Katharina Schaebbicke, Heiko Seeliger & Sophie Repp (Universität zu Köln)

Negative polarity items in different negative contexts in German: two explorative experiments

Negative polarity items (NPIs) are frequently used in psycholinguistic and linguistic experimental studies – either to study the phenomenon itself or to use them as a diagnostic for other linguistic phenomena such as propositional vs. non-propositional negation or bias in questions. However, there are two problems when it comes to building theories on NPI licensing on the one hand, and using NPIs in experimental studies on the other. First, (theoretical) research of NPIs often seems to be limited to the investigation of very few NPIs, particularly English *any*, *ever*, and *yet*, or German *jemals* 'ever', *sonderlich* 'particularly' and *(keine) müde Mark* 'no red cent'. However, there is a vast landscape of NPIs in different shapes and forms, and it is an open question if they all can be accounted for by theories that are based on the analysis of only a few token NPIs. Second, the use of NPIs in experimental studies usually is not backed up by (control) studies exploring the precise licensing conditions of the NPIs, particularly in German. If we do not know exactly under which conditions a given NPI is licensed, said NPI cannot function as a reliable diagnostic, leading to possible confounds in the experimental design.

The goal of the present study is to provide a better empirical insight into the landscape of German NPIs. It builds on and complements corpus studies which have provided a database of German NPIs through (semi)automated extraction of NPI candidates from corpora: the CoDII database (Sailer/Trawiński 2006a/b, Trawiński/Soehn 2008, Trawiński et al. 2008). We present data from two acceptability rating experiments, where we explored the licensing conditions of a large set of German NPIs. The goal of **Experiment 1** was a classification of 60 German NPIs taken from the CoDII database as superstrong, strong, and weak NPIs, which is a classification suggested by Zwarts (1986, 1993) and van der Wouden (1994). According to Zwarts and van der Wouden, superstrong NPIs are only licensed by classical negation (*not*), while strong NPIs are also licensed in anti-additive environments (with licensers such as *nobody*), and weak NPIs are additionally licensed in downward-entailing environments (with licensers such as *few*). Because this classification doesn't capture the fact that some NPIs can also appear in questions, we extended this classification by the category 'superweak NPI', which builds on Giannakidou (1997, 1998, 2002, 2011), according to whom the nonveridicality of questions may license some NPIs. Furthermore, we tested if a nonveridical lexical operator may also license NPIs in German. The experiment had six conditions: There were six types of target sentences hosting the NPI such that the sentence either provided one of the five licensing conditions just mentioned or no licensor, as a control condition, see the table below. Each target sentence was preceded by a short context story that made the intended meaning of the target sentence highly plausible (not shown in the table).

	Condition	Licenser	Target sentence (NPI: *sonderlich* 'particularly')
[1]	Antimorphic	*nicht* 'not'	*Peter mag die Aufgabe **nicht** sonderlich.* Peter likes the task not particularly
[2]	Anti-additive	*kein x* 'no x'	***Kein** Mitarbeiter mag die Aufgabe sonderlich.* no colleague likes the task particularly
[3]	Downward entailing	*kaum ein x* 'hardly an x'	***Kaum** ein Mitarbeiter mag die Aufgabe sonderlich.*
[4]	Nonveridical lexical	*vielleicht* 'maybe'	***Vielleicht** mag Peter die Aufgabe sonderlich.*
[5]	Nonveridical nonlexical	polar question	**Q** [*Mag Peter die Aufgabe sonderlich?*]
[6]	Positive assertion	none	*Peter mag die Aufgabe sonderlich.*

Katharina Schaebbicke, Heiko Seeliger & Sophie Repp (Universität zu Köln)

The predictions were as follows. We expected sentences with superstrong NPIs to receive high acceptability ratings in the antimorphic condition [1] and low ratings in the other conditions; sentences with strong NPIs should receive high ratings in conditions [1] and [2]; sentences with weak NPIs should receive high ratings in [1]-[3]; sentences with superweak NPIs should receive high ratings in [1]-[4]. If nonveridicality in general, rather than just Q, can license NPIs in German, the ratings for conditions [4] and [5] should be the same. In the control condition [6] all NPIs should receive low ratings. Note that each NPI was tested in only one lexicalization, due to the large number of NPIs, so that there were 360 items altogether. The items were distributed across 6 lists and subdivided into 18 questionnaires, so that each participant saw 20 target items, as well as 20 fillers. We tested 360 monolingual German speakers (164 m, 235 f, 1 d; aged 16-85, m=29), resulting in 20 data points per NPI per condition. Participants were asked to judge the acceptability of the experimental items on a 7-point Likert scale with endpoints labelled *very unnatural* and *very natural*. Since Exp. 1 was exploratory, we did not conduct inferential statistics, but did a cluster analysis to see which NPIs pattern together (R package `cluster`, Maechler et al. 2019). We identified seven distinct clusters (CL), see Figure below. CL1 was overall unacceptable, suggesting the NPIs in CL1 are antiquated or generally infrequent. CL2 had high ratings in all conditions including the positive assertions, suggesting the NPIs in CL2 may not be NPIs. CL3 shows the licensing pattern that is expected for superstrong NPIs: acceptability ratings are high only in condition [1]. The licensing pattern in CL4 is surprising: the NPIs in CL 4 were rated equally high in conditions [1] and [5], and low in all other conditions. This pattern is at odds with our predictions: if an NPI is licensed in a question, it is expected to be licensed in downward entailing and anti-additive environments, too. This finding was explored further in Exp. 2. In CLs 5 and 6 all conditions but [4] and [6] received high acceptability ratings. The difference between CL5 and CL6 is that CL6 received better ratings overall. CL6 may be classified as a cluster of superweak NPIs. The classification of CL5 isn't quite clear. Interestingly, both CL5 and CL6 show that the nonveridical operator *vielleicht* cannot license German NPIs, suggesting that in German, questions have their independent licensing power: nonveridicality alone cannot license NPIs, at least not if the operator is *vielleicht*. CL7 does not show a clear pattern: condition [1] received the highest ratings and the other conditions received mostly medium ratings. The classification of CL7 remains unclear.
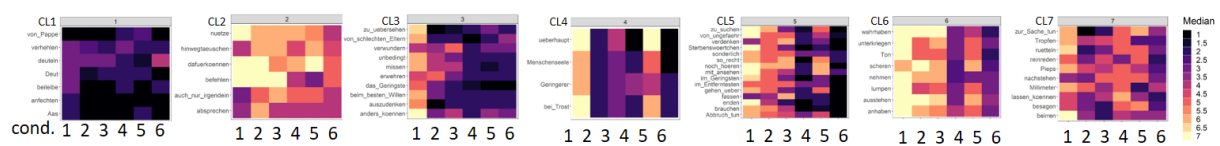
Experiment 2 explored the surprising finding concerning CL4. Although great care was taken in Exp. 1 that the questions were interpreted as information questions in the context story, some of the questions may have been interpreted as rhetorical questions, e.g. *Ist Philipp bei Trost?* (rhetorical: 'Is Philipp insane?'). Exp. 2 tested whether participants interpreted the questions that received a high acceptability rating in Exp. 1 as rhetorical, which might be a reason for the high rating (e.g. van Rooy 2003). In Exp. 2, 82 monolingual German speakers (64 f, aged 19-43, m=22) judged the acceptability of the questions from Exp.1, and additionally they judged on a 7-point Likert scale how likely they thought it was that the person asking the question knew its answer already, which we took as a measure of rhetoricity. The results indicate that rhetoricity did not contribute much to high acceptability for questions. In Exp. 2, out of 22 questions with a high acceptability rating (median $\geq$ 5), only 8 received a high rhetoricity rating (median $\geq$ 5), while 7 were rated as distinctly non-rhetorical (median $\leq$ 3). There only was a weak correlation

between acceptability and rhetoricity ($r = 0.23$). Interestingly, the findings from Exp. 2 are at odds with a generalization in the theoretical literature on NPIs in rhetorical questions: van Rooy (2003) suggests that only minimizer type NPIs can turn a question into a rhetorical question. In Exp. 2, out of the 8 questions with high ratings for acceptability and rhetoricity, only one NPI is a minimizer type NPI. This requires closer scrutiny in future research.

Overall, the two experiments show that several different licensing patterns can be distinguished for German NPIs. Importantly, not all of them match the predictions based on semantic theory, and thus the generalizations that are familiar from the literature. These new observations must be explored in future theoretical work, especially the issues of nonveridicality and questionhood need closer scrutiny. For experimental studies, the findings imply that very careful pre-testing for the selection of NPIs is necessary when using them as diagnostic.