

Loanword Typology

**Investigating lexical borrowability in the
world's languages**

Martin Haspelmath & Uri Tadmor
(coordinators)

Goal:

Find **universals of lexical borrowing**, to allow **predictions** about the kinds of words that are borrowed, based on what we know about the sociohistorical circumstances of contact.

Why?

Because we are interested in the **prehistory** of human populations, and hypotheses about **genealogical relationships** and protolanguages are often confounded by lexical borrowing.

Methods

- We study lexical borrowing patterns in a **world-wide sample** of languages (hence: “typology”)
- We examine a fairly **large number of lexical meanings** (1460 meanings), i.e. not just “basic vocabulary”.

More specific goals

- Assess the “**borrowability**” of different (classes of) lexical meanings, i.e. the likelihood that a given meaning is expressed by a borrowed word
- Go beyond a vague concept of “**basic vocabulary**”, often defined in a purely extensional way (e.g. “the words on the Swadesh list”)

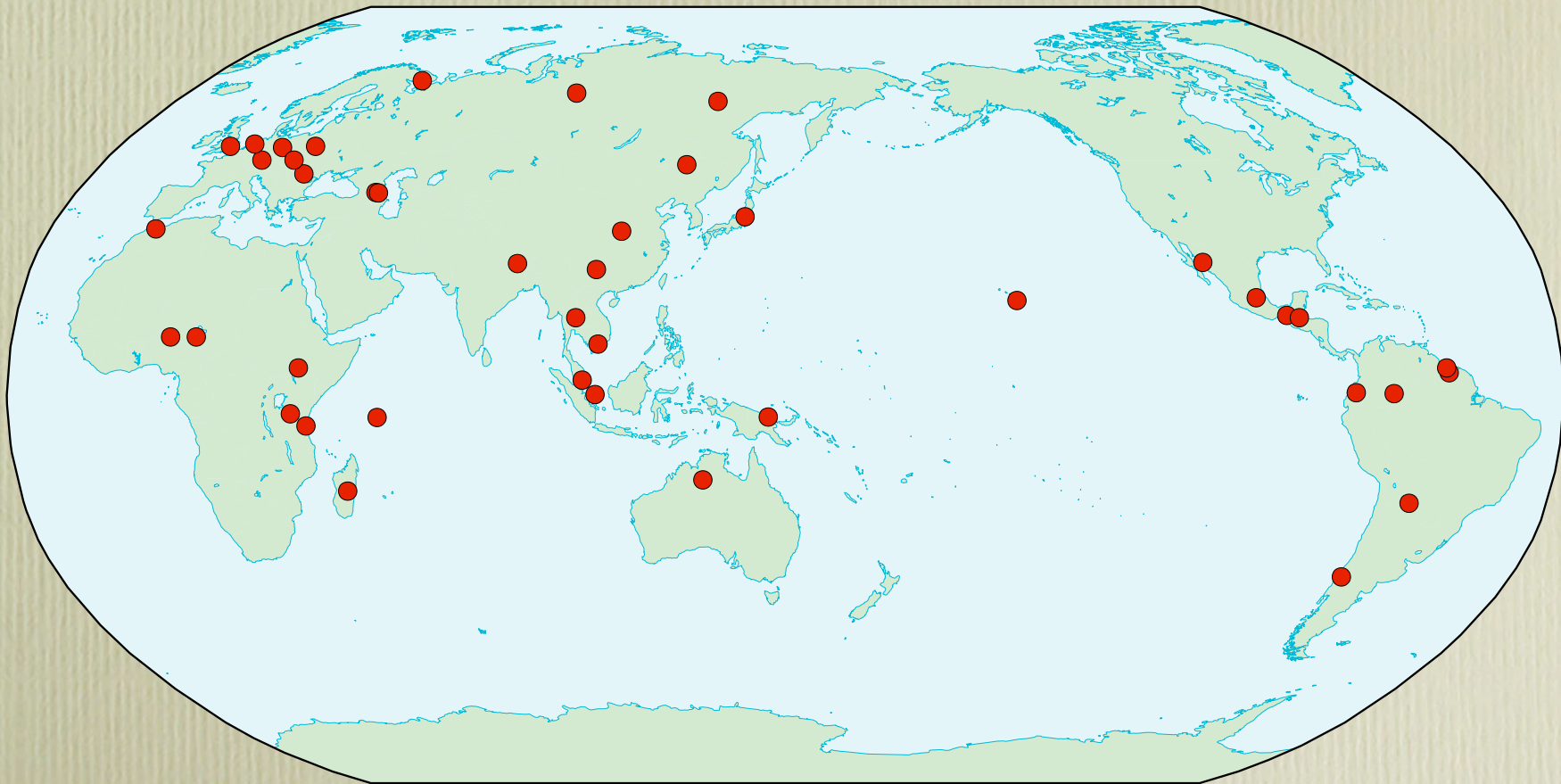
Implementation

- Data for each language are contributed by a **language expert** (or team of experts) who knows the language and (what is known about) its history well.
- The language experts provide **extensive lexicographic data** on the meanings of the sample. They give the counterparts for each meaning in their language, say whether it is a loanword, and if so, what its source is.

Motivating the contributors

- They get a **publication**, or rather two publications:
 - lexical dataset, as part of the electronically published **Loanword Typology lexical database**
 - prose chapter in a **book** publication
 - (Haspelmath, Martin & Tadmor, Uri (eds.) 2009. *Lexical borrowing in cross-linguistic perspective*. 2 vols. Berlin: Mouton de Gruyter.)
- They got the chance to meet at **workshops** and discuss the set-up of the project.

Loanword Typology project languages



Languages from Africa and northern Eurasia

Afro-Asiatic	Cushitic	Iraqw
		Gawwada
	Berber	Tarifit
	Chadic	Hausa
Niger-Congo	Bantoid	Swahili
Nilo-Saharan	Saharan	Kanuri
Finno-Ugrian	Saami	Kildin Saami
Nakh- Daghestanian	Lezgic	Archi
	Tsezic	Bezhta
Altaic	Turkic	Sakha (= Yakut)
	Tungusic	Oroqen

Languages from Europe

Indo-European	Germanic	Yiddish
		Old High German
		English
		Dutch
	Romance	Romanian
		Seychelles Creole
	Slavonic	Lower Sorbian
	Indic	Romani

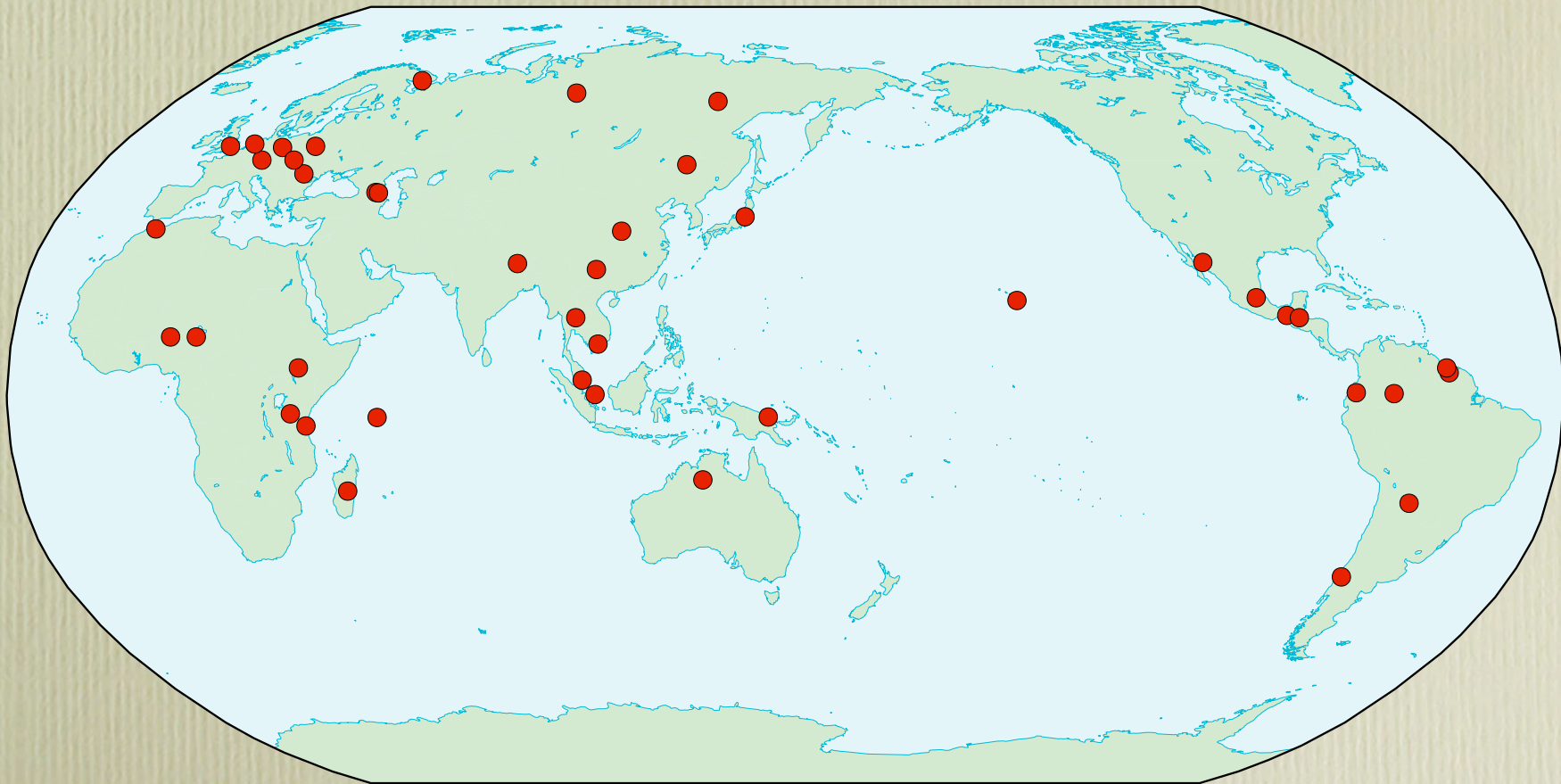
Languages from Asia and Oceania

Sino-Tibetan	Himalayish	Manange
	Sinitic	Mandarin Chinese
Hmong-Mien		White Hmong
Mon-Khmer	Vietic	Vietnamese
	Aslian	Ceq Wong
Tai		Thai
Japanese		Japanese
Austronesian	Western Malayo-Polynesian	Indonesian
		Malagasy
	Oceanic	Takia
		Hawaiian
Pama-Nyungan		Gurindji

Languages from the Americas

Uto-Aztecan		Yaqui
Mayan		Tzotzil
		Q'eqchii'
Oto-Manguean		Otomí
Cariban		Kali'na
Quechua		Imbabura Quechua
Nadahup		Hup
Matacoan		Wichí
Araucanian		Mapudungun

Loanword Typology project languages



Contributors

MPI-EVA members	Susanne Michaelis & Katrin Muhme, Bernard Comrie & Madjid Khalilov, Brigitte Pakendorf & Innokentij Novgorodov, Uri Tadmor, Søren Wichmann, Jeff Good, Patience Epps
linguists from Leipzig and surroundings	Ari Awagana, Doris Löhr, Ekkehard Wolff, Rosemarie Lühr, Hauke Bartels, Michael Rießler, Kristine A. Hildebrandt, Thekla Wiebusch
regular MPI-EVA visiting scientists	Mauro Tosco, Edward Vajda, Malcolm Ross, Patrick McConvell, Zarina Estrada, Alejandra Vidal, Lucía Golluscio
others	Thilo Schadeberg, Maarten Mous, Maarten Kossmann, Anthony Grant, Moshe Taube, Nicoline van der Sijs, Viktor Elsík, Marina Chumakina, Li Fengxiang & Lindsay Whaley, Titima Suthiwan, Mark Alves, Martha Ratliff, Christopher K. Schmidt, Nicole Kruspe, Sander Adelaar, 'Oiwi Parker Jones, Cecil Brown, Ewald Hekking & Dik Bakker, Jorge A. Gómez Rendón, Odile Renault-Lescure

The lexical questionnaire


- List from IDS (**Intercontinental Dictionary Series**, Carl Darling Buck/Mary Ritchie Key), augmented and adapted
- **1460** lexical meanings
- Original plan: one-to-one relationship between meanings and words:

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

The LWT Database

- Also allows many-to-many relationships:

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



- Includes information not only on loanword status and loanword source, but also on analyzability, age, and other word properties
- Implemented in FileMaker Pro and distributed to the contributors

Most loanword-friendly languages (top 7)

	<i>total words</i>	<i>not borrowed</i>	<i>borrowed</i>	<i>% borrowed</i>
1. Selice Romani	1538	541	930	60
2. Tarifiyt Berber	1690	758	818	48
3. Romanian	2113	1162	839	40
4. English	1504	838	583	39
5. Saramaccan	1039	490	353	34
6. Ceq Wong	900	532	284	32
7. Imbabura Quechua	1243	839	373	30

Most loanword-friendly languages (next 7)

	<i>total words</i>	<i>not borrowed</i>	<i>borrowed</i>	<i>% borrowed</i>
8. Japanese	1858	1167	536	29
9. Indonesian	1869	1177	550	29
10. Sakha	1416	911	372	26
11. Yaqui	1391	1012	368	26
12. Takia	1126	309	285	25
13. Swahili	1773	1207	444	25
14. Mapudungun	1262	950	275	22

Most loanword-resistant languages (bottom 7)

	<i>total words</i>	<i>not borrowed</i>	<i>borrowed</i>	<i>% borrowed</i>
28. Hup	995	865	105	11
29. Gawwada	993	785	98	10
30. Gurindji	1031	282	93	9
31. Seychelles Creole	2016	1788	169	8
32. Manange	1009	883	72	7
33. Ket	965	872	71	7
34. Mandarin Chinese	2115	1885	25	1

Most borrowable word meanings

[LÅNORD]

Känguru mest lånat

De lånord som etablerar sig i ett språk är inte vilka ord som helst. *Känguru*, *tidning* och *vin* är alla bland tio-i-topp. Substantiv är den ordklass som lånas mest, tillsammans med andra innehållsord.

Prepositioner, konjunktioner och andra så kallade funktionsord lånas nästan inte alls.

Sällsynta ord lånas mer än vanliga ord. Basordförrådet, det vill säga ord för grundläggande mänskliga aktiviteter och behov, lånas minst. Till basförrådet hör också ord för kroppsdelar och familjeförhållanden. Vanliga verb som *se* och *höra* är ovanliga lån.

– De vanligaste lånen är av västerländskt ursprung,



MEST LÅNADE ORD

1. känguru
2. kaffe
3. te
4. vin
5. tidning
6. siden
7. kamel
8. oliv
9. socker
10. soldat

MINST LÅNADE ORD

1. att gå
2. du
3. i går
4. svart
5. rygg
6. näsa
7. tunga
8. att döda
9. revben
10. ögonfrans



berättar professor **Martin Haspelmath** som leder ett stort internationellt projekt om lånord vid Max Planck-institutet i Leipzig.

– Det är inte så förvånande. Med ny teknik kommer också nya termer.

I Martin Haspelmaths projekt ingår ett fyrtiotal språk över hela världen. Resultaten är fortfarande preliminära men tyder på att 70 procent av orden

Most borrowable word meanings

<i>no.</i>	<i>meanings</i>	<i>lgs</i>	<i>words</i>	<i>loans</i>	<i>% loans</i>
3,88	the kangaroo	14	14	14	1
5,78	the olive	18	19	19	1
23,185	the motor	27	28,5	27,5	0,96
3,78	the camel	22	22	21	0,95
23,91	the coffee	28	31	29	0,93
23,11	the television	25	28	26	0,92
23,44	the bank (financial)	25	28	26	0,92
6,25	the silk	23	27	25	0,92
23,57	the plastic	24	24	22	0,91
6,95	the soap	30	31	28	0,90

Least borrowable word meanings

Meanings with *no* or at most one loanword counterpart:

he, she, it, he/she/it, you (singular), we, we (inclusive), we (exclusive), in, where?, which?, why?, this, behind, up, what?, how?, who?, that, not, here, there,

the married woman, the younger sister, the stepmother, the older sibling, the brother, the younger brother, the grandson, the married man, the granddaughter,

the mouth, the heel, the udder, the tooth, the breast, the nose, the molar tooth,

the fire, the yolk, the men's house, the body louse, the digging stick (=yamstick), the head louse, the thatch, the fly, the night, the nit, the itch,

three times, the day after tomorrow, the day before yesterday, next, yesterday, in front of,

wide, slow, black, raw, stinking, loud, bitter, alone, drunk

to light, to hear, to listen, to howl, to fart, to be hungry, to milk, to spin, to break, to blow, to carry under the arm, to bring, to stand, to be thirsty, to show, to lie down, to run, to make, to play, to carry, to go, to flee, to drop, to fall, to chop, to ask(2), to walk, to follow, to rise, to grasp, to find, to drip, to hunt, to climb, to meet, to have, to float, to carry on shoulder, to lose, to cut down, to chew, to go out, to go or return home, to trap, to hollow out, to spit

Comparison with the Swadesh list as a list of hard-to-borrow meanings

- As expected, many of the meanings on the Swadesh list are indeed not readily borrowed:
 - While the (simple) words of the Loanword Typology database (all 1460 meanings) have a **34%** chance of being borrowed,
 - the words corresponding to meanings on the Swadesh list (211 meanings) only have a **14.5%** chance of being borrowed.

Comparison with the Swadesh-200 list (1)

(the 13 most **borrowing-resistant** Swadesh meanings)

<i>no.</i>	<i>meaning</i>	<i>lgs</i>	<i>simple words</i>	<i>borrowed</i>	<i>% borrowed</i>
2,92	you (singular)	36	37	0	0
24,06	wide	36	33	0	0
10,38	to blow	35	29,5	0	0
15,65	the fire	36	28,4	0	0
24,07	this	35	25	0	0
2,93	he/she/it	32	24,5	0	0
12,15	to stand	35	24,3	0	0
17,66	where?	36	22	0	0
15,41	to hear	36	20,6	0	0
12,012	in	31	20,3	0	0
2,94	we	30	19,8	0	0
16,26	the body louse	31	18,8	0	0
8,28	the digging stick	21	5,7	0	0

Comparison with the Swadesh-200 list (2)

(the 13 most **borrowing-prone** Swadesh meanings)

	<i>meaning</i>	<i>lgs</i>	<i>simple words</i>	<i>borrowed</i>	<i>% borrowed</i>
17,52	because	35	26	15,5	0,60
1,32	the sea	34	25	13,2	0,53
3,11	the animal	36	27,2	14,5	0,53
1,1	the world	34	25,6	12	0,47
13,107	to count	36	21	8,5	0,40
5,71	the fruit	34	25,3	10	0,39
8,6	the tree	36	24,5	8,8	0,36
10,71	the road	35	30	10,5	0,35
17,13	to think	34	21,5	7,5	0,35
1,77	the ice	36	30,8	10	0,32
16,73	right ('correct')	35	24,8	8	0,32
1,53	the moon	36	31,7	9,8	0,31
1,76	the snow	29	22,3	7	0,31

Towards a Leipzig List of borrowing-proof meanings

- The 200 most borrowing-resistant meanings in the Loanword Typology (LWT) database are significantly less borrowable than the Swadesh-list meanings:
 - On average, they have a **5%** chance of being borrowed;
 - 42 word meanings are never borrowed in the LWT database;
 - Item #200 on the Leipzig List has a **10%** chance of being borrowed, still lower than the Swadesh average.
 - Only 78 words are both on the Swadesh list and on the Leipzig List

Loanwords by semantic field

<i>semantic field</i>	<i>% borrowed</i>	<i>semantic field</i>	<i>% borrowed</i>
Modern world	0,70	Cognition	0,28
Religion and beliefs	0,43	Speech and language	0,28
The house	0,42	Quantity	0,25
Clothing and grooming	0,41	Time	0,25
Law	0,41	Motion	0,25
Food and drink	0,39	Emotions and values	0,24
Agriculture and vegetation	0,37	Physical world	0,22
Social and political relations	0,35	Spatial relations	0,20
Warfare and hunting	0,34	The body	0,15
Animals	0,34	Kinship	0,15
Basic actions and technology	0,30	Sense perception	0,14
Possession	0,30	Function words	0,11

Thank you