Department of English and American Studies

Nikolaus Ritt
[STRASSE]
[ORT]

T +43 4277 42424
M +43 650 327 0160
Nikolaus.ritt@univie.ac.at
https://anglistik.univie.ac.at/staff/staff/detail-seite/user/rittn8/inum/1063/backpid/30750/

To whom it may concern

Vienna, 28/07/23

External examiner's review of Alona Kononenko-Szoszkiewicz' doctoral thesis

Exploring (mor)phonotactic patterns: A comparative study of selected Germanic and Slavic languages

The core of the submitted thesis is made up by four research articles, three of which have already been published, and one accepted, in peer reviewed academic journals or edited volumes. They are integrated into the present thesis as appendices. The thesis also features a substantial introduction, which (a) describes the research area explored in the thesis as well as the specific theoretical frameworks on which the author has based her phonological and morphological analyses, (b) provides an overview and classification of extant research applying corpus linguistic methods, and (c) introduces the submitted research articles by means of short summaries. Finally, the thesis provides a summarizing discussion of the main findings of the four studies, which also acknowledges some of their limitations, and suggests directions for possible further research. Appendices contain the research articles themselves as well as some additional material (such as a list of languages covered in extant corpus linguistic research), and, importantly, statements describing the applicant's contributions to the three research articles that are co-authored.

Before giving a detailed assessment of the individual parts of the thesis, I make clear at the outset that the thesis fulfils, in all respects, the requirements for receiving an unambiguous pass grade, even though it does not quite achieve excellence. This verdict will be justified in the following appreciative discussions of the submitted research articles, the introduction, and the discussion.

I begin my review with research article 4, because is the only one which the candidate has published without co-authors, because it exemplifies best where the strengths of the submitted thesis lie, and because it also allows one to pinpoint some of the issues that that might have been dealt with in some more depth and that the candidate may find worth pursuing in her further career as a researcher.

Article 4 explores the morphonotactics of Ukrainian by providing an exhaustive survey of word-initial consonant clusters, by analysing them in terms of their markedness (or preferability), and by using them as empirical evidence in order to test predictions that have been made in extant literature on the interaction of phonology and morphology in the domain of (mor)phonotactics. Her explorative and descriptive work is corpus based, and through the application of appropriate queries on the General regionally Annotated Corpus of Ukrainian, Kononenko manages to extract the full set of initial consonant clusters attested in that corpus. Since this has never been undertaken before, and also

because Ukrainian phonology is generally understudied, this achievement breaks new ground and fills an obvious research gap. In a very reader friendly manner, the paper provides a full list of identified clusters together with word type and token frequencies, as well as an example item with its English translation. That she makes her data accessible in this manner show how well she has internalized emerging standards of open-access-open data-research. In short, the paper reports and describes a set of interesting and so far undocumented data reliably and by exploiting contemporary methods of digital corpus analysis. In that regard it can be evaluated as a genuine and original contribution to phonological research and the linguistics of Slavic languages.

The paper also pursues a theoretical agenda, however, and in this respect its merits lie more in raising interesting questions than in settling them, and in demonstrating the complexities of the issues at hand. The paper is theoretically grounded in Natural Linguistics and attempts to test two predictions that are derived from theories of phonological markedness, or rather preferability, and from theories concerning the interaction of phonology and morphology in the domain or word-level phonotactics. The first prediction is, in Kononenko's words, that "morphonotactic clusters [i.e. clusters that come about through morphological operations, and that typically span morpheme boundaries] are expected to be less preferred than phonotactic ones [i.e. morpheme internal clusters]". The second prediction is that "the degree of cluster preferability is [intended to be read as *should be*] directly proportional to frequency".

She tests these predictions against her data by measuring the preferability of clusters in terms of the auditory distances between their constituents and the vowels next to them. For that purpose, she employs an online tool devised by the research group of Katarzyna Dziubalska-Kołaczyk that calculates it automatically, i.e. it puts out both a number corresponding to the assumed degree of preferability of a cluster, as well as a binary measure by which clusters are categorized as either preferred or dispreferred. Frequency is established both on the type level (i.e. as the number of attested cluster types), and on the token level (i.e. as the number of attestations of each cluster type).

As her data show, about two thirds of phonotactic double cluster types are categorized as preferred, while among morphonotactic clusters, the opposite is the case. This is correctly interpreted as confirming prediction 1, but only if that hypothesis is interpreted as a prediction about the likelihood of types to fall into the categories of preferred clusters on the one hand, and dispreferred ones on the other (which is actually not completely clear from Kononenko's formulation of the prediction). Among triple clusters, the data show very much the same thing: again, about two thirds of the phonotactic cluster types come out as preferred, while more than two thirds of all morphonotactic cluster types come out as dispreferred. All in all, however, Kononenko's interpretation is plausible.

As to the second prediction, however, the data are somewhat ambiguous, and Kononenko does not fully succeed at their interpretation. On the one hand, and only for triple clusters, she addresses the issue on the type level, reading prediction 2 as saying one ought to find more preferred clusters than dispreferred ones. The fact the opposite seems to be the case, is not interpreted as a clear falsification of the hypothesis, however, but as a partial confirmation, and this interpretation is justified with reference to the high token frequency of preferred cluster types in the morphonotactic subset. While the observation may certainly be factually correct, adducing token frequency in the context of a type level analysis is not entirely consistent. Kononenko then addresses the same issue on the token level as well. Here she tests the prediction that there ought to be a correlation between the degree of preferability of clusters as measured by the Net Auditory Distance Calculator (i.e. not their categorical membership in the preferred or dispreferred set) and their token frequency. Impressively (albeit not fully transparently), Kononenko establishes that correlation by means of a regression analysis and displays her results in two scatterplots. For a reason that I do not fully understand, Kononenko calculates correlations not only for the whole set of attested clusters, but also separately for the

categorically preferred and the categorically dispreferred clusters. Anyway, her analysis yields significant but weak correlations only for the dispreferred subset among double clusters, and for the complete set of triple clusters. Kononenko then interprets these results as partially confirming the prediction that preferability and token frequency ought to correlate. While this interpretation appears plausible for triple clusters, I don't think it can be fully upheld for double clusters, particularly since the significant correlation that Kononenko observes is weakly negative (!) rather than positive, as would have been predicted.

On the whole, therefore, the theoretical agenda pursued in the article is not fulfilled quite as successfully as its descriptive and explorative one. While it is a good thing, of course, that rigorous quantitative methods are applied to test interesting hypotheses proposed in the relevant literature, and while it is also good to try and use advanced statistical tools for the purpose, the application of these methods and tools could have profited from a bit more transparency and precision.

Another general point to be made about the theoretical embedding of Kononenko's work, is that she is firmly based in the theory of Natural Linguistics, and has incorporated not only its basic tenets, but displays an excellent familiarity with the specific sub-theories that are relevant for the research agenda she pursues in the submitted thesis. This is particularly true of Katarzyna Dziubalska-Kołaczyk's beats-and-bindings phonology, as well as of the theory of morphonotactics proposed by Dresser and Dziubalska-Kołaczyk. Thus, her work is theoretically well supported, and she adheres to the principles of the theories she subscribes to both faithfully and rigorously.

As I will show in the rest of my review, the other studies are in principle very comparable to the first one. They are all very strong in terms of description, they all explore uncharted territory and represent genuine and original contributions to research in the phonology of Slavic and Germanic languages. They are all characterized by a strict adherence to the basic assumptions of Natural Linguistics and display an extraordinary familiarity with the way in which that theory addresses questions in the field of phonotactics and the interaction of phonology with morphology. When they go beyond exploration and description, they either develop plausible hypotheses (research article 1), or test hypotheses developed in the subfields that the thesis contributes to. They do so rigorously and sometimes with advanced statistical methods (although they are not always quite appropriately applied). In the following I briefly discuss the remaining chapters in order of their appearance in the thesis.

Research article 1, entitled, *Morphological richness, transparency and the evolution of morphonotactic Patterns* discusses the emergence of consonant clusters and particularly morphonotactic ones from a diachronic perspective. It provides an exhaustive survey of the mechanisms by which consonant clusters may come about and illustrates them with examples from a number of different languages from a variety of Indo-European families. The article is impressive for a number of reasons. On the one hand, for the breadth of languages covered, and on the other hand for the large variety of relevant factors that it identifies, and about whose interaction it formulates clear and plausible hypotheses. It refrains from oversimplifications, and remains responsibly tentative in its conclusions, which is the best strategy for an article of this kind, whose main purpose is to identify relevant factors in the emergence and the historical stabilization of sound patterns that count as universally marked or dispreferred. The article is also well structured, clearly distinguishing between an introductory part surveying the phenomena to be accounted for, and a theoretical part in which explanatory models of the ways in which morphological and phonological preferences interact to give rise to morphonotactic clusters on the one hand, and phonotactic ones on the other. The article complements the diachronic perspective with arguments and evidence from language acquisition, and from typology, demonstrating the specific role that differences in morphological richness play in the establishment of consonant

clusters in the (mor)phonotactics of specific languages. It represents a valuable contribution to the field and a good starting point for any phonologist interested in the diachrony of cluster phonotactics.

Research article 2, *Main Differences Between German and Russian (Mor)phonotactics*, approaches consonant clusters from a contrastive perspective. Its agenda is both descriptive and explanatory. As to its descriptive side, it is based on corpus evidence. For German, the vast Austrian Media Corpus has been queried, and for Russian a much smaller corpus extracted from a Russian-Polish dictionary. For the questions addressed the relatively small size of the latter (about 37000 words) does not matter, since at least type diversity is well represented in dictionary data. The paper is, like all others, descriptively impeccable, and provides an exhaustive survey of word-initial and word-final clusters. For each cluster, its preferability score (once again based on Dziubalska-Kołaczyk's Calculator) is reported, as well as it categorical attribution to the set of preferred or dispreferred clusters, and its status as a phonotactic, morphonotactic, or mixed cluster. Words exemplifying the clusters are also provided. Since the main agenda of the paper is descriptive, the helpful tables make up a large part of it, and much of the actual text summarizes the findings that are also reported in the tables. In terms of explanations, the paper accounts for the differences between Russian in German mostly with reference to differences in the inventories of consonantal affixes: Russian has purely consonantal prefixes that German lacks, and this provides a plausible reason why Russian is richer than German in initial clusters. For suffixes, the opposite is the case. This is why the inventory of Russian initial clusters is richer than that of German, while German has more final clusters than Russian. The facts reported in the paper are no doubt correct, if not altogether surprising. Interestingly, the paper also shows that the asymmetry in the distribution of clusters does not only pertain to morphonotactic ones, as would be expected, but seems to be mirrored – at least weakly – in the distribution of phonotactic, i.e. morpheme-internal ones as well. This raises the interesting question whether this correlation has a causal motivation or whether it is accidental. Due to the primarily descriptive agenda of the paper, this question is not really taken up, but that it has been raised is valuable enough.

Research article 3, *German phonotactic vs. morphonotactic obstruent clusters: a corpus linguistic analysis*, focusses on German and deals specifically with the difference between morphonotactic and phonotactic obstruent clusters (where obstruent clusters include also such that comprise sonorants as well, but because they are considered typical of German only clusters with at least two obstruents are focussed on). In contrast to the other three contributions, the article contains a theoretical introduction that describes its central tenets (typically only referred to in the other articles) in some detail, which will be very helpful for a wider linguistic readership. Its empirical basis is the vast Austrian Media Corpus, from which words containing the relevant clusters have been extracted. The paper provides a survey of all quadruple and triple clusters that occur word-initially and word-finally. For all clusters it provides example words, and identifies them as phonotactic, morphonotactic or both. Sometimes, interesting etymological information or information about dialectal variants is provided. The clusters are then also reported in table form, listing the number of word types in which they appear their token frequencies, and the type-token ratios. For triple clusters, also their preferability as calculated by Dziubalska-Kołaczyk's Calculator is reported, and they are categorized as preferred or not. The reason why the same is not done for quadruple clusters is that the Calculator cannot yet handle them. A similarly pragmatic decision has been taken regarding word medial clusters. Since the interface of the Austrian Media Corpus does not allow their easy extraction, they have been dealt with only summarily. Albeit not based on quantitative corpus evidence, their discussion strikes me as plausible in all respects. – The paper also includes an attempt to establish correlations between all the variables that have gone into the analysis (i.e. type frequency, token frequency, Net Auditory Distances between the consonants and the neighbouring vowels, preferability, morphonotactic vs. phonotactic) by means of a factor analysis. The analysis was certainly carried out correctly but does not strike me as the most appropriate method, since some of the variables depend on one another. Thus, it would have been

surprising if the factor analysis did not show a correlation between the different auditory differences and the preferability of the clusters between the three do not vary independently (in fact preferability follows from the former two). Thus, this paper is most similar to research paper 1, in that it is descriptively impeccable. It also provides plausible accounts of the collected data but shows some room for improvement as far as the application of quantitative methods of analysis and the interpretation of their results are concerned.

Before coming to my overall conclusion, I turn to the introduction provided in the thesis, to the discussion, and the conclusion. Basically, they are all well written, but actually add little to the research articles themselves. The introduction briefly sketches key elements of Dressler's approach to morphonotactics and his scalar classification of clusters even though the papers themselves make little reference to that classification. It also sketches beats-and-binding phonology and the principles on which Dziubalska-Kołaczyk's Auditory Distance Calculator works, but in order to gain a real understanding of them, one would have to go to the original works. The introduction also includes a bibliometric survey of research done in the area of corpus linguistics, which is interesting but does not obviously relate to the research articles except for the fact that three of them also apply corpus linguistic methods. Finally, the introduction provides summaries of the four research articles themselves. While I am not saying that the introduction does not do its job, it would have been nice had it been used, for example, for contextualising the submitted work in the wider field of research of phonotactics outside the school of Natural Linguistics. – *Mutatis mutandis*, what applies to the introduction, applies to the discussion and conclusion as well. Essentially, they are summaries of the central articles. As such, they do a good job but providing a little bit of meta-theoretical reflection would certainly not have done much harm either. The concept of Net Auditory Distance, for example, which figures centrally in all of the four papers, is clearly a concept in development and has not yet unfolded its full potential. Some of the findings produced in the papers would have lent themselves perfectly for reflecting, for example, in what way the operationalization of the concept might be modified. As already said, of course, the introduction, discussion, and conclusion do a decent job as they are, but together with some of the methodological issues observed in some of the papers, their rather minimalist character prevent me from classifying the thesis as excellent.

As already made clear at the outset and as, I hope, has become evident in my discussion, the submitted thesis represents a solid piece of research. Its strengths lie in description, and in exploring genuinely new ground, empirically speaking, in the domain of consonant cluster phonotactics. That it meets academic standards in all respects does not need to be highlighted, as all of the papers in the thesis have passed rigorous peer review. For the reasons that I have also given in my discussion above, and although I will not propose to assess the submitted thesis as excellent, it strikes me as solid and good and certainly meets all requirements for earning the candidate a doctorate.

Vienna, July 28th, 2023