SPCL meeting, Tampere

20 June 2017

Asymmetry in path coding: Creole data support a universal trend

SUSANNE MARIA MICHAELIS

Universität Leipzig & Max Planck Institute for the Science of Human History (Jena)

susanne.michaelis@uni-leipzig.de

1. Introduction

- **Universal trend:**

In all languages, MOTION-FROM constructions ('I come **from** Leipzig') are **longer** than MOTION-TO constructions ('I go **to** Leipzig'), or at least not shorter. (Michaelis 2013: 323)

(Length refers to the flagging of the reference object 'Leipzig' via adposition or case 'from'/'to')

- **Explanation**

MOTION-FROM constructions are universally much **rarer** in discourse than MOTION-TO constructions.

In general, rarer constructions/forms in a grammatical opposition tend to be longer than their more frequent counterparts (Haspelmath 2008), e.g. singular vs. plural, present tense vs. future tense (s. already Greenberg 1966).

The reason for this is a general **efficiency principle**: More **surprising** meanings need more coding than **predictable/expected** meanings.

Since MOTION-FROM meanings are rarer in discourse, they are more **surprising** and need more coding than MOTION-TO meanings, which are more frequent and therefore more **predictable/expected** meanings and therefore coded with less segmental material.

—> *from* vs. *to*

- **Diachrony**

Such functional-adaptive explanations have a diachronic component (Bybee 1988): Since the current system is often rigidly conventional, the adaptive forces must have been active in earlier diachronic change.

The different diachronic stages are often difficult to reconstruct for languages with a long history.

Most grammatical change processes often stretch over long periods of time and are hard to document, even in languages where we have good historical records.

BUT: **Creoles are a good test case** for tracing functional-adaptive change processes as some of these contact languages have evolved within a short time frame of only a few generations to maximally 500 years.

During language change processes, the asymmetric coding patterns must have fossilized due to the different frequency patterns.

In this way, languages are **functionally adapted** to speakers and hearers.

The mechanism is frequency-based language change (s. also Stolz et al. 2014: 307ff.).

https://research.uni-leipzig.de/unicodas/grammatical-universals/

Nikolai-Lab Leipzig

ERC Project "Grammatical Universals" at Leipzig University, plus guests (Nikolaistrasse 10)

Grammatical Universals

Usage-based explanation of universal coding asymmetries in grammar

An ERC-funded project (officially called "Form-frequency correspondences in grammar", or FormGram), headed by Martin Haspelmath (2015-2020). For more information on the project, see this page: Grammatical Universals Project.

2. MOTION-TO/MOTION-FROM

Path/ orientation	**MOTION-TO**	**MOTION-FROM**
	allative / GOAL	**a**blative / SOURCE
English	***to*** *Leipzig*	***from*** *Leipzig*
French	***à*** *Leipzig*	***de*** *Leipzig*
German	***nach*** *Leipzig*	***von*** *Leipzig*
Seychelles Creole	*ø* *Leipzig*	*ø* *Leipzig*

- intransitive constructions with a telic motion event

• different local referents may require different constructions even within a single language (cf. Stolz et al. 2014):

I go **ø** home
I go **to** London
I go **to the** hospital/**to** hospital

Je vais **à** Paris
Je vais **chez** Sibylle
Je vais **au** consulat (**à+le**)

Ich fahre **nach** Aachen
Ich gehe **zur** Schule (**zu+der**)
Ich gehe **zu** Oma

– named places
– frequent places
– humans

3. Some MOTION-TO/MOTION-FROM correspondences

Flagging of the reference object	language	MOTION-TO (named place)	MOTION-FROM (named place)
equally long (same flagging)	Saramaccan	***a***	***a***
equally long (different flagging, same number of segments)	Norvegian	***til***	***fra***
FROM longer < TO marker plus another marker	Rumanian	***la***	***de la***
	Cebuano	***sa***	***gikan sa***
FROM longer: independent marker	Lithuanian	***į***	***iš***
	Indonesian	***ke***	***dari***
	Hawaiian	***i***	***mai***
	Hmong	***rau***	***los ntawm***
	Yoruba	***si***	***lati***
FROM longer, TO zero	Swahili	***ø***	***kutoka***
	Hausa	***ø***	***daga***
potential counter example	Arabic	*'ilaa*	*min*
counter example	Maori	***ki***	***i***

- cross-linguistic data from 5 families (named places):

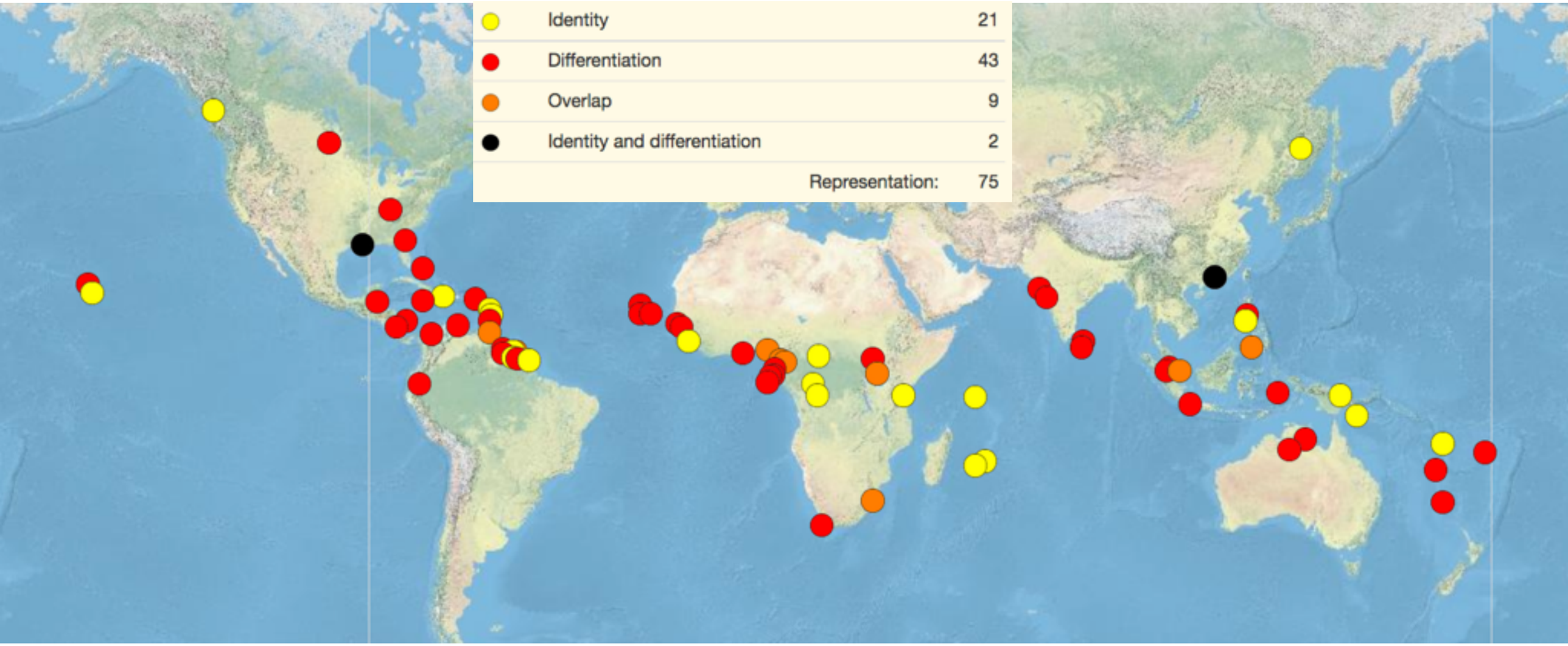
- universal tendency:

> MOTION-FROM is coded equally long as or longer than MOTION-TO.
>
> (ecxept for Arabic and Maori)

- s. also typological data in Stolz et al. 2014 (116 languages) regarding zero marking in path constructions: if zero-marked MOTION-FROM, then also zero-marked MOTION-TO (2014: 280f.).

4. Motion-to and motion-from (Michaelis & *APiCS* Consortium 2013)

- local reference object: frequent goals/sources, e.g. woods, beach, church, market



(1) Krio (English-based; Finney 2013)

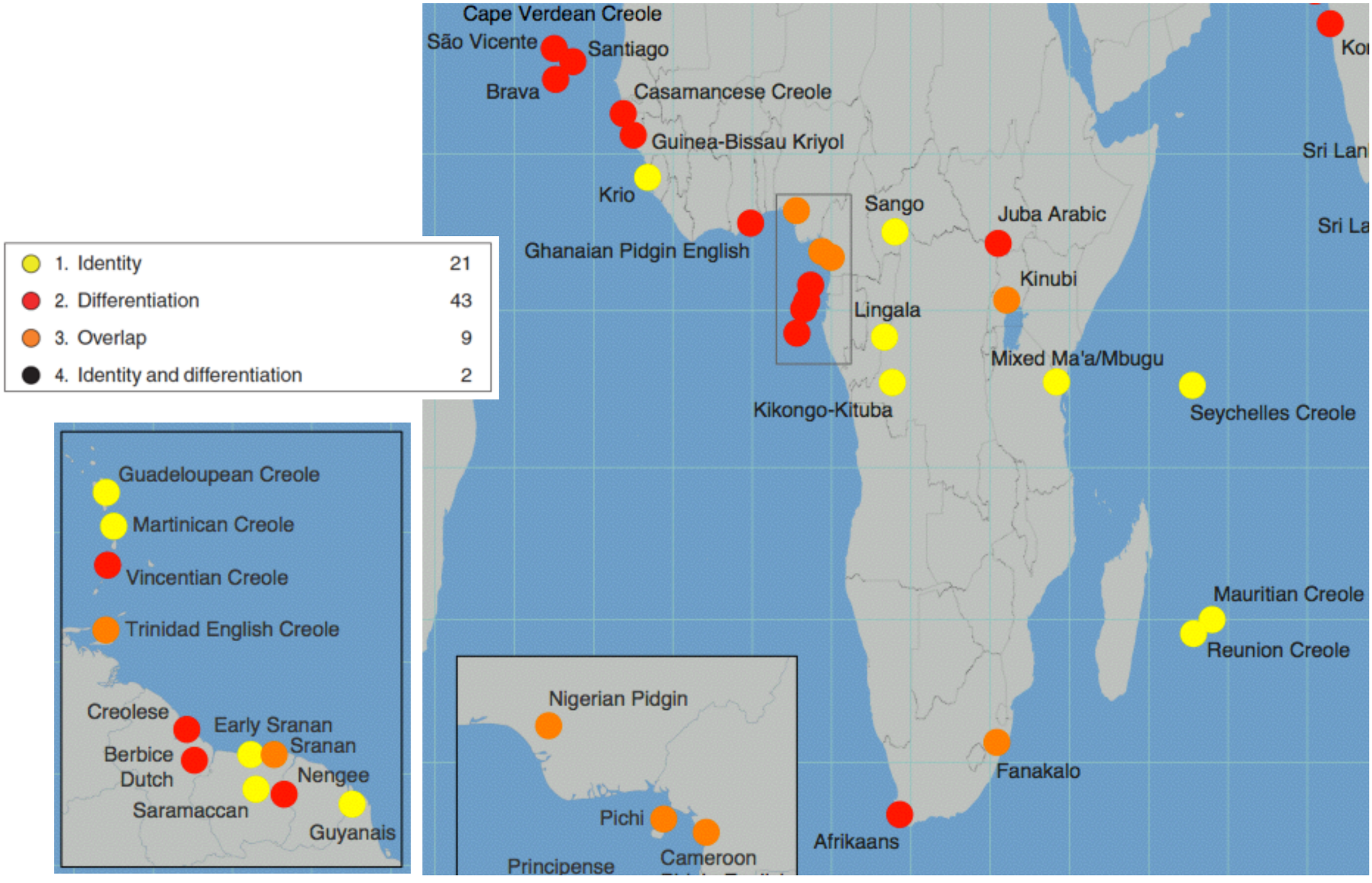
MOTION-TO

a. *a di go* ***na*** *di makit*
1SG PROG go LOC ART market
'I am going to the market.'

MOTION-FROM

b. *a jɛs kɔmɔt* ***na*** *di makit*
1SG just come LOC ART market
'I just came back from the market.'

Motion-to and motion-from (Michaelis & *APiCS* Consortium 2013)



Motion-to and motion-from (Michaelis & *APiCS* Consortium 2013)

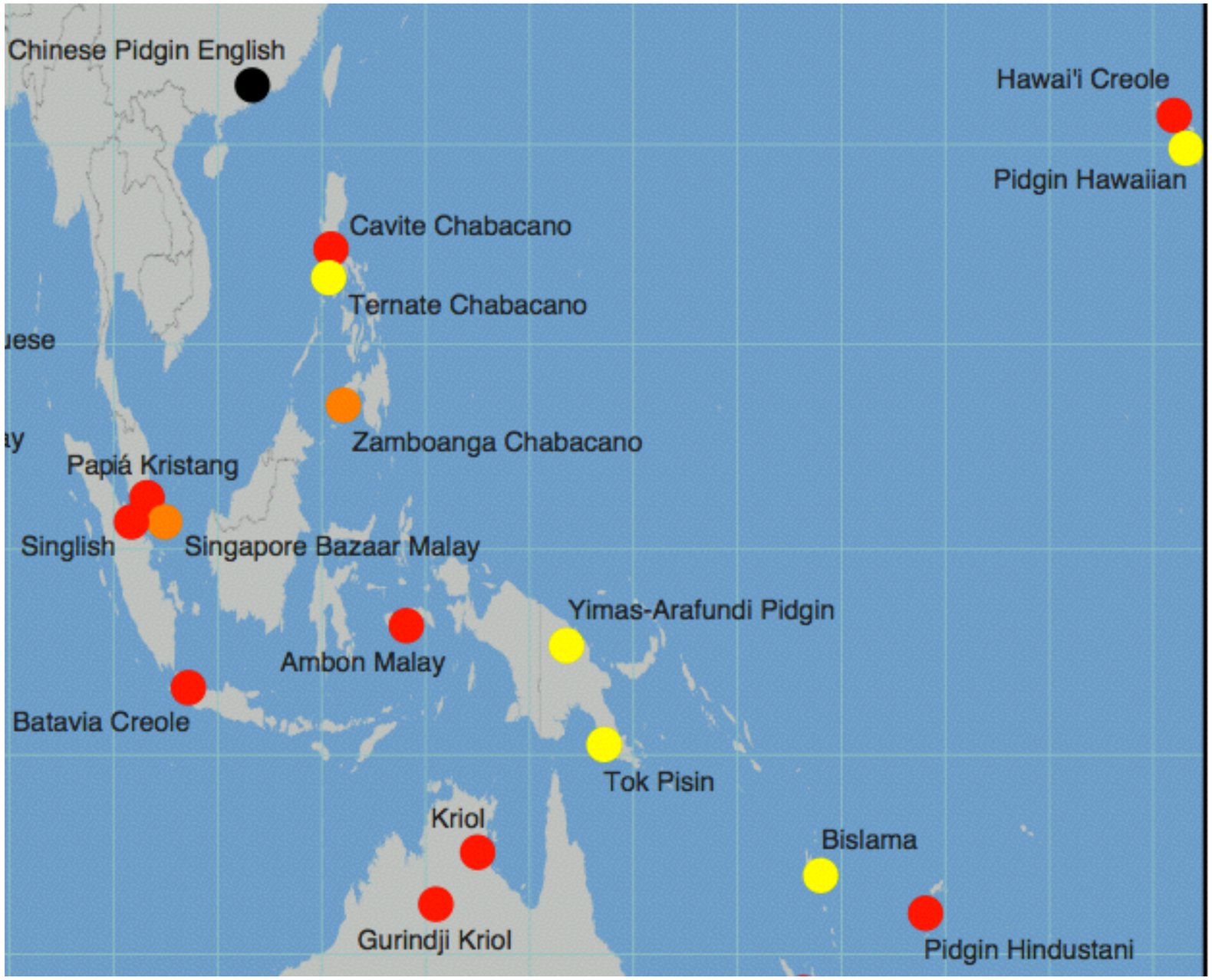
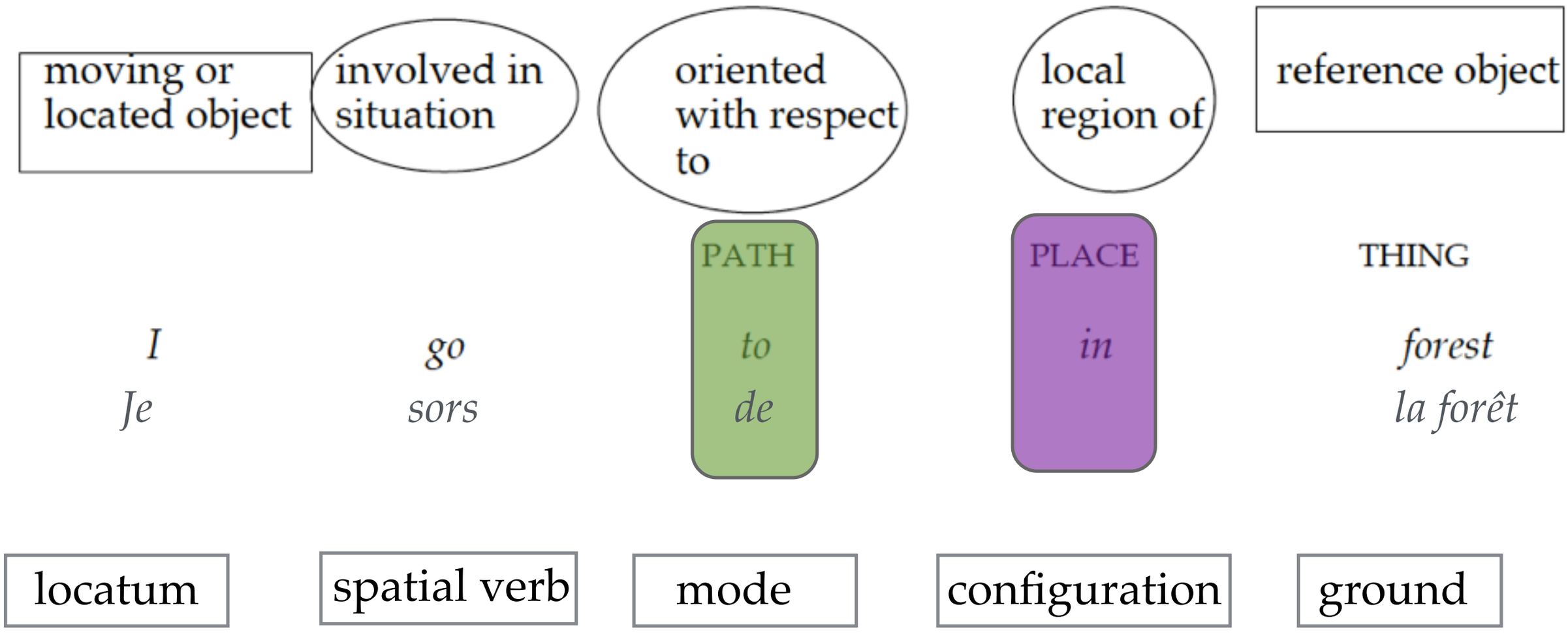


Figure 1. Structure of a local situation (Lehmann 1992: 629; Jackendoff 1983: 161ff.)

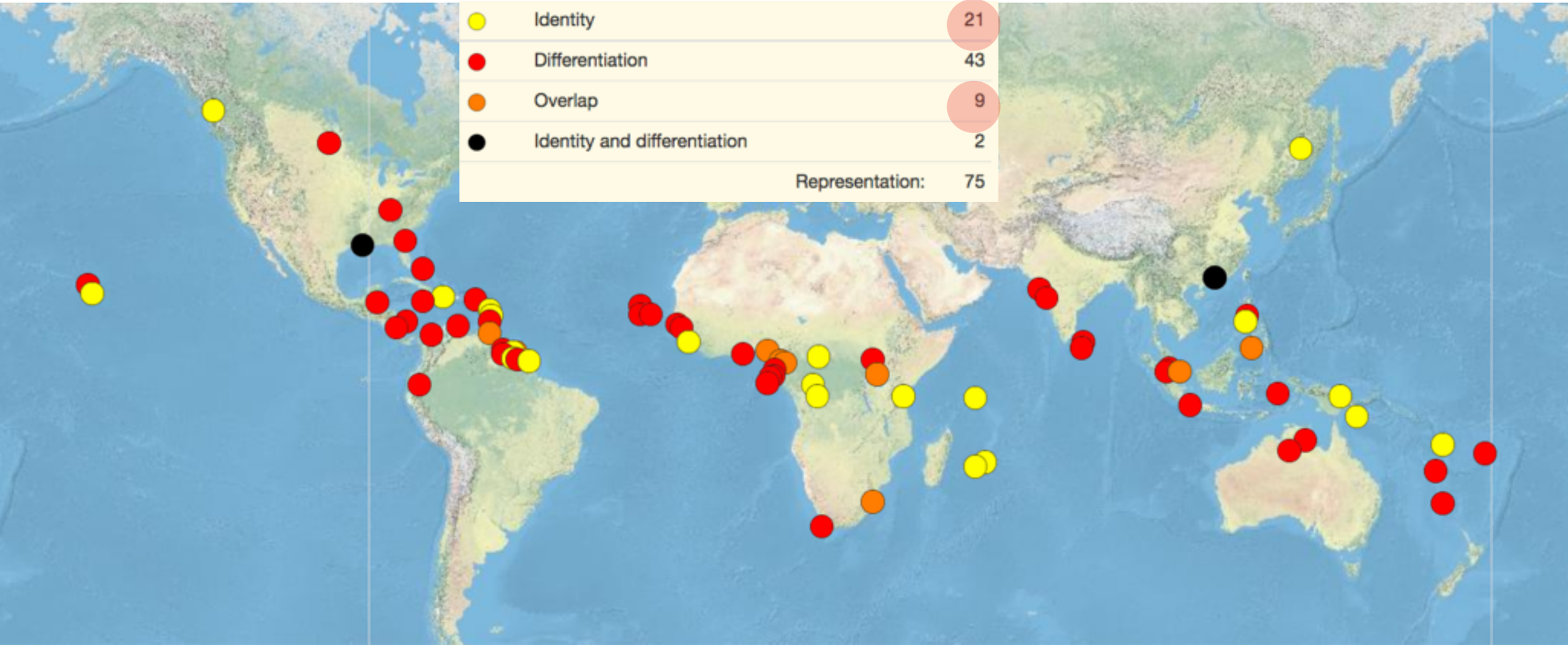


(Stolz et al. 2014)

Krio

a	*di go*		***na***	*di makit*
1SG	PROG go		LOC	ART market

Motion-to and motion-from (Michaelis & *APiCS* Consortium 2013c)



(1) Krio (English-based; Finney 2013)

MOTION-TO

a. *a* *di* *go* ***na*** *di* *makit*
1SG PROG go LOC ART market
'I am going to the market.'

MOTION-FROM

b. *a* *jɛs* *kɔmɔt* ***na*** *di* *makit*
1SG just come LOC ART market
'I just came back from the market.'

Identity-Pattern

MOTION-TO MOTION-FROM

(2) Seychelles Creole (Michaelis & Rosalie 2013)

a. *mon al* ***dan*** *bwa*
1SG go in forest
'I go into the forest.'

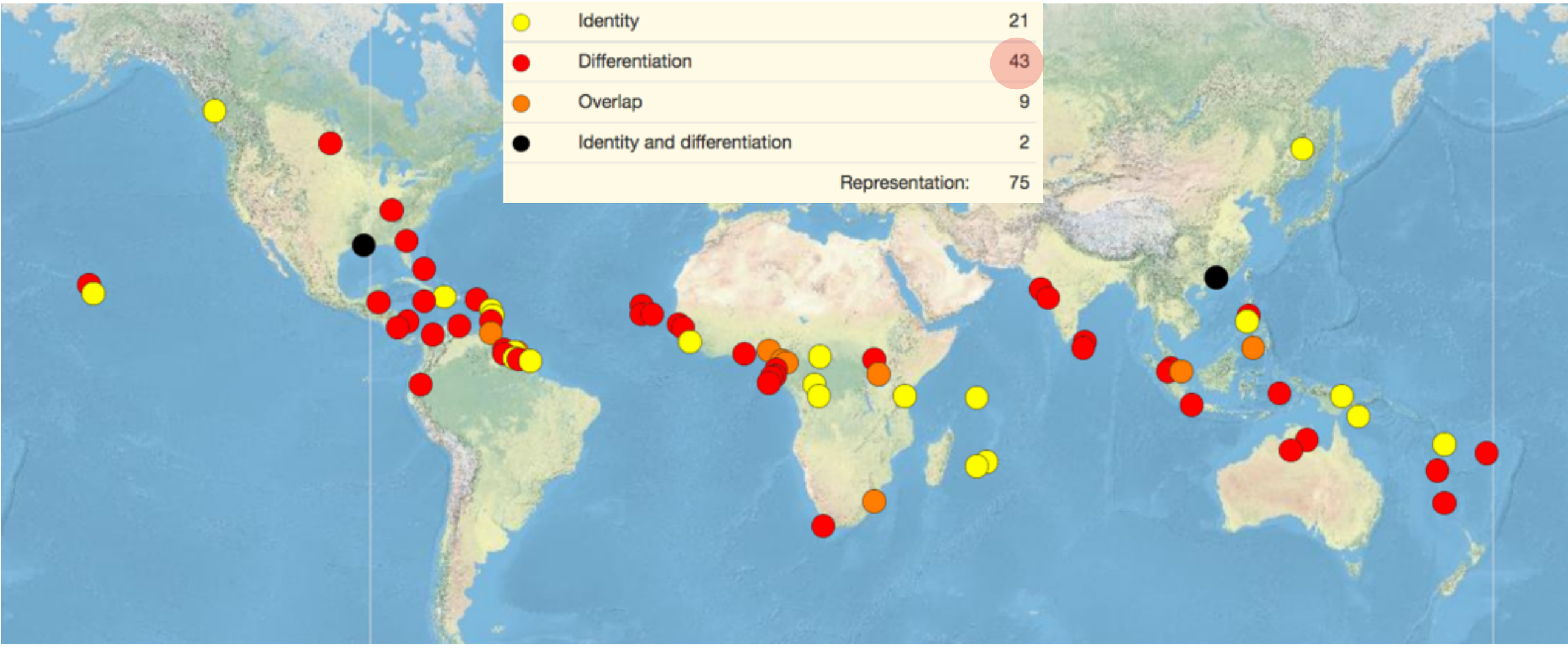
b. *mon sorti* ***dan*** *bwa*
1SG come.from in forest
'I come out of the forest.'

(3) Ternate Chabacano (Sippola 2013)

a. *Mótro ta andá ayá* ***na*** *merkádo.*
1PL IPFV go there LOC market
'We go to the market.'

b. *A vine lotru* ***nah*** *Isla di Muluccas [...].*
PFV come 3PL LOC island of Moluccas [...]
'They came from the Moluccan Islands [...].'

Motion-to and motion-from (Michaelis & *APiCS* Consortium 2013c)



(4) Palenquero (Schwegler 2013)

a. *I tan labá* ***aí*** *loyo.*
I go wash there creek
'I am going to wash (clothes) at the creek.'

b. *I ta miní* ***ri*** *loyo.*
I PROG come from creek
'I am coming from the creek.'

Differentiation

MOTION-TO

MOTION-FROM

(5) Principense (Maurer 2013)

a. *N we* ***ø*** *fya.*
1SG go market
'I went to the market.'

b. *N vika* ***fo*** *fya.*
1SG come come.from market
'I came from the market.'

(6) Tayo (Ehrhart & Revis 2013)

a. *nu ale* ***ø*** *lamisjo*
1PL go Mission
'We are going to the Mission.'

b. *[...] nu vja* ***ⁿde*** *partu-la*
1PL come PREP everywhere-DEM/DEF
'(...) we come from everywhere.'

Creoles with differentiation (🔴)

three subtypes	language	MOTION-TO	MOTION-FROM
two different adpositions	Palenquero	*aí*	*ri*
	Batavia Creole	*nu*	*di*
	Bahamian Creole	*to*	*from*
	Sri Lankan Malay	*na*	*ring*
TO: optional adposition FROM: obligatory adposition	Creolese	*(a)*	*from*
	Papiamentu	*(na)*	*di; for di*
	CVC Santiago	*ø; (pa; na)*	*di; d'*
TO: ø FROM: obligatory adposition/serial verb construction	Jamaican	*ø*	*frahn*
	Casamancese	*ø*	*di*
	Papiá Kristang	*ø*	*di*
	Tayo	*ø*	*^{n}de*

Length differences in the flagging of MOTION-TO/FROM

Language	MOTION-TO	seg	MOTION-FROM	seg
Palenquero	*aí*	2	*ri*	2
Batavia Creole	*nu*	2	*di*	2
Bahamian Creole	*to*	2	*from*	4
Sri Lankan Malay	*na*	2	*ring*	3
Creolese	*(a)*	(1)	*from*	4
Papiamentu	*(na)*	(2)	*di; for di*	2; 4
CVC Santiago	*ø; (pa; na)*	0; (2; 2)	*di; d'*	2; 1
Jamaican	ø	0	*frahn*	4
Casamancese	ø	0	*di*	2
Papiá Kristang	ø	0	*di*	2
Tayo	ø	0	*^{n}de*	2

Diachronic pathways in coding motion-to/ motion-from in creole languages

- Ideally we would compare each creole with its lexifier(s) and substrate(s).

- I haven't done a systematic comparison of all the contributing languages for all creole languages yet, especially with respect to the different coding patterns relating to different reference objects (named places, frequent locations, humans)

- BUT: It seems that in many creolization scenarios, the substrate pattern prevails over the lexifier pattern (e.g. all French/Macro-Sudan, French/Bantu bi-clans that mark MOTION-TO/FROM identically, Michaelis 2008; 2017; as for the African substrates s. Creissels 2006, Wälchli & Zuñiga 2006)

• But irrespectively of the question whether the creole pattern can be traced back to the lexifier or substrate clan (or is a complete innovation), my main question here is:

Does the universal tendency of the coding asymmetry in terms of matter also hold for pidgins and creoles which have evolved through extremely accelerated language change processes?

- Yes: All pidgins and creoles in *APiCS* show the predicted coding asymmetry:

--> MOTION-FROM constructions are **longer** than MOTION-TO constructions, or at least not shorter.

BUT: in contrast to the lexifier language, often new markers are grammaticalized.

- There are two scenarios:

(i) M**atter differs:** The creole shows the same pattern as the lexifier (and maybe the substrate?), but has innovated markers.

(ii) M**atter and pattern differ:** The creole has both, a different pattern from the lexifier and innovated markers (or no markers).

(i) same pattern as in lexifier, but innovated markers:

Language	TO	FROM	lexifier
Palenquero	*aí*	*ri*	Spanish: a/de
Batavia Creole	*nu*	*di*	Portuguese: a/de
Belizean Creole	*a*	*fa*	English: to/from
Norf'k	*ap, daun, out*	*fram*	English: to/from

(ii) different pattern from the lexifer, plus innovated markers that can be optional or zero in MOTION-TO:

language	TO	FROM	lexifier
Papiamentu	*(na)*	*di; for di*	Spanish/Portg. *a/ de*
Creolese	*(a)*	*from*	English: *to/from*
Guinea-Bissau Kriyol	ø	*di*	Portuguese: *a/de*
Papiá Kristang	ø	*di*	Portuguese: *a/de*
Principense	ø	*fo*	Portuguese: *a/de*
Tayo	ø	*ⁿde*	French: *à/de*
Jamaican	ø	*frahn*	English: *to/from*

Variability/optionality seems to be frequency-driven, cf. Kouwenberg 2013 on Papiamentu *na* (LOC):

"The complement of *bai* 'go' can be introduced by *na* 'LOC', or can be a bare noun, in the case of a named place (...) or in the case of a type of errand (go to school, go home, go to hospital)."

- two diachronic pathways in creole languages:

(a) MOTION-TO marker (more frequent member of the opposition) can be innovated, but is then often optional.

(b) MOTION-FROM marker (less frequent member of the opposition) is reinforced/innovated, but never optional or zero (against some overt MOTION-TO marker):

–> *for di* in Papiamentu < Portg. *fora*, Span. *fuera* 'outside'

(compare Italian *da* 'from' < Latin *de ab* 'of from')

–> *depi* < French *depuis* 'since, from' in the Bhojpuri-influenced variety of Mauritian Creole (Kriegel et al. 2009) against identity pattern in Mauritian Creole.

–> *fu* < Engl. *from; uit* < Dutch *uit* 'out of, from' in Dutch-influenced Sranan (Yakpo 2017) against the earlier identity pattern in Sranan.

• Compared to other grammatical oppositions (SG vs PL, dependent vs independent possessive forms (my Gießen talk), these two processes are exactly the two main pathways to yield coding asymmetries.

• Having grammaticalization scenarios in the back of our minds, one could imagine that the **shortening of the high-frequent member** of the opposition (motion-to markers) is the default process to cause an asymmetry.

• BUT the data also suggest the alternative scenario: the **lengthening/reinforcement of the rarer member** of the opposition (motion-from markers), which ultimately causes the coding asymmetry.

- Even though many creole patterns with different marking of MOTION-TO/MOTION-FROM may at first sight look quite similar to their lexifier languages, it is clear that structurally we are dealing with completely new systems, due to optional and zero-markers.

- But still: the outcome of the restructuring processes are again asymmetric coding systems:

  MOTION-FROM constructions are **longer** than MOTION-TO constructions, or at least not shorter – in line with the universal coding tendencies predicted by the form-frequency correspondences hypothesis (Haspelmath 2008 and related work).

Frequency of use data from corpora

• The idea is that in the course of hundreds (and thousands of years) of language change processes, differences in frequency patterns of constructions incrementally cause the cristallization of grammatical coding asymmetries.

--> form–frequency correspondences

There are clear **functional pressures** due to the communicative setting between speaker and hearer:

- The speaker does not have to code **more frequent** and therefore **more predictable meanings** as costly since s/he anticipates that the hearer will easily decode the meaning anyway.

- But **rarer** and **less preditable meanings** must be coded with more segments in the signal to give the hearer enough information to properly reconstruct the meaning.

• Are there differences in the frequency of use of the two meanings MOTION-TO and MOTION-FROM?

• We should find these different usage frequencies in any language because the coding asymmetry is assumed to be universal and therefore applicable and falsifiable in every language.

• Do speakers of individual languages express MOTION-TO more often than MOTION-FROM?

--> this would be the prediction!

(i) American English (COCA, 520 million words)

reference object	MOTION-TO	MOTION-FROM	
US	7.206	3.081	
India	1.537	1.116	
school	13.932	3.673	
church	2.938	381	
(home	310.359	5.334	(at-rest plus TO))
hospital	526	197	
the hospital	5.023	1.421	

"The corpus contains more than 520 million words of text (20 million words each year 1990-2015) and it is equally divided among spoken, fiction, popular magazines, newspapers, and academic texts." http://corpus.byu.edu/coca/

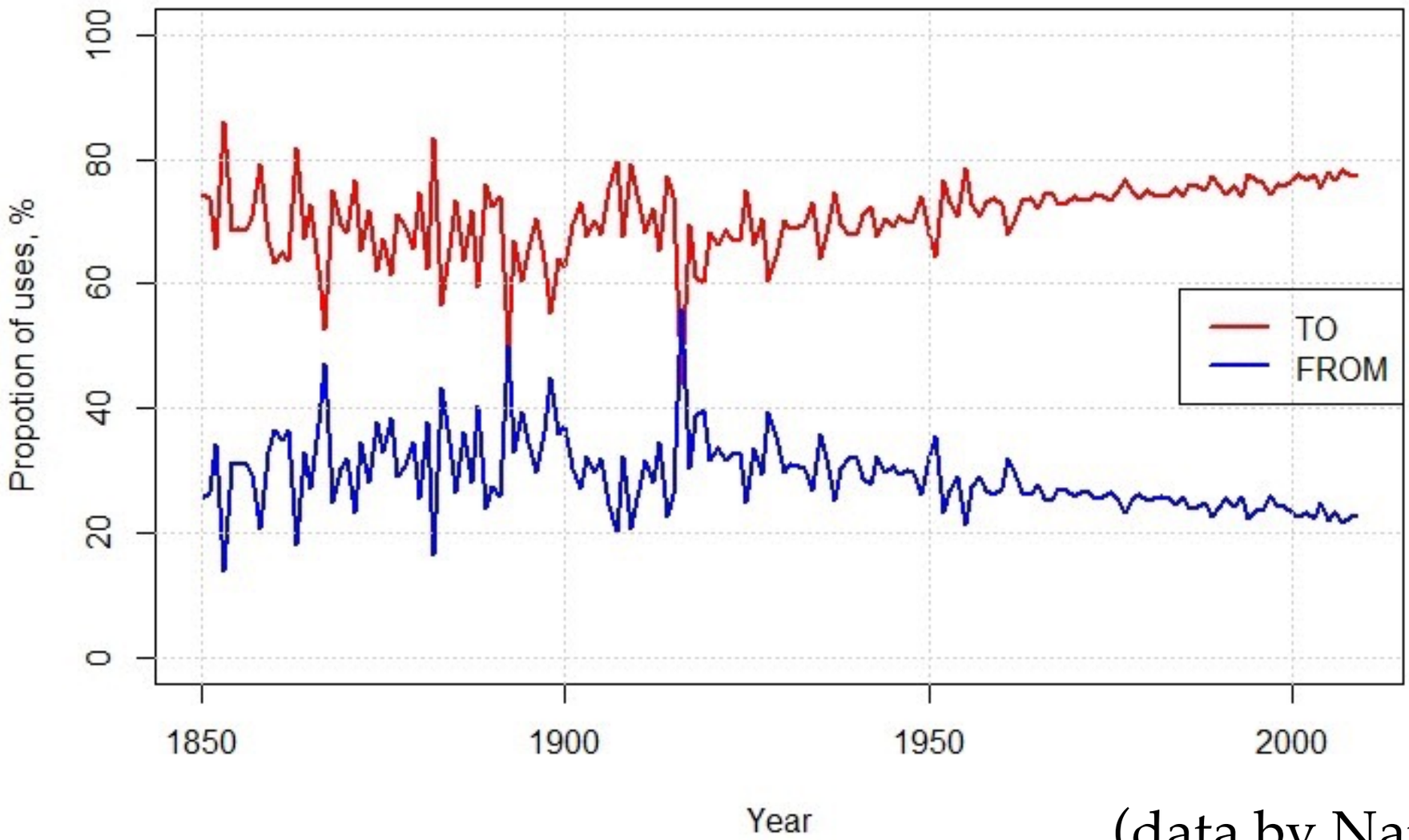
(ii) Russian (Russian National Corpus, 300 million words)

reference object	MOTION-TO	MOTION-FROM
hospital	4355	1423
school	5199	1692
market	2058	462
home	46704	6045
church	4449	1243
America	2675	891
Israel	768	179
India	962	358
Russia	8388	4182
Moscow	25975	9942
Europe	1879	725

(data by Natalia Levshina)

(iii) Modern Hebrew: 'to Jerusalem' vs. 'from Jerusalem'

- diachronic development in Google Books Ngrams
- MOTION-TO *le* vs. MOTION-FROM *min*



(data by Natalia Levshina)

Conclusion (1)

- MOTION-TO and MOTION-FROM constitute an instance of a grammatical coding asymmetry.

- According to the form-frequency correspondences MOTION-FROM constructions are longer as or at least not shorter than MOTION-TO constructions.

- Corpus data from different languages show that there is indeed a clear difference in frequency of MOTION-TO and MOTION-FROM constructions.

Conclusion (2)

• These figures fit nicely with the hypothesis that speakers highlight the less frequent and therefore less predictable meaning: MOTION-FROM constructions are longer.

• In the course of language change processes, we see these coding preferences fossilizing into coding asymmetries.

• Independently of the origin and degree of innovation in the creole patterns and independently of the fact that creoles are the outcome from heavy restructuring processes, creole languages support the universal trend.

SPCL meeting, Tampere

20 June 2017

Asymmetry in path coding: Creole data support a universal trend

SUSANNE MARIA MICHAELIS

Universität Leipzig & Max Planck Institute for the Science of Human History (Jena)

susanne.michaelis@uni-leipzig.de

Literatur

Creissels, Denis. 2006. Encoding the distinction between location, source and direction: a typological study. In: Hickmann, Maya & Stéphane Robert (eds.), *Space in languages*, 19-28, Amsterdam: John Benjamins

Finney, Malcolm Awadajin. 2013. Krio structure dataset. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Haspelmath, Martin, Andreea Calude, Michael Spagnol, Heiko Narrog & Elif Bamyacı. 2014. Coding causal-noncausal verb alternations: A form-frequency correspondence explanation. *Journal of Linguistics* 50(3). 587–625.

Jackendoff, Ray. 1983. *Semantics and Cognition*, Cambridge, Massachusetts: MIT press.

Jaeger, T. Florian & Buz, Esteban, Signal reduction and linguistic encoding. Draft

Kriegel, Sibylle, Ludwig, Ralph & Henri, Fabiola. 2009. Les rapports entre créole et bhojpouri à Maurice: Contact de langues et actes identitaires. In Hookoomsing, Vinesh, Ludwig, Ralph & Schnepel, Burkhard (eds.), *Multiple identities in action: Mauritius and some Antillean parallelisms*, 203-252. Frankfurt am Main: Peter Lang.

Kurumada, Chigusa & T. Florian Jaeger. 2015. Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language* 83. 152–178. (doi:10.1016/j.jml.2015.03.003)

Lehmann, Christian. 1992. Yukatekische lokale Relatoren in typologischer Sicht. In: *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung (ZPSK)*, 45: 626-641.

Lüdtke, Helmut. 1979. Auf dem Weg zu einer Theorie des Sprachwandels. In: Lüdtke, Helmut (ed.) Kommunikationstheroetische Grundlagen des Sprachwandels, Berlin: De Gryuter, 182-252.

Maurer, Philippe. 2013. Principense structure dataset. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Michaelis, Susanne, Philippe Maurer, Martin Haspelmath & Magnus Huber (eds.). 2013. *The atlas of pidgin and creole language structures*. Oxford: Oxford

Michaelis & *APiCS* Consortium 2013a. Going to named places. In: Michaelis et al. (eds.). 2013, 314-7.

Michaelis & *APiCS* Consortium 2013b. Coming from named places. In: Michaelis et al. (eds.). 2013, 318-21.

Michaelis & *APiCS* Consortium 2013c. Motion-to and motion-from. In: Michaelis et al. (eds.). 2013, 322-5.

Prescod, Paula. 2013. Vincentian Creole structure dataset. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Smith, Ian R. 2013. Sri Lanka Portuguese structure dataset. In: Michaelis, Susanne Maria & Maurer, Philippe & Haspelmath, Martin & Huber, Magnus (eds.) *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Stolz, Thomas & Lestrade, Sander & Stolz, Christel. 2014. *The crosslinguistics of zero-marking of spatial relations*, Berlin: DeGruyter/Mouton.

Wälchli, Bernhard & Zúñiga, Fernando. 2006. Source-goal (in)difference and the typology of motion events in the clause. *Sprachtypologie und Universalienforschung* (*STUF)* 59. 284-303.

Zipf, George K. 1935. The psycho-biology of language. Boston, MA: Houghton Mifflin.