

## Reviews

Tanja Anstatt, Anja Gattnar, and Christina Clasmeier, eds. *Slavic languages in psycholinguistics: Chances and challenges for empirical and experimental research*. Tübingen: Narr Francke Attempto, 2016. 315 pp. [Tübinger Beiträge zur Linguistik, 554.] ISBN 978-3-8233-6969-1.

Reviewed by Olga Parshina

This book is a result of the workshop on empirical psycholinguistic methods, “Slavic languages in the black box”, which aimed to create a space for discussion of methodological problems in the field of Slavic psycholinguistics. The current volume goes beyond simple discussion. In 12 different contributions, it 1) offers an overview of the existing experimental designs, online and offline methods of investigation as well as various tasks that are employed in modern Slavic psycholinguistics research, 2) presents the key issues associated with these designs and methods, and, crucially, 3) suggests possible solutions to overcome the challenges. Below I briefly outline the content, summarize the takeaway message and provide a short review for each contribution of the volume. I conclude with a general evaluation of the book.

In the first paper of the volume, Barbara Mertins discusses several online and offline methods used inside and outside psycholinguistic research. The chapter starts with a brief classification of experimental methods as online (e.g., eye-tracking, elicitation), offline (e.g., surveys), and true online methods (e.g., EEG, fMRI), along with an outline of the potential benefits and drawbacks of using one technique over another. Next, Mertins concentrates on the evaluation of several methods that she and her colleagues employed in her language-production research: 1) elicitation, 2) memory tasks, 3) eye-tracking, 4) speech onset times, and 5) preference/grammatical judgment tasks. Finally, Mertins presents three experimental studies (authored by Mertins and other colleagues) that implemented either one technique or a combination of these methods in language-production studies, enabling the reader to observe the application of the techniques in experimental contexts. In the overview of each study, Mertins provides detailed information as well as a critical evaluation of the design, materials, randomization procedures, and general protocol (e.g., sample size, language background of the participants, stimulus length,

amounts of fillers, coding procedure, etc.). Crucially, for each study, the contributor also points out benefits and issues that should be considered before planning to use one of the described methods in combination with a specific aspect of study design (e.g., elicitation method and intercultural suitability of the materials). The chapter also places an interesting focus (in study 2) on the importance of using a combination of linguistic and non-linguistic (e.g., memory) tasks to investigate the effects of language on cognition (i.e., language leads to differences in thinking). However, the link between these two types of tasks (especially the memory task) is not clear. It would be useful to provide researchers with more arguments for the necessity of adding non-linguistic tasks to the experiments. In general, the article, although likely not intended for this purpose, might serve as an excellent introduction for graduate students and early-career researchers to various psycholinguistic methods and advantages and caveats of the designs with the emphasis on language-production research.

Chapter 2 (by Roumyana Slabakova) shifts the focus of discussion from details of the experimental design to issues of the inconsistency of results caused by the variability in linguistic judgments of native Russian speakers. First, Slabakova introduces the results of her study (2004) that examines how native Russian speakers interpret telicity based on the perfectivity of the verb. Specifically, the goal of study 1 is to confirm that perfectivity of verbs in Russian (as an example of a Slavic language) dictates the (non-)quantization of the objects they refer to, an association known as Event-Object Homomorphism. Counter to expectations, the findings of the study indicated that Russian native speakers marked sentences with perfective verbs as having two possible interpretations (as opposed to one, as was expected) almost half of the time (49%). Slabakova argues that other factors than perfectivity of the verb come into play, e.g., free word order in Russian and Information Structure associated with the word order. In the second part of the contribution, Slabakova takes word order and Information Structure as points of discussion and demonstrates again based on a previous study (Cho and Slabakova 2014) that native speakers of Russian unexpectedly accepted the focused object (i.e., object carrying new information) in the preverbal position, although the Focus is typically located post-verbally. In the final section of the chapter, Slabakova suggests that such variability in judgments may be due to the fact that some grammatical meanings in Russian are underspecified (vary as a function of semantics, word order, context, intonation, Information Structure, etc.). As a result, Russian speakers show sensitivity to this variation, as is evidenced by flexibility in their linguistic judgments. In concluding remarks, Slabakova suggests that in addition to 'typical' considerations of the experimental design, such diversity of internal linguistic factors that affects speaker's interpretation should be taken into account before choosing a method for a study (e.g., online methods will not be sensitive to speaker's interpretations).

Taken together, this chapter a) provides a plausible theory of why there is high variability in interpretations produced by native speakers of Russian and thus creates a platform for future research to test the predictions of this theory and b) serves as a caution for researchers in the second language acquisition field who compare interpretations of L2 speakers to the assumed homogeneous monolingual baseline.

Chapter 3 focuses on the advantages and disadvantages of using a corpus-based approach in linguistic research as well as statistical methods appropriate for the analysis of corpus-based data. The authors, Dagmar Divjak, Antti Arppe, and Harald Baayen, discuss how Tense, Aspect, and Mood (TAM) markers affect the processing of synonymous verbs expressing the meaning 'try' in Russian. First, the authors review findings from a previous study in which Divjak and Arppe (2013) trained a polytomous logistic regression model to predict which of six synonymous verbs will fit into a sentence based on several variables that describe the properties of these verbs. The results indicated that for more frequent synonyms TAM marking is a strong predictor for the choice of 'try' verbs. Second, in a self-paced reading task the authors investigate whether reading times of these synonymous verbs are affected by the probability of TAM marking on the verb. The series of mixed-effect generalized linear-regression models indicated the lack of significant effect on the verb reading time, which the authors suggested was due to several factors, including the assumption violation of the linear relationship between the effects. As a solution, Divjak, Arppe, and Baayen used a generalized additive mixed model that can correctly estimate effects that are non-linear in nature. This model confirmed that TAM marking plays a significant role in online processing by native speakers but in an unexpected way: participants slowed down when reading verbs with a highly likely TAM marking. The explanation they suggest is the sudden change of surprisal. Without a previous context, participants were going through words quickly until they encountered the verb that tied everything together, allowing information integration. The chapter is valuable from both theoretical and methodological perspectives. First, the results of the study warn against exclusively using lemmas as predictors; one should take word forms into account, especially when dealing with morphologically rich languages. Second, the authors go into great detail in explaining each step of the analyses, including data preparation, variable coding, and reasons for adding each variable to the structure of the mixed-effect models, thus making this chapter extremely informative for researchers who plan to use these statistical methods in corpus-based research.

In Chapter 4 Anja Gattnar discusses challenges that occur when designing materials for cross-linguistic research, including inner-Slavic studies. Taking verbal aspect as an example, Gattnar provides a detailed description of how similarity among languages does not make the task of design transfer easier

but on the contrary, leads to difficulties not predicted by researchers. First, taking a previously conducted eye-tracking study as a basis, Gattnar discusses the differences between German and Russian that made the identical design transfer impossible. Among these differences, Gattnar mentions the mismatch in a number of control sentences in Russian (as the language has two verbal aspects) vs. German, the different number of syllables in Russian verbs compared to German, differences in word order, and the way the two languages express (in)definiteness with bare nouns. In the other two studies, which both used self-paced reading technique to investigate aspectual processing, Gattnar and colleagues faced the challenge of translating materials from the original study in Russian to another Slavic language—Czech. Researchers found that the problem was not trivial. The languages differ in word frequencies (e.g., names), connotations, verbal constructions, and aspect usage itself. In a final section Gattnar maintains that although it could seem counter-intuitive, the design transfer might be easier for languages with core differences in grammar. For example, crucial experimental elements are located in different sentential positions in the non-aspect German language and the aspect-bearing Russian language, removing the pressure to make the translation of other sentence elements ideal. In Slavic languages, on the other hand, these differences are peripheral, as all words in the sentence prior to or following the target construction have to match in frequency, length, connotation, etc., which is unfortunately not feasible. Gattnar provides two possible ways to deal with the challenge: 1) translating materials as close to the original as possible and 2) adapting the materials. The first suggestion, while it allows researchers to maintain design, leads to the possible necessity of reformulating the hypotheses, since the translation most likely will lead to differences in results (e.g., different reading times due to increased word length). The latter solution, while enabling researchers to keep their hypotheses, makes it almost impossible to compare results statistically due to the number of variables that should be considered in translation (e.g., frequency, length, predictability, connotations, etc.). In general, the chapter provides a detailed account of the challenges that researchers are likely to face when designing a cross-linguistic comparative study of the Slavic languages. The task might be difficult, but these studies are definitely needed to address an important limitation on existing cross-linguistic research, i.e., its bias toward Romance-based languages.

Chapter 5 by Anastasia Makarova describes two experiments that, as in previous chapters, examine the usage of aspectual morphology by native Russian speakers. In this paper, however, experiments are concerned with the distribution of and motivation for the use of affixes associated with attenuative and semelfactive Aktionsarten in Russian. Relevant to the broad topic of the book, the chapter focuses on the methodological challenges related to the selection of stimuli for the two cloze-task corpus-based experiments. In

experiment 1, which investigated the distribution of prefixes on attenuative verbs, stimuli were chosen from the Russian National Corpus, and the task was to add the most fitting prefix to the verb in the sentence. In experiment 2, which examined the prefix/suffix distribution of the morphological marking for semelfactives, the stimuli consisted of constructed contexts and nonce-verbs to elicit the whole verb form (to reduce the bias for prefix or suffix and to avoid memory retrieval of existing verbs). In sum, based on the example of studies that focus on two very similar phenomena (two types of *Aktion-sarten*) in Russian, the chapter describes the necessity for modifications in the methodology as well as statistical analysis for each research question. The takeaway message from this contribution is that in morphologically rich Slavic languages such as Russian, even closely related linguistic phenomena should be approached with methodological scrutiny. The assumption that if the first experimental design works well for examining the first phenomenon, then it would transfer to the similar second phenomenon is not valid.

In chapter 6 Denisa Bordag reviews studies with several experimental paradigms that use reaction time as a dependent variable to investigate various morphological phenomena in Czech. The first two studies are concerned with the processing of inflected verbs and use lexical-decision and repetition-priming paradigms. Besides discussing the stimuli requirements in the experiments (e.g., word frequency match) and experimental findings, Bordag notes that these studies might be the only available psycholinguistic research in Czech comprehension. In the next sections, Bordag reviews two studies (Bordag and Pechmann 2008, 2009) that employed picture-word interference-paradigm to investigate the representation and processing of such grammatical features as gender, declensional class of nouns, and the conjugational class of verbs. The studies are valuable as they add more data from rarely investigated languages to (dis)confirm psycholinguistic theories that are assumed to apply across languages. For example, Bordag discusses the Split Morphology Hypothesis in respect to the processing of inflected verbs and the Hierarchical feature selection mechanism in relation to grammatical feature processing. Crucially, these studies inform theories by exploring phenomena (e.g., declensional classes of nouns) that are typically absent in frequently explored languages such as English. In general, the chapter can serve as an inspiration for researchers on languages that have a short history of psycholinguistics research. It shows that one can conduct highly impactful and novel studies employing relatively simple and inexpensive designs.

Chapter 7, by Elena Dierker, overlaps with chapter 2 as it describes a series of studies that aim to explore cases of doubt in grammaticality judgments, cases when two or more grammatical variations are accepted as correct by native Russian speakers. In this chapter, however, the primary focus is not on the internal reasons for variability in grammatical judgments (in case and animacy categories), but on the results as a function of the experimental task.

Specifically, Dierer describes several experiments that used grammaticality-judgment tasks with (scale 1–5) and without (thermometer judgments) endpoints as well as questionnaires that required respondents to put the words under consideration into the grammatically required form based on the syntactic frames of the sentences. Furthermore, these experiments have additional tasks: to improve the forms that are considered ungrammatical. The goal of the additional tasks was to investigate whether they affect the numerical judgements of the sentences. While none of the tasks yielded statistically different results, the findings still provide valuable insights into the effectiveness of the grammaticality-judgment method. First, Dierer concludes that there is no difference in using scales with or without endpoints. Second, it was only additional tasks that revealed some curious findings: some native speakers considered assumed deviations as codified forms. Finally, the comparisons of results from the grammaticality-judgment tasks and questionnaires showed that some judgments were made accidentally, presumably due to lack of attention. To summarize, the chapter further solidifies conclusions from several contributions in this volume: 1) the optimal solution is to use multiple methods and tasks in experimental design (see also chapter 1 and 10) and 2) linguistic judgments of native speakers are not always unified and in many cases are graded: one form that is nearly unacceptable for one native speaker may be perfectly fine for another.

Chapter 8 (Julija Nigmatulina, Olga Raeva, Elena Riechakajnen, Natalija Slepokurova, and Anatolij Vencov) further builds upon the necessity of using a combination of experimental methods and careful selection of experimental materials in psycholinguistic research, here from the perspective of spoken-word recognition in spontaneous speech. Nigmatulina and colleagues start the chapter with a detailed description of the steps they undertook and challenges they faced in creating the materials for their experiments, which constitute a newly developed corpus of spontaneous Russian. The corpus includes both orthographic and phonetic transcriptions of various radio news and TV shows. Next, the authors discuss the findings of the experiment that used a dictation task to investigate the processing of reduced wordforms in spontaneous speech. They note that the dictation task, although it revealed several important aspects of using asemantic vs. semantic stimuli in the task as well as confirming expected frequency and type-frequency effects (i.e., frequency of the form is the strongest predictor of the homophone preference in asemantic condition, see also chapter 3 of the volume), it also allows for factors that cannot be controlled by the researcher (e.g., orthographic mistakes or lapses by participants, or particular strategies that a participant follows when she is writing down the words). Finally, the authors discuss the results of two other studies, which employed cloze-test and estimation of naturalness of speech methods with the purpose of confirming the influential role of context in natural speech processing that was established in the dictation task ex-

periments. The chapter concludes with a list of methodological principles that the authors recommend following when planning spoken-word recognition research. This list, however, can be generalized to any area of psycholinguistic research and can serve as a checklist in an attempt to increase the external validity of any laboratory experiment.

Chapter 9, by Christina Clasmeier, Tanja Anstatt, Jessica Ernst, and Eva Belke, looks further into the challenges researchers face when conducting spoken-word comprehension research. The authors discuss particular difficulties of choosing stimuli for a visual-world paradigm experiment that investigates differences in bilingual mental lexicon between languages from different family branches—German and Russian (see also chapter 4). First, the authors discuss problems in measuring and matching the word frequencies. Not only is it extremely difficult to select items that match in their frequencies cross-linguistically (besides having a phonological overlap in the onset), but it is also hard to establish the frequency within one language, as different dictionaries provide different results. In addition, there is no guarantee that the dictionary or corpus-based frequency list adequately represents the frequencies of the word in the participants' mental lexicon. As a solution, the authors suggest a method of collecting subjective frequencies from the participants of the study. Next, the authors describe in great detail the procedure of picture selection, which resulted in multiple sets of pre-tests and stimuli exclusion. Finally, they provide a thorough description of how they measured the phonetic distances among stimuli words in languages with drastically different phonetic systems in order to be able to choose target stimuli with the highest phonetic overlap. The chapter presents an example of a thorough approach to stimuli selection, where researchers made every attempt to consider variables that can affect the results of the study. The discussion of the possible ways to deal with word-frequency challenges is especially useful for anyone conducting research with bilingual populations, and heritage speakers in particular. As of now, there is no objective test for establishing word frequencies in such populations.

Chapter 10, by Barnhard Brehmer, Tatjana Kurbangulova, and Martin Winski, continues the topic of Slavic heritage languages and discusses the most reliable method of assessing lexical proficiency in this population. In a study with heritage speakers of Russian and Polish (dominant German), they tested four different methods—picture naming, semantic mapping, translation, and verbal fluency—of evaluating lexical abilities in both dominant and heritage languages. Based on the results of cluster analysis and correlational analysis, the authors found that in the dominant language, German, the results of the tasks did not yield significant correlations, which likely means that these tasks tap into different dimensions of lexical knowledge (e.g., active vs. passive vocabulary size). For heritage languages, however, a positive correlation was established and among the four tasks, the translation tasks

yielded the most consistent results in relation to the average scores from other assessments. The authors conclude that, although the translation task might be the best option for designs with limited time resources, the combination of several methods is the most reliable way to assess the lexical proficiency of heritage speakers comprehensively. This contribution along with chapter 9 presents an invaluable source of information for researchers who conduct studies on heritage languages, as it is extremely challenging to establish proficiency levels in heritage language speakers due to the wide range of individual variability in language abilities among these speakers. The heritage-language research community, therefore, is in desperate need of establishing some reliable assessment tools that can be used in experimental settings; these chapters present the first steps in this direction.

Jan Patrick Zeller, Gerd Hentschel, and Esther Ruigendijk in chapter 11 discuss what online methods such as event-related potentials (ERPs) can contribute to the knowledge of code-switching (CS) between two closely related languages, i.e., Russian and Belarusian. The chapter starts with a brief overview of the code-switching phenomenon and the specific type of Belarusian-Russian bilingualism. Next, the authors lay out the main goals for the study, which is to investigate 1) whether CS between two structurally close languages is different or similar to CS between languages that are semantically, syntactically, and phonetically distant, and 2) whether the direction of the switch matters. The section is followed by a CS literature overview, method, and results of the study conducted with young Belarusian-Russian bilinguals. Zeller, Hentschel, and Ruigendijk found that similar to results of studies exploring less related languages, there is an early negativity effect that suggests two separate subsystems of the mental lexicon. Curiously, the authors also report two effects that have not been observed before in CS studies: 1) the late right frontal negativity present when code-switching from Russian to Belarusian, and 2) absence of late positivity—a platform for future hypothesis testing in studies with closely related languages. In general, the chapter serves as an example of an online investigation of psycholinguistic phenomena in closely related Slavic languages and the challenges that come up as a result of this relatedness, e.g., creating stimuli and recruiting participants. In this respect, the chapter cross-references chapter 4 of this volume (by Anja Gattner). The closer the languages structurally, the harder it is to create an ideal experimental design.

The volume concludes with chapter 12, by Jakub Jehlička, which investigates whether the previous findings of reduced mental-rotation effect in users of sign language will hold for deaf users of Czech Sign Language as compared to native hearing Czech speakers (control group). The study adopts the design from research conducted with hearing participants speaking American English and American Sign Language participants (Emmorey, Klima, and Hickok 1998). In this volume, however, Jehlička presents only preliminary re-

sults for Czech hearing participants and compares them to the findings from the original study. Jehlička reports that in the condition with no rotation both groups (American English and Czech hearing participants) perform on par. The effect of mental rotation in the rotation condition, however, is much less pronounced in Czech participants than in American counterparts. Jehlička suggests that such differences may be a result of experimental design modifications, in particular changes in the presentation order of the stimuli, an interstimulus interval that was based on the production duration of the respective stimulus by a native Czech Sign Language speaker as well as subject pool composition—there were more women than men in the experiment, which is a relevant factor for a mental rotation effect study. The chapter concludes with Jehlička's remarks on the need for specific task designs when working with special populations, the need for more replication studies and especially for those that produce cross-linguistic comparisons of the same psycholinguistic phenomenon.

## Conclusion

The main goal of the book is to thoroughly overview methodological challenges and specifics of psycholinguistic studies in Slavic languages. In my opinion, the volume not only successfully accomplishes this goal but also surpasses it, as each chapter offers valuable advice and possible solutions to overcome the challenges. As such, the book will especially appeal to researchers conducting psycholinguistic experiments with Slavic languages. Although each chapter covers very distinct topics and various aspects of methodological issues, the volume gives a coherent outline of general issues that researchers deal with when designing and running an experiment: 1) choosing an appropriate method, 2) selecting or creating stimuli, 3) applying appropriate statistical tests, 4) managing cross-linguistic differences and similarities, 5) handling the cultural and individual differences of the participants, and 6) testing linguistic effects in special populations of speakers.

Another strength of the book is that it can serve not only as a resource for experimental designs but also as an inspiration for new ideas and theories—many chapters present readers with curious and yet unexplored questions in the field of Slavic psycholinguistics. Overall, this is a valuable contribution to the literature and should be read by all researchers in Slavic psycholinguistics.

## References

- Bordag, Denisa and Thomas Pechmann. (2008) "Grammatical gender in speech production: Evidence from Czech". *Journal of psycholinguistic research* 37(2): 69–85.

- Bordag, Denisa and Thomas Pechmann. (2009) "Externality, internality, and (in)dispensability of grammatical features in speech production: Evidence from Czech declension and conjugation". *Journal of experimental psychology: Learning, memory, and cognition* 35(2): 446.
- Cho, Jacee and Roumyana Slabakova. (2014) "Interpreting definiteness in a second language without articles: The case of L2 Russian". *Second language research* 30(2): 159–90.
- Divjak, Dagmar and Antti Arppe. (2013) "Extracting prototypes from exemplars: What can corpus data tell us about concept representation?" *Cognitive linguistics* 24(2): 221–74.
- Emmorey, Karen, Edward Klima, and Gregory Hickok. (1998) "Mental rotation within linguistic and non-linguistic domains in users of American sign language". *Cognition* 68(3): 221–46.
- Slabakova, Roumyana. (2004) "Effect of perfective prefixes on object interpretation: A theoretical and empirical issue". *Cahiers linguistiques d'Ottawa* 32: 122–42.

Moscow, Russia 101000  
3 Krivokolenny pereulok  
HSE University  
Center for Language and Brain, aud. 311.  
oparshina@hse.ru