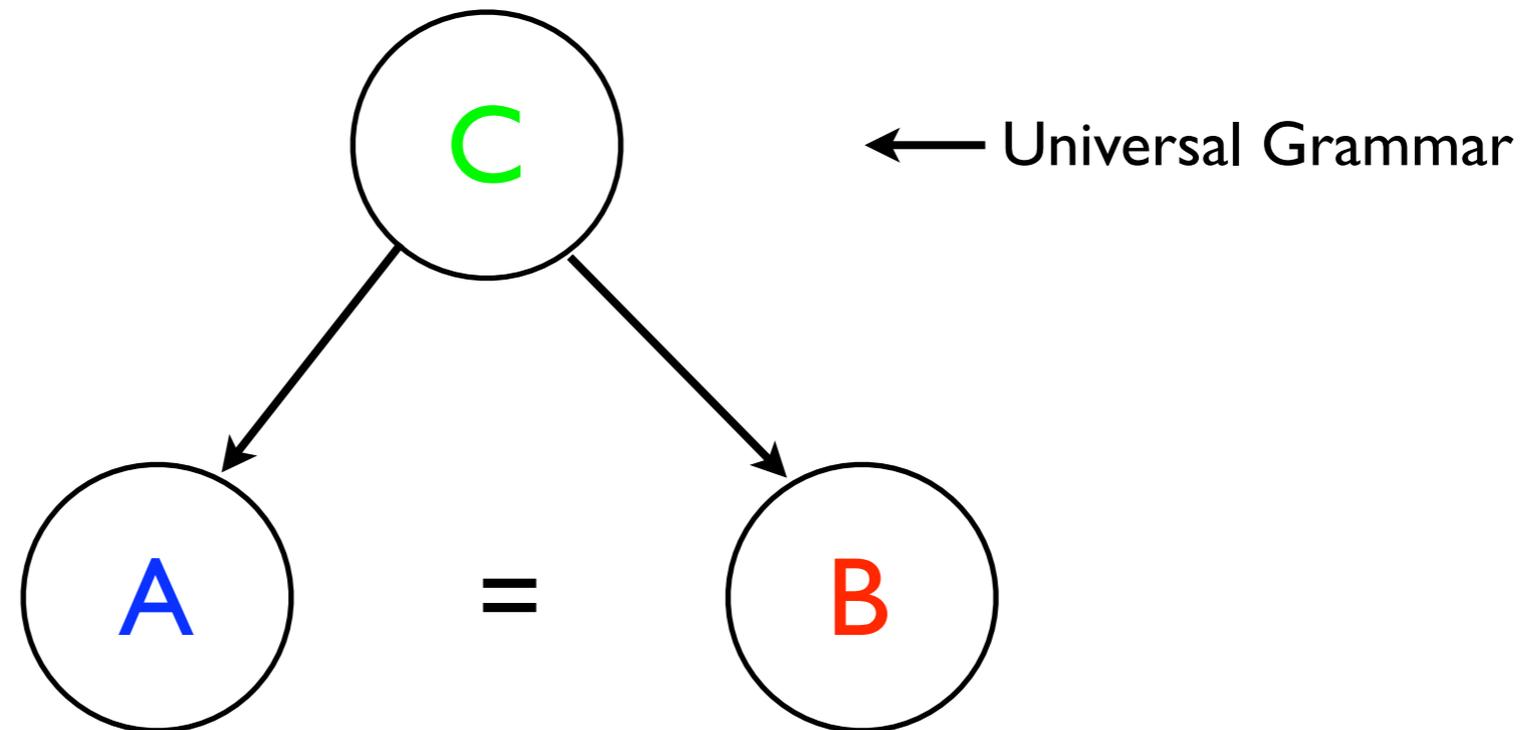
# universal vs. cross-linguistic (categories/meanings/etc.)

- **universal** = occurring in all languages
- **cross-linguistic** = universally available, potentially occurring in any language

# strict universalism

- There is a substantial set of **universal/cross-linguistic** meanings (perhaps innately given) that can be instantiated in all languages. Sometimes words have more than one meaning (ambiguity), and sometimes there are several words for a meaning (synonymy), but these are anomalies.
- In grammar: categories such as "noun", "dative", "infinitive", "tense", "topic" are widely thought to be universally available, i.e. they are cross-linguistic categories (Newmeyer 2007).
- Research programme: Find the universal meanings and categories (Chomsky 1965: "substantive universals of UG"), then find further generalizations.

# strict universalism

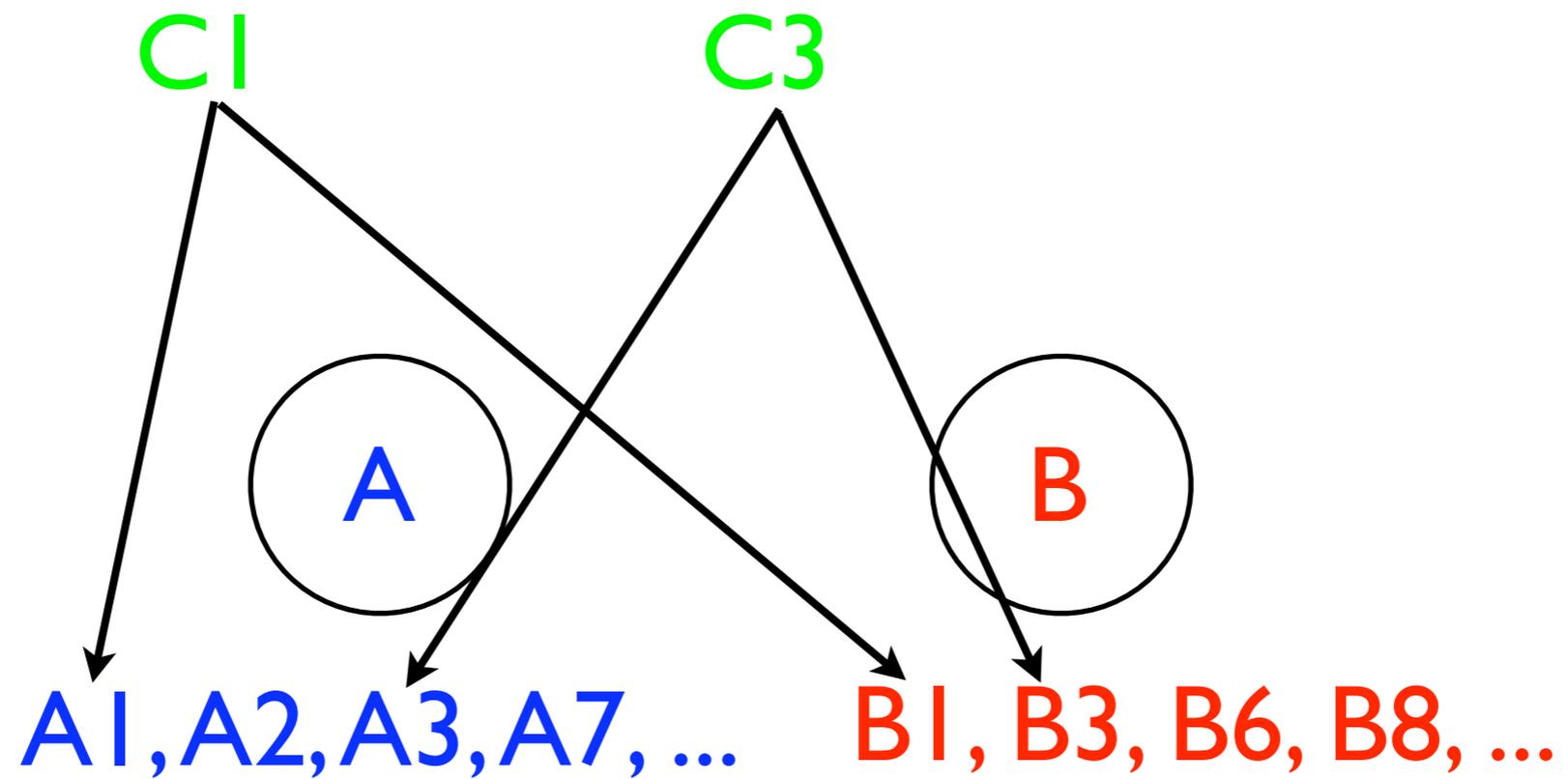


A and B are **equated**, and both **instantiate** the universal/cross-linguistic category C.

# low-level universalism

- There is a small set of semantic primitives out of which all actual meanings can be put together. Many meanings are highly complex, consisting of many dozens of semantic components. (Wierzbicka, Goddard, Natural Semantic Metalanguage approach)
- In grammar: Bickel (2010) thinks that at a lower level, grammatical phenomena are commensurable across languages. Thus, while macro-categories such as "coordination", "subordination", and "cosubordination" are not universal, lower-level distinctions such as "illocutionary scope", "negative scope", "tense marking", are universal.
- Research programme: Find the low-level universal meanings and categories, then find further cross-linguistic generalizations.

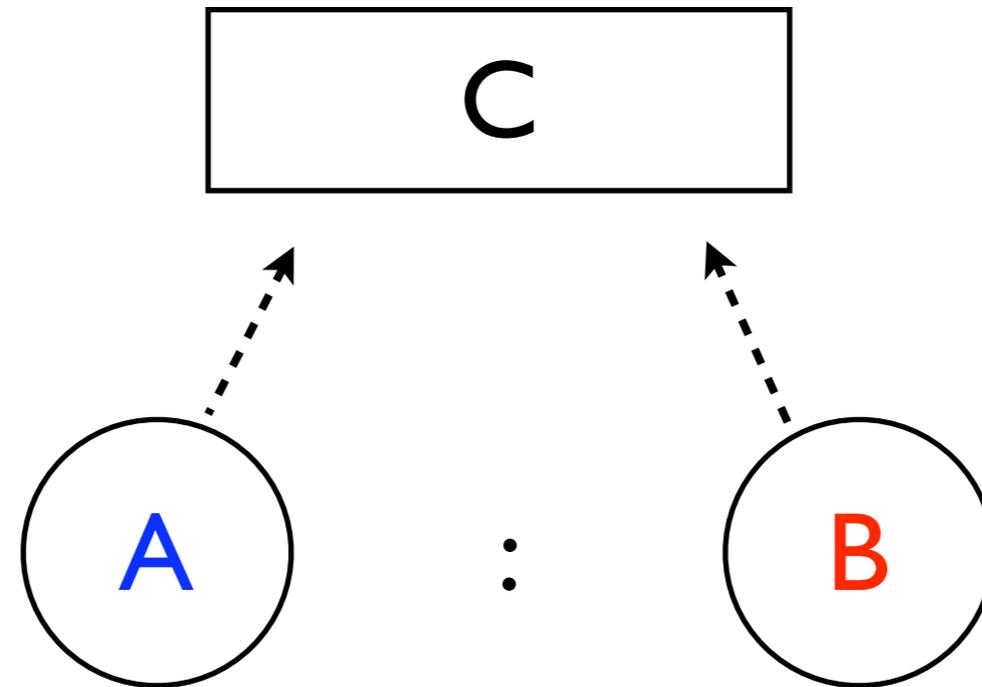
# low-level universalism



# particularism

- There are no absolute limits on the semantic variation across languages. Patterns of coexpression often recur, but in principle, each language has its own meanings.
- In grammar: Each language has its own categories (Boas 1911). Categories tend to be similar across languages, but there are no absolute limits (Dryer 1997, Croft 2001, Lazard 2006, Haspelmath 2007).
- Research programme: Find cross-linguistic generalizations **despite the nonexistence of cross-linguistic categories.**

# particularism



A and B are related to a *tertium comparationis* C, a **comparative concept**.

# low-level universalism and particularism

- In practice, there is no big difference between low-level universalism and particularism:
- The presumed low-level universal categories are very hypothetical and not stable, and at least for broad cross-linguistic comparison, they are often unsuitable.
- Thus, as long as we acknowledge rampant cross-linguistic variation, we can leave aside the difference between the two views.

# Loanword Typology

## (2004-2009)

- Guiding questions:
- What kinds of borrowings are common, what kinds are unusual? Under what circumstances? What is the direction of borrowing?
- (Also as a resource for historical linguistics, to distinguish cognates from loans, cf. English *hand*/German *Hand*; Italian *viola*/English *viola*)

# Loanword Typology

## (2004-2009)

- research method: compare the borrowing behaviour of 41 languages, by looking at comparable samples of words (about 1500 words per language), identifying the loanwords
- but how do we compare words across languages?
- for practical purposes (and perhaps somewhat naively), we started with a meaning list of 1460 meanings, based on the **Intercontinental Dictionary Series (IDS)**
- the IDS is basically like an expanded Swadesh list, except that it has 1300 word meanings instead of 207

# Intercontinental Dictionary Series

## The Intercontinental Dictionary Series

*Founding Editor:*

Mary Ritchie Key (University of California, Irvine)

*General Editor:*

Bernard Comrie (Max Planck Institute for Evolutionary Anthropology, Leipzig)

**Purpose:** The purpose of the IDS is to establish a database where lexical material across the continents is organized in such a way that comparisons can be made. Historical studies, comparative, and theoretical linguistic research can be based on this documentation. This is a long term cooperative project that will go on for the next generation or so and will involve linguists all over the world. It is aimed towards international understanding and cooperation. This is a pioneering effort that will have global impact. The purpose also contributes to preserving information on the little-known and "non-prestigious" languages of the world, many of which are becoming extinct.

**Rationale:** Information on languages of the world is scattered over all the continents and islands and published in dozens of languages and scripts. There is need of a database where one can find comparable material to formulate hypotheses and test and validate those theories. For example, theories on intercontinental connections have been proposed on the basis of the distribution of 'sweet potato' and yet there is no single source, where words with this meaning can be found in many languages. Good quantitative and statistical

- <http://lingweb.eva.mpg.de/ids/>



[IDS Main Page](#)

[Simple Browsing](#)

[Advanced Browsing](#)

[Download Data](#)

[Technical Background](#)

# Intercontinental Dictionary Series

entry_id	trans_english	Chatino, Zacatepec [Phonemic]	Finnish [Phonemic]	Rotuman [Phonemic]	Yanomámi [Phonemic]
110	body	t <sup>y</sup> uk <sup>w</sup> i ɲaʔā-; tɪɲaʔā-	ŗuumis	foro	pei ya; pei a
120	skin, hide	kihī-	iho; nahka	ʔuli	pei si
130	flesh	kunāʔ-	liha	tiko	pei yāhi
140	hair (head)	kičāʔ-	hiukset; tukka	leva	pei het <sup>h</sup> ānāu
142	beard	sū-	paŗta	kum-kumu; kum-kumi	pei kaweiki
144	body hair	slu-	ihokaŗwa	leva	pei kōi
145	pubic hair		hæpūkaŗwat	lev-leva ??	pei weši
146	dandruff	šiyēʔ	hilse	ʔül tat ne pelʔuta	
150	blood	tnē-	weŗi	toto	ĩyā-ĩyā-pə
151	vein, artery	sta tnē-	laskimo; waltimo	uaua	māt <sup>h</sup> ō-yāhi
160	bone	tih <sup>y</sup> ā-	luu	sui	pei ũ
162	rib	tih <sup>y</sup> ā siʔ-	kūlkiluu	sui vakuaka	pei repokosi
170	horn	yka ke-	saŗwi	ʔipesi	pei yōrāki

# Loanword Typology

About 60,000 words online (<http://wold.livingsources.org/>):



WORLD LOANWORD DATABASE (WOLD)

parpaare Gawwada    فلفل Arabic    بلبل Persian    \*pippari Middle Indic    πικέρι Greek

pillipili Swahili    सिपलि Sanskrit

Home    Vocabularies    Languages    Meanings    Authors    Newsblog  
Contact    Help

book | WOLD project pages | legal

## Welcome to the World Loanword Database (WOLD)

The World Loanword Database (WOLD), edited by [Martin Haspelmath](#) and [Uri Tadmor](#), is a scientific publication by the [Max Planck Digital Library](#), Munich (2009).

It provides **vocabularies** (mini-dictionaries of about 1000-2000 entries) of 41 languages from around the world, with comprehensive information about the loanword status of each word. It allows users to find **loanwords**, **source words** and **donor languages** in each of the 41 languages, but also makes it easy to compare loanwords across languages.

Each vocabulary was contributed by an expert on the language and its history. An accompanying book is being published by Mouton de Gruyter ([Loanwords in the World's Languages: A Comparative Handbook](#), edited by Martin Haspelmath & Uri Tadmor).

The World Loanword Database (WOLD) 2009 consists of vocabularies contributed by 41 different authors or author teams. When citing material from the database, please cite the corresponding vocabulary (or vocabularies).

The database can be accessed by language, by meaning, by author, or by reference.

Search WOLD for meanings, languages, words, authors, references ...

×

powered by [Google™](#)

### WOLD News

#### Hello world!

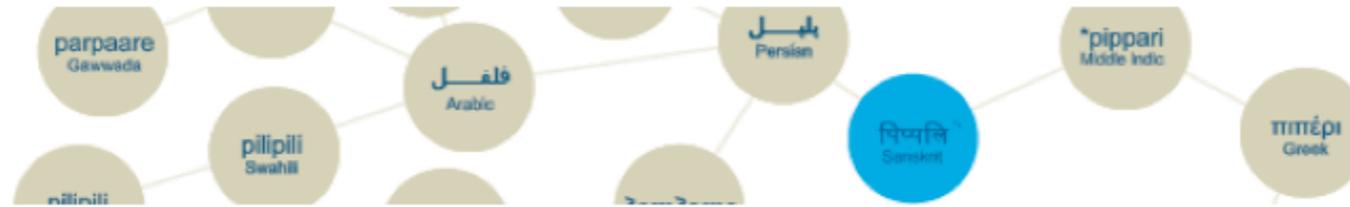
by wold - Jan 26, 2010

Today saw the release of the World Loanword Database (WOLD). This database was collected during the Loanword Typology Project ...

# Loanword Typology

The 1460 comparison meanings are given online with some statistics:

WORLD LOANWORD  
DATABASE (WOLD)



[Home](#)

[Vocabularies](#)

[Languages](#)

[Meanings](#)

[Authors](#)

[Newsblog](#)

[Contact](#)

[Meanings by Semantic Field](#) | [All Meanings](#) |

## Meanings: Semantic Field The body

LWT code ↓ [help]	Meaning [help]	Semantic category [help]	Borrowed score [help]	Age score [help]	Simplicity score [help]	Representation [help]
4.11	the body	Noun	0.19	0.85	0.94	51
4.12	the skin or hide	Noun	0.11	0.90	0.92	49
4.13	the flesh	Noun	0.15	0.89	1.00	44
4.14	the hair	Noun	0.06	0.88	0.92	48
4.142	the beard	Noun	0.17	0.83	0.93	45
4.144	the body hair	Noun	0.12	0.87	0.87	40
4.145	the pubic hair	Noun	0.17	0.81	0.80	43
4.146	the dandruff	Noun	0.26	0.78	0.82	40
4.15	the blood	Noun	0.10	0.90	1.00	47
4.151	the vein or artery	Noun	0.26	0.82	0.79	54
4.16	the bone	Noun	0.08	0.88	0.97	44

# Loanword Typology

For each comparison meaning, the counterparts from all 41 languages are listed:

## Meaning 4.142: the beard

**Semantic field:** The body

**Semantic category:** Noun

**Borrowed score :** 0.17

**Age score :** 0.83

**Simplicity score :** 0.93

### Counterpart words in the World Loanword Database

Voc. ID ↓	Vocabulary	Word Form	Original Script	Borrowed status	Borrowed score
1	Swahili	<i>ndevu</i>		5. no evidence for borrowing	0.00
2	Iraqw	<i>daamóot</i>		1. clearly borrowed	1.00
3	Gawwada	<i>puše</i>		5. no evidence for borrowing	0.00
4	Hausa	<i>géemùu</i>		5. no evidence for borrowing	0.00
5	Kanuri	<i>njìshì</i>		4. very little evidence for borrowing	0.25
6	Tarifiyt Berber	<i>aɣəšmia</i>		5. no evidence for borrowing	0.00
6	Tarifiyt Berber	<i>tmāt</i>		5. no evidence for borrowing	0.00
7	Seychelles Creole	<i>labarb</i>		5. no evidence for borrowing	0.00
8	Romanian	<i>barbă</i>		5. no evidence for borrowing	0.00
9	Selice Romani	<i>sakálla</i>		1. clearly borrowed	1.00
10	Lower Sorbian	<i>broda</i>		5. no evidence for borrowing	0.00
11	Old High German	<i>bart</i>		5. no evidence for borrowing	0.00
12	Dutch	<i>baard</i>		5. no evidence for borrowing	0.00
13	English	<i>beard</i>		5. no evidence for borrowing	0.00
14	Kildin Saami	<i>seam</i>	СЯМ	5. no evidence for borrowing	0.00

# Mismatches

- but **there is no true homosemasy**: counterparts of lexical meanings are often inexact counterparts
- the relationship between the comparison meanings and words is not one-to-one, but many-to-many

# Mismatches

(1) one-to-one meaning-word relationships

comparison meaning	French word
'book'	<i>livre</i>
'father'	<i>père</i>
'house'	<i>maison</i>
'cat'	<i>chat</i>

(2) one-to-many meaning-word relationships

comparison meaning	German words
'bone'	<i>Knochen, Gräte</i>
'go'	<i>gehen, fahren</i>
'time'	<i>Zeit, Mal</i>

# Mismatches

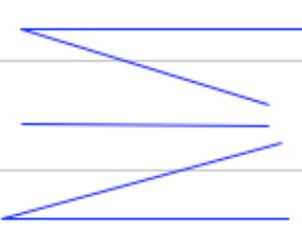
## (3) many-to-one meaning-word relationships

comparison meanings	German word
'go', 'drive'	<i>fahren</i>
'elect', 'choose', 'dial'	<i>wählen</i>
'pillow', 'cushion'	<i>Kissen</i>

- Note: Being a counterpart of several comparison meanings does **not** imply being **polysemous**.

## (4) many-to-many meaning-word relationships

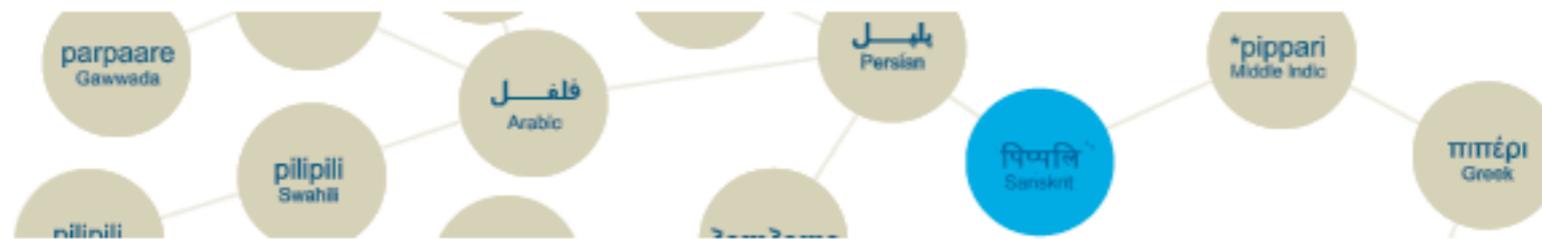
comparison meaning	German word
'go'	<i>gehen</i>
'drive'	<i>fahren</i>
'ride'	<i>reiten</i>



# World Loanword Database

A word that corresponds to two comparison meanings:

WORLD LOANWORD  
DATABASE (WOLD)



[Home](#)

[Vocabularies](#)

[Languages](#)

[Meanings](#)

[Authors](#)

[Newsblog](#)

[Contact](#)

[Help](#)

[Archi](#)

[description](#)

[abbreviations](#)

[citation](#)

[RDF+XML](#)

## *kavāra*

a word from Vocabulary [Archi](#)

**Word form:** *kavāra*

**LWT meaning(s):** [the paper](#)  
[the letter](#)

**Word meaning:** paper, letter

**Analyzability:** unanalyzable

**Age :** No information

## Loanword Information

**Borrowed status :**

1. clearly borrowed

**Source words:** *kavāz* 'paper, letter, document',  
from Persian

**Effect:**

Insertion

**Saliency:**

No information

**Contact**

**situation:**

Persian contact

# World Loanword Database

A comparison meaning with two counterpart words in a language:

## Meaning 4.142: the beard

Semantic field: [The body](#)

Semantic category: Noun

Borrowed score : 0.17

Age score : 0.83

Simplicity score : 0.93

### Counterpart words in the World Loanword Database

Voc. ID ↓	Vocabulary	Word Form	Original Script	Borrowed status	Borrowed score
1	Swahili	<i>ndevu</i>		5. no evidence for borrowing	0.00
2	Iraqw	<i>daamóot</i>		1. clearly borrowed	1.00
3	Gawwada	<i>puše</i>		5. no evidence for borrowing	0.00
4	Hausa	<i>géemùu</i>		5. no evidence for borrowing	0.00
5	Kanuri	<i>njìshì</i>		4. very little evidence for borrowing	0.25
6	Tarifiyt Berber	<i>aɣəšmia</i>		5. no evidence for borrowing	0.00
6	Tarifiyt Berber	<i>tmāt</i>		5. no evidence for borrowing	0.00
7	Seychelles Creole	<i>labarb</i>		5. no evidence for borrowing	0.00
8	Romanian	<i>barbă</i>		5. no evidence for borrowing	0.00
9	Selice Romani	<i>sakálla</i>		1. clearly borrowed	1.00
10	Lower Sorbian	<i>broda</i>		5. no evidence for borrowing	0.00
11	Old High German	<i>bart</i>		5. no evidence for borrowing	0.00
12	Dutch	<i>baard</i>		5. no evidence for borrowing	0.00
13	English	<i>beard</i>		5. no evidence for borrowing	0.00
14	Kildin Saami	<i>seam</i>	сям	5. no evidence for borrowing	0.00

# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

exact counterpart

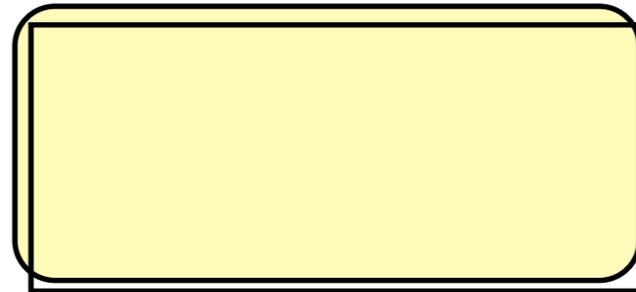


← comparison  
meaning

# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

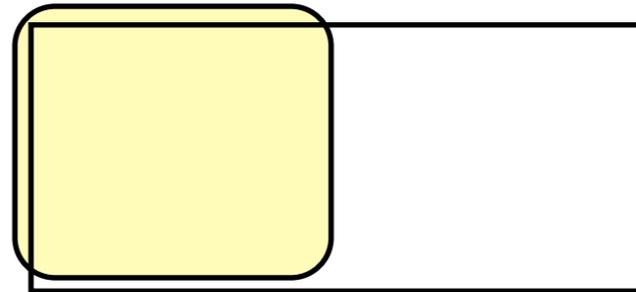
**exact counterpart**



# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

exact counterpart  
**sub-counterpart**



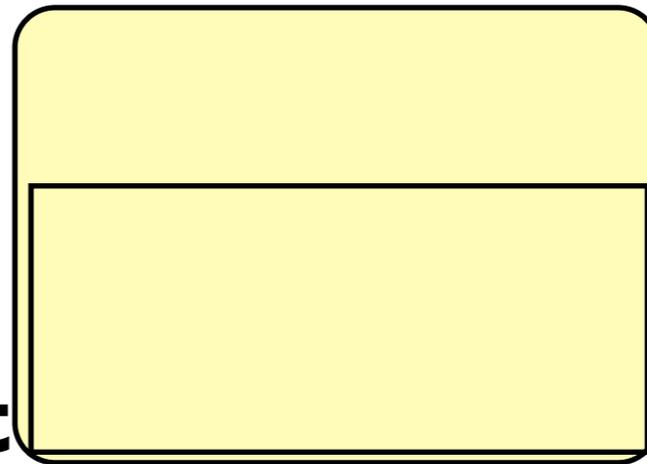
# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

exact counterpart

sub-counterpart

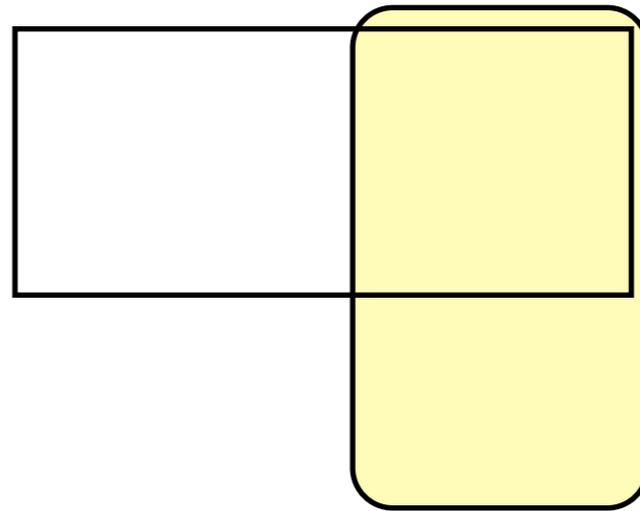
**super-counterpart**



# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

exact counterpart  
sub-counterpart  
super-counterpart  
**para-counterpart**



A final possibility is:

no counterpart

# More mismatches

Moreover, we added four kinds of relationships between meanings and words:

exact counterpart  
sub-counterpart  
super-counterpart  
para-counterpart



A final possibility is:

**no counterpart**

# How to compare words

- Thus, the comparative database does not assume that all languages have words corresponding to the meanings on the meanings list. Neither strict universalism nor low-level universalism is presupposed.
- The World Loanword Database is fully compatible with particularism: Different languages have words with different meanings, and it is not necessary for any two meanings to be fully identical across languages.

# How to compare words

- All that is required is the possibility of **matching** word meanings with **comparison meanings**, i.e. establishing word counterparts (of some sort) of the comparison meanings.
- How are the comparison meanings selected?
- Answer: Not really on a principled basis. Loanwords used IDS, IDS was based on Buck (1949), and Buck (1949) was based on Indo-European languages. Quite possibly, the list is strongly biased toward IE languages.

# How to compare words

- The Indo-European bias does not matter so much here, because the main purpose was to get a comparable representative sample of the words from different languages – the precise meanings did not play a crucial role.
- In general, comparison meanings are selected intuitively, on the basis of an initial hypothesis based on a few languages.

# Selecting comparison meanings

- Take Hjelmslev's example again:

English	German	Danish	French	Spanish
tree	Baum	træ	arbre	árbol
wood	Holz		bois	leña
woods	Wald	skov	forêt	madera
forrest				bosque
				selva

- Danish: just two meanings
- add German: three meanings
- add French: four meanings
- add Spanish: five meanings

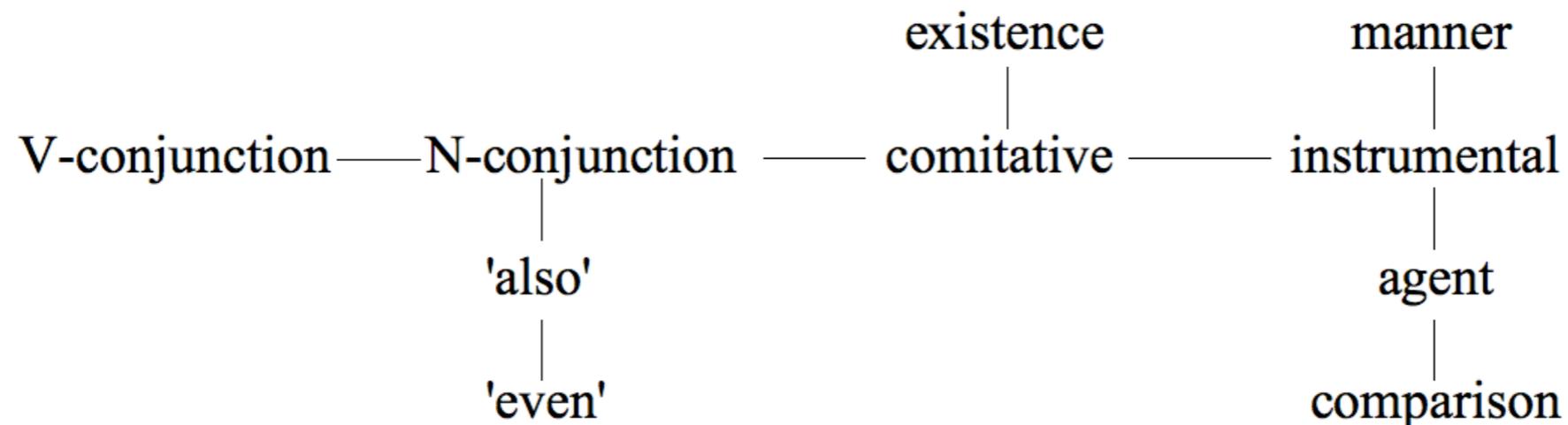
# Selecting comparison meanings

- This method may not seem satisfactory, but there is no alternative: To compare languages, we have to start somewhere – it would be hopeless to try to derive the concepts for comparison (**comparative concepts**) from extralinguistic considerations. The more languages go into the determination of the comparative concepts, the better.
- Gilbert Lazard (2005, 2006): The comparative concepts are "arbitrary conceptual frameworks": intuitive, but explicit and clearly formulated notions at the initial stage of research (but intuition plays no role at subsequent stages).

# Comparison via semantic maps

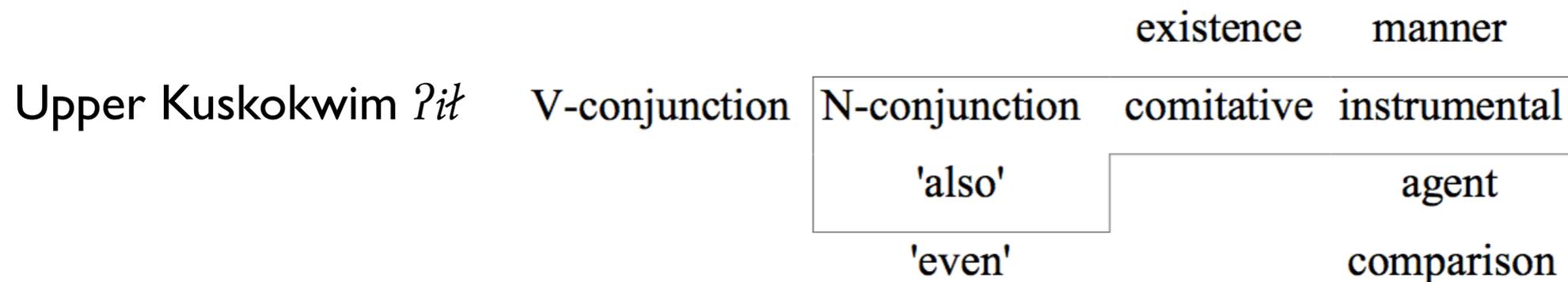
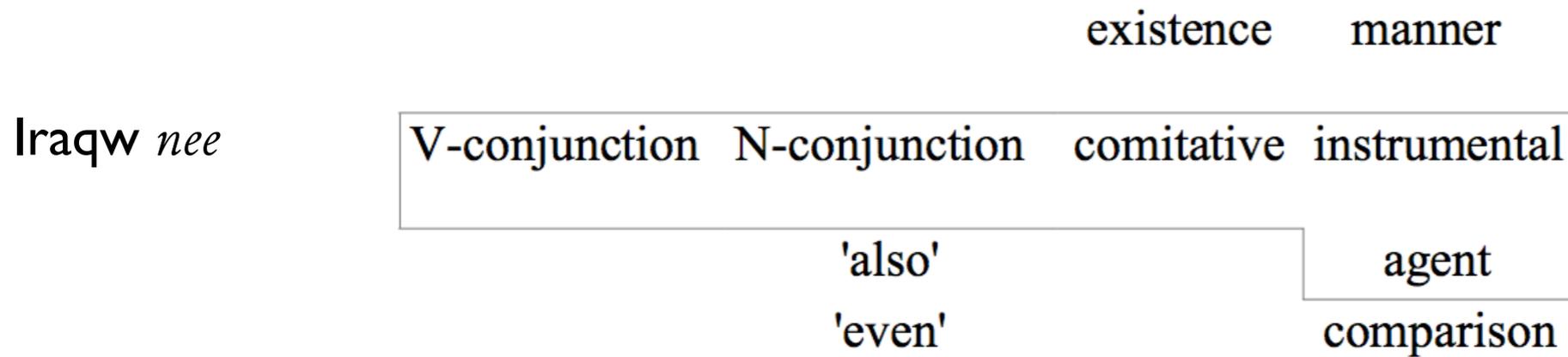
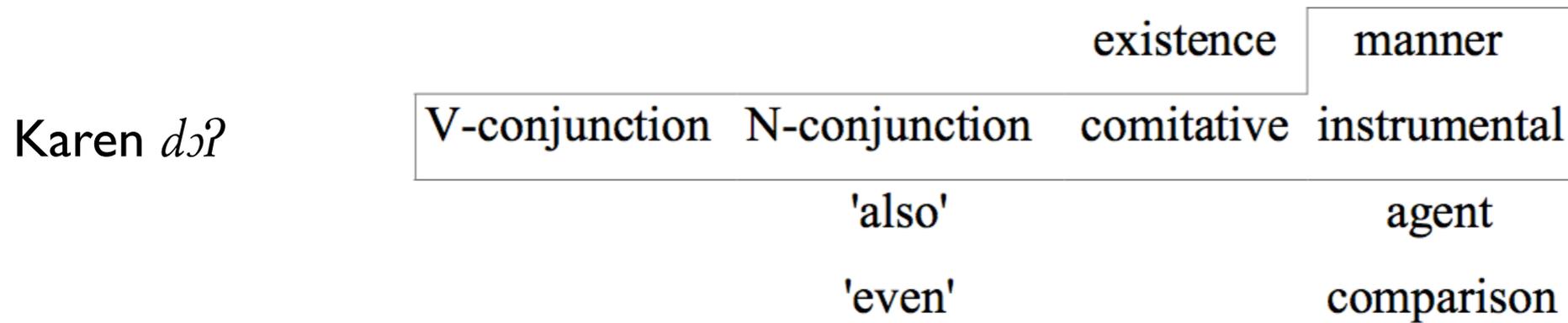
- Just like the comparison meanings in the Loanwords database, the nodes on a semantic map for comparison of grammatical meanings are **comparative concepts**.
- The nodes are **not** cross-linguistic concepts or universal concepts

An example map (Haspelmath 2004), concerning expressions meaning ‘and’, ‘with’, etc.:



# Comparison via semantic maps

Examples of marker distributions from various languages:



# Comparison via semantic maps

Examples of marker distributions from various languages:

			existence	manner
Dargi <i>-ra</i>	V-conjunction	N-conjunction	comitative	instrumental
		'also'		agent
		'even'		comparison

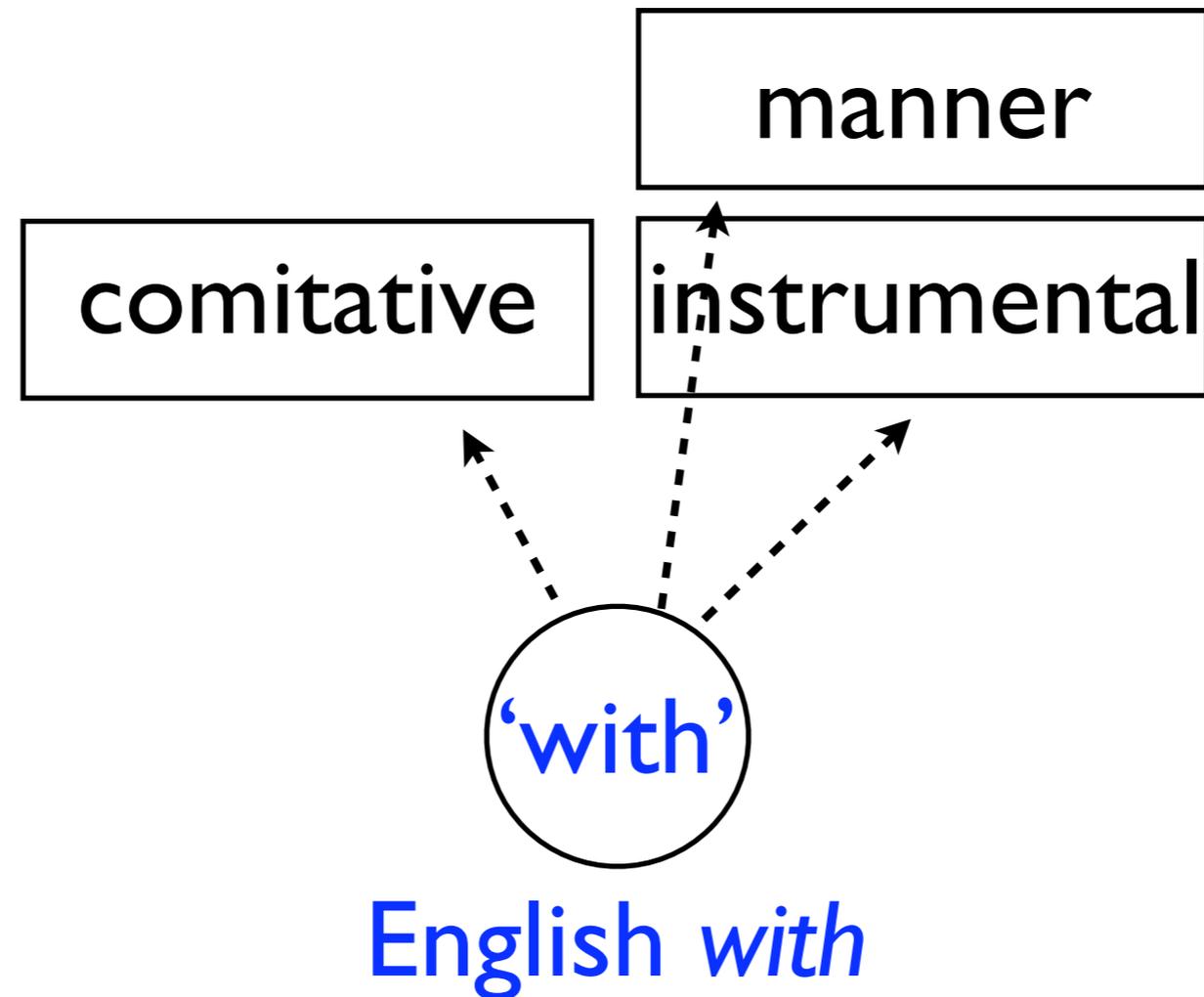
			existence	manner
Riau Indonesian <i>sama</i>	V-conjunction	N-conjunction	comitative	instrumental
		'also'		agent
		'even'		comparison

			existence	manner
English <i>with</i>	V-conjunction	N-conjunction	comitative	instrumental
		'also'		agent
		'even'		comparison

# Comparison via semantic maps

- Just like the comparison meanings in the Loanwords database, the nodes on a semantic map for comparison of grammatical meanings are **comparative concepts**, created by a linguist for the purposes of comparison.

# English *with*/'with'



The claim is that English *with* matches the three comparison meanings, not that it instantiates three cross-linguistic meanings.

# Semantic maps with universal concepts?

- Croft (2001:364): semantic maps show us "the geography of the human mind, which can be read in the facts of the world's languages in a way that the most advanced brain scanning techniques cannot ever offer us." (Cf. Croft's term "**conceptual space**")
- Thus, Croft implies that the nodes on the semantic map are cross-linguistic or universal concepts.

# Semantic maps with universal concepts?

- a letter from California:

“I am a graduate student in ... I'm working with ... and investigating the language and thought interface on the subject of causality.

As part of my investigation, I've been searching for a semantic map, akin to yours on indefinite pronoun use, for causative morphology and have so far come up empty-handed. My supervisor suggested that if one existed, you would likely know of it. Have you ever seen any such?

This is in service of **determining the conceptually primitive elements of causation**, as reflected in language, before comparing it to what is known about the psychology of causation. Any suggestions you had would be terrifically appreciated.

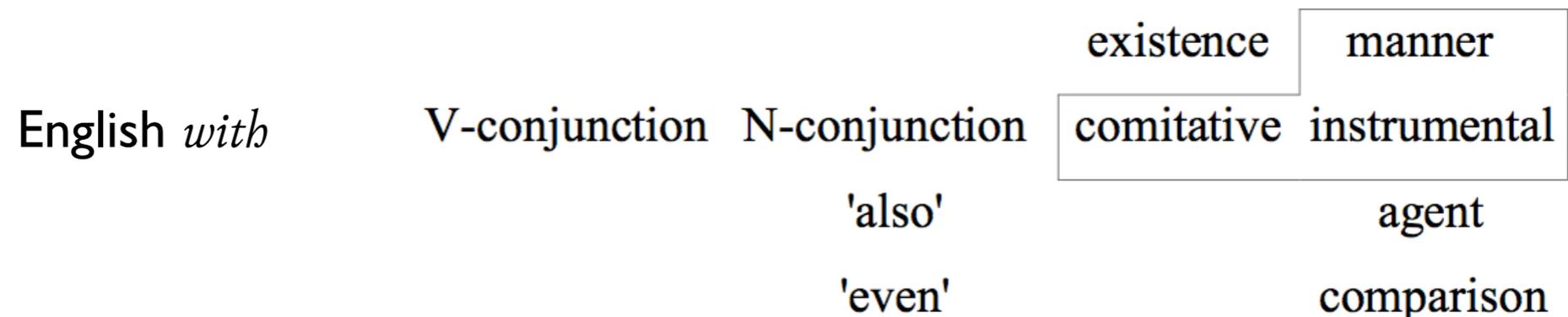
Thanks very much;  
....”

# Semantic maps with universal concepts?

- But the choice of nodes on a semantic map is necessarily **arbitrary** to some extent (as noted earlier).
- The meanings of a language-particular item can never be captured fully by describing its distribution over the nodes on the map.

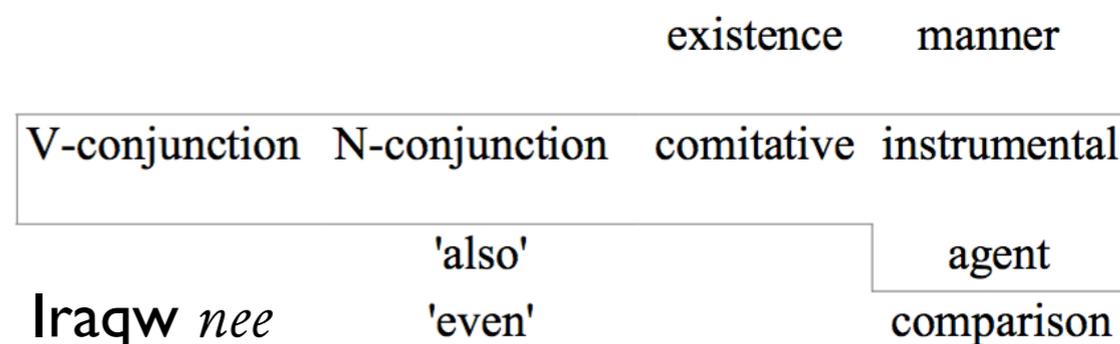
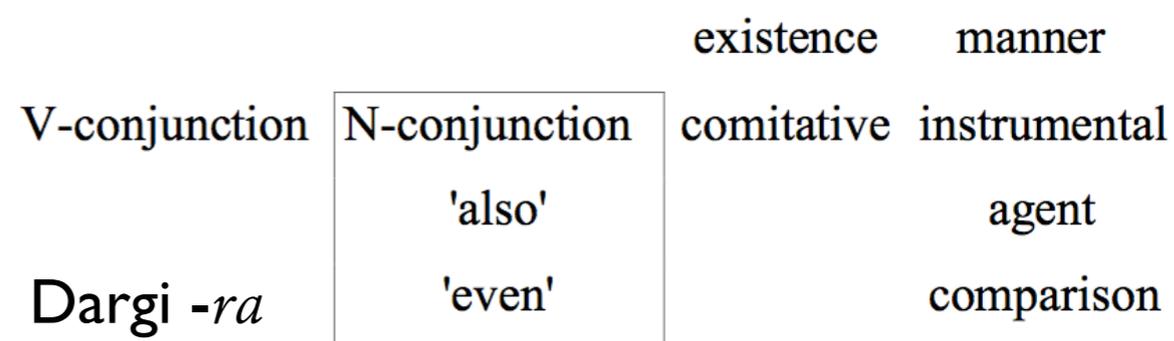
# Semantic maps with comparative concepts

- For example, the meaning of English *with* cannot be captured by saying that it expresses ‘comitative’, ‘instrumental’ and ‘manner’ – its properties are **much** richer.
- Moreover, within English, comitative and instrumental readings, and instrumental and manner readings cannot be strictly separated. The word *with* may well be **monosemous**.
- Again, as with lexical meanings: Being a counterpart of several nodes does **not** imply being **polysemous**.



# Semantic maps with comparative concepts

- Dargi *-ra* is almost certainly monosemous – ‘and’, ‘also’ and ‘even’ are semantically very similar and can readily be described as contextual variants of each other.
- Even more clearly, the two nodes “V-conjunction” and “N-conjunction” are not semantically distinct, but only syntactically.



- Thus, **nodes on semantic maps cannot be equated with universal concepts.**

# Etic vs. emic notions

- Comparative concepts are **etic** notions (not part of the language system)
- Language-particular meanings are **emic** notions, part of the language system
- (Etic vs. emic: introduced by Kenneth L. Pike, 1912-2000)

# Etic vs. emic notions

- Back to Kayardild sibling terms:

	♂ referent		♀ referent	
	♂ speaker	♀ speaker	♂ speaker	♀ speaker
(elder)	thabuju	kularrind	kularrind	yakukathu
(younger)	duujind	kularrind	kularrind	duujind

Evans (2011:510): “A disadvantage of concentrating on the etic is that it overlooks obvious elegances of characterization that appear once one gives emic formulations. ... Focusing on Kayardild *kularrind*, an etic characterization can merely note that it occurs in four cells, as shown. But this overlooks the more elegant characterization that can be given emically, namely, that it means ‘opposite sex sibling’ (i.e. brother of a female or sister of a male).”

# Etic vs. emic notions

- More generally, very often the meaning of a word does not lend itself to characterization in terms of etic notions.
- Each word is a unique way of cutting up human experience.
- To describe the meaning of a word, a language-particular perspective must be adopted.

# Comparative concepts vs. descriptive categories

- Quite generally, **comparative concepts** should not be confused with **descriptive categories** of particular languages (Haspelmath 2010, in *Language*).
- **Comparative concepts** are created by comparative linguists for the purposes of comparing languages, as a kind of artificial auxiliary language.
- **Descriptive categories** are the categories of the languages themselves.
- Speakers and non-comparativists do not need comparative concepts.

# Comparative concepts vs. descriptive categories

- Comparative concepts are often confused with cross-linguistic/universal categories of the generative type.
- “Cross-linguistic/universal categories” are used both for language-particular description and for comparison.
- But this doesn't work:
  - Languages can only be described sensibly with language-specific categories
  - Comparison cannot be based on the diverse descriptive categories

# Comparative concepts in grammar, based on meaning and form

- Haspelmath 2007 (“Pre-established categories don’t exist”), following Stassen (1985) and Croft (2003):
  - Morphosyntactic comparison must be meaning-based (“For morphosyntactic comparison to be possible, we must hold the meaning constant – at least this must be universal.”)
  - Thus, there is no comparison without **homosemasy**
- But form sometimes plays a crucial role in defining comparative concepts (Newmeyer 2007, Rijkhoff 2009)
- e.g. ergative case = marker of transitive agent when the intransitive subject is marked in a different form

# Confusion between comparative concepts and descriptive categories

- Haspelmath 1995 (“The converb as a cross-linguistically valid category”)
- converb = “a nonfinite verb form whose main function is to mark adverbial subordination”
- converb: a term coined by Gustav John Ramstedt in Helsinki in 1903 (for Mongolian)

Modern Greek

*I kopéla ton kítakse xamojel-óndas.*  
the girl him looked [smile-CONV]  
‘The girl looked at him, smiling.’

Hungarian (de Groot 1995:292)

*János hallgat-va a zené-t tanul-t.*  
János [listen-CONV the music-ACC] study-PAST  
‘János studied listening to the music.’

# Confusion between comparative concepts and descriptive categories

- Problems with “converb” for describing languages:
  - some languages have forms that can be used both adnominally and adverbially, e.g. Italian:
    - a. *Arrivato a casa presto, Giorgio si mise a leggere il giornale.*  
'Having got home early, Giorgio started reading the newspaper.'
    - b. *L'uomo arrivato ieri è mio fratello.*  
'The man who arrived yesterday is my friend.'
  - some languages have forms that can be used both adverbially and as complement clauses, e.g. Lezgian:
    - a. *Am ajwandi-k gazet k'el-iz aqwaz-nawa.*  
he:ABS balcony-SUB [newspaper read-CONV] stand-PERF  
'He is standing on the balcony, reading a newspaper.'
    - b. *Ada-z ktab k'el-iz k'an-zawa.*  
he-DAT [book read-INF] want-prs  
'He wants to read a book.'

# Confusion between comparative concepts and descriptive categories

- Problems with “converb” for describing languages:
  - some languages have forms that can be used both adverbially and for coordination, e.g. Turkish:
    - a. *Otur-up mu konuş-tu-lar?*  
sit-CONV INTERROG speak-PST-3PL  
'Did they speak (while they were) sitting?'
    - b. *Gel-ip gör-sün mü?*  
[come-AND] see-OPT.3SG INTERROG  
'Shall he come here and see?' (not: 'Having come here, shall he see?')

We do not want to say that in all these cases, there is homonymy or even polysemy.

# Confusion between comparative concepts and descriptive categories

- Problems with “converb” as a comparative concept (or as a cross-linguistic category):
  - neither “nonfinite” nor “adverbial” are well-defined notions in general terms
  - “nonfinite” means ‘without tense and subject person-number markers’, but many languages lack tense and person-number markers to begin with
  - “adverbial subordinate” seems to mean simply ‘subordinate but not relative clause or complement clause’

# Solution:

Separation between descriptive categories  
and comparative concepts

- Descriptive categories are just applied at the language-particular level (and capitalized, e.g. “the Hungarian Converb”)
- Comparative concepts are just applied at the cross-linguistic level – languages are described in their own terms
- In this perspective, language description is not dependent on language comparison, and vice versa
- Both disciplines are conceptually independent, but they need the other’s research results

# Conclusion

- Homosemasy is of course indispensable for comparative semantics and morphosyntax
- But languages differ semantically (just as they differ in form) without definite limits, so comparison cannot be based on universal/cross-linguistic meanings (or forms)
- Thus, comparison must be via **comparative concepts**
- Homosemasy does not mean that two meanings of different languages can be **equated**, only that they match the same comparison meaning.

# Conclusion

- Comparison must be via comparative concepts, as in (a), not as in (b)

