

On Developing a *Human Diachronic Simulation Paradigm (HUDSPA)*

Remus Gergel, Saarland University *

Formal Diachronic Semantics: Budapest, November 10-11, 2022

*I am indebted to my Saarbrücken core team for great support with parts of the enterprise, the German Research Foundation (DFG) for funding parts of the team, and a special thank you to Edgar Onea & Simon Dampfhofer, U. Graz, cf. *ELM 2* paper.



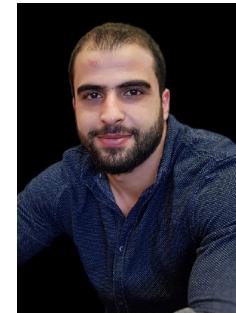
Saarbrücken core team



Maike Puhl



Martin Kopf



Mohammad Babli

A big thank you also to our student assistants for their help, especially Dana Rebar, Maryam Rajestari, and Nicolaie Dascalu.

Overview

1. HUDSPA to get started
2. Puzzles of cognates (ELM 1 paper, 2021)
3. Presuppositions in a sociolinguistically enriched context (ELM 2, forth.)
4. Annotations and beyond (project work decomposition)
5. Experiments on inverse iteratives

1. HUDSPA to get started

Diachronic difficulties in semantics

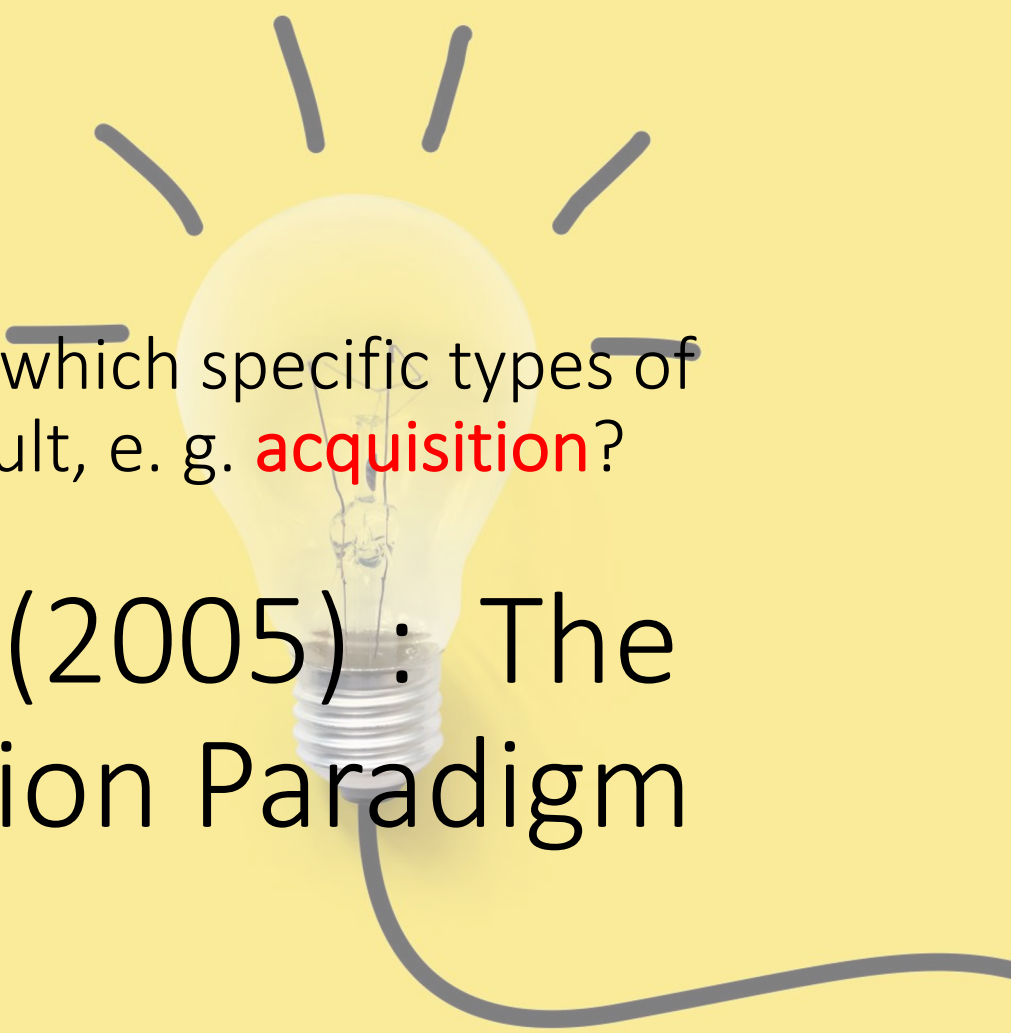
- Despite insight from **corpus** studies, answers to diachronic questions: often hard to come by, not always satisfactory for the variationist semanticist; cf. e. g. Deal (2020) for discussion
- Recent takes : use **experimental** methods to help elucidate **some diachronic/typological issues** (e. g. Gergel & Stateva 2014, Zhang, Piñango & Deo 2018, Fedzechkina & Roberts 2020, Fuchs & Piñango 2021, among others)

More issues:

- Simply looking for cases in the **present** to somehow **explain the past** will **not** suffice for many diachronic questions (many interesting changes from the past aren't obviously reproduced/detected 'live' in the present).
- Unlike e.g. in the physical reality beyond phonetic research (sound change), templates for, say, a near-complete inventory of meanings + the ways in which they can(not) develop is hard to establish.

Break it down:

- Two types of **key** problems :
 - ***Not enough data*** - e. g. to validate paths of change or fine-grained semantic/interfaced-based analyses;
 - ***Too much data*** - e. g. to process it adequately, with the required contextual details, etc.



What about other cases in which specific types of data extraction are difficult, e. g. **acquisition**?

Gleitman et al. (2005): The Human Simulation Paradigm

From Gleitman et al. to diachrony

- If accessible adults are good enough as consultants to test certain hypotheses about children, then
- accessible adults should also be good enough as consultants to test certain hypothesis about language change undergone by other adults
→ **reproductive experimental conditions**
- Crux: be clear about what you reproduce from the possible original linguistic environment of change (and as clear as possible about the many things you don't)

Semantic change (and its reproduction) as re-learning strategies

- The fact that **children** have a plethora of quick learning strategies in meaning is well-known (prosodic scanning, fast mapping, exclusion of irrelevant alternatives etc.)
- **Adults**, however, are not all that bad either. No limit on learning phrasal meanings (i. e. no classical ‘critical period’), if the bottleneck of the functional glue with its meanings is in place (cf. Slabakova’s (2012) work on L2)

Building up first quick intuitions

- Consider Austrian German ***sich-ausgehen*** (SAG, ‘reflexive+go.out’)

(1) *Ein Kaffee geht sich vor dem Termin aus.*

a coffee goes itself before the appointment out

‘There is enough time for a coffee before the appointment.’/‘We can still have a coffee before the appointment.’ etc.

Gergel & Kopf-Giammanco 2021 (*Can. J. Ling*) – in a nutshell:

SAG= sufficiency construction in Austrian G. with presuppositional restrictions

- Try to imagine: what could an SAG mean, for those Federal German speakers whose grammars lack this form-meaning correspondence?

Intuitions below the acceptability threshold?

(2) Context: Dominica can see that Martina's windows are lit and thinks:

#Es geht sich sicherlich aus, dass Martina zu Hause ist.

it goes itself surely out that Martina to home is

Intended: 'Martina must be home.'

Not only are several modal flavors (including epistemic ones as the one just contextualized above) excluded for SAGs in Austrian German varieties.

Federal German speakers also seem to find such intended readings degraded compared to the form-meaning pairings that are licit in Austrian German.

2. Puzzles of cognates

(ELM 1 paper – Gergel, Kopf-Giammanco, Puhl 2021)

E. *even* / G. *eben*: quick background

- *eben* did **not** develop the **scalar additive** particle meaning of *even*
- only non-cognates of *even* are used as scalar additives in German, e. g. *selbst, sogar, ...* (cf. Eckardt 2001, Eckardt & Speyer 2014)
- only few contexts left in Present-day usage of the two languages in which the two items can still mean similar things (e. g. *even surface*)

E. *even* / G. *eben*: quick background

- G. *eben* has, among other meanings, a **particularizer** one.
- Traugott (2006) identifies particularizer meanings as precursors of the scalar additive in the history of English.
- Does a word like *eben* show any **comparative propensity** of attaining *even* meanings when we try to simulate aspects of a change?

Eben manipulated as English *even*

- 12 target items
- 3 item sets with each set consisting of 4 items and respectively licensing readings of *sogar* ('even'), *nur* ('only'), and *auch* ('too/also')
- two cues to activate speakers to such readings:
 - context to clarify the intended meaning;
 - instruction to treat the examples as spoken by a non-mainstream community
- Task: rate acceptability in context on a 7 point scale
- initially 71 consultants, students of English; after usual exclusions 810 original data points

Eben manipulated as English *even*

e.g. context: *Last week we had a big party.*

target: EBEN Mary, who usually stays at home, showed up.

Letztes Wochenende hatten wir eine große Party.

Eben Maria, die sonst immer zuhause bleibt, ist gekommen.

völlig akzeptabel
im Kontext



gar nicht akzeptabel
im Kontext

Kommentar: _____

Outcome

	<i>sogar</i> -‘even’	<i>nur</i> -‘only’	<i>auch</i> -‘also/too’
Mean	5.17	4.34	4.62
Median	6	5	5
SD	1.46	1.6	1.83

- *eben* was rated better for manipulated *even* meanings
- partially similar result when trying to approximate the modal particle meaning of *doch* via *though* (with English speakers)
- issues remain, but a highly preliminary result

3. Presuppositions in a sociolinguistically enriched context

(*ELM 2* – Gergel, Puhl, Dampfoher, & Onea forth.)

Phrasing the issue (simply)

- Is there a predilection of presuppositional **LOSS** vs **GAIN**?

Why so simplistically?

A) Similar things have been insightfully investigated in other domains – e.g. **morphosyntax** for a long time and in different guises (cf. loss of inflectional morphology, increase in word-order rigidity etc.)

B) Such considerations are not immune to the area of meaning either, if we take a closer look: cf. loss of implicated meaning (**implicatures**) and gain in **conventionalized** meaning (new semantic entries etc.)

C) **Exploratory** line of study, but with some broader implications.

....

What's at stake now

- Exploratory study: lexical item shifting between the meanings of **BOTH / ALL**
- Assume: words like *both* are universals and presuppose the cardinality of their restrictor is two (cf. Heim & Kratzer 1998's discussion for view assumed)

CENTRAL QUESTION:

Will participants find it **easier**

to re-learn an item they had learned as meaning **BOTH** with the altered meaning **ALL** (notation: **both**→**all**) and thus to potentially **lose** a PSP or rather

to re-learn the opposite way, as towards only later incorporating the restricted cardinality (notation: **all**→**both**) and thus to potentially **gain** a PSP?

Method – in a nutshell

- 25 native speakers of German (11m/14f) with mean age 23.1 (SD 3.2) from Austria (conducted in Graz), split into two groups, which determined whether they would learn a **nonce word** *gure* in the meaning BOTH or ALL during training (successful tests after training)
- Spoken stimuli for practical reasons produced in a version of the **Saarland** dialects -- remote and little prominent variant of Mosel-Franconian (from the perspective of the South-Eastern **Austrian** region in which the study was conducted)
- Subsequent exposure to **contexts leading to a reinterpretation** towards the respective other meaning

Key characters and related premise

Character	Stands for	Phase
Non-native person (w.r.t. dialect)	Learner of language, introduces initial stimuli	training phase
Old person	Old stage of language	training phase
Young friend F who had been abroad	Old stage of language	test phase
Local young person S	New stage of language	test phase

Premise: Language has changed w.r.t. the meaning of *gure*

Method – training

- Participants asked to imagine visiting a fictitious **community** (German diaspora in the US) **guided by a native speaker** who studies with them
- Participants were taught the word *gure* by viewing images on a computer screen and listening to sentences containing the target non-word describing the situation, spoken by a non-native person (w.r.t. the dialect); an old person would then tell the participants whether the sentence was true in the situation presented
- If the sentence was not true, the old person in addition provided a reason why it was false. After three training items each, participants were asked to rate the truth of five sentences themselves (on a binary scale). After each judgement, they received written feedback from the older speaker whether their choice was correct.

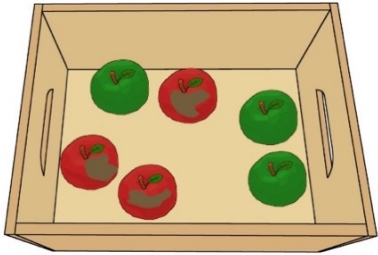
Method – specifics on testing (I)

- Participants were asked to imagine visiting a reunion of **younger** members of the community.
- Two characters of importance here: their **friend F**, who having been abroad for some time, is not up to date with current language developments (within the younger members of the community), and a high prestige competent **local speaker S**.
- In this context, participants are faced with examples showing that **gure** is used by S precisely in the respective opposite meaning of what they learned from the old person (i.e. **both** → **all** or **all** → **both**)
- Participants were then asked to rate their agreement for the sentence in the newly presented situation on a scale from 1 to 10.

Method –specifics on testing (II)

Someone utters: **Gure** red apples are rotten.

Task: Rate acceptability on a scale from 1 to 10.



all → **both**:

S: That's not right.

“Gure” is something my grandma would say in this case!

both → **all** :

F: Didn't she see the third apple?

S: Why? She said gure. She was right.

[translated from orig. examples in German for ease of presentation; nonce-word *gure* identical]

Results

- insight into the speed of learning the new usage of *gure* in the younger/prestige community
- judgments of items containing those fillers whose meaning had not changed compared to the training phase were as expected, i.e., they did not change significantly compared to judgments during training
- Interest: order in the presentation of the items and the group variable (reflecting the BOTH→ALL vs. ALL→BOTH)
- Based on judgment values and reaction times, **losing** the presupposition turned out to be the significantly faster process overall (replication essentially verified in a second experiment, N=24, with additional checking of presuppositional status)

4. Annotations and beyond

DFG project *Decomposing Decomposition*

- Goal : **annotate** as many potential decompositional items as possible from the Penn-Helsinki corpora of English to better understand them and their paths of change
- Issue: **how** to do it?
- **Expert annotation** by project members and trained student assistants with discussion of, until agreement about, divergent cases
- Follow-up: Is there a way to do this differently/independently, e.g. under certain experimental conditions?

Crowd-sourcing issues

- Crowd-sourcing judgments - doable on contemporary data.
- Historical ones?
- Problem: older grammars will most likely not be intuitively or immediately accessible for quick judgments.
- However, *some* sensitivity to the data can still reasonably expected to be available, depending on background of consultants – plus their goals and motivation.

Towards informed crowd-sourcing

- Expert annotators also lack direct L1 experience with earlier stages of the language.

Consider historical stages of English:

- **Students of English** in historical/contrastive **lectures** share an interest in understanding the language, including some of its earlier stages
- Add a minimal training for subjects: **1p-instructions**; training to achieve some minimal goals; then let the actual annotation begin.
- Both training and actual annotation with corpus contexts provided.
- 328 instances of *again* were collected as part of work for credit.

Data points collected: instances of *again*

– 17 th cent:	1,086
– 18 th cent:	969
– 19 th cent:	1,264

3,319 data points

Results: majority votes

Degrees to which the the gold standard was approximated

	17 th c.		18 th c.		19 th c.		all	
	N	%	N	%	N	%	N	%
rep	51	94.1	56	89.3	69	92.8	176	92.0
res_ct	56	67.0	36	77.8	29	82.8	121	74.0
other	1	100.0	8	87.5	11	81.8	20	85.0
all	112	78.1	102	83.8	114	86.8	328	82.9
C's κ	112	0.6	102	0.7	114	0.72	328	0.68

Improvement?

- Rather than simply taking a majority vote, different types of weighting can be performed (cf. Kopf & Gergel, ms. – based on metrics following the literature; cf. Aroyo & Welty (2013ff), Dumitrache et al. (2018))
- Quality metrics such as unit-quality scores, worker-unit agreement, worker-worker agreement
- At a next step, an unsupervised classification can be conducted (KMeans clustering, Pedregosa et al. 2011)

Scores based on updated classification

	17 th c.		18 th c.		19 th c.		all	
	N	%	N	%	N	%	N	%
rep	51	94.1	56	87.5	69	88.4	176	89.8
res_ct	56	75.0	36	80.6	29	89.7	121	80.2
other	1	100.0	8	87.5	11	90.9	20	90.0
all	112	81.2	102	83.8	114	87.3	328	84.1
C's κ	112	0.65	102	0.7	114	0.73	328	0.7

Takeaway from informed crowd-sourcing

- The experiment on informed crowd sourcing shows that attaining a **reasonable performance** is possible if not always easy;
- quality difference for a **subset of the data** – cf. repetitive readings;
- within the window of observation, **diachronic distance** was not crucial (17th vs. 19th c.)
- **more** distant times of observation (e.g. Middle or Old English) would require increased training, more resources, etc.

Implications?

- from a practical perspective: approach *could* be used e. g. to delegate readings that are easier to get, so that an expert team can concentrate more on the 'difficult' readings, to select further expert annotators etc.
- NB: The interest here has not been on how to *substitute* expert annotation (to be clear: our group does not have such plans for the current project),
- but experimentally on how much can be attained with a lower but still above-the-average level of expertise, i. e. approximating intuitions.
- Further outlook: e.g. **what** would it take to get less experienced crowd-workers than the informed ones we have had (i) properly trained; (ii) in a position to engage in a meaningful way with contexts from actual historical texts?

5. Inverse iteratives

Terms and usual suspects

- **Iteratives:** so-called re domain (adverbs like *again, wieder* etc.)
- Includes: repetitive, restitutive, ... (Dowty 1979, von Stechow 1996, Beck 2005, Pedersen 2004, Zwarts 2019, a. m. o.)
- Usual diachronic development observed for the most researched items: from counterdirectional/ restitutive **towards the repetitive readings** (cf. e. g. Fabricius-Hansen 2001, Beck & Gergel 2015, Gergel & Beck 2015 for discussions)

Terms (cont'd)

- Observation of **actual changes** from last year's conference (Gergel, Bablli, Puhl – FoDS 6 Cologne): some iteratives such as the adverb *nochmal* (Saarland version) or the Arabic adverb *thaniyaten* have undergone systematic developments the other way around, i. e. from **repetitive** → **restitutive** (further granularity on these items too)
- Call such items **inverse iteratives**.

Experiments on inverse iteratives

Questions:

- What are relevant **conditions** for growing iteratives in the respective varieties observed?
- If certain conditions are suspected, can they be **reproduced experimentally** (to an extent, as always!) in similar varieties/with similar items?

Restitutive *nochmal*

- *nochmal* ('once more', with variants) can have restitutive readings in Saarland dialects (unlike standard German)
- is there a more general **syntactic sensitivity** in such developments?
- try to see if **placing the adverb** high/low (in linear terms: front/late) is a facilitating factor in obtaining the reading for speakers whose grammars are not assumed to natively have provided it
- speakers whose background (residence and place of birth) was distinct from Saarland region have eventually been recruited online

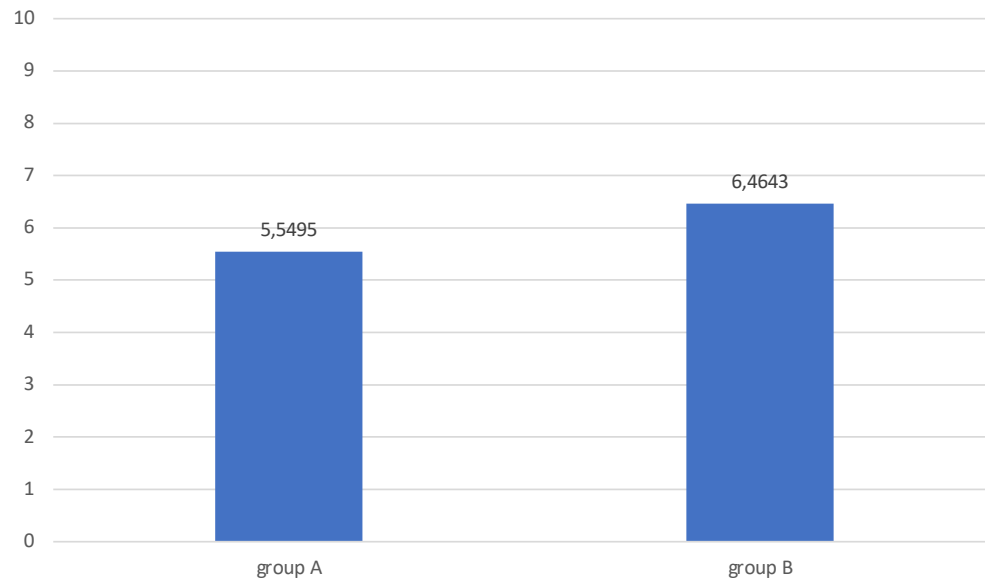
Quick exp. background *nochmal/once more*

- 10 target items (res. contexts only!), 12 filler items, 2 attention fillers
- N=40 (for standard German), reduced to 30 (16+14) after exclusion of Saarland dialect, non-native speakers, attention problems...

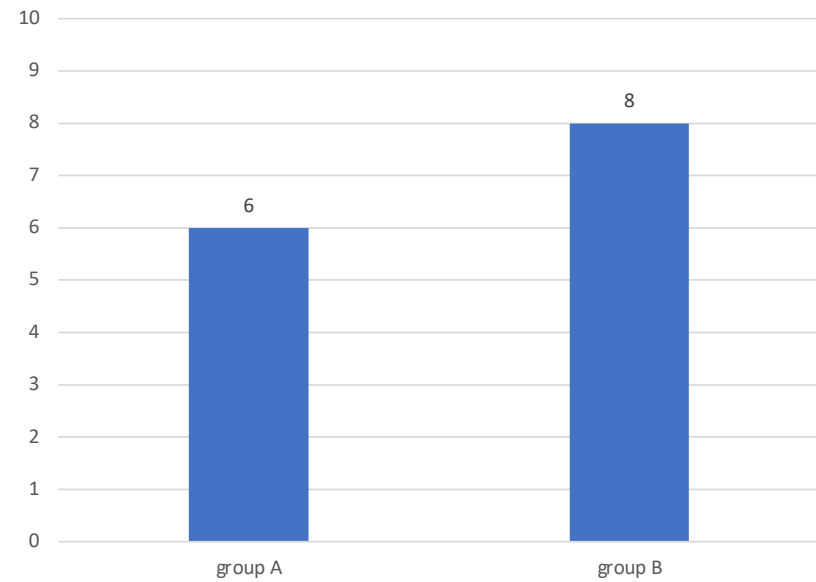
Position <i>nochmal/once more</i>	
high/front	low/late
Group 1	Group 2
<i>... weil er nochmal den Zaun weiß streichen will.</i>	<i>... weil er den Zaun nochmal weiß streichen will.</i>
<i>He once more painted the fence white.</i>	<i>He painted the fence white once more.</i>

Mean and median ratings

Mean Ratings by Group



Median Ratings by Group



Interim summary + connections

- **low** positions are more likely to be re-interpreted as **restitutive**
- this indicates that the development has the potential of being syntax-sensitive
- this is similar to *again's* repetitive cousin *once more* (N=40, GB speakers),
- but different e. g. from Old English restitutive *eft*, which could be found even in the pre-field (Gergel 2017);
- there was no sociolinguistic incentive in this case and no indication to take into account a non-standard variety – plain judgments.

Overall

HUDSPA:

- doable in multiple ways; builds on adult interpretive flexibility based on natural-language intuitions in appropriate morphosyntax+contexts;
- helps check plausibility of actual changes by experimentally controlling relevant alternatives (related languages/dialects; related items);
- simulates aspects of change by placing consultants (to different degrees!) in the shoes of **participants of change**.

NB:

- the program is at a forming stage;
- multiple additional studies and especially several **controls** still to be conducted (e. g. on non-PSP counterparts, but also on multiple other aspects);
-

Selected references

Deal, A. R. (2020). Comments on diachronic formal semantics. *Formal Approaches to Grammaticalization*, LSA New Orleans. || **Eckardt**, R. 2006. *Meaning Change in Grammaticalization*. OUP. || **Fedzechkina**, M., & G. **Roberts**. 2020. Learners sacrifice robust communication as a result of a social bias. <https://doi.org/10.31219/osf.io/usfhz>. || **von Fintel**, K. & L. **Matthewson**. 2008. Universals in semantics. *The Linguistic Review* 25: 139–201. || **Fuchs**, M. & M. M. **Piñango**. 2021. Language variation and change through an experimental lens: contextual modulation in the use of the Progressive in three Spanish dialects. *Spanish Socio-Historical Linguistics : Isolation and Contact*, Benjamins. || **Gergel**, R. 2020. Sich ausgehen: actuality entailments and further notes from the perspective of an Austrian German motion verb construction. In *Proceedings of the Linguistic Society of America*, M. Fuchs & J. Phillips (eds). LSA Publications 5: 5-15. || **Gergel**, R., M. **Kopf-Giammanco** & M. **Puhl** 2021. Simulating semantic change: a methodological note. In *Experiments in Linguistic Meaning (ELM)* 1. || **Gergel**, R., M. **Puhl**, S. **Dampfhofer**, E. **Onea** forth. The rise and particularly fall of presuppositions: evidence from duality in universals. ELM 2 Proceedings.

Selected references (cont'd.)

Gleitman, L., K. Cassidy, R. Nappa, A. Papafragou & J. Trueswell. 2005. Hard words. *Language Learning and Development* 1. 23–64. || **Kopf, M. & R. Gergel.** 2022. Informed crowd-sourcing for annotating semantic variation and change. Ms. Saarland University. || **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.** (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830. || **Slabakova, R.** 2012. Meaning in the second language acquisition. C. Maienborn, K. v. Heusinger & P. Portner (eds). *Semantics: An International Handbook of Natural Language Meaning*, 2753–2775. De Gruyter Mouton. || **Zhang, M., M. Piñango & A. Deo.** 2018. Real-time roots of meaning change: Electrophysiology reveals the contextual-modulation processing basis of synchronic variation in the location possession domain. *40th Annual Conference of the Cognitive Science Society*, 2783–2788.

App. 1. - Further extensions

- phenomena, phenomena, phenomena...
- data **types** in terms of classical methods of experimentation (cf. e. g. eye tracking for PSP experimenting, etc. etc.)
- use even more **channels** of communication (in addition to audio, writing, imaging, contexts, ...)
- more **interactive** scenarios (while keeping cognitive burden and especially costs manageable)
- developing principled **linking** procedures to actualized changes
- consider going from recipients to **actors** of change (machine simulation, interactive games etc.)?

App. 2 – Redoing the Arabic inverse iterative?

- quick background: *thaniyaten* has a long history of mostly repetitive, but in the meantime also other readings, including restitutive ones;
- so called mixed-antecedent contexts existed from early on (*raise his head/lower his head/raise his head again*; Gergel, Puhl, Babli FoDS 6)
- different task here: find an adverb that is only repetitive in current Arabic, and check its acceptability with repetitive/restitutive/mixed antecedents
- *marahnten aukhra* ('once another') seems to be a candidate; run with N=30 (3 exclusions) in the three versions (intra-subject in this case); rating of acceptability in context 1-7 (1 best!)
- mean ratings: Repetitive 2.3; Restitutive 3.5; Mixed: 2
- Small sample and additional controls required; but a certain (presumably priming) effect familiar from the corpus data seems to be reproduced.

App. 3 – PSPs and cyclicity

The two possible tendencies

- of losing presuppositional inferences
- increasing triggers

may not be as antagonistic as first thought.

Lose the inference, then reinforce it with a newly recruited item.

→ Classical situation of a cyclical development (Gergel 2022, Ms. UdS elaborates on the cyclicity aspect w.r.t. psps – available on request).