

Simulating semantic change: A methodological note

Remus Gergel, Martin Kopf-Giammanco, & Maike Puhl*

Abstract. The current work discusses the Human Diachronic Simulation Paradigm (HUDSPA), a method to experimentally probe into historical meaning change set up to (i) scan for configurations similar to attested alterations of meaning but in (typically, but not necessarily, related) languages or varieties which did not actualize the change(s) under investigations; (ii) measure the reactions of native speakers in order to ascertain the verisimilitude as well as the particular semantic and pragmatic properties of the items scrutinized. Specifically, the present paper discusses the relative propensity of a particularizer (German *eben*) to be interpreted with comparatively high confidence as a scalar additive particle such as *even* and of a concessive item like English *though* to be interpreted similar to a modal particle along the lines of German *doch*.

Keywords. diachronic semantics; additives; modal particles; experimental semantics; experimental pragmatics

1. Introduction. Diachronic and fieldwork semantics both model natural language variation. However, their standard methods of empirical verification vary considerably. Sometimes, they are even viewed as not (yet) fully compatible. For instance, Deal (2020) considers variable-force modals in synchronic and diachronic semantics and raises questions about diachronic conclusions (e.g. when variable-force semantics is suggested based on a sample of 72 Old English examples). Our present goal is not to engage with particulars of Old English modality (cf. Cournane 2017, Gergel 2016, 2017, Yanovich 2006, Solt & Umbach 2019 for broader discussions of natural-language modality from different perspectives with relevance to historical studies). But the more general point raised by Deal holds and needs to be addressed systematically. Diachronic semantics is constrained in multiple respects and to some serious extent this appears to be due to its intrinsic nature, which seems to run counter to methods of inquiry used in, say, modern cross-linguistic semantics. (Ultimately, we do not think that such a putative incompatibility is a necessary conclusion, as we will see.) Hence, regardless of the origins of possible empirical dissonances and difficulties in diachronic semantics, a continuous refinement of the empirical methods that are used in this branch seems to us to be imperative and useful.¹

*We thank the organizers, reviewers, and audience of ELM 1 for providing a fruitful and encouraging forum of discussion for our ideas. Authors: Remus Gergel, Universität des Saarlandes (remus.gergel@uni-saarland.de) & Martin Kopf-Giammanco, Universität des Saarlandes (martin.kopf@uni-saarland.de) & Maike Puhl, Universität des Saarlandes (maike.puhl@uni-saarland.de)

¹The value of diachronic studies themselves is not at issue here or in any of the studies mentioned and it should be evident, but we offer - in oversimplified fashion - an argument to readers less familiar with the field. Diachronic data is transitional data. That is, such data not only provides descriptions of, e.g. a well-studied language (variety) A and an exotic language (variety) B based on a parametric or otherwise postulated discriminating modelling criterion, but it is concerned with an old language (variety) O and a new language (variety) N that has developed over time from O. Typically this requires a more constrained theoretical modelling, as an explanation for how to plausibly and factually reach from O to N is by default required.

We wish to emphasize that diachronic semantics has not been static over recent years, but it has made considerable progress; see e.g. Deo (2015) for an overview, to which we only add a few relevant points before narrowing down further to our current point. Importantly, clear theoretical programs exist (cf. e.g. Fintel 1995, Eckardt 2006) and likewise specific corpus studies, for instance even to make sense of conflicting analyses that could not be previously solved synchronically (Beck & Gergel 2015, Gergel & Beck 2015) as well as studies that connect somewhat broader typological views with methods as refined as electrophysiological measurements (cf. Zhang et al. 2018). When not all the data relevant for semantic change is available, researchers have moreover capitalized on the interfaces of the semantic component, e.g. with structure or with pragmatic processes (cf. Gianollo 2018, Traugott 2006, to name just one example for each possibility in this case as well). Furthermore, the lack of extensive corpus data for low-frequency phenomena can sometimes be partially compensated when more methods are amassed (see Gergel & Kopf-Giammanco forth. for an example and discussion). And nonetheless, when it comes to the necessary details of virtually any semantic inquiry, there are impasses that appear to be often insurmountable from the perspective of historical linguistics if one takes fieldwork variationist semantics as a term of comparison. It is not just the lack of negative data that poses a well-known problem. Receiving graded judgments in appropriate and detailed contexts from native speakers to test the validity of both proposed trajectories and causal chains in actual changes can be just as critical as the lack of such data can be frustrating. Such issues motivate a more general question: How can a semantic path of change (and especially under the inclusion of typologically less well-trodden ones) be established given the impossibility of eliciting contextualized judgments, of receiving comments from interviewed speakers, etc.?

Some further venues are conceivable to compensate such deficits and offer partial answers to the question raised. For example, a possible solution to the main problem is to analyze current changes in progress (e.g. D'Arcy 2007), ideally such that they resemble changes that have occurred in the past. However, there are by far not enough detectable changes in progress to match the numerous interesting meanings that have arisen historically. Therefore, we will support the use of experimental semantics as a bridge to cross the gap between semantic fieldwork and diachrony drawing from other cases of semantic development with disadvantaged extraction of speaker intuitions, namely the earliest stages of language acquisition (cf. e.g. Gleitman et al. 2005). Specifically, we discuss in this paper two experiments to support the hypothesis in (1) (Gergel 2020: 13):

(1) Human Diachronic Simulation Paradigm (HUDSPA)

Humans confronted with new meaning-form pairings modeled after an attested semantic change will react similarly when they are placed in conditions that resemble those of the actual change (e.g. via a cognate that is similar but did not undergo the transformation investigated).

The idea is to confront speakers with meanings that happened in a related language or variety but not in their own and to compare them to meanings that did not develop in either language or variety. Following HUDSPA, we hypothesize that the meanings that were targeted in the actual change will perform better than other meanings that could have developed from the same semantic domain. We thus seek to replicate relevant parts of historical processes under testable conditions.

In other words, we target and sort out suboptimal (“ungrammatical”) but relevant judgements with a potential for change in pertinent environments. Speakers in historical change situations also end up with choosing form-meaning pairings that are originally nontransparent on the meaning that will later conventionalize. But such choices are less random than they may appear. The hypothesis we support in this contribution is precisely that they can be replicated to a degree in lab conditions.

In section 2, we will present two experiments in which we confronted German speakers with contexts in which the meaning of English *even* is conveyed, by utilizing a cognate (German *eben*), and English speakers with contexts in which a modal-particle meaning modeled after German *doch* is utilized for the word *though*. Section 3 provides a detailed discussion of the results and section 4 concludes and provides an outlook of how HUDSPA experiments can be refined further.

2. Two experiments. Our first experiment targets the development of English *even*, simulated from the perspective of German *eben*. German *eben* did not develop a meaning such as (Modern) English *even*. German only uses noncognates of *eben* for additives of improbability. The meaning of *eben* can be approximated to the meaning of *even* in contexts such as *an even surface*.

Our second experiment targets the German discourse particle *doch* through the prism of English *though*. We paid attention to syntax, e.g. by using final *though* in view of relevant factors (Van Kemenade 2019). Similar to above, *though* did not develop a presuppositional meaning as *doch* (Grosz 2014), regardless of syntax.

In both experiments we used two cues to activate speakers to such readings: one is context to clarify the intended meaning; the other is the instruction to treat the examples as spoken by some non-mainstream German (and English) community and to grade the naturalness of the examples encountered w.r.t. to the context given. From an earlier study, we had confirmation that speakers can reliably assign meanings in rich contexts to sentences which they find otherwise unacceptable (cf. e.g. Gergel 2020, Gergel & Kopf-Giammanco forth. for discussions).

2.1. *Eben* MANIPULATED AS ENGLISH *even*. In this experiment, a questionnaire with 12 target items and 13 filler items was used. The target items consisted of 3 item sets with each set consisting of 4 items and respectively licensing readings of *sogar* (‘even’), *nur* (‘only’), and *auch* (‘too/also’). In place of *sogar*, *nur*, and *auch*, the items featured *eben* — cf. Fig. 1 for an example item. All items consisted of a context description (*Letztes Wochenende hatten wie eine große Party*; ‘Last weekend we had a big party’; Fig. 1 top) and a target sentence (*Eben Maria, die sonst immer zuhause bleibt, ist gekommen*; ‘EBEN Mary who usually stays at home showed up’; boldfaced in Fig. 1) as well as a comment section. Subjects were asked to rate the target sentences based on a 7-point scale ranging from ‘fully acceptable in context’ (7 pts) to ‘not at all acceptable in the context’ (1 pt). In the comment section, subjects were encouraged to suggest improvements should they find certain expressions odd.

We collected data from 71 subjects, all of them undergraduate students in the English department of Saarland University, yielding 810 observations (excluding 42 missing values from the ratings). We excluded non-native German speakers. Additionally, we manually categorized the comments provided by the subjects as to their suggestions for improving the target sentences. The criterion here was that the subjects suggested replacing *eben* with *sogar/nur/auch*. If subjects suggested supplementing *eben* with *sogar/nur/auch*, commented on an unrelated issue (or did not provide a comment at all) their rating was not considered for this analysis. This criterion was cru-

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: _____

Figure 1: Example item - Experiment 1; *eben* manipulated for *even/sogar*

cial because *eben* can be used in connection with *sogar/nur/auch* but is interpreted as a discourse particle rather than with the targeted meaning (cf. Repp 2013). Based on this manual categorization, we had 199 observations (53 for the *sogar* condition, 94 for *nur*, 52 for *auch*) for further analysis.

In descriptive terms, the three conditions were rated as in table 1.

	<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

Table 1: Mean and median ratings of Experiment 1

For statistical analysis, we relied on the R software (R Core Team 2019) and the lme4-package (Bates et al. 2015) for R. In a first step, we transformed the ratings into norm scores² and fit the data into a random slope model with ‘NormScore’ as a function of ‘condition’ (i.e. the 3 levels: *sogar*, *nur*, *auch*), allowing for different slopes per subject: $\text{NormScore} \sim \text{condition} + (1 + \text{condition} | \text{subject})$ (cf. Bates et al. 2015, R Core Team 2019). The estimate for the *sogar*(‘even’)-level is 0.222 and the slope for the *nur*(‘only’)-level is -0.561, for *auch*(‘also/too’) -0.382. In a second step and in order to obtain a *p*-value, we conducted a likelihood ratio test, pitching the full model against a null model (i.e. without the factor of interest, ‘condition’). The three levels of the factor condition affected the transformed ratings ($\chi^2(2) = 13.221, P = .0013$) lowering them by 0.561 for the *nur*-level and by 0.382 for the *auch*-level. This comparison suggests that the variability in the data collected is not random but can be explained by the three levels of the experiment.

2.2. *Though* MANIPULATED AS GERMAN *doch*. In the second experiment, manipulating final *though* as *doch*, an online questionnaire was used with 12 target items (joined by 14 fillers) with 4 target items per condition, where the respective readings approximated three different types of particles: *doch*, *ja*, *wohl* (cf. Zimmermann 2011 for an overview of the untranslatable material and Puhl & Gergel forth. for a discussion on the meaning contribution of final *though*).

As an approximation, the modal particle *ja* marks an utterance *p* as uncontroversial because

²Normscore transformation was performed in order to account for inter-subject-variability and achieve more normally distributed values.

p is already in the Common Ground (CG). According to Repp (2013), *ja* fulfils a RETRIEVAL function, meaning that the speaker's use of *ja* instructs the addressee to *retrieve* a proposition p from the CG (Repp 2013). This proposition p is not under consideration at the time of utterance, meaning that it is not entailed or implicated by the immediately preceding utterance (Repp 2013). Consider (2):

- (2) Ich gehe heute nicht spazieren. Es regnet ja.
 I go today not walk. It rains PRT.
 'I'm not going for a walk today. It is raining, you know.'

In (2), the weather is not entailed or implicated by the speaker's decision not to go for a walk. The hearer is assumed to be aware of the weather - it is part of the CG - but it is not being considered at the time of utterance.

The particle *doch* is similar to *ja* in that it also instructs the hearer to retrieve a proposition from CG. The difference between *ja* and *doch* is that *doch* also signals a contrast. Following Repp (2013), this contrast lies between the proposition p in *doch(p)* and a proposition q ($= \neg p$) in the CG.

- (3) Ich gehe heute nicht spazieren. # Es regnet doch.
 I go today not walk. # It rains PRT.
 'I'm not going for a walk today. # But it is raining, you know.'
- (4) A: I'm going for a walk now.
 B: Es regnet doch.
 It rains PRT.
 'But it's raining, as you know'

The use of *doch* in (3) is infelicitous because there is no contrast between not going for a walk and bad weather³. In (4), the use of *doch* is felicitous. It is assumed that both speakers A and B are aware of the weather. *Doch* signals this and instructs A to retrieve this information from CG. A's decision to go for a walk is at odds with the fact that people tend to go for walks in good weather⁴. This contrast between bad weather and going for a walk licenses the use of *doch* and, at the same time, makes the use of *ja* infelicitous, see (5).

- (5) A: I'm going for a walk now.

³Felicitous readings of (3) are possible depending on intonation and an assumed hearer. If "Es regnet doch" is the answer to an implied "why not?", (3) is acceptable because it highlights that the hearer is not currently considering the weather. However, if contrasted with (2) and keeping intonation constant, (3) is infelicitous.

⁴Note that if A has a habit of going for walks in the rain and B knows this, a dialogue such as (i) is perfectly acceptable.

- (i) A: I'm not going for a walk today.
 B: Es regnet doch.
 It rains PRT.
 'But it's raining, as you know'

B's utterance can be paraphrased as: "Why are you not going for a walk? It's raining and you like walking in the rain".

B: # Es regnet ja.
 # It rains PRT.
 ‘# It’s raining, you know.’

The modal particle *wohl* has no overlap with either *ja* or *doch*. It is an epistemic marker signaling that the speaker is not fully committed to the utterance, but merely assumes the utterance to be true (e.g. Zimmermann 2004, 2011), as in (6).

(6) A knows that B wanted to go for a walk.
 A: Es regnet. Dann geht sie wohl nicht spazieren.
 It rains. Then goes she PRT not walk.
 ‘It’s raining. I assume she’s not going for a walk’

In (6), A does not know whether B ended up going for a walk despite the weather but can only assume that this is not the case.

Trying to reproduce felicitous readings of *ja*, *doch* and *wohl* required the use of slightly longer and only dialogic contexts, compared to the first experiment. Given that the particles do not have counterparts in English, the experiment included two tasks, the first one consisting of a training section and asking if the meaning from the contextual clues was understood. The answer to this task was given through a slider ranging from 1 (‘very hard to understand’) to 101 (‘very easy to understand’). Subjects were also asked to provide a paraphrase of what they assume is meant by this sentence. Given that the language in which this experiment was conducted, English, lacks the particles, it could not be expected to have the same precision in the additional comments as in experiment 1. The second task was a forced-choice yes/no slider to test if the item was actually understood, i.e. whether final *though* conveyed the intended meanings of *doch*, *ja* and *wohl*, respectively. See figure 2 for an example item of the second task.

Jack and Jill are preparing for a party. Peter has accepted his invitation.

Jill: It’s too bad that Peter isn’t coming.
 Jack: **Peter’s coming, though.**
 Jill: Ah, yes, I remember.

Does Jack think that Jill should have remembered that Peter is coming to the party?



Figure 2: Example item - Experiment 2; *though* manipulated for *doch*

40 native speakers of English participated in this experiment, but due to the inclusion of attention-testing fillers, only 36 were considered. These attention-testing fillers included specific instructions in the context sentence about where to move the slider regardless of how good the supposed target sentence was, e.g. “Move the slider all the way to the left”. See table 2 for the descriptive statistics of this experiment. While the sentences seemed easy to understand (Task 1),

the intended meanings were not captured reliably (Task 2).

Task 1	<i>doch</i>	<i>ja</i>	<i>wohl</i>
Mean	95.40	92.47	92.15
SD	10.48	18.53	17.53
Task 2			
Mean	84.45	59.94	23.73
SD	25.33	38.34	34.07

Table 2: Mean ratings of Experiment 2

Comments were analyzed by assigning each comment a category which best fits the content of the comment. 14 categories in total included Paraphrase (target sentence without *though*), Reminder (*doch*), Common ground (*ja* and *doch*) and Assumption (*wohl*). Other important categories are Reason/Explanation (most common) and Concessive (use of *though* in PDE). For *doch*, almost 40% of the comments fell into the categories Reminder and Common Ground, which closely resemble the meaning of the particle. For *ja*, the most common category was Explanation/Reason (38%), which does not capture the meaning of the particle *ja*. Better suited categories, such as Common Ground or Knowledge, add up to less than 2%. The most common category of comments for *wohl* was also Explanation/Reason with 71%. Again, this does not capture the intended meaning of *wohl*. Assumption, which best fits *wohl*, received 18%.

We conducted Exact Wilcoxon signed rank tests for each pair (*doch-ja*, *doch-wohl*, *ja-wohl* for Tasks 1 and 2, and *doch1-doch2*, *ja1-ja2*, and *wohl1-wohl2*). In Task 1, *doch* was rated significantly higher than *wohl* ($p = 0.002$) but there are no significant differences between *doch* and *ja* ($p = 0.093$), and *ja* and *wohl* ($p = 0.740$). *Doch* readings were rated significantly higher in Task 2 than *ja* and *wohl*, and *ja* was rated significantly higher than *wohl* ($p < 0.001$ in all three cases). All three target readings showed significant differences between Tasks 1 and 2 (*doch*: $p = 0.004$, *ja*: $p < 0.001$, *wohl*: $p < 0.001$).

3. Discussion. The experiments show that the meanings of the cognates were interpreted more appropriately than competitors. Both the *doch* meaning of *though* and the *even* meaning of *eben* were captured significantly more reliably than the meanings of their competitors. The discourse particle meaning of *doch* seems calculable from the relationship between the currently available concessive component of final *though*, which was reflected in the comments, which is close in meaning to the presupposition of contrast in *doch*. Both *doch* and *though* are possible in concessive contexts, see (7) and (8).

- (7) A: I'm going for a walk.
B: It's raining, though.
- (8) A: I'm going for a walk.
B: Es regnet doch.
It rains PRT.
'But it's raining, as you know.'

In both (7) and (8), the contrast between bad weather and going for a walk is signaled by *though* and *doch*, respectively.

Nevertheless, with only 40% of the comments falling into the categories of Reminder and Common Ground, which most accurately describe the meaning and usage of the MP *doch*, the differences between the cognates *though* and *doch* remain apparent. While both *doch* and *though* convey the notion of contrast, they differ in their conditions of use due a key point of difference: *doch* includes a RETRIEVAL function while *though* does not. *Doch* can be used to signal a reminder. Consider (9) and (10).

(9) A: It's too bad that Tim isn't coming to the party.
 B: Er kommt doch.
 He comes PRT.
 'He's coming, remember?'

(10) A: It's too bad that Tim isn't coming to the party.
 B: He's coming, though.

Both (9) and (10) are felicitous. However, in (9), the implicature arises that A should have known (or did know at some point) that Tim was coming to the party. In (10), no such implicature arises. This implicature is easily cancellable, as in (11).

(11) A: It's too bad that Tim isn't coming to the party.
 B: Er kommt doch.
 He comes PRT.
 'He's coming, remember?'
 A: Really? I didn't know that.

In the experiment, this reminder function of *doch* was reinforced. Participants appear to have identified the notion of contrast that is present in *doch* and *though* but seem to have struggled with the RETRIEVAL component of *doch*.

The additive case may seem more surprising. However, if we consider that German *eben* can have e.g. a meaning similar to what Traugott (2006) identifies as a particularizing focus modifier reading (PFM; for Early English *even*), as in (12), then we can explain the significantly higher acceptability ratings for the items where *eben* was manipulated for *even*. Traugott describes such a reading of *even* as precursor (Stage II of a 3-stage development) towards the modern one. Beaver & Clark (2008) characterize particularizers as typically non-scalar focus operators. They propose (i) that their use indicates that a speaker has provided an indication of being in a position to answer the (possibly implicit) *Current Question* (CQ) and (ii) that specific particularizers possibly provide additional information. Along these lines, German *eben*_{PFM} as in (12) could be viewed as – aside from Beaver & Clark's (i) – indicating that, among the alternatives, there is exactly one possible candidate to answer the CQ (*Who did Peter meet?*) and that the focused individual is salient. For scalar additives, Beaver & Clark note that they state that the strongest true answer to the CQ is weaker than expected – in other words, the prejacent is the most improbable proposition from a set of alternatives.⁵ Given the possible availability of *eben* as in (12) in the subjects' grammars, they might have had an easier time accommodating a scale of (im)probability rather than for items where *eben* was manipulated for *also/too* and *only*. Specifically, the salience of the entity singled

⁵Strength is by default characterized by entailment; however, see Beaver & Clark (2008; 254ff) for details.

out by *eben*_{PFM} might have been responsible for participants' higher acceptability for its use in *even*-contexts. (We surmise, however, that in appropriate settings, salience might also offer a considerable bias for *only*-readings, even if the *eben*-readings have been found significantly the most acceptable ones in our experiment. We will return to this below.)

- (12) Peter hat letzte Woche im Krankenhaus Maria kennengelernt. Eben [(diese) Maria]_F hat Peter has last week in.the hospital Maria gotten to know PRT this Maria has er heute beim Einkaufen getroffen.
he today at shopping met
'Last week, Peter and Maria met at the hospital for the first time. Today at the grocery store, he ran into exactly that Maria.'

In turning to *eben* manipulated for *only*, we want to point out that the *only*-meaning can be associated with the use of *eben* only in varieties of Austrian German, cf. (13). Subjects evaluated in the experiment reported here were all native speakers of Federal German.

- (13) Ich habe so viel zu tun, aber ich habe eben zwei Hände.
I have so much to do, but I have PRT two hands.
'I have so much to do, but I have only two hands.' (Only varieties of Austrian German; adapted for standard orthography)

As noted above, the *only*-sentences were "guessed" correctly more often (94 vs. 53 times for *even*- and 52 times for *also/too*-sentences; via the improvement task). Nonetheless, recall that acceptability was rated significantly highest and most appropriately on the *even* readings.

We think, the set of facts we have so far is less recalcitrant than it may appear at first glance not only from an intuitive perspective (as salience could possibly be made to play an important role to different degrees in both *even* and *only* readings) or from a variationist perspective (as some dialects did in fact develop some *only* readings of the particularizer). If we take a step back and consider the broader picture from the point of view of semantic theory, it is also not too surprising that the two types of readings may compete in a close race historically (e.g. when contexts of change are indeterminate or biased one way or another). Beaver & Clark suggest that scalar additives (as *even*) and exclusives (e.g. *only*) are close-by in a certain sense and represent pragmatic opposites. While scalar additives amount to stating that the strongest true answer is stronger than expected, for exclusives the strongest true answer to the CQ is weaker than expected. In (14), the prejacent *Mary and Phil came to the party* is the strongest true answer to the CQ *Who came to the party?*. The upper bound placed on the strength of possible answers to the CQ by *only* is where its truth conditional impact originates: Any possible answer with more individuals than Mary and/or Phil showing up to the party is not true. It seems that contextual clues pertaining to truth conditions in the experimental items provided participants in the particular set-up with more solid ground for identifying the intended meaning of *eben*⁶.

⁶In designing our experimental items, we heavily relied on the upper bound that *only* triggers by using e.g. examples as in (i) (translated from G to E for ease of representation):

- (i) Context: The seminar this past term was a lot of work. Students had to submit practice sheets every week. Although they tried their best, just before the end of term it was too much for everybody:

(14) Only Mary and Phil came to the party.

(15) Even Mary came to the party.

Even (cf. (15)) does not have an upper bound that causes a truth conditional effect (any stronger alternatives are not necessarily excluded) but it conversely pushes the upper bound of what is expected to be the strongest true proposition. It is *even*'s capacity to surprise which we suspect to have given it an edge over *only* in terms of acceptability with their respective meanings forced onto *eben* — especially, as noted above, in face of *eben*'s particularizer function and its condition that the focus element in its clause be salient. This last point also supports the core hypothesis in HUDSPA that when appropriately contextualized, originally nontransparent form-meaning pairs can be accommodated and conventionalized along an actual trajectory of semantic change.

4. Summary and outlook. To sum up, HUDSPA at this point shows convergence towards the actually developed meaning if the speakers' grammars are properly factored out. This is a minimal but crucial result towards more refined investigations of change. While, for instance, several currently popular game-theoretic approaches have a similar goal of simulating paths of change on the basis of rational tools, they do so at times by stipulating (often rather abstract) costs and benefits, so that in principle nearly any course could be attained. HUDSPA, by contrast, constrains the course of change appropriately, by using as its primary sources solely natural-language intuitions, which can further be probed into experimentally and theoretically. From a broader perspective, we think HUDSPA should not be all that surprising. It generalizes a certain perspective on uniformitarianism (cf. Walkden 2019 for a rich historical discussion even if without a semantics excursus) and the well-known idea from several branches of linguistics including sociolinguistics and language acquisition that the present is worth considering also to explain aspects of the past. What we take to be just as worthwhile is a starting attempt to raise such questions in controlled experimental environments for the area of semantic change.

While the results of the experiments above seem to support HUDSPA, they can only be regarded as initial findings. Refining experimental design based on HUDSPA is the goal of follow-up research. We mention here only a few ways how we envisage the paradigm can be refined in future work. A first extension entails the incorporation of training tasks, which could be additionally tested (also via interactive experimental design, targeting different types of memory storage, etc.).

A second way to refine the insight obtained is to test if results differ depending on whether or not participants received the instruction that they are encountering a non-standard variety. To some extent, this could be viewed as paralleling a putative dichotomy of contact-induced, i.e. external vs. endemic changes. But notice that from the perspectives of speakers who have not yet adopted a given semantic change, contact with progressive speakers with regards to the change in question is in practice almost always the case (even when they belong to the same linguistic community in other respects).

A third controlling step would be to rule out additional possible biases ranging from less obvious lexical semantic facts to relevant phonological biases. The clearest cut in this area would

Target: Out of all students, EBEN Mary submitted her homework on time.

seem to be offered by the use of nonce words, but in such a case too, possible associations with relevant words known by participants would have to be controlled. (E.g. if a nonce word is more similar to an existing relevant word than competitors, then it could still have a starting advantage.) The usage of nonce words would clearly reduce the "etymological burden" in design, but notice that this by and of itself does not automatically offer an improved insight. Speakers in actual change situations quite often in fact take the previous meaning as a starting package and build on it through interactive processes. However, a controlled design could target nonce words that have been introduced (and crucially: trained) in very specific ways, so that only those features will figure prominently that are relevant to the experimental task.

Last but not least, we have only illustrated a minimal amount of variation in the methods used for HUDSPA for practical reasons; there is naturally no a priori reason to constrain either the technical battery of methods or the range of applications (a quick extension could be, for instance, to look beyond naturalistic L1 semantic changes and also incorporate the potential of different types of bilingual or L2 extensions.) The major restriction remains, just as much as the potential that we see, that a close investigation of the actual critical contexts of change (as opposed to possibly too broad generalizations) may be one of the safest ways to plausibly simulate semantic change.

References

- Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01).
- Beaver, David & Brady Clark. 2008. *Sense and sensitivity. how focus determines meaning*. Chichester, UK: Wiley-Blackwell.
- Beck, Sigrid & Remus Gergel. 2015. The diachronic semantics of English *again*. *Natural Language Semantics* 23,3. 157–203.
- Cournane, Ailís. 2017. In defence of the child innovator. In Éric Mathieu & Robert Truswell (eds.), *Micro change and macro change in diachronic syntax*, 10–24. Oxford: OUP.
- D’Arcy, Alexandra. 2007. Like and language ideology: Disentangling fact from fiction. *American Speech* 82. 386–419.
- Deal, Amy-Rose. 2020. Comments on diachronic formal semantics (as compared to formal semantic fieldwork). Position paper given at the workshop Formal Approaches to Grammaticalization. LSA Annual Meeting, New Orleans, January 5, 2020.
- Deo, Ashwini. 2015. Diachronic semantics. *Annual Review of Linguistics* 1. 179–197.
- Eckardt, R. 2006. *Meaning Change in Grammaticalization. An Enquiry into Semantic Reanalysis*. Oxford: OUP.
- Fintel, Kai von. 1995. The formal semantics of grammaticalization. In *Proceedings of NELS 25 2: Papers from the Workshops on Language Acquisition & Language Change*, 175–189. UMass, Amherst.
- Gergel, Remus. 2016. Modality and gradation: Comparing the sequel of developments in ‘rather’ and ‘eher’. In Elly van Gelderen (ed.), *The linguistic cycle continued*, 319–350. Amsterdam/Philadelphia: John Benjamins.
- Gergel, Remus. 2017. Dimensions of variation in Old English modals. In Ana Arregui, Maria Luisa

- Rivero & Andres Salanova (eds.), *Modality across syntactic categories*, 179–207. Oxford: Oxford University Press.
- Gergel, Remus. 2020. Sich ausgehen: Actuality entailments and further notes from the perspective of an Austrian German motion verb construction. In *Proceedings of the LSA workshop "Formal approaches to grammaticalization"*, vol. 5(2), 5–15. New Orleans: Linguistic Society of America.
- Gergel, Remus & Sigrid Beck. 2015. Early modern english *again*: a corpus study and semantic analysis. *English Language and Linguistics* 19,1. 27–47.
- Gergel, Remus & Martin Kopf-Giammanco. forth. ‘Sich ausgehen’: On modalizing go-constructions in Austrian German. *Canadian Journal of Linguistics* .
- Gianollo, Chiara. 2018. *Indefinites between Latin and Romance*. Oxford: OUP.
- Gleitman, Lila, Kimberly Cassidy, Rebecca Nappa, Anna Papafragou & John Trueswell. 2005. Hard words. *Language Learning and Development* 1. 23–64.
- Grosz, Patrick. 2014. German *doch*: An element that triggers a contrast presupposition. In *Proceedings of the 46th annual meeting of the Chicago Linguistic Society*, 163–177.
- Puhl, Maike & Remus Gergel. forth. Final *though*. In Remus Gergel, Ingo Reich & Augustin Speyer (eds.), *Particles in German, English and Beyond*, Amsterdam: Benjamins.
- R Core Team. 2019. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <https://www.R-project.org/>.
- Repp, Sophie. 2013. Common ground management: Modal Particles, Illocutionary Negation, and VERUM. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond expressives: Explorations in use-conditional meaning*, 231–274. Leiden, The Netherlands: Brill.
- Solt, Stephanie & Carla Umbach. 2019. Comparison via *eh*. Paper presented at Sinn & Bedeutung 24, University of Osnabrück.
- Traugott, Elizabeth. 2006. The semantic development of scalar focus modifiers. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English Handbooks in Linguistics*, 335–359. Malden et al.: Blackwell.
- Van Kemenade, Ans. 2019. Discourse particle *then* in English: Clause structure and discourse organisation. Paper presented at Particles in German, English, and beyond, Saarland University.
- Walkden, George. 2019. The many faces of uniformitarianism in linguistics. *Glossa: A Journal of General Linguistics* 4(1). 52. 1–17.
- Yanovich, Igor. 2006. Old English **motan*, variable-force modality, and the presupposition of inevitable actualization. *Language* 92(3). 489–521.
- Zhang, Muye, Maria Mercedes Piñango & Ashwini Deo. 2018. Real-time roots of meaning change: Electrophysiology reveals the contextual-modulation processing basis of synchronic variation in the location-possession domain. In C. Kalish, M Rau, J. Zhu & T.T. Rogers (eds.), *Proceedings of the 40th annual conference of the Cognitive Science Society*, 2783–2788. Austin, TX: Cognitive Science Society.
- Zimmermann, Malte. 2004. Zum “Wohl”: Diskurspartikeln als Satztypmodifikatoren. In *Linguistische Berichte 199*, 253–286. Hamburg: Helmut Buske Verlag.
- Zimmermann, Malte. 2011. Discourse particles. In Paul Portner, Claudia Maienborn & Klaus von Stechow (eds.), *Semantics Handbücher zur Sprach- und Kommunikationswissenschaft*,

HSK 33.2, 2011–2038. Berlin: Mouton de Gruyter.